International Journal of Interactive Multimedia and Artificial Intelligence

June 2020, Vol. VI, Number 2
ISSN: 1989-1660

unir LA UNIVERSIDAD EN INTERNET

"Cyber hygiene, patching vulnerabilities, security by design, threat hunting and machine learning based artificial intelligence are mandatory prerequisites for cyber defense against the next generation threat landscape. "

James Scott

INTERNATIONAL JOURNAL OF INTERACTIVE MULTIMEDIA AND ARTIFICIAL INTELLIGENCE
ISSN: 1989-1660 –VOL. 6, NUMBER 2

Dr. Ke Ning, CIMRU, NUIG, Ireland
Dr. Monique Janneck, Lübeck University of Applied Sciences, Germany
Dr. S.P. Raja, Vel Tech University, India
Dr. Carina González, La Laguna University, Spain
Dr. Mohammad S Khan, East Tennessee State University, USA
Dr. David L. La Red Martínez, National University of North East, Argentina
Dr. Juan Francisco de Paz Santana, University of Salamanca, Spain
Dr. Yago Saez, Carlos III University of Madrid, Spain
Dr. Guillermo E. Calderón Ruiz, Universidad Católica de Santa María, Peru
Dr. Moamin A Mahmoud, Universiti Tenaga Nasional, Malaysia
Dr. Madalena Riberio, Polytechnic Institute of Castelo Branco, Portugal
Dr. Juan Antonio Morente, University of Granada, Spain
Dr. Manik Sharma, DAV University Jalandhar, India
Dr. Elpiniki I. Papageorgiou, Technological Educational Institute of Central Greece, Greece
Dr. Edward Rolando Nuñez Valdez, Open Software Foundation, Spain
Dr. Juha Röning, University of Oulu, Finland
Dr. Paulo Novais, University of Minho, Portugal
Dr. Sergio Ríos Aguilar, Technical University of Madrid, Spain
Dr. Hongyang Chen, Fujitsu Laboratories Limited, Japan
Dr. Fernando López, Universidad Internacional de La Rioja - UNIR, Spain
Dr. Jerry Chun-Wei Lin, Western Norway University of Applied Sciences, Norway
Dr. Mohamed Bahaj, Settat, Faculty of Sciences & Technologies, Morocco
Dr. Abel Gomes, University of Beira Interior, Portugal
Dr. Víctor Padilla, Universidad Internacional de La Rioja - UNIR, Spain
Dr. Mohammad Javad Ebadi, Chabahar Maritime University, Iran
Dr. José Manuel Saiz Álvarez, Tecnológico de Monterrey, México
Dr. Alejandro Baldominos, Universidad Carlos III de Madrid, Spain

Editor's Note

The International Journal of Interactive Multimedia and Artificial Intelligence - IJIMAI (ISSN 1989 - 1660) provides an interdisciplinary forum in which scientists and professionals can share their research results and report new advances on Artificial Intelligence (AI) tools or tools that use AI with interactive multimedia techniques. Already indexed in the Science Citation Index Expanded by Clarivate Analytics, within the categories "Computer Science, Artificial Intelligence" and "Computer Science, Interdisciplinary Applications", during the next month the journal will be listed in the 2019 Journal Citation Reports [1]. Again, given this great milestone, the IJIMAI Editorial Board reiterates its appreciation for their support to authors, reviewers and readers.

The present regular issue starts with two articles related to one of the most relevant problems nowadays, which is the COVID-19 pandemic. Over the past years, there have been great advancements in health, but the evidence is that it remains a challenge to deal with pandemics and to achieve global health. IJIMAI has always reserved a space for heath topics [2] [3] [4] and in the last years, a Special Issue on Big Data and e-health [5] or a Special Issue on 3D Medicine and Artificial Intelligence [6] were published. As Mochón and Baldominos state [7], "from a global perspective, a clear statement can be made: Artificial Intelligence can have an immense positive impact on societies... AI is turning into a key player at the time of diagnosing diseases at an early stage or developing new medicines and specialized treatment". Being aware of this, a great number of researchers and scientific entities are focusing the efforts in this field and, specifically on the current world pandemic, as the researchers involved in the first two articles of this regular issue.

The first article by Dur-e-Ahmad and Imran [8] proposes the use of a SEIR model to estimate the basic reproduction number R_0, obtaining an accurate prediction of the pattern of the infected population with data from some of the most affected countries by COVID-19 at the time of the research. An interesting finding is the identification of the most significant parameter values contributing to the estimation of R_0. The next article, by Saiz and Barandiaran [9], targets to quick detection of COVID-19 in chest X-ray images using deep learning techniques. They use a merged dataset that includes pneumonia images, obtaining a robust method able to distinguish between COVID-19 and pneumonia diseases.

The next work by Shikha, Gitanjali, and Kumar [10] proposes a hybrid content-based image retrieval system (CBIR) to target limitations of those systems based on a single feature extraction, traditional inefficient machine learning approaches or lacking semantic information. They propose a system that extracts color, texture and shape features, uses an extreme learning machine classifier and relevant feedback to capture the high-level semantics of an image. The experiments report that the proposal outperforms other state-of-art related CBIR solutions.

In the field of affecting computing, Huang et al. [11] describe a three-dimensional space model valence-arousal-dominance (VAD) based on the theory of psychological dimensional emotions. Specifically, they study the clustering and evaluation of emotional phrases. The work proposes a VAD based model, develops a rule-based inference system using fuzzy perceptual evaluation, and introduces dimensional affective based VAD clustering called VADdC, taking successfully application on a dataset that has been acquired from an online questionnaire system.

Clustering methods are also used in the next study by Fyad, Barigou, and Bouamrane [12]. They are applied to the analysis of genes. Their process consists of grouping data (gene profiles) into homogeneous clusters using distance measurements. Although various clustering techniques are applied with this objective, there is no consensus for the best one. Therefore, this paper describes the comparison of seven clustering algorithms against the gene expression datasets of three model plants under salt stress.

The next papers relate to the *Internet of Things* (IoT). This is supported by the Radio Frequency Identification (RFID) technology. RFID networks usually require many tags and readers and computation facilities, having limitations in energy consumption. Thus, for saving energy, networks should operate and be able to recover in an efficient way. The first work by Rathore, Kumar and García-Díaz [13], enlarges the RFID network life span through an energy-efficient cluster-based protocol used together with the Dragonfly algorithm, managing complex networks with reduced energy consumption.

The second paper about the IoT by Balakrishna et al. [14] targets the unification of streaming sensor data generated by the IoT devices and the automatic semantic annotation of the data. They present an Incremental Clustering Driven Automatic Annotation for IoT Streaming Data (IHC-AA-IoTSD) using SPARQL to improve the annotation efficiency. The approach is tested on three health datasets and compared with other state-of-art approaches finding encouraging results.

Next work by Ríos-Aguilar, Sarría and Pardo [15] proposes a mobile information system for class attendance control using Visible Light Communications (VLC). This system allows the automatic clocking in and clocking out of students through their mobile devices, so that lectures do not spend time in attendance management. A proof of concept has been developed, setting up a testbed representing a real world classroom environment for experimentation, showing the viability of the system.

Nowadays, there is a variety of speech processing applications which suffer from background noise distortions. Saleem, Khattak and Verdú [16] present a comprehensive review of different classes of single-channel speech enhancement algorithms in the unsupervised perspective in order to improve the intelligibility and quality of the contaminated speech. A taxonomy based review of the algorithms is presented and the associated studies regarding improving the intelligibility and quality are outlined. Objective experiments have been performed to evaluate the algorithms and various problems that need further research are outlined.

Next paper by Hurtado et al. [17] targets the problem of automatically detecting large homogenous groups of users, which is a very useful task in recommender systems like those focused on e-commerce and marketing. These authors use clustering methods based on hidden factors instead of ratings to make a virtual user that represents the set of users of a group. The approach outperforms the state-of-the-art baselines, specifically improving results when it is applied to very sparse datasets.

Nowadays, there is a search of alternative energy sources due to the lack of conventional ones and the pollution caused. Fuel cells are considered promising sources, specifically the proton exchange membrane fuel cell (PEMFC) can be a suitable solution for various applications. Sultan et al. [18] propose to apply the Tree Growth Algorithm (TGA), an optimization technique, to extract the optimal parameters of different PEMFC stacks. Four case studies of commercial PEMFC stacks under various operating conditions are used to validate

DOI: 10.9781/ijimai.2020.05.007

the solution and to compare with other optimization techniques, showing better results for the TGA-based approach.

The following article goes back to the education field, Villagrá-Arnedo et al. [19] describe a student's performance prediction system based on support vector machines. The aim is to help teachers diagnose students and select the better moments for teacher intervention. The system exploits the time-dependent nature of student data, producing by weekly predictions which are shown in progression graphs that have the potential for giving early insight into student learning trends.

These days there is a high demand of continuous intelligent monitoring systems for human activity recognition in different contexts. Deep learning techniques arise again in this issue with their application in a new two-stage intelligent framework for detection and recognition of human activity types inside premises. Verma et al. [20] propose this framework to recognize human single-limb and multi-limb activities in real-time using video frames. A Random Forest classifier is used to distinguish input activities into human single-limb and multi-limb. Then, a two 2D Convolution neural network classifier is used to recognize the activities. The experiments done shows a high accuracy in real time recognition of the activity sequences.

This regular issue closes with an article of García-Holgado, Marcos-Pablos and García-Peñalvo [21] which is of interest for the whole scientific community. Although there are different methods for systematic reviews of literature to address a research question, there is no a method to undertake a systematic review of research projects, which are not only based on scientific publications. Therefore, their work provides the guidelines to support systematics reviews of research projects following the method called Systematic Research Projects Review (SRPR). The proposal represents a solid base to define future research projects, providing a method to identify the gaps in previous research projects, to identify the results of other projects that can be reused, and to prove the innovation of the new projects.

Dr. Elena Verdú

References

[1] R. González-Crespo and E. Verdú, "Editor's Note," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp.4-5. 2019.

[2] E. Verdú, "Editor's Note," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 5, pp.4-5. 2019.

[3] F. Mochón. "Editor's Note," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 4, pp.4-7. 2019.

[4] R. González-Crespo, "Editor's Note," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 3, pp.4-7. 2018.

[5] F. Mochón and C. Elvira, "Editor's Note," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 4, no. 7, pp.4-6. 2018.

[6] J.L. Cebrian Carretero, "Editor's Note," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 4, no. 5, pp.4. 2017.

[7] F. Mochón and A. Baldominos-Gómez, "Editor's Note," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 6, pp.4-5. 2019.

[8] M. Dur-e-Ahmad and M. Imran, "Transmission Dynamics Model of Coronavirus COVID-19 for the Outbreak in Most Affected Countries of the World," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no. 2, pp. 7-10, 2020.

[9] F. A. Saiz and I. Barandiaran, "COVID-19 Detection in Chest X-ray Images using a Deep Learning Approach," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no. 2, pp. 11-14, 2020.

[10] B. Shikha, P. Gitanjali, and D. Pawan Kumar, "An Extreme Learning Machine-Relevance Feedback Framework for Enhancing the Accuracy of a Hybrid Image Retrieval System," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no. 2, pp. 15-27, 2020.

[11] W. Huang, Q. Wu, N. Dey, A. S. Ashour, S. J. Fong, and R. González Crespo, "Adjectives Grouping in a Dimensionality Affective Clustering Model for Fuzzy Perceptual Evaluation," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no. 2, pp. 28-37, 2020.

[12] Houda Fyad, Fatiha Barigou, and Karim Bouamrane, "An Experimental Study on Microarray Expression Data from Plants under Salt Stress by using Clustering Methods," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no. 2, pp. 38-47, 2020.

[13] P. S. Rathore, A. Kumar, and V. García-Díaz, "A Holistic Methodology for Improved RFID Network Lifetime by Advanced Cluster Head Selection using Dragonfly Algorithm," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no. 2, pp. 48-55, 2020.

[14] S. Balakrishna, M. Thirumaran, V. Kumar Solanki, and Edward Rolando Núñez-Valdez, "Incremental Hierarchical Clustering driven Automatic Annotations for Unifying IoT Streaming Data," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no. 2, pp. 56-70, 2020.

[15] S. Rios-Aguilar, I. S. M. Mendivil, and M. B. Pardo, "NFC and VLC based Mobile Business Information System for Registering Class Attendance," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no. 2, pp. 71-77, 2020.

[16] N. Saleem, M. I. Khattak, and E. Verdú, "On Improvement of Speech Intelligibility and Quality: A Survey of Unsupervised Single Channel Speech Enhancement Algorithms," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no. 2, pp. 78-89, 2020.

[17] R. Hurtado, J. Bobadilla, A. Gutiérrez, and S. Alonso, "A Collaborative Filtering Probabilistic Approach for Recommendation to Large Homogeneous and Automatically Detected Groups," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no. 2, pp. 90-100, 2020.

[18] H. M. Sultan, A. S. Menesy, S. Kamel and F. Jurado, "Tree Growth Algorithm for Parameter Identification of Proton Exchange Membrane Fuel Cell Models," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no. 2, pp. 101-111, 2020.

[19] C. J. Villagrá-Arnedo, F. J. Gallego-Durán, F. Llorens-Largo, R. Satorre-Cuerda, P. Compañ-Rosique and R. Molina-Carmona, "Time-Dependent Performance Prediction System for Early Insight in Learning Trends," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no. 2, pp. 112-124, 2020.

[20] K. K. Verma, B.M. Singh, H. L. Mandoria, and P. Chauhan, "Two-Stage Human Activity Recognition Using 2D-ConvNet," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no. 2, pp. 125-135, 2020.

[21] A. García-Holgado, S. Marcos-Pablos, and F. J. García-Peñalvo, "Guidelines for performing Systematic Research Projects Reviews," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no. 2, pp. 136-144, 2020.

TABLE OF CONTENTS

OPEN ACCESS JOURNAL

ISSN: 1989-1660

The International Journal of Interactive Multimedia and Artificial Intelligence is covered in Clarivate Analytics services and products. Specifically, this publication is indexed and abstracted in: *Science Citation Index Expanded, Journal Citation Reports/ Science Edition, Current Contents®/Engineering Computing and Technology*.

Transmission Dynamics Model of Coronavirus COVID-19 for the Outbreak in Most Affected Countries of the World

Muhammad Dur-e-Ahmad[1]*, Mudassar Imran[2]

[1] University of Waterloo, Ontario, Canada & Sigma Business Analytics and Technology Solutions (Canada)
[2] Center for Applied Mathematics and Bioinformatics, Department of Mathematics Gulf University for science and technology (Kuwait)

Received 14 March 2020 | Accepted 1 April 2020 | Published 2 April 2020

unir
LA UNIVERSIDAD EN INTERNET

Abstract

The wide spread of coronavirus (COVID-19) has threatened millions of lives and damaged the economy worldwide. Due to the severity and damage caused by the disease, it is very important to fore-tell the epidemic lifetime in order to take timely actions. Unfortunately, the lack of accurate information and unavailability of large amount of data at this stage make the task more difficult. In this paper, we used the available data from the mostly affected countries by COVID-19, (China, Iran, South Korea and Italy) and fit this with the SEIR type model in order to estimate the basic reproduction number R_0. We also discussed the development trend of the disease. Our model is quite accurate in predicting the current pattern of the infected population. We also performed sensitivity analysis on all the parameters used that are affecting the value of R_0.

Keywords

Coronavirus COVID-19, Stability, Sensitivity Analysis,Statistical Inference, Basic Reproductive Number.

DOI: 10.9781/ijimai.2020.04.001

I. Introduction

Coronavirus, commonly known as COVID-19, was first detected in Wuhan City, Hubei Province, China. Initially, it was linked to a live animal market but is now spreading from person-to-person in a similar way to influenza, via respiratory droplets from coughing or sneezing on a continuum pattern [2], [11]. From human to human transition, the toll of the infested is rising almost exponentially at this early stage. The time between exposure and symptom onset is typically seven days but may range from two to fourteen days [11]. Due to its severity and spread in many countries, a global health emergency warning was issued by the World Health Organization on January 30, 2020 [8].

The number of infected cases of coronavirus (COVID-19) has skyrocketed since the first announcement on December 31, 2019, in China [1]. As of March 9, 2020, more than 111,000 cases have been confirmed positive from 111 countries and territories around the world and 1 international conveyance (the Diamond Princess cruise ship harbored in Yokohama, Japan [1], [8]).

Since the beginning of the outbreak, besides laboratory work, significant effort is also currently going into quantitative modeling of the epidemic. Particularly noteworthy in this connection is the time delay model [5], the prediction model [6], [10] and a model involving the basic reproduction number [7], [9]. Although these models discuss important aspects of disease dynamics, however, in the current scenario, a more detailed diagnostics based model which also depends on the real existing data of active cases is important to monitor the pattern of the epidemic. In this article, we used the data of active cases from the top four mostly effected countries from the outbreak, China (with 80,739 cases), South Korea (with 7,478 cases), Italy (with 7375 cases) and Iran (with 7,161 cases) [8], during the periods of January 23 and March 5, 2020. We used an SEIR model to fit the data and compute the basic reproduction number using the next generator operator method proposed in [12]. We also checked the sensitivity of R_0, based on various parameter values used in our data-driven model by computing the partial rank correlation coefficient (PRCC).

* Corresponding author.
E-mail address: mdureahm@asu.edu

II. Model Formulation

We base our study on a deterministic ordinary differential equations (ODE) epidemic model in which the population size is divided into four mutually exclusive compartments. The total population at any time instant t, denoted by $N(t)$, is the sum of individual populations in each compartment that includes susceptible individuals $S(t)$, exposed individuals $E(t)$, infected individuals $I(t)$, and recovered individuals $R(t)$, such $N(t) = S(t) + E(t) + I(t) + R(t)$.

Since there is no vertical transmission of the infection, we assume that all newborns are susceptible. The population of susceptible individuals is generated at a constant recruitment rate π and diminishes following effective contact with an infected individual(s) at rate λ. The susceptible population also decreases at a natural death rate μ_1. The differential equation governing the dynamics of the susceptible population is given as,

$$\frac{dS}{dt} = \pi - (\lambda + \mu_1)S \quad (1)$$

Where

$$\lambda = \frac{\beta(\eta_1 E + \eta_2 I)}{N}$$

The parameter β denotes the affective contact rate and η_i, $i \in \{1,2\}$, represents the modification parameters accounting for the relative infectiousness of individuals in the *E* and *I* classes. It may represent the effectiveness of adopting social distances.

The exposed population is generated after susceptible acquire infection at a rate λ The population in this class diminishes when individuals become asymptomatically infected or get infected at a rate α_1. Exposed individuals also die a natural death rate μ_1. Thus,

$$\frac{dE}{dt} = \lambda S - (\alpha_1 + \mu_1)E \tag{2}$$

The population for asymptomatically infected individuals, generated at a rate α_1 decreases when individuals in this compartment recover at rate κ_1. This class also diminishes when individuals die at a natural death rate μ_1 or disease-induced death rate μ_2. The corresponding differential equation is therefore given as,

$$\frac{dI}{dt} = \alpha_1 E - (\kappa_1 + \mu_1 + \mu_2)I \tag{3}$$

Finally, the population of recovered individuals is generated when various measures are used, such as hospitalization, quarantine, medication and other precautions. Individuals recover at a rate κ_1. This class also decreases when recovered individuals die naturally at a rate μ_1. Since we are investigating the transmission dynamics of the current epidemic season, we assume that every recovered individual gain immunity for the rest of the season. The differential equation for recovered individuals is therefore given as,

$$\frac{dR}{dt} = \kappa_1 R - \mu_1 R \tag{4}$$

The description of all the model variables and parameters is given in Table I and Table II.

TABLE I. Descriptions of Model Variables

Values	Descriptions
N(t)	Total human population
S(t)	Population of susceptible humans
E(t)	Population of exposed humans
I(t)	Population of infected humans
R(t)	Population of recovered humans

TABLE II. Description and Values of Model Parameters

Parameters	Description	Values
π	Recruitment rate of humans	100 - 1000
μ_1	Natural death rate of humans	0.003 - 0.01
μ_2	Disease-induced death rate of infected individuals	0.025 - 0.035
κ_1	Recovery rate of infected humans	0.3
α_1	Progression rate from exposed to infected class	0.18 - 0.22
β	Effective contact rate	0.4 - 0.45
η_1, η_2	Modification parameter for relative infectiousness	0.4 - 1

III. Basic Reproduction Number

The basic reproduction number R_0 is the number of individuals infected by a single infected individual during the infectious period in the entire susceptible population. To find its relations, we adapted a next-generation matrix approach [12]. Using differential equations associated with the exposed *E* and infected *I* compartments as stated below, we compute a function *F* for the rate of new infection terms entering, and another function *V* for the rate of transfer into and out of the exposed and infected compartments by all possible means depicted in the model.

$$E' = \lambda S - (\alpha_1 + \mu_1)E$$
$$I' = \alpha_1 E - (\kappa_1 + \mu_1 + \mu_2)I$$

The matrices *F* (for the new infection terms) and *V* (of the transition terms) are given by,

$$F = \begin{pmatrix} \beta\eta_1 & \beta\eta_2 \\ 0 & 0 \end{pmatrix} \text{ and } V = \begin{pmatrix} \alpha_1 + \mu_1 & 0 \\ -\alpha_1 & \kappa_1 + \mu_1 + \mu_2 \end{pmatrix}$$

Reproductive ratio R_0 is then computed as $R_0 = \rho(FV^{-1})$ where ρ is the spectral radius (the maximum eigenvalue of the matrix) and FV^{-1} is the next generator matrix. This leads to the following expression

$$R_0 = \frac{\beta\eta_1}{\alpha_1 + \mu_1} + \frac{\alpha_1\beta\mu_2}{(\kappa_1 + \mu_1 + \mu_2)(\alpha_1 + \mu_1)} \tag{5}$$

The epidemiological significance of the basic reproductive ratio R_0 - which represents the average number of new cases generated by a primary infectious individual in a population is that the corona pandemic can be effectively controlled by reducing the number of high-risk individuals. This can be done either by decreasing peoples' contact with infected individuals or by using effective vaccination. Since the vaccination is not available at this point, so the only option is quarantine. The ideal case is to bring the threshold quantity R_0 to a value less than unity for the case of disease elimination.

A. Data Source

The epidemic data used in this study were mainly collected by the WHO during the current outbreak (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/). Besides this, there are many other sources where data is available, such as the Centre for disease control of China (http://www.nhc.gov.cn/xcs/yqtb/list-gzbd.shtml), European CDC (ECDC) (https://www.ecdc.europa.eu/ en/geographical-distribution-2019-ncov-cases) and Coronavirus COVID-19 Global Cases by Johns Hopkins CSSE (https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6). Here we focus on the data of mostly effected countries including China, South Korea, Iran, and Italy from the period of January 23 through March 5, 2020.

B. Parameter Values

Based on recent studies, the mean infection period for the epidemic is seven days, but in some cases, it may also be prolonged up to 14 days [8], [13]. Therefore, we used an infected progression rate α_1 to be in between 0.18-0.22 and disease recovery rate κ_1 to be 0.3. Thus the incubation period is approximately calculated as $\frac{1}{\alpha_1} + \frac{1}{\kappa_1} \approx 7.6$ *days*. Since the mechanism of disease spread is almost similar in all countries, effective contact rate β for all the simulations is kept in the range of a small interval of (0.4, 0.45). The rest of the parameter values are adjusted to obtain the best fit for the data. For the data fitting, we employ ordinary least squares (OLS) estimation outlined in [15] to estimate the parameter β by minimizing the difference between predictions of Model and the epidemic data. All the simulations are run using MATLAB, and ODE45 suite from the MATLAB routines for the model integration.

IV. Results and Discussion

The prevalence of a disease in any population can be determined by the threshold quantity R_0, as given by equation (5). Since our model is deterministic in nature, the only sources of uncertainty are the model parameters and the initial conditions. Therefore, slightly different parameter values were used to fit the four different data sets of severely infected countries such as China, Iran, South Korea, and Italy. Fig. 1(a) shows the fit for the data taken from China for the period between January 23 and February 29, 2020. The value of β we used is 0.45 and the estimation of basic reproduction number R_0 is 2.7999. The first active case in Iran was surfaced on February 20, hence the only data available was from February 20 to March 5, 2020, at this point. This data was used to fit the model shown in Fig. 1(b). The value of β used for this data was 0.42 and the estimation of R_0 was 2.7537. Likewise, the first case registered both in South Korea and Italy, was on February 18, 2020, therefore the data taken from February 18 through March 5, 2020, is used for the model fitting. The values of β used for South Korea and Italy are 0.45 and 0.4 and the estimation of basic reproduction numbers is 2.8931 and 2.7875 respectively. It is clear that at this early stage of the disease when the contact rate is closer to 0.4, the value of R_0 was in the proximity of 2.8.

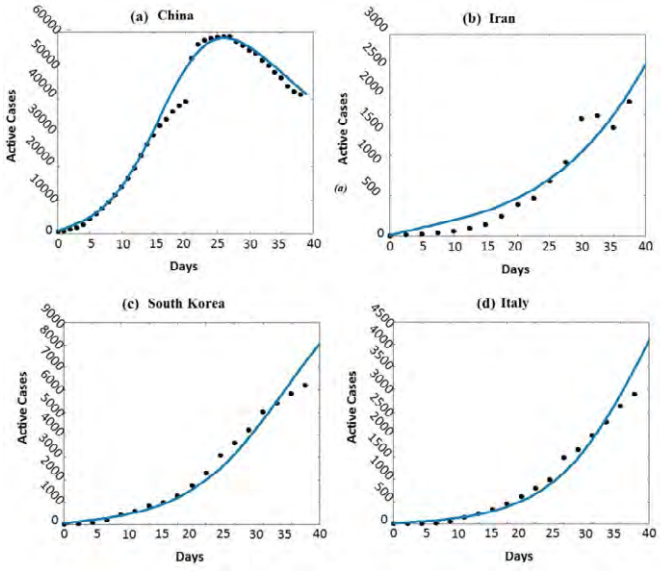


Fig. 1. Data fitting: The trajectories of the model with the real-time data for the most affected countries from COVID-19.

Next, we assessed the sensitivity of values of R_0 to the uncertainty in the parameter values. For this purpose; we identify crucial model parameters by computing partial rank correlation coefficient (PRCC) which is a measured impact of each input parameter on the output, i.e., R_0. PRCC reduces the non-linearity effects by rearranging the data in ascending order, replacing the values with their ranks and then providing the measure of monotonicity after the removal of the linear effects of each model parameter keeping all other parameters constant [14]. The horizontal lines in Fig. 2 represent the significant range of correlation, i.e., |PRCC| > 0.5, for the parameter used in our model. This analysis suggests that the most significant parameters are the contact rate β and η_1 and they are directly associated with disease progression parameter α_1. Hence, these parameters should be estimated with precision to accurately capture the dynamics of infection. Further, these values need to be adjusted accordingly to control the spread of an epidemic.

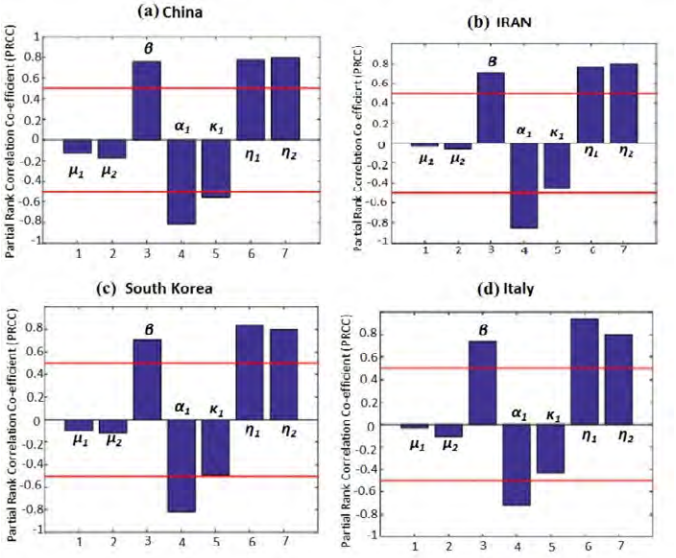


Fig. 2. Sensitivity Analysis of the basic reproduction number R_0.

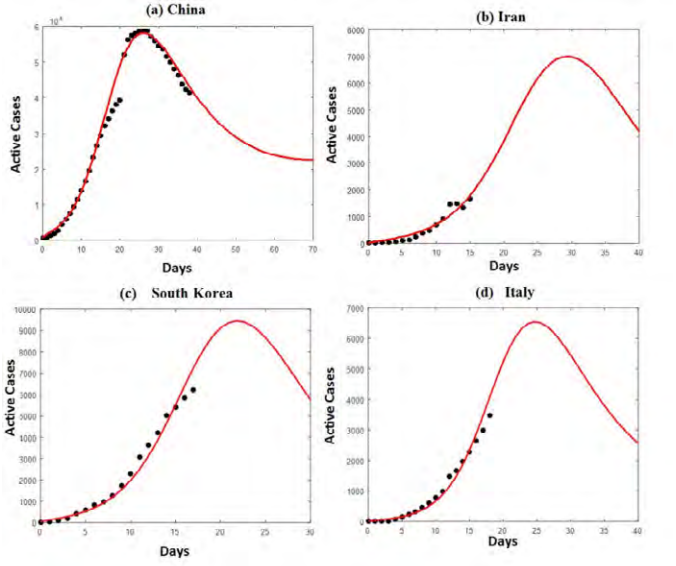


Fig. 3. Time Series: the trajectories of the model with extended time to forecast the epidemic future. The parameter values are the same as for Fig. 1 for the most affected countries from COVID-19.

Using the same parameter values discussed above and integrating the system for an extended time, we can also estimate the epidemic forecast relying on the current scenario and available data. A decline in the number of active cases is already visible in China, as depicted in Fig. 3(a), while the rest of the countries still have to see its peak. For instance, the disease will attain its peak in thirty days for Iran, twenty-two days for South Korea and twenty-five days for Italy. Thus, we can hope that by the end of March, a decline in active cases will be quite visible provided the same scenario exists and by using active measures for quarantine. The situation can even get better if some vaccination also becomes available.

V. Conclusion

Although, the estimation of R_0 for the case of COVID-19 has been discussed earlier, see for instance [3] and [4], for the first time the cases of most affected countries were discussed separately. We used their data to fit in our model for the parameter estimation and disease forecast.

From our model, we have identified the most significant parameter values contributing to the estimation of R_0. Further, these estimations can also suggest precautionary measures that need to be considered for disease control, such as effective measures of quarantine. Further, based on our model, we can also predict the possible future about the spread of epidemics in the current circumstances.

References

[1] Data source John Hopkins University. https://systems.jhu.edu/research/public-health/ncov/.

[2] Matteo Chinazzi et al:: The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. March 2020. Science, DOI: 10.1126/science.aba9757.

[3] Jonathan M Read, Jessica RE Bridgen, Derek AT Cummings, Antonia Ho, Chris P Jewell: Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. January 2020 medRxiv preprint. doi:https://doi.org/10.1101/2020.01.23.20018549.

[4] Simon James Fong, Gloria Li, Nilanjan Dey, Rubén González Crespo and Enrique Herrera-Viedma: Finding an Accurate Early Forecasting Model from Small Dataset: A Case of 2019-nCoV Novel Coronavirus Outbreak: March 2020. International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 6, No 1. DOI: 10.9781/ijimai.2020.02.002.

[5] Chen Y, Cheng J, Jiang Y, Liu K.: A Time Delay Dynamical Model for Outbreak of 2019-nCoV and the Parameter Identification. Preprint 2020; arXiv:2002.00418.

[6] Liang Y, Xu D et al.: A Simple Prediction Model for the Development Trend of 2019-nCoV Epidemics Based on Medical Observations. Preprint 2020; arXiv:2002.00426.

[7] Zhou T, Liu Q, Yang Z, et al. Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. Preprint 2020; arXiv:2001.10530.

[8] WHO 2019-nCoV situation reports. https://www.who.int.

[9] Igor Nesteruk: Statistics-Based Predictions of Coronavirus Epidemic Spreading in Mainland China. February 2020, Innov Biosyst Bioeng, vol. 4, no. 1, 13–18 DOI: 10.20535/ibb.2020.4.1.195074.

[10] Ye Liang, Dan Xu, Shang Fu, Kewa Gao, Jingjing Huan, Linyong Xu, Jia-da Li: A Simple Prediction Model for the Development Trend of 2019-nCov Epidemics Based on Medical Observations. February, 2020 Quantitative Biology , Populations and Evolution: arXiv:2002.00426v1 [q-bio.PE]

[11] Zhou, P., Yang, X., Wang, X. et al.: A pneumonia outbreak associated with a new coronavirus of probable bat origin. February 2020, Nature, https://doi.org/10.1038/s41586-020-2012-7.

[12] P. van den Driessche, and J. Watmough, Reproduction Numbers and Sub-Threshold Endemic Equilibria for Compartmental Models of Disease Transmission. 2002, Mathematical Biosciences, Vol. 180 , pp. 29{48}.

[13] Stephen A. Lauer et al.: The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. MARCH 10, 2020. Annals of Internal Medicine, DOI: 10.7326/M20-0504

[14] M. A. Sanchez, and S. M. Blower, Uncertainty and Sensitivity Analysis of the Basic Reproductive Rate. 1997, American Journal of Epidemiology, Vol. 145 , pp. 1127-1137.

[15] A. Cintron-Arias, C. Castillo-Chavez, L. M. A. Bettencourt, A. L. Lloyd, and H. T. Banks, The Estimation of the Effecctive Reproductive Number from Disease Outbreak Data. 2009, Mathematical Biosciences and Engineering, Vol. 6, pp. 261-282.

Muhammad Dure Ahmad

Dr. Dure Ahmad graduated from Arizona State University, USA. His main research is in Mathematical Biology and Data Science. Currently, he is working as a senior consultant in data science and mathematical modeling in medicine. Previously, he also served as a faculty in various universities including the University of Toronto, PSU, and the University of New Brunswick, Canada.

Mudassar Imran

Dr. Mudassar Imran graduated from Arizona State University, USA. His main research is in the subject of Dynamical systems and its application to Mathematical Biology. He is an associate professor at Gulf University for science and technology, Kuwait.

COVID-19 Detection in Chest X-ray Images using a Deep Learning Approach

Fátima A. Saiz*, Iñigo Barandiaran

Vicomtech Foundation, Basque Research and Technology Alliance (BRTA), Donostia – San Sebastián (Spain)

Received 10 April 2020 | Accepted 29 April 2020 | Published 30 April 2020

Abstract

The Corona Virus Disease (COVID-19) is an infectious disease caused by a new virus that has not been detected in humans before. The virus causes a respiratory illness like the flu with various symptoms such as cough or fever that, in severe cases, may cause pneumonia. The COVID-19 spreads so quickly between people, affecting to 1,200,000 people worldwide at the time of writing this paper (April 2020). Due to the number of contagious and deaths are continually growing day by day, the aim of this study is to develop a quick method to detect COVID-19 in chest X-ray images using deep learning techniques. For this purpose, an object detection architecture is proposed, trained and tested with a public available dataset composed with 1500 images of non-infected patients and infected with COVID-19 and pneumonia. The main goal of our method is to classify the patient status either negative or positive COVID-19 case. In our experiments using SDD300 model we achieve a 94.92% of sensibility and 92.00% of specificity in COVID-19 detection, demonstrating the usefulness application of deep learning models to classify COVID-19 in X-ray images.

Keywords

COVID-19, Deep Learning, Object Detection, X-ray.

DOI: 10.9781/ijimai.2020.04.003

I. Introduction

THE new SARS-CoV-2 coronavirus, which produces the disease known as COVID-19, kept the whole world on edge during the first months of 2020. It provoked the borders close of many countries and the confinement of millions of citizens to their homes due to infected people, which amounts to 868,000 confirmed cases worldwide at this moment (April 2020). This virus was originated in China in December 2019. From March 2020, Europe was the main focus of the virus sprout, achieving more than 445,000 infected people.

China, with a total of 3,312 deaths and more than 81,000 infected people, has managed to contain the virus almost three months after the start of the crisis in December 2019. Italy, which surpassed the Asian country in death toll on March 2020, became the most affected country, in number of deceased is followed by Spain, with more than 10,000 dead based on a report made on April 2020. This number was constantly growing. There were different studies that predicted the growth of the curves of infections, based on different parameters such as exposed, infected or recovered human's number. These studies allowed to get an idea of the transmission dynamics that could occur in each country [1] [2].

The origin of the outbreak is unknown. The first cases were detected in December 2019. The clinical characteristics of COVID-19 include respiratory symptoms, fever, cough, dyspnea, and viral pneumonia [3] [4]. The main problem of these symptoms is that there are virus-infected asymptomatic patients.

* Corresponding author.
E-mail address: fsaiz@vicomtech.org

The test to detect the COVID-19 is based on taking samples from the respiratory tract. It is carried out by a health care professional at home, generally when the case study is asymptomatic or symptoms are mild, or in a health center or hospital, if the patient is admitted for a serious condition. Carrying out as many tests as possible has shown to be the key tool to stop the virus in countries like Germany or South Korea. Spain was not able to carry out so many tests, therefore it is important to research and develop alternative methods to perform these tests in a quick and effective way.

AI and radiomics applied to X-Ray and Computed Tomography (CT) are useful tools in the detection and follow-up of the disease [5] [6]. As stated in [7], conspicuous ground grass opacity lesions in the peripheral and posterior lungs on CT images are indicative of COVID-19 pneumonia. Therefore, CT can play an important role in the diagnosis of COVID-19 as an advanced imaging evidence once findings in chest radiographs are indicative of coronavirus. AI algorithms and radiomics features derived from Chest X-rays would be of huge help to undertake massive screening programs that could take place in any country with access to X-ray equipment and aid in the diagnosis of COVID-19 [8] [9].

The current situation evokes the necessity to implement an automatic detection system as an alternative diagnosis option to prevent COVID-19 spreading among people. There are different studies that apply machine learning for this task, such as Size Aware Random Forest method (iSARF) that was proposed by [10], in which subjects were categorized into groups with different ranges of infected lesion sizes. Then a random forest-based classifier was trained with each group. Experimental results show that their proposed method yielded an accuracy of 0.879 under five-fold cross-validation, a sensitivity of 0.907 and a specificity of 0.833.

Deep learning techniques are also used in order to achieve better results than using more traditional machine learning approaches. One of the most used approach in image classification is the use of convolutional neural networks (CNNs). This type of models are used in different studies for COVID-19 detection in medical images like in [11], in their study the authors propose a CNN model trained with a randomly selection of image regions of interest (ROIs), achieving a 85.2% of accuracy, 0.83 of specificity and 0.67 of sensitivity. Other example of the results that can be obtained using CNNs is presented by [12]. They propose the COVID-Net CNN network obtaining 92.4% of accuracy, 80% of sensibility and 88.9% of specificity.

A method called COVIDX-Net is presented by [13], COVIDX-Net includes seven different architectures of deep convolutional neural network models, such as a modified version of Visual Geometry Group Network (VGG19) and the second version of Google MobileNet. Each deep neural network model is able to analyze the normalized intensities of the X-ray image to classify the patient status either negative or positive COVID-19 case. Their experiments evaluation achieves f1-scores of 0.89 and 0.91 for healthy and COVID-19 detection respectively.

The results shown in mentioned works demonstrate that deep learning techniques are useful for the virus detection, and that improve the obtained metrics using more traditional machine learning approaches [11] [12] [13].

The main contribution of our paper is focused on the improvement of the detection accuracy of COVID-19, by proposing a new dataset that combines COVID-19 and pneumonia images to make more stable predictions and by applying image processing that allows image-standardization and also improves model learning.

II. Method

We propose to use a deep convolutional neural network specialized for object detection along with a new dataset composed of COVID-19 and pneumonia images. Both are publicly available on GitHub [14] and Kaggle [15] respectively. The chest X-ray or CT images that are available in GitHub belong to COVID-19 cases. It was created by assembling medical images from public available websites and publications. This dataset contains 204 COVID-19 X-ray images. On the other hand, the Kaggle dataset was created for a pneumonia detection challenge. The images have bounding boxes around diseased areas of the lung. Samples without bounding boxes are negative and contain no definitive evidence of pneumonia. Samples with bounding boxes indicate evidence of pneumonia.

We propose a new dataset by merging COVID-19 and pneumonia images to obtain a wider and diverse one. The fact of having pneumonia images in the training dataset supposes an extra advantage, due to normal pneumonia and COVID-19 have similar appearance in chest X-ray images. This dataset merge allows to get a robuster model that is able to better distinguish between those diseases. Another advantage of this merge is the fact of enlarging the train dataset, because the COVID-19 images are not abundant at the time of writing this paper. This merge does not enlarges COVID-19 image set but improves detection quality because of the similarity between pneumonia and COVID-19. Train with pneumonia images gives an extra knowledge to the model in order to not confuse COVID-19 with pneumonia, being more effective and stable in disease detection.

We split the images in train and test sets, dividing all the data in a balanced way, meaning that all samples of each class in the training sets are well-balanced, in order to avoid biased results. For this purpose, even though we have a large number of pneumonia and normal images, we compose a dataset of 1500 images.

We compose the dataset as follows: we select 104 COVID-19 images, 205 health lung images and 204 pneumonia images for training, and 100 COVID-19 images, 444 health lung images and 443 pneumonia images for testing. In the Kaggle dataset there are more samples of pneumonia and normal images, but we select only 205 for training with the purpose of having a balanced dataset. In the test stage, we add more pneumonia and normal images to demonstrate the robustness of the model giving no false positives in the detection of COVID-19. Summing up, we use a train set composed of 513 images and a test set of 887 images in total.

In the next step, training images are labeled using a XML annotation file based on Pascal VOC format [16]. Each sample in the dataset is an image with ground truth bounding boxes for each object in the images as shown in the Fig. 1.

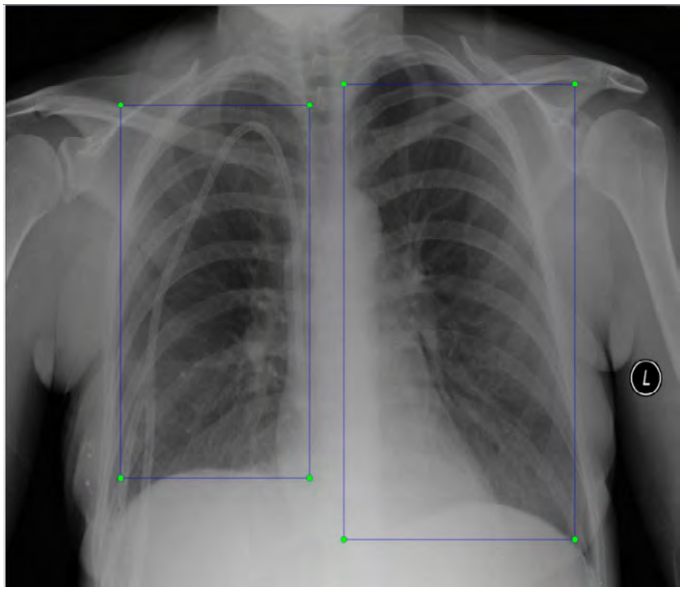


Fig. 1. Created bounding boxes of normal regions in a chest X-Ray image.

III. Model Architecture

The selection of the used architecture is based on the good results obtained with CNNs in the state-of-the-art works for COVID-19 image classification, and the good results obtained in other similar tasks with this kind of architecture [11] [12] [13]. We used the same network architecture as proposed in [17], based on Single Shot Multibox Detector (SSD). This architecture is optimized for detecting objects in images using a single deep neural network. This approach discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes.

Experimental results on different remarkable datasets confirm that SSD has comparable accuracy to methods that utilize more than one architecture for detecting objects being much faster, while providing a unified framework for both training and inference. Compared to other single stage methods, SSD has much better accuracy, even with a smaller input image size [17].

We use VGG-16 [18] as the base network for performing feature extraction in this architecture. This model is also based on Fast R-CNN. During training, we have multiple boxes with different sizes

and different aspect ratios across the whole image. SSD finds the box that has more Intersection-Over-Union (IoU) compared with the ground truth. A detailed description of the layers architecture of SDD network is shown in [17].

Our main goal is to obtain a more robust model to various input object sizes and shapes. Therefore, during the SSD training a data augmentation step is performed. This process is composed by the following operations applied to every image in the dataset:

- Use the entire original input image.
- Sample a patch so that the minimum overlap with the objects is 0.1, 0.3, 0.5, 0.7, or 0.9. The size of each sampled patch is a percentage between [0.1, 1] of the original image size.
- Randomly sample a patch.

During the inference, SDD uses 8732 boxes for a better coverage of object location. After the inference step, a set of boxes representing the detected objects are given along with the respective label and score, as shown in Fig. 2.

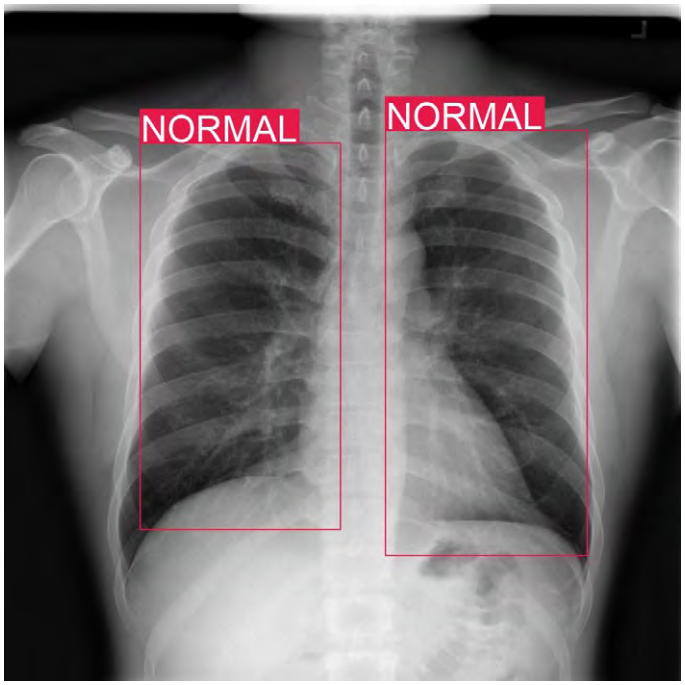


Fig. 2. Detection boxes of SSD model inference output detecting normal lungs.

One of the advantages of this model is the possibility of precise object localization, which is not the case in previous works [11] [12] [13].

IV. Experiments and Results

The first experiment made in this work, is the contrast adjustment of each image in the dataset. This adjustment is necessary because the exposure time in X-Ray images can be different between acquisitions. All the images of the dataset are from different hospitals around the world, so the image acquisition settings and conditions are different in each place. In X-Ray images, an adjustment in the voltage spike results in a change in the contrast of the radiography. Exposure time, which refers to the time interval during which x-rays are produced, is also a factor that affects the contrast of the obtained image [19].

In order to get image similarity between the dataset, Contrast Limited Adaptive Histogram Equalization (CLAHE) [20] is applied. This is a transformation that aims to obtain a histogram with an even distribution for an image. That is, there is the same number of pixels for each level of gray in the histogram of a monochrome image. As cited in [20], in X-ray imaging, when continuous exposure is used to obtain an image sequence or video, usually low-level exposure is administered until the region of interest is identified, so reducing the radiation applied to the patients. As a drawback, images with low signal-to-noise ratio are obtained. In this case, and many other similar situations, it is desirable to improve image quality by using some type of image enhancement such as histogram equalization algorithms. An example of the application of this image operation is shown in Fig. 3.

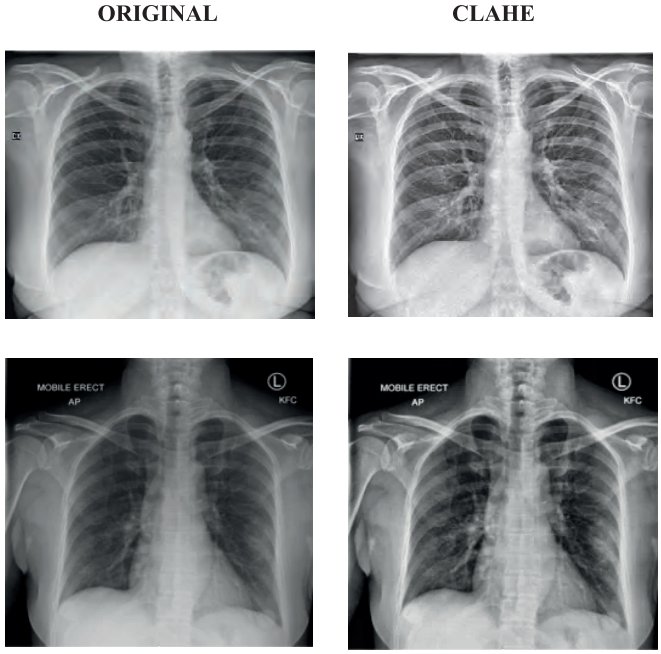


Fig. 3. Comparison between original image and CLAHE applied images.

The differences obtained in the detection applying or not CLAHE in the train and test datasets are shown in TABLE I. As the table shows, the fact of applying this pre-processing increases notably the detection accuracy in health and infected lungs.

We load VGG-16 weights trained on ImageNet. There are many works that evaluate the accuracy improvement using transfer learning, specially in small datasets [21] [22]. We apply these weights because though lower layers learn features that are not necessarily specific to this dataset, this action improves the detection accuracy and the sensibility and specificity metrics. The key idea is to take advantage of a trained model for similar images and adapt a base-learner to a new task for which only a few labeled samples are available.

The last experiment evaluates the detection accuracy obtained in the detection of COVID-19. For this purpose, we set apart two sets of images, one with non-COVID-19 and the other one with COVID-19 positives. We obtain different metrics that can be seen in TABLE II. Another metric that we take in care is the inference time that we obtain running the model on a GPU, achieving 0,13s per image.

TABLE I. Obtained Results in Image Classification Applying or Not CLAHE in the Dataset.

Image class	CLAHE	Total images	True detection	Accuracy
Normal	No	887	827	93.24%
Normal	Yes	887	842	**94.92%**
COVID-19	No	100	83	83.00%
COVID-19	Yes	100	92	**92.00%**

TABLE II. Obtained Metrics Values

Metric	Operation	Value
Sensibility	842 / (842+45)	0.9492
Specificity	92 / (92+8)	0.9200

V. Conclusion

This study demonstrates the useful application to detect COVID-19 in chest X-ray images based on image pre-processing and the proposed object detection model.

The proposed merged dataset using pneumonia images allows getting a more robust model that is able to distinguish between COVID-19 and pneumonia diseases. With the histogram equalization operation, we can get a normalized dataset that helps to model training step. It also improves the normal image detection and minimizes the false positives rate.

With our proposed method, we achieve a 94.92% of sensibility and 92.00% of specificity in COVID-19 detection. The detection accuracy obtained using this architecture and the proposed dataset improves the results described in [11] [12] [13]. These results demonstrate that object detection models trained with more images of similar diseases and applying transfer learning, combined with CLAHE algorithm for image normalization, could be successful in medical decision-making processes related with COVID-19 virus diagnosis.

References

[1] M. Dur-e-Ahmad and M. Imran, "Transmission Dynamics Model of Coronavirus COVID-19 for the Outbreak in Most Affected Countries of the World," *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. In Press, no. In Press, pp. 1-4, 2020.

[2] S. J. Fong, N. D. G. Li, R. Gonzalez-Crespo and E. Herrera-Viedma, "Finding an Accurate Early Forecasting Model from Small Dataset: A Case of 2019-nCoV Novel Coronavirus Outbreak," *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 6, no. 1, pp. 132-140, 2020.

[3] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. M. Leung, E. H. Y. Lau, J. Y. Wong and others, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *New England Journal of Medicine,* 2020.

[4] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang and Z. Peng, "Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China," *JAMA,* vol. 323, pp. 1061-1069, 3 2020.

[5] S. Chauvie, A. De Maggi, I. Baralis, F. Dalmasso, P. Berchialla, R. Priotto, P. Violino, F. Mazza, G. Melloni and M. Grosso, "Artificial intelligence and radiomics enhance the positive predictive value of digital chest tomosynthesis for lung cancer detection within SOS clinical trial," *European Radiology,* p. 1–7, 2020.

[6] G. Chassagnon, M. Vakalopoulou, N. Paragios and M.-P. Revel, "Artificial intelligence applications for thoracic imaging," *European Journal of Radiology,* vol. 123, p. 108774, 2020.

[7] F. Song, N. Shi, F. Shan, Z. Zhang, J. Shen, H. Lu, Y. Ling, Y. Jiang and Y. Shi, "Emerging 2019 Novel Coronavirus (2019-nCoV) Pneumonia," *Radiology,* vol. 295, pp. 210-217, 2020.

[8] J. C. L. Rodrigues, S. S. Hare, A. Edey, A. Devaraj, J. Jacob, A. Johnstone, R. McStay, A. Nair and G. Robinson, "An update on COVID-19 for the radiologist-A British society of Thoracic Imaging statement," *Clinical Radiology,* 2020.

[9] J. Wu, J. Liu, X. Zhao, C. Liu, W. Wang, D. Wang, W. Xu, C. Zhang, J. Yu, B. Jiang and others, "Clinical characteristics of imported cases of COVID-19 in Jiangsu province a multicenter descriptive study," *Clinical Infectious Diseases,* 2020.

[10] F. Shi, L. Xia, F. Shan, D. Wu, Y. Wei, H. Yuan, H. Jiang, Y. Gao, H. Sui and D. Shen, *Large-Scale Screening of COVID-19 from Community Acquired Pneumonia using Infection Size-Aware Classification, arXiv preprint arXiv:2003.09860,* 2020.

[11] S. Wang, J. M. Bo Kang, X. Zeng and M. Xiao, "A deep learning algorithm using CT images to screen for Corona Virus Disease COVID-19)," *medRxiv,* 2020.

[12] L. Wang and A. Wong *COVID-Net A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest Radiography Images, arXiv preprint arXiv:2003.09871* ,2020.

[13] E. E.-D. Hemdan, M. A. Shouman and M. E. Karar, *COVIDX-Net A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images, arXiv preprint arXiv: arXiv:2003.11055,* 2020.

[14] J. P. Cohen, P. Morrison and L. Dao, *COVID-19 Image Data Collection, arXiv preprint arXiv: arXiv, arXiv:2003.11597,* 2020.

[15] *RSNA Pneumonia Detection Challenge. Kaggle. [online]* Available at: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge, Accessed 29 April 2020.

[16] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision,* vol. 88, p. 303–338, 2010.

[17] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu and A. C. Berg, "SSD Single Shot MultiBox Detector," *Lecture Notes in Computer Science,* p. 21–37, 2016.

[18] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *CoRR,* vol. abs1409.1556, 2014.

[19] H. Jansen, Radiología dental. Principios y técnicas., Mc Graw Hill, 2002.

[20] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement," *Journal of VLSI signal processing systems for signal, image and video technology,* vol. 38, p. 35–44, 2004.

[21] H.-W. Ng, V. D. Nguyen, V. Vonikakis and S. Winkler, "Deep learning for emotion recognition on small datasets using transfer learning," in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, 2015.

[22] Q. Sun, Y. Liu, T.-S. Chua and B. Schiele, "Meta-transfer learning for few-shot learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

Fátima A. Saiz

Fátima Aurora Saiz Álvaro studied Engineering in Geomatics and Topography the University of the Basque Country during 2012-2017, being the first of its promotion. She obtained the title with her final degree project "Applications of programming languages to artificial intelligence, complex geophysical calculations and augmented reality" valued with honors, joining in the Talentia program of the Bizkaia Provincial Council. She completed the Master in Visual Analytics and Big Data of the Faculty of Engineering of the International University of La Rioja during 2017-2018. She obtained the title with her work "Study of architectures for the extraction and exploitation of superficial defects data using Deep Learning techniques". Since October 2017 she has been part of the Vicomtech staff as a researcher in the Department of Industry and Advanced Manufacturing, working in the field of cognitive vision and Deep Learning.

Íñigo Barandiaran

Iñigo Barandiaran received his PhD in Computer Science from the University of the Basque Country in the fields of computer vision and pattern recognition. He worked as a researcher in the group of "Artificial intelligence and computer science" at the University of Basque Country. Since 2003 he is a researcher at Vicomtech, participating and leading various R&D national and international projects, in various sectors such as biomedicine and industry. He has several scientific publications in conferences and journals in areas such as image analysis, or pattern recognition. He is currently the director of the Industry and Advanced Manufacturing department at Vicomtech.

An Extreme Learning Machine-Relevance Feedback Framework for Enhancing the Accuracy of a Hybrid Image Retrieval System

B. Shikha[1]*, P. Gitanjali[2], D. Pawan Kumar[2]

[1] Department of ECE, DCRUST, Murthal & UIET, Kurukshetra University, Haryana (India)
[2] Department of ECE, DCRUST, Murthal, Sonepat, Haryana (India)

Received 20 March 2019 | Accepted 2 October 2019 | Published 20 January 2020

Abstract

The process of searching, indexing and retrieving images from a massive database is a challenging task and the solution to these problems is an efficient image retrieval system. In this paper, a unique hybrid Content-based image retrieval system is proposed where different attributes of an image like texture, color and shape are extracted by using Gray level co-occurrence matrix (GLCM), color moment and various region props procedure respectively. A hybrid feature matrix or vector (HFV) is formed by an integration of feature vectors belonging to three individual visual attributes. This HFV is given as an input to an Extreme learning machine (ELM) classifier which is based on a solitary hidden layer of neurons and also is a type of feed-forward neural system. ELM performs efficient class prediction of the query image based on the pre-trained data. Lastly, to capture the high level human semantic information, Relevance feedback (RF) is utilized to retrain or reformulate the training of ELM. The advantage of the proposed system is that a combination of an ELM-RF framework leads to an evolution of a modified learning and intelligent classification system. To measure the efficiency of the proposed system, various parameters like Precision, Recall and Accuracy are evaluated. Average precision of 93.05%, 81.03%, 75.8% and 90.14% is obtained respectively on Corel-1K, Corel-5K, Corel-10K and GHIM-10 benchmark datasets. The experimental analysis portrays that the implemented technique outmatches many state-of-the-art related approaches depicting varied hybrid CBIR system.

Keywords

Color Moment, Extreme Learning Machine, Gray Level Co-occurrence Matrix, Relevance Feedback, Region Props Procedure.

DOI: 10.9781/ijimai.2020.01.002

I. Introduction

With the tremendous advancements in the domain of smart-phones and various image capturing devices, there has been a great revolution in the field of image processing. Nowadays, various social media platforms are more-and-more utilized for sharing these captured images. Hence, it has led to the creation of many massive image repositories [1]. The credit for an enhancement in digital images also owes to different advanced satellite and industrial based cameras, usage of the web, divergent portable devices which are used for image capturing and storing. In these gigantic databases or repositories, the tasks of exploring, browsing, indexing and retrieving are too strenuous. Manual searching and retrieval of these images from massive databases is very time consuming and is also prone to human delusions [2]. Therefore, for the trouble-free management and retrieval of digital images from these huge databases, an efficient system is required and the key to this problem is Content-based Image retrieval (CBIR). This system performs an effective image retrieval based on various image attributes like; texture, shape, color, spatial information, etc.

Generally, an image retrieval process is divided into two forms: Text-based and Content-based. Images are retrieved on the basis of text annotations in a text-based retrieval system. But, it suffers from many disadvantages like human annotation errors and moreover the usage of synonyms, homonyms, etc. lead to an inaccurate image retrieval [3].

In order to overcome these limitations, the best solution is CBIR systems. In these systems, feature vectors are created for each type of extracted feature which in turn represents the different attributes of an image. Similar feature vectors are created for the complete database also. Lastly, similarity matching is done based on the obtained feature matrix of the given input image and feature vectors obtained from the total images in the repository and afterward final results are achieved. Therefore, for the evolvement of a highly effective CBIR system, a low dimensional feature vector is one of the prime requirements.

A. Motivation

In literature, numerous retrieval systems, particularly for images, have been developed which are focused entirely on the single feature of an image like color, texture and shape. But, these systems are incapable of describing the images with complex appearance. Therefore, for the representation of complex images, an ideal combination of features is required which are finally transformed into a precise and single feature vector. This single feature vector contains additional comprehensive information related to an image and works magnificently as compared to the feature vector based on a solitary technique. Therefore, to

* Corresponding author.
E-mail address: shikpank@yahoo.com

represent the complicated images, a perfect amalgamation of basic image features (color, texture and shape) is required. Classification of massive datasets is also a desired and imperative process in order to obtain precise and accurate results. Different machine learning based classifiers have been used significantly for this purpose. But, they lack the features of a neural-network-based system. In the present era, the use of Deep-learning has a significant contribution in the image processing domain. The working of these deep structures depends primarily on the total number of hidden layers as well as the number of neurons in those layers. So, the drawbacks of these customary classifiers can be removed by using a classifier based on a neural network. This feed-forward single hidden layer classifier encapsulates the capabilities of both Deep learning by using hidden layer neurons and of machine learning by performing a more accurate classification task. Finally, to remove the gap between the basic image features and lofty human perception, particularly called as Semantic gap, Intelligent techniques are required. So, in this implementation, our main objective is the formation of a hybrid CBIR system by using best techniques for color, texture and shape retrieval. Then, the formation of a feature vector based on these combined features and by using a feed-forward single hidden layer neural network for the purpose of an efficient classification. Lastly, utilization of an intelligent technique which captures high-level semantics of an image.

B. Related State-of-the-Art Work

In the past decade, many techniques have been used in CBIR systems to extract the features of an image. CBIR is an efficient system which utilizes features of an image and thereby retrieves the required results by using these extracted features. Extraction of features is indeed the foremost phase in a majority of CBIR systems. General features like color, shape, texture, spatial information etc. can be extracted during this extraction process. However, the extraction of multiple or hybrid features is the demand of the current era. Among these hybrid features, color is the most distinctive cue that humans can visualize very promptly. Texture describes the discernible patterns of an image. Among the prominent features of an image, the shape is also considered as an important feature that describes the characteristics of an image. Fadaei et al. [2] propose a hybrid retrieval system which utilizes only color and texture attributes. For the extraction of color features, Dominant color descriptor (DCD) has been used. In order to extract texture features, wavelet and Curvelet features have been utilized. Finally, Particle swarm optimization (PSO) has been used which selects the optimum features of these techniques for an ideal combination.

Fusion of color and shape features have also been utilized in research. Color coherence vector (CCV) has been used as a color extraction technique. Various shape parameters have been used to define and extract the connected parts of the region or boundary of an image [3]. Numerous hybrid systems have been deployed in literature like that of Pavitra et al. [4] who develop a fusion based technique in which Color moment is utilized for the extraction of color features as well as a filter to select some specified images based on the value of a moment. Further, Local binary pattern (LBP) and Canny edge detector have been used to denote texture and edge information of an image.

Another hybrid system which utilizes the memetic algorithm has also been developed. In this system, color is extracted using a basic RGB color space model. For the shape analysis, the median filter has been used and finally, Gray level co-occurrence matrix (GLCM) has been employed to extract features related to texture [5]. For similarity calculation, memetic algorithm, which is the combination of Genetic and Great deluge algorithms, has been used. One more fusion system in which Statistical moments and 2D histograms have been used for color extraction and texture analysis has been carried out by using GLCM [6].

The basic necessity of any CBIR system is the brilliant selection of image parameters needed for extraction of features [7]. Color and edge directivity descriptor (CEDD) is utilized for the extraction of both color as well as texture features while a 2nd level of Discrete wavelet transform (DWT) is employed for the analysis of shape features. Lastly, in order to classify the images, Support vector machine (SVM) classifier has been used. A CBIR system developed in different levels of the hierarchy has been developed by Pradhan et al. [1]. Here, adaptive tetrolet transform is used to extract the textual information. Edge joint histogram and color channel correlation histogram have been used respectively to analyze shape and color features related to an image. This system is realized in the form of a three-level hierarchical system where the highest feature among the three is depicted at every level of the hierarchy.

There are primarily two main domains in which features of an image can be classified. One is the frequency domain and another is the spatial domain. A CBIR system associated with features pertaining to these two domains has been developed by Mistry et al. [8]. In this system, various spatial domain techniques like color auto-correlogram, HSV histogram, and color moments have been used. Under the frequency domain, methods like Gabor wavelet transform and SWT moments have been employed. Also, binarized statistical image features in combination with color and edge directivity descriptor have been used for further enhancing the performance of the system.

Indexing is considered as a prominent technique in order to lessen the memory requirements and to save the execution time. A hybrid CBIR system based on the technique of indexing is developed by Guo et al. [9]. This system utilizes Ordered-Dither Block Truncation Coding (ODBTC) in order to index color images. Color distribution and contrast of an image is represented by Color co-occurrence feature (CCF) and information regarding edges is given by Bit pattern feature (BPF).

Prashant et al. [10] develop a hybrid retrieval system which uses Local binary pattern (LBP) as a texture extraction technique. Here, the LBP calculation is based on the different scales which captures more prominent features of an image as compared to single-scale LBP. Lastly, Gray level co-occurrence matrix (GLCM) has been used efficiently for the computation of feature vectors. Chandan Singh et al. [11] describe a CBIR system where the Color histogram is utilized as a color descriptor. A Color histogram is a graphical characterization of pixels in an image. In addition to Color histogram, Block variation of local correlation coefficients (BVLC) and Block difference of Inverse Probabilities (BDIP) are adopted for texture extraction.

Dominant color descriptor (DCD) is considered as a vital color descriptor used in CBIR where course partitions are created by the division of a space utilized for color analysis. Each partition has a partition center and its percentage. Integration of DCD, Gray level co-occurrence matrix (GLCM) and Fourier descriptors is deployed for the constitution of a hybrid CBIR system [12]. Relevance feedback is also considered as an important intelligent technique which retrieves relevant images of interest based on the feedback obtained by the user. A hybrid system based on a single iteration of relevance feedback with the utilization of Non-dominated Sorting Genetic Algorithm with an exploitation algorithm has been designed by Miguel et al. [13].

Relevance feedback can also be used in combination with many meta-heuristics. Sequential forward selector (SFS) meta-heuristic is utilized in combination with relevance feedback using a single iteration. Many distance metrics have also been tested and analyzed here [14].

A color image retrieval system has been developed in two levels by using Color moment, Edge histogram descriptor (EHD) and Angular radial transform (ART) specifically for the extraction of color, texture and shape attributes of an image respectively. Color moment has been

used at the first level of the retrieval process and then followed by texture and shape descriptors in the second level [15]. An image retrieval system which utilizes only color and shape features by using RGB color model space and Edge directions has also been described [16].

Color co-occurrence matrix (CCM) computes the probability of occurrence of a pixel amongst a specific pixel and its neighbors which in turn is depicted as the extracted feature of an image. Difference between pixels of scan pattern (DBPSP) is based on the difference calculation between two pixels and its conversion into probability. These techniques, CCM and DBPSP, have been used for the extraction and analysis of color and texture attributes and finally, a hybrid system is formed [17]. Multiple features can be combined efficiently for the evolution of an effective image retrieval system. Pandey et al. [18] describe a retrieval system where Bi-cubic interpolation (BCI) is deployed for an initial pre-processing of an image. For the extraction of color features, color coding (CC) is used. Gray level difference method (GLDM) and Discrete wavelet transform (DWT) have been employed for texture feature extraction. Finally, Hu moments have been analyzed for shape feature extraction.

For the analysis and extraction of texture features Discrete cosine transform (DCT) [19] has also been used. Discrete wavelet transform has also been added to enhance the efficiency of the system. Then, the difference between these two techniques is computed and re-ranking of images is done. Lastly, for the extraction of color attributes from an image, Color coherence vector (CCV) has been employed. CBIR is also known by the name of Query by image content (QBIC) because the user's query is given in the form of an image. A different type of hybrid retrieval system is proposed in which a combination of shape and texture features is used. For shape extraction, Fourier descriptors have been used and Radial Chebyshev moments [20] have been used to analyze texture features. K-means Clustering has also been used to augment the classification accuracy of the system.

The extraction of features from an image can be done globally and locally. Global feature extraction is based on the whole image while local extraction is devised to be used for a specified region of an image. Based on this global and local feature extraction, a CBIR system is proposed specifically for shape feature extraction. In this system, Angular radial transform (ART) is used for global feature extraction while Histograms of spatially distributed points (HSDP) [21] is utilized for local content extraction.

Apart from these hybrid systems which are based on a certain feature extraction technique, Machine learning has also contributed fabulously in the domain of image retrieval. Many machine learning based classification algorithms have been successfully utilized in this area. Support vector machines (SVM), Naive Bayes, Random forest etc. are some of the common classifiers categorized under machine learning. A hybrid CBIR system described by Pradnya et al. [22] consists of color correlogram for color extraction and the combination of Gabor and Edge histogram descriptor (EHD) for texture feature extraction. SVM has also been used to obtain a precise classification accuracy. Segmentation based Fractal Texture Analysis (SFTA) can also be employed as a texture analysis technique [23]. Again, for classification purpose SVM classifier has been used.

In the current era, the focus of the researcher community has been shifted from machine learning to Deep learning. Many deep learning techniques have been used for the extraction of image features. Arun et al. [24] describes a Hybrid deep learning architecture (HDLA) which is capable of generating sparse representations used for reducing the semantic gap. This model uses Boltzmann machines in the upper layers and Softmax model in the lower levels. Another deep learning technique which is specified as deep belief network (DBN) [25] has also been utilized for feature extraction and classification in hybrid CBIR system.

Extreme learning machine (ELM) can be considered as a type of Deep learning network because it uses a neural network as its hidden layer. This network is a type of feed-forward neural system and has a solitary hidden layer and has excelled results as compared to other machine learning based classifiers [26]. Kaya et al. [27] describes a texture based hybrid retrieval system which utilizes two texture extraction techniques. For classification accuracy of butterfly images, ELM classifier has also been deployed. In the modern era, Convolutional neural networks (CNN) have also been utilized for the various functions related to image analysis [28].

In literature, the discussed hybrid systems lack in performance due to the utilization of either one or two visual attributes of an image. Due to this, the left-out and un-analyzed attribute cannot contribute to the formation of a final feature vector which is the desired condition for the retrieval of complex images. Moreover, the described systems in literature do not provide information concerning frequency relating to co-occurring of local patterns of an image. But, the proposed system is an intelligent fusion descriptor which contributes to producing a highly effective and accurate system by including spatial information and prominent interconnection among pixels of an image. The information concerning the shape attribute of an image is also included by analyzing some of its prominent parameters. The proposed system also includes human semantic information by using an intelligent technique of Relevance feedback.

C. Main Contributions

A predominant requirement of an effectual CBIR system is that the system should evaluate all the three visual media attributes i.e texture, shape and color in order to form an accurate retrieval system. CBIR systems based on a single feature extraction lack the description of complex images. Moreover, the usage of traditional machine learning based classifiers is deficient of desired retrieval accuracy. Lack of semantic information also reduces the performance of the system. To address these issues, an effective and novel CBIR is proposed where color is extracted using color moment, texture analysis is done using Gray level co-occurrence matrix (GLCM) and different region props are used for the extraction of shape attributes. To get an exact classification accuracy, an Extreme learning machine (ELM) classifier is used. Finally, to capture the high-level semantics of an image, Relevance feedback is used following various rounds to reformulate the training of an ELM classifier.

The remaining organization of this paper is given in the following way: Before the description of the implemented technique and various utilized techniques have been discussed in section II, designated as Preliminary section. Proposed work has been given in section III. Experimental simulation and analysis have been given in section IV and finally, in section V, the conclusion with future trends has been given.

II. Preliminaries

Various techniques used in the creation of the proposed system are described in this section.

A. Color Moment

In image retrieval systems, many techniques have been used for the extraction of color features. Among these, color histogram [29] is the conventional method of color feature extraction. Though it is very simple and is invariant to scale and angle rotation but it does not convey any spatial information regarding an image. Researchers have also used Color coherence vector (CCV) as a color feature descriptor but the feature vector produced by this method is of high dimensionality which is against the basic requirements of any CBIR system. Dominant color descriptor (DCD) also suffers from the lack of complete spatial

information and moreover, if the obtained feature vector is compact, then it is utilized feasibly, otherwise vague results can be produced.

Color auto-correlogram (CAC) has a high computation time and cost and it is very sensitive to noise also. Therefore, based on these facts and conclusions, the color moment has been chosen as an effective color feature extraction technique. It is robust, fast, scalable, consumes less time and space. Color moments are metrics which signify color distributions in an image. According to probability theory, its distribution can be effectively characterized by its moments. Color moment can be computed for any color space model. Different moments specify diverse analytical and statistical measures. This color descriptor is also scale and rotation invariant but it includes the spatial information from images [30].

If I_{ij} specifies the i_{th} color channel and j_{th} image pixel, the number of pixels in an image are N, then index entries associated with the particular color channel and region r is given by first color moment, which signifies average color in an image denoted as:

$$Mean(E_{r,i}) = \frac{1}{N}\sum_{j=1}^{N} I_{ij} \quad (1)$$

The next moment is given by Standard deviation which is obtained from a color distribution of a particular image and is defined as the square root of the variance. It is given as:

$$Standard\ deviation(\sigma_{r,i}) = \left(\frac{1}{N}\sum_{j=1}^{N}(I_{ij} - E_{r,i})^2\right)^{\frac{1}{2}} \quad (2)$$

The third moment is called as skewness and signifies how asymmetric is the color distribution. It is given as:

$$Skewness\ (S_{r,i}) = \left(\frac{1}{N}\sum_{j=1}^{N}(I_{ij} - E_{r,i})^3\right)^{\frac{1}{3}} \quad (3)$$

Fourth moment is denoted by Kurtosis and it signifies the color distribution shape, emphasizing particularly on tall or flat shape of the color distribution.

B. Gray Level Co-occurrence Matrix

In order to withdraw features related to textural patterns, various second-order statistical methods have been utilized. Among these, Gray level Difference matrix (GLDM) is error free in the calculation and is based on the difference between gray level pixel pairs. But, its main drawback is that with the change in gray level variance, it also becomes variant. On the other hand, Gray level run length matrix (GLRLM) is based on counting the number of runs which in turn depicts gray levels in an image. This technique suffers from meagerness to represent the pattern of an image and also its computational involvement is high. For analyzing a signal in time-frequency zone, initially Fourier based transforms like Discrete cosine transform (DCT) and Discrete fourier transform (DFT) were in trend. But, these traditional methods are deficient, as they do not convey the local information of an image. Also, the images regenerated by these techniques are of deprived standard, especially at the edges because of the presence of high-frequency bristly components.

In the wavelet domain, both Gabor wavelet and Discrete Wavelet Transform are highly prominent. But, due to a large dimension of feature vectors, Gabor wavelet takes more time in image analysis. Discrete wavelet transform (DWT) also suffers from many disadvantages like ringing near discontinuities, variance with shift, the lack of directionality of decomposition functions, etc. Local binary pattern (LBP) is also considered as an important tool to extract texture information from an image. The main issues associated with this technique are the production of very long histograms which slows down the process of image recognition, sensitivity to noise and the effect of the center pixel is sometimes not included. Based on these conclusions, the best second-order statistical texture feature extraction technique is Gray level co-occurrence matrix (GLCM) which has many dominant attributes like: (1) Rotation invariance (2) Diverse applications (3) Simple implementation and fast performance (4) Numerous resultant (Haralick) parameters (5) Precise inter-pixel relationship.

It is an analytical technique used to withdraw texture features for the retrieval and classification of different images. The spatial correlation enclosed by the pixel duplet in any given image is also computed by this geometrical method. A co-occurrence matrix denoted by $P_{dis}\ (m,n)$ specifying grey levels, contains information about two pixels: Gray level content m is denoted by the first pixel and content n is denoted precisely by the second pixel which further is separated by a distance denoted by dis. These specifications are chosen according to a specific angle. Matrices produced by this technique give the gray level spatial frequencies which further gives association among pixels that are adjacent and have distinct distances amidst them [31].

Therefore, the description of GLCM is as follows:

$$P_{dis,\Theta}(m,n) = P_r(I(p_1) = m \wedge I\ (p_2) = n \wedge \|p_1 - p_2\| = d_{is} \quad (4)$$

In equation (4) I is the given image and the positional information in the same image I is given by p_1 and p_2, the probability is denoted by P and Θ denotes the range of different angle directions given by 0º, 45º, 90º and 135º. Therefore, a GLCM image is represented by d as its vector used for movement, δ as its radius and θ as orientation. A generalized GLCM matrix can be represented by Fig. 1.

Gray Tone	0	1	2	3
0	# (0,0)	# (0,1)	# (0,2)	# (0,3)
1	# (1,0)	# (1,1)	# (1,2)	# (1,3)
2	# (2,0)	# (2,1)	# (2,2)	# (2,3)
3	# (3,0)	# (3,1)	# (3,2)	# (3,3)

Fig. 1. A Generalized GLCM.

Therefore, for a given specific test image with gray tone values, a GLCM matrix is formed, which is the spatial co-occurrence dependence matrix.

Prominent four types of GLCM feature parameters which are subjected to be used for the extraction [32] of the textual content of an image and are denoted by:

$$Contrast = \sum_m \sum_n [P(m,n) * (m-n)]^2 \quad (5)$$

The disparity between the topmost and the bottom most conterminous pixel sets is given by Contrast intensity.

$$Correlation = \frac{\sum_m \sum_n (m-u_x) * (n-u_y) * P(m,n)}{\sigma_i \sigma_j} \quad (6)$$

The correspondence between a reference pixel and its adjoining pixels in an image is diagnosed by making use of Correlation. It considers the mean and standard deviation of a matrix by encapsulating both the row and column of that particular matrix.

$$Homogeneity = \sum_m \sum_n \frac{P(m,n)}{1+|m-n|} \quad (7)$$

In the spatial domain, the proximity among gray levels in an image is defined by the term homogeneity.

$$Energy = \sqrt{\sum_m \sum_n P(m,n)^2} \quad (8)$$

Energy of a texture denotes the cyclic consistency of gray level allocation in an image.

C. Region-props Process

Shape is also an important visual attribute which can be used to depict the information related to an image. Shape features of an object provide valuable information about the identity of the same object. These features can be categorized into two types: based on its region and based on its boundary [33]. Information about an object's internal regions is depicted by region-based descriptors while boundary based shape descriptors are based on the usage of boundary information of an object. Fourier descriptors are among the popular techniques of boundary-based shape descriptors. They are robust, contain perceptual characteristics but information about local features is not present because only magnitudes of the frequencies are present in Fourier transform and location information is missing.

Curvature scale space (CSS) is another shape descriptor which analyzes the boundary of an object as a 1D signal and finally represents the signal in scale space. The major issue with this technique is the superficial projections on the shape of an image. Angular radial transform (ART) produces a high dimensional feature vector which causes a hindrance in the performance of a CBIR system. Image moments, Zernike moments, Hu moments, Canny edge detector are also among the prominent shape and edge descriptors but suffer from some main functioning issues like prior image normalization to obtain scale invariance, more time consumption, computational complexity, etc. Properties of image regions are efficiently measured by using region props which measure the number of connected components in an image. Many types of region props such as Center of gravity (Centroid) [3], Mass, Dispersion, Eccentricity, Axis of least inertia, Hole area ratio, etc. can be used to find the shape related information in an image. In this paper, some of the region props are used to find the largest connected component. They are:

Mass: It is the total number of pixels present in one class. It is given as:

$$\sum_{mn} h(m,n) \quad (9)$$

$$\text{Where } h = \begin{cases} 1 \; if \; s(m,n) \in C \\ 0 \; if \; s(m,n) \notin C \end{cases}$$

Centroid: The center value of all the pixels is denoted by Centroid and is also known as the center of mass. C denotes the cluster, h specifies mask over the same cluster C over image $S(m,n)$. A Centroid is given by:

$$y_C = \frac{\sum_{mn} m*h(m,n)}{mass} \quad (10)$$

$$\text{And } z_C = \frac{\sum_{mn} n*h(m,n)}{mass} \quad (11)$$

In these equations (y_c, z_c) are the co-ordinates of the centroid.

Mean: It is defined as the average value of all pixels and is denoted by:

$$\mu_C = \frac{\sum_{mn} I_{mn} * h_C(m,n)}{mass} \quad (12)$$

Variance: It is an analysis which measures the distance of the spread from an average value of given random numbers.

$$\sigma_C^2 = \frac{\sum_{mn} (I_{mn} - \mu_C) * h_C(m,n)}{mass} \quad (13)$$

Dispersion: Dispersion is defined as the total distance from the centroid to every class present in an image. It is given as:

$$Dispersion = \sum_j dist(O_{D,} O_{j,D}) \quad (14)$$

Where $dist(O_D, O_{j,D})$ gives information related to the distance metric, centroid of class D is O_D, centroid of region j of class D is O_{jD}.

D. Extreme Learning Machine

To procure a precise accuracy during the classification of the image dataset, many types of classifiers have been used in the field of image retrieval. Support vector machines (SVMs) [34] are amongst the most widely used classifiers due to their ease of access. But, they are less accurate and take more training time. Moreover, no multi-class SVM is directly available. Naive's Bayes also suffers from various issues like: To deal with continuous features initially, a binning procedure is required to be adopted to translate those features into discrete features. Data scarcity, data assumptions are also some of the issues related to this classifier. Random forest, KNN are also used as classifiers but lack in accurate classification due to some issues. Therefore, an Extreme learning machine (ELM) classifier based on neural network is chosen which has many advantages as compared to these classifiers. The training speed of an ELM is very swift and also it has a rationalized performance based on its operation. ELM is well known as a single-layer feed-forward neural network (SLFN), designed to be used for classification and regression applications. It is a supervised learning model based on labeled data. It was initially introduced by Huang at el. [26]. ELM has been successfully utilized in many research problems like pattern recognition, classification, fault identification, over-fitting, etc. In this system, the input weights are selected without any conscious decision and by using a specified analytical technique, its output weights are decided. This algorithm indeed is based on a single hidden layer and the total number of nodes present in this layer is a principal parameter to be decided. ELM has many advantages in contrast to many related traditional techniques like: least human involvement, [35] swift learning speed, complimentary universal capability, convenience to use, varied kernel functions, etc. [36].

If there are M different samples denoted by (X_j, t_j), where $X_j = [x_1, x_2, x_3 \ldots . x_n]^T \in R^n \times R^m$ (where j=1, 2…… M).

Then, it is a SLFN with $\widetilde{M}$ hidden nodes and f(x), where f(x) is an activation function, and can be represented as:

$$\sum_{j=1}^{\widetilde{M}} \beta_i f(u_j x_k + v_j) = o_k \quad (15)$$

In the above equation, the weight vectors, which connect the input nodes to the j^{th} hidden nodes, are given by $u^j = [u^{j1}, u^{j2}, u^{j3} \ldots . u^{jn}]_T$ and $\beta^j = [\beta^{j1}, \beta^{j2}, \beta^{j3} \ldots . \beta^{jn}]_T$ represents the connecting vectors between the j^{th} hidden nodes and the nodes depicting outputs. The value describing threshold for the various hidden nodes is given by v^j whereas u^j and x^k is the inner product. The basic diagram of ELM is shown in Fig. 2.

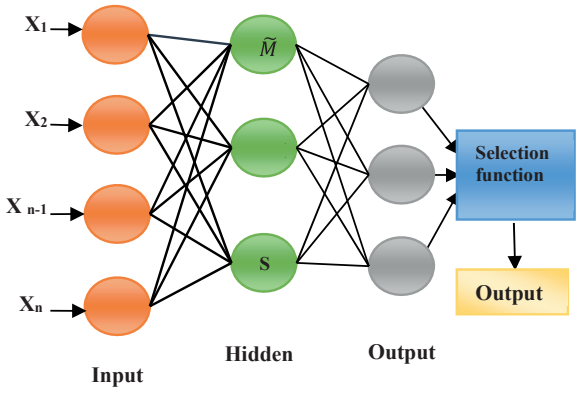


Fig. 2. A basic form of an Extreme learning machine.

The Equation (15) can be briefly re-written as:

$$H = \beta O \quad (16)$$

Where

$$H = \begin{bmatrix} f(u_{1,}x_{1,}v_1) & \cdots & f(u_{\widetilde{M},}x_{1,}v_{\widetilde{M}}) \\ \vdots & \vdots & \vdots \\ f(u_{1,}x_{M,}v_1) & \cdots & f(u_{\widetilde{M},}x_{M,}v_{\widetilde{M}}) \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\widetilde{M}}^T \end{bmatrix} \text{ and } O = \begin{bmatrix} O_1^T \\ \vdots \\ O_1^T \end{bmatrix}$$

Here, H represents the output matrix of the hidden layer. Also, the H matrix becomes invertible, if the total number of given samples $\widetilde{M}$ becomes equal to hidden node parameters H. However, the learning attributes of ELM, u_j, v_j and the hidden nodes can be allocated randomly, in the absence of any input data, therefore the output weights β of a linear system can be calculated by implementing least square technique as follows:

$$\hat{\beta} = H^{\dagger}T \tag{17}$$

Where Moore-Penrose conception in one of its versions is given by $H^{\dagger}$. So, from equation (17), we can see that the calculation of the output weights is done by using a straightforward mathematical equation, thereby avoiding any lengthy procedures. Thus, the algorithm of ELM can be summarized in three steps, which are as follows:

(1) Assigning of hidden node parameters u_j and v_j, where i=1,2……. $\widetilde{M}$.

(2) Calculation of H, i.e matrix of a hidden layer by using an activation function.

(3) Output weights calculation by using $\beta = H^{\dagger}T$

Thus, ELM transforms a complex problem into a simpler and linear function. Fast speed, more accuracy and many more advantages contribute to making this technique more sophisticated and precise as compared to many other customized methods.

E. Relevance Feedback

It is a strategy which helps to refine a particular image based on the feedback obtained by the user. To search the system with massive database images, text or a combination of images annotated with text can be used as a query. Then, a set of relevant images are obtained based on a specific query image. These retrieved images are analyzed by the user and finally, the query image is refined by using Relevance feedback which selects the best-matched images, based on some common features. This process works iteratively until the desired results are obtained or the user gets satisfied. This intelligent technique can also be used in combination with many other concepts like Support vector machines (SVM), Neural networks with different training algorithms [37], Deep learning, machine learning, etc. [38].

The query input given by the user can be broadly classified into three types: The first category consists of a system in which a query image is composed of only keyboard text letters. This technique has some limitations like polysemy, synonymy, homonymy, etc. So, finding the desired images based on the user's intention is a major issue of concern. A query can also be given in the form of an image, which is the second medium of inputting a query image. This technique has removed many ambiguities, which were present in the traditional method of the query by text. Also, this method has gained vast popularity in recent times due to its numerous applications in image processing. Relevance feedback can be considered as the third category of providing a query image, indeed through the iterative refinement of a user's query image. The three basic ways of a query refinement are as follows:

(1) Extension of Query: In this technique, the neighboring images of an actual query image are also included in it, based on the feedback obtained by the user. Thus, in a way, an expansion of an original query image is done.

(2) Query Re-Weighting: This method enhances the weights of some prominent attributes of an image and simultaneously reduces the weights of some un-important attributes. In this way, a query becomes more refined.

(3) Movement of Query: A query is moved close to the required images by the adjustment in the attributes of a utilized distance function.

In this paper, to make the hybrid textural system more effective, Relevance Feedback is utilized. It works on the relevant images obtained after classification by ELM. It works on the refinement process until the satisfactory results are obtained.

III. Proposed Methodology

In the proposed work, a unique image retrieval system has been described which is an amalgamation of texture, color and shape features depicting an image. Color moment has been utilized for color feature extraction while Gray level co-occurrence matrix (GLCM) and varied region props have been adopted respectively for texture and shape feature extraction. These three techniques together present a great combination. Color moment effectively captures spatial information of a particular image and is also invariant to rotation, scale and angle. Similarly, GLCM has highly factual results and takes very little computation time in its execution. For the shape feature extraction, Mass, Centroid, Mean, Variance and Dispersion parameters have been calculated. These parameters effectively describe the shape of an object, specified in the given image. Hence, our proposed framework consists of an amalgamation of these three techniques. For classification accuracy, the incorporation of an Extreme Learning Machine (ELM) has also been done whose training images have been reformulated or updated based on the condition of relevance feedback. For the detailed explanation of the proposed work, this approach has been divided into two major subsections: The first section explains the working of both the main phases of the system, i.e., Training stage and the Testing stage. The other stages related to the proposed model have been given in another subsection.

A. Training and Testing Stage

The stage related to training of the system is basically concerned with the training of an ELM model. Every neural network is required to be trained through the total images in the database. In this phase, the three types of features are withdrawn from the whole dataset and three independent feature vectors are formed. These three feature vectors are normalized to form a hybrid feature vector (HFV). To train the ELM model, this HFV is given as an input to it and different categories of a dataset are formed and gets trained as a classifier. Normalization brings the feature dimensions into a common range. Here, minimum-maximum normalization is used and is denoted by:

$$Obtained\ normalized\ features = \frac{y - maximum}{minimum - maximum} \tag{18}$$

Where y is the value of a particular feature, *minimum* is the bottom value of every single feature vector and *maximum* is the highest value of every single feature.

Decimal scaling and Z-score are also among the prominent Normalization techniques. In Decimal scaling, the normalization is achieved by moving the decimal point of input values. But, it is generally based on calculating the least and the largest value of given data, which is difficult in some cases. Moreover, if the assumption about these values is done, then also the results could be impermissible. The main disadvantage of Z-score Normalization is that it always assumes a normal distribution. But, if this condition is not met, then vague

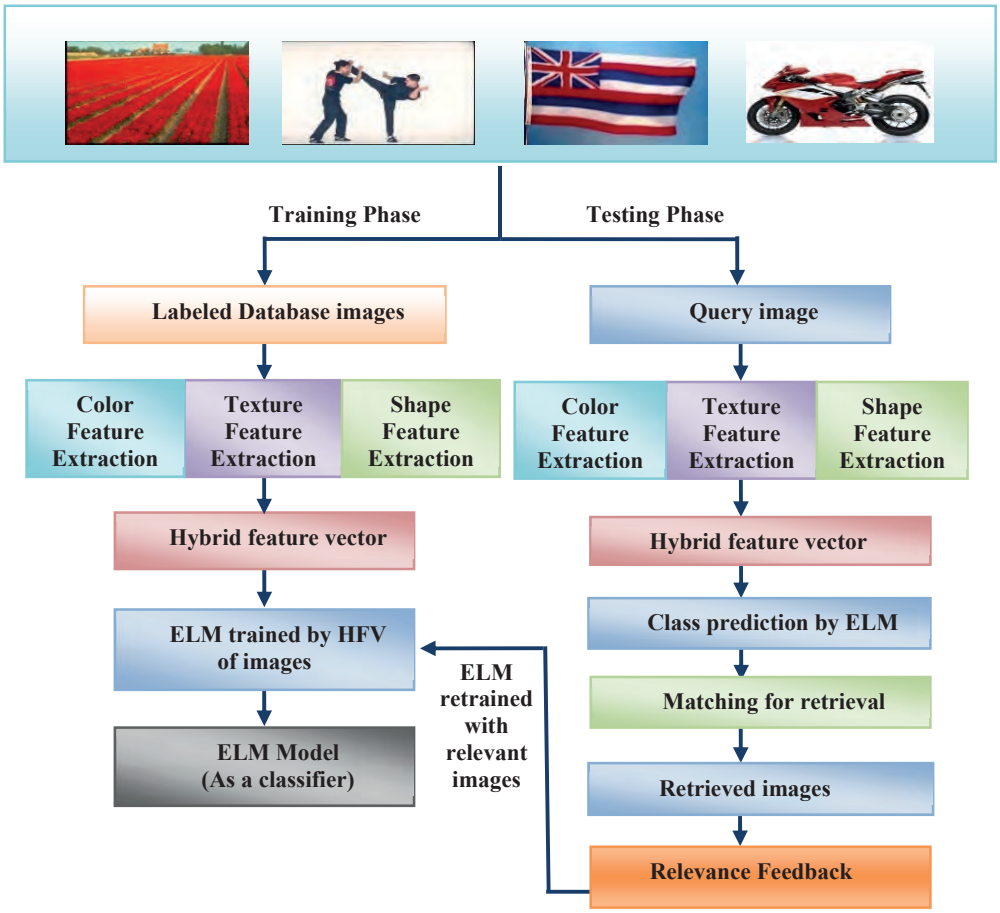


Fig. 3. Architectural framework of the proposed system.

results are produced. Therefore, on the basis of these conclusions, Min-Max Normalization is preferred because the correspondence among all the data values is conserved, without any bias introduction. The basic architectural implementation is given in Fig. 3.

The main aim of the testing stage is concerned with the testing of the proposed system by using a pre-trained ELM model with a specific query image and obtaining the desired top N images.

B. Prominent Stages During the Working of the Proposed System

The various prominent stages during the working of the proposed system are as follows:

1. Feature Extraction Stage

In this stage, texture, shape and color features are withdrawn from a query image using GLCM, region props process and color moment respectively and three respective feature vectors are obtained. Again, a hybrid feature vector (HFV) is finally obtained by combining the three independent feature vectors by using the process of normalization.

2. Classification Stage

The second important stage of working is based on classification. Here, the obtained HFV is applied in the form of an input to a pre-trained ELM model. This ELM model has been initially trained in the training phase with the complete specific database.

To train the ELM model, different types of activation functions can be used and here, Radial basis function (RBF) is utilized. RBF is a type of multi-layer perceptron (MLP) which can use one or more hidden layer besides input and output layer. RBF has only a single hidden layer of neurons where each neuron of the hidden layer calculates the RBF function. These hidden nodes project the lower dimension of feature vector to a higher dimension. The output nodes contain the classifying neurons. The number of neurons in the output layer are equal to the number of classes of the dataset. The output layer nodes decide to which class, the input feature vector should lie. If the output of first node = 1 and all other nodes = 0, then it means that the query image belongs to first category.

The output of this classification step is the images with meticulous categorization and class prediction by an ELM classifier. Various working parameters of the proposed system are given in Table I. Here, 9 features of color moment are used as input, 44 of GLCM and Region props contain 5 features. To form a hybrid system, these features are added and 58 input features are obtained. For, Corel-1K dataset, the output neurons are 10, 50 for Corel-5K and so on. Single layer feed forward neural network is used here as it contains a single hidden layer and it is a feed forward neural network. Radial basis function (RBF) is utilized as an activation function where 100 denotes the number of neurons in the hidden layer.

TABLE I. Various Working Parameters of the Proposed System

Number of input features	9 (Color), 44 (Texture), 5 (Shape)
Fused features	9+44+5=58
Number of Output (Classes)	10 (Corel-1K), 50 (Corel-5K), 100 (Corel-10K), 20 (GHIM-10)
ELM Kernel Type	Radial basis function (RBF)
Kernel Parameters	100
Algorithm	Feed-forward single hidden layer neural network

3. Similarity Matching Stage

After an accurate class prediction by an ELM network, the whole dataset images are successfully classified into varied categories. Now, based on a query image, a similarity matching is done between a given query image and the respective category to which it belongs. These categories are formed as a result of an ELM classification. After the results of similarity calculation are obtained, the resultant images are arranged in increasing order based on the utilized distance metric. A result of zero with regard to distance metric exhibits accurate resemblance between two images. Many types of distance metrics are utilized for the purpose of calculating similarity. Some of the prominent distance metrics which are used in similarity calculation are given under:

$$Distance_{Euclidean} = \sqrt{\sum_{j=1}^{n}(|I_j - D_j)^2|} \quad (19)$$

$$Distance_{Manhattan} = \sum_{j=1}^{n}|I_j - D_j| \quad (20)$$

$$Distance_{Minkowski} = [\sum_{j=1}^{n}(|I_j - D_j)^{1/P}|] \quad (21)$$

Here, I_j denotes the input query image and D_j depicts all database images.

4. Relevance Feedback Stage

The main aim of this step is to encompass the user feedback to check the relevancy of the retrieved images after classification by ELM. The images retrieved after classification are divided into two groups: Relevant and Non-Relevant based on the feedback obtained by the user. These set of Relevant images are again considered to improve and reformulate the classification procedure of ELM. The novelty of the proposed approach is that the obtained set of predicted semantics by ELM is improved to a great extent by using two iterations of relevance feedback. This process rejects the non-relevant images based on user's feedback and finally, based on these refined images, final top 10 images are retrieved. This process enhances the accuracy of the proposed system to a significant value.

IV. Experimental Methods and Results

To analyze the retrieval efficiency of the implemented technique, various benchmark datasets for CBIR system has been utilized. There are a diversity of images present in these databases which have been used successfully in many retrieval techniques. All the experiments have been performed on Windows 10 operating system (OS) by utilizing version R2017a of MATLAB with core i3 processor, 4 GB RAM and 64-bit windows.

1st Dataset: Corel-1K: The initial database analyzed for experimentation is Corel-1K which is composed of 1000 different images. It has 10 categories which are Africa, Buildings, Bus, Beach, Dinosaur, Elephant, Flowers, Horse, Food and Mountains. 100 images are present in each category and each image bears the size of either 256 × 384 or 384 × 256. (http://wang.ist.psu.edu/docs/related/)

2nd Dataset: Corel-5K: The next dataset is Corel-5K and consists of 5000 images. There are 50 categories in this dataset and again every group has 100 images. Some of the group types of this dataset are butterfly, cards, vegetables, dogs, judo, etc. The size of each image is either 256 × 384 or 384 × 256. (http://www.ci.gxnu.edu.cn/cbir/)

3rd Dataset: Corel-10K: The third dataset is Corel-10K and there are 10,000 images in this dataset. This database has different groups of 100 images and every set has further 100 images in it. Every image bears the size of either 256 × 384 or 384 × 256. Some of the images of this dataset belong to cars, dolls, lamp-posts, flags, etc. category. (http://www.ci.gxnu.edu.cn/cbir/)

4th Dataset: GHIM-10: The last database for the experimental analysis is GHIM-10. It also consists of 10,000 images with a total of 20 categories. Each category has 500 images in it consisting of bikes, car, aeroplane, grasshopper, etc. Every image size is of either 300 × 400 or 400 × 300. (http://www.ci.gxnu.edu.cn/cbir/)

For the formation of a query or input image, each and every single image of all the databases is utilized. If the obtained resultant images correspond to the concordant native category of the input image then, the proposed system has effective results and the retrieval is considered as a successful retrieval. The role of Extreme learning based-Relevance feedback framework contributes to enhancing and upgrading the retrieval efficiency of the implemented technique. Few of the sample images from all the datasets is shown in Fig 4.

Fig. 4. Sample images from all datasets.

A. Methods

In CBIR systems, the capability of a particular system can be concluded with respect to many evaluation parameters [39]-[40]. Precision and Recall are the most well-known evaluation metrics. These are defined using the given equations:

$$Precision(P_i) = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ images\ retrieved} \quad (22)$$

$$Recall(R_i) = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ relevant\ images\ in\ the\ database} \quad (23)$$

Here, the total recovered images are 10 while the number of relevant images of a particular dataset depends on the number of images present in each category of that dataset. Corel-1K, Corel-5K and Corel-10K, has 100 images as relevant while in the case of GHIM-10, it is 500.

B. Results

Results based on the retrieval of desired images are obtained by taking each and every image from all the datasets i.e Corel-1K, Corel-5K, Corel-10K and GHIM-10 as a query or an input image. Then, Color moment, GLCM and various region props are used for the withdrawal of color, texture and shape attributes and the hybrid feature vector is formed by combining independent feature vectors of the three techniques. The same procedure applies to all database images also. After feature extraction, the pre-trained ELM model performs the accurate class prediction and finally, a similarity is calculated between the entered input image and the classified images of the complete dataset. Similarity calculation is done by utilizing three distance metric techniques. The results in terms of Average precision on all the employed datasets, obtained after classification by ELM using the three distance metrics are shown in Table II.

TABLE II. AVERAGE PRECISION USING THREE DISTANCE METRICS

Dataset	Euclidean	Manhattan	Minkowski
Corel-1K	92.1	88.45	85.3
Corel-5K	79.6	73.6	69.9
Corel-10K	72.4	68.7	65.3
GHIM-10	89.95	83.2	78.62

As evident from Table II, the average precision obtained by using Euclidean distance metric outperforms the other distance metrics. Since Manhattan distance is a distinctive case of Minkowski distance, it produces innumerable false negatives and does not yield accurate results. Euclidean distance metric is based on weighted and normalized attributes and has speedy computational performance. Therefore, the Euclidean distance metric gives precise results and is being used here.

The average precision of the proposed system is obtained in the form of three phases. In the first phase, precision is calculated only after the fusion of texture, color and shape parameters. In the second phase, the results are obtained from the combination of hybrid features and ELM classifier and, in the last phase, the results of the total proposed system are obtained which is the combination of hybrid attributes, ELM and Relevance feedback. Fig. 5 plot shows the Average precision vs Datasets plot for all the four databases in the stepping of three levels.

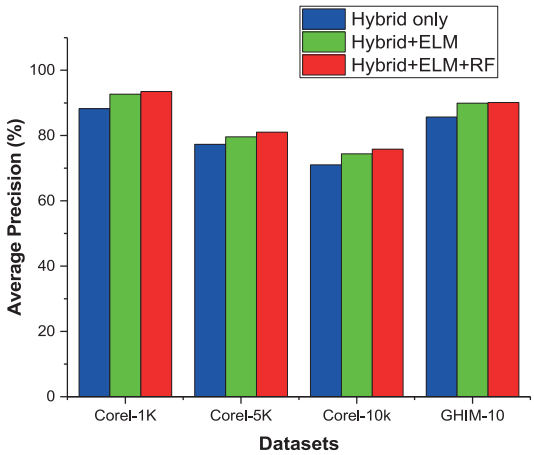


Fig. 5. Average Precision of the proposed system depicting three levels of functioning.

From Fig. 5, it is clear that the average precision of the system increases as more and more intelligence is being added to the proposed system. With the increase in intelligence of the system, more high-level semantics are captured and the system becomes more efficient.

Precision and Recall are the prominent measures to check the effectiveness of any particular CBIR system. Therefore, Precision vs Recall curves of the implementation by varying the number of retrieved images from 10 to 50 on all the four datasets is shown in Fig. 6 and 7.

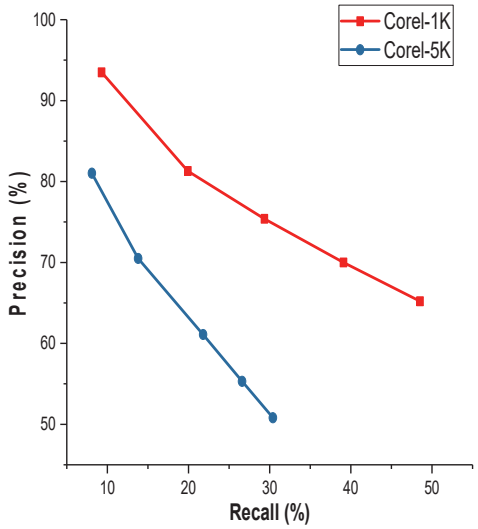


Fig. 6. Precision Vs Recall plot on Corel-1K and Corel-5K datasets.

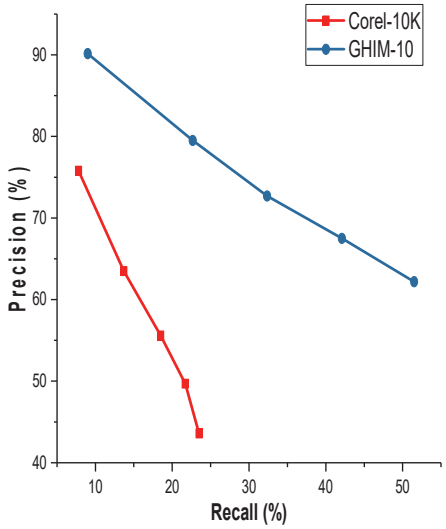


Fig. 7. Precision Vs Recall plot on Corel-10K and GHIM-10 datasets.

From Fig. 6 and 7, it can be seen that with the rise in the number of images retrieved, the value of precision decreases while recall increases. Separate Graphical user interface (GUI's) have been designed for each of the four datasets. Top ten images are retrieved through each utilized dataset. The GUI's depicting the retrieval results for each of the four datasets based on a specific query image are shown in Fig. 8 (a-d).

From Fig 8, it can be concluded that the top ten images are retrieved from the desired category of the dataset which in-turn belongs to the native category of the input image. Thus, the gross accuracy of the implemented technique is undoubtedly outstanding as compared to the other state-of-the-art techniques based on intelligent and hybrid feature retrieval. The accuracy of the presented system is also given in Table III.

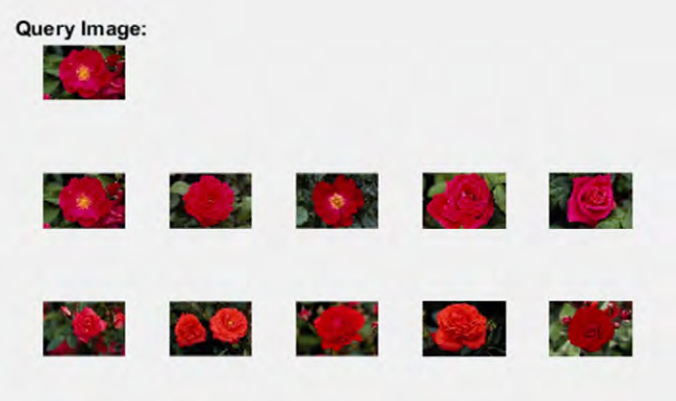


Fig. 8(a). Retrieval results from Corel-1K dataset.

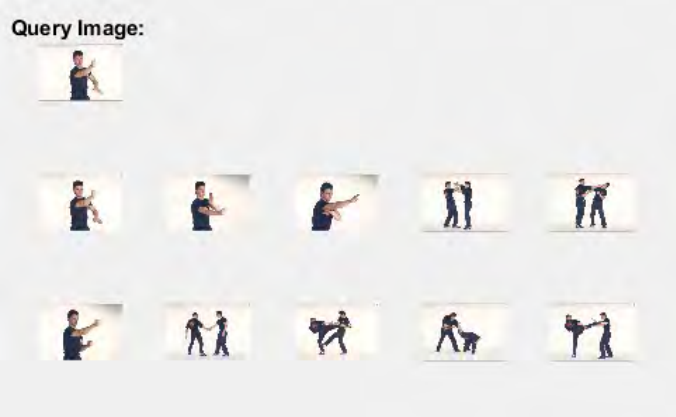


Fig. 8(b). Retrieval results from Corel-5K dataset.

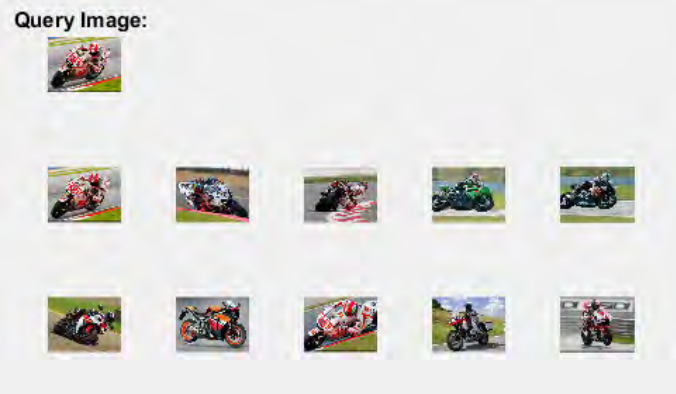


Fig. 8(c). Retrieval results from Corel-10K dataset.

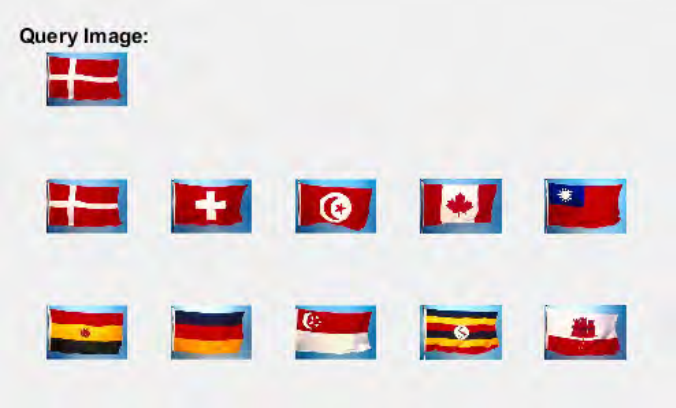


Fig. 8(d). Retrieval results from GHIM-10 dataset.

TABLE III. Accuracy of the Proposed System

Dataset	Accuracy (%)
Corel 1K	96.5
Corel 5K	94.67
Corel 10K	94.85
GHIM-10	99.02

The obtained accuracy can also be verified by checking the diagonal elements of the generated confusion matrix during classification by ELM classifier. Confusion matrix of one of the four datasets (GHIM-10) is also given in Table IV.

TABLE IV. Confusion Matrix of Ghim-10 Dataset

496	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	
2	496	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
1	0	495	0	1	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	
0	0	0	498	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	498	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
1	1	0	0	1	495	0	0	1	1	0	0	0	0	0	0	0	0	0	0	
2	0	1	0	1	0	496	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	2	0	0	0	0	0	497	0	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	1	1	0	1	0	497	0	0	0	0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	0	500	0	0	0	0	0	0	0	0	0	0	
0	1	0	0	1	0	1	0	1	0	496	0	0	0	0	0	0	0	0	0	
0	0	1	0	1	0	1	1	0	0	0	496	0	0	0	0	0	0	0	0	
0	1	0	0	0	0	0	0	2	2	0	0	495	0	0	0	0	0	0	0	
3	1	0	0	0	0	1	0	2	0	0	0	0	493	0	0	0	0	0	0	
1	1	1	0	2	0	1	0	0	0	0	0	0	0	494	0	0	0	0	0	
1	0	0	0	0	2	0	1	1	0	0	0	0	0	0	495	0	0	0	0	
0	0	0	0	1	0	0	0	2	2	0	0	0	0	0	0	495	0	0	0	
0	2	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	494	0	0	
1	0	1	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	494	0	
1	0	0	0	0	2	0	0	2	2	0	0	0	0	0	0	0	0	0	493	

From the diagonal elements of the confusion matrix, it can be seen that out of 500 images present in 20 categories of the dataset, a handful of images are present in every single group, which corresponds to an accuracy of 99.02%.

C. Comparison of the presented system with the related techniques

The implemented system has been initially compared to many state-of-the-art related techniques regarding average precision obtained on Corel-1K, Corel-5K and Corel-10K datasets. In comparison, the main considerations are:

The majority of the hybrid systems are based only on the extraction of either one or two attributes of an image in spite of all the three basic visual attributes. Due to this, those systems lack in the recognition of complex and large dataset images.

The methods which are used for feature extraction by the related hybrid systems are deficient in one or the other ways as compared to the proposed system. For eg., the color histogram is used for color extraction but it lacks the spatial information and, moreover, two different images can produce the same histograms.

In order to classify the images, generally Support vector machine (SVM) has been utilized but it has less classification accuracy as

TABLE V. COMPARISON OF THE PROPOSED SYSTEM WITH THE STATE-OF-THE-ART TECHNIQUES

Database	Semantic Name	Average Precision (%)							
		Ref. [2]	Ref. [5]	Ref. [7]	Ref. [8]	Ref. [1]	Ref. [9]	Ref. [10]	Proposed
Corel-1K	Africa	72.4	81	68	74.5	95	81	84.7	84.5
	Beach	51.5	66	65	69.3	60	92	45.4	98
	Building	59.55	78.75	80	85.1	55	79	67.8	79
	Bus	92.35	96.25	90	95.4	100	93	85.3	86
	Dinosaur	99	100	100	100	100	99	99.3	100
	Elephant	72.7	70.75	80	83.3	90	79	71.1	100
	Flower	92.25	95.75	85	98	100	99	93.3	99
	Horse	96.6	98.75	95	94.2	100	80	95.8	100
	Mountain	55.75	67.75	75	75.9	75	85	49.8	84
	Food	72.35	77.25	85	92.6	100	88	80.8	100
	Average	76.5	83.225	82.3	77.8	87.5	87.5	77.3	93.05
Corel-5K	Average	55.25	68.6	64.5	59.5	72.56	75.4	58	81.03
Corel-10K	Average	49.58	59.98	57.54	52.45	65.45	68.9	72	75.8

compared to Extreme learning machine (ELM) employed in the presented system.

To capture the high-level semantic features of an image, Relevance feedback has been used in the proposed system but the related systems lack this concept of human intelligence.

These comparisons have been given in Table V where the precision of the implemented system outshines the other state-of-the-art compared techniques and performance plot has been given in Fig. 9.

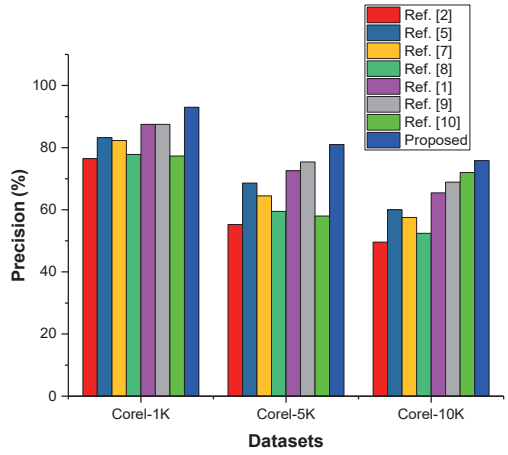


Fig. 9. Comparison plot of the proposed technique with the related techniques on Corel-1K, Corel-5K and Corel-10K dataset.

Again from the comparison of GHIM-10 dataset, it can be concluded that the proposed system has enhanced and accurate results as compared to many related techniques based on this database. The average precision of the proposed system obtained on GHIM-10 dataset is shown in Table VI and its comparative plot in Fig. 10.

TABLE VI. COMPARISON OF THE PROPOSED SYSTEM BASED ON GHIM-10

Database	**Average Precision (%)**			
GHIM-10	Ref. [11]	Ref.[25]	Ref. [39]	Proposed
	76.99	73.7	57.51	**90.14**

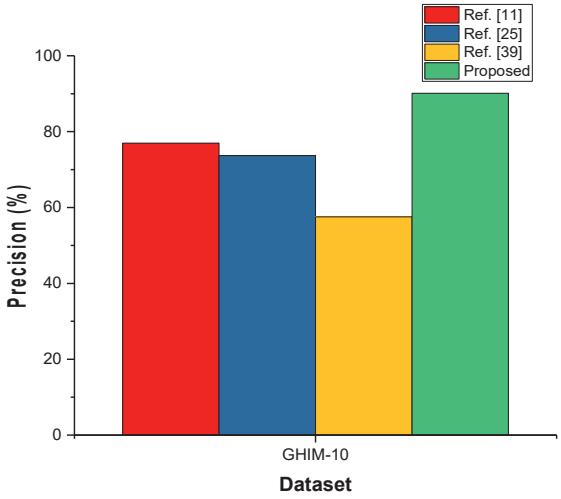


Fig. 10. Comparison plot of the proposed technique with the related techniques on GHIM-10 dataset.

The comparison of the proposed system in terms of Recall is also presented in Table VII for Corel-1K, Corel-5K and Corel-10K datasets. Table VIII gives the comparative analysis for GHIM-10 dataset for the Recall parameter. These recall results are based on the retrieval of 20 images.

TABLE VII. COMPARISON OF THE PROPOSED SYSTEM BASED ON COREL-1K, COREL- 5K AND COREL-10K

Database	**Average Recall (%)**				
	Ref. [1]	Ref. [5]	Ref. [7]	Ref. [8]	Proposed
Corel-1K	17.50	16.64	16.60	7.58	**19.92**
Corel-5K	14.67	13.72	14.45	5.33	**16.54**
Corel-10K	12.34	11.99	13.65	4.65	**13.75**

TABLE VIII. COMPARISON OF THE PROPOSED SYSTEM BASED ON GHIM-10

Database	**Average Recall (%)**			
GHIM-10	Ref. [11]	Ref.[39]	Ref. [25]	Proposed
	22.33	1.38	25.65	**28.98**

Thus, the proposed method has superior results both in terms of Precision and Recall, in contrast to many state-of-the-art techniques and can work accurately and precisely on both small and large datasets.

D. Time Performance Analysis

In order to increase the accuracy of the proposed system, time performance analysis is an important parameter to be considered. Here, analysis of time is done during training and testing the model. This time analysis during training phase is divided into Feature extraction time and ELM training time while during testing phase it is based on the testing time of the complete proposed model. The time analysis for all the four utilized datasets is given in Table IX.

TABLE IX. Time performance Analysis of the Proposed System

Dataset	Time (in seconds)
Corel 1K	Feature extraction time = 348.64 ELM training time = 0.1781 Total training time = 348.81 Testing time = 0.389
Corel 5K	Feature extraction time = 595.733 ELM training time = 1.3029 Total training time = 597.03 Testing time = 0.955
Corel 10K	Feature extraction time = 2845.00 ELM training time = 245.67 Total training time = 3090.67 Testing time = 4.342
GHIM-10	Feature extraction time = 2543.00 ELM training time = 205.78 Total training time = 2748.78 Testing time = 3.044

From Table IX, it can be concluded that as the used dataset becomes more and more complicated i.e. number of images increases, some more time is utilized for the total training of the model but the testing time is much less. Thus, the proposed system is very effective in testing both smaller and larger image datasets.

V. Conclusion and Future Work

This paper describes a novel and an efficient technique for Content-based image retrieval (CBIR) system which is focused on the formation of a hybrid feature vector (HFV). This HFV is formed utilizing the independent feature vectors of three visual attributes of an image, namely texture, shape and color which are extracted by using Gray level co-occurrence matrix (GLCM), region props procedure employing varied parameters and color moment respectively. The proposed system is the combination of these three techniques which has many advantages as GLCM has precise inter-pixel and inter-pattern relationship, as compared to many basic texture extraction methods. Color moment captures spatial information of an image and is also invariant to scale, angle and rotation. Shape parameters can be used to detect the connected components in an image. These hybrid features are applied to an Extreme learning machine (ELM) deputing as a classifier which is a feed-forward neural network having one hidden layer. After that, to retrieve the higher level semantic attributes of an image, Relevance feedback is used in the form of some iterations based on the user's feedback. This extreme learning based-Relevance feedback framework helps in the evolution of an intelligent and modified system for learning and classification. Four benchmark datasets have been tested on the proposed system with respect to Precision, Recall and Accuracy. The average precision for the presented implementation is 93.05%, 81.03%, 75.8% and 90.14%, respectively, on Corel-1K, Corel-5K, Corel-10K and GHIM-10 datasets, which is significantly larger than that of many state-of-the-art related methods of hybrid CBIR system. The proposed work does not consider the information of the desired region of an image but is based on an entire image. Therefore, our future work will concentrate on the Region of interest (ROI) of an image by using local as well as global information of an image by using Deep learning techniques for feature extraction. Internet of Things (IoT) will be used for the online creation and transfer of database images.

References

[1] J. Pradhan, S. Kumar, A. K. Pal, and H. Banka, "A hierarchical CBIR framework using adaptive tetrolet transform and novel histograms from color and shape features," *Digit. Signal Process. A Rev. J.*, vol. 82, pp. 258–281, 2018.

[2] S. Fadaei, R. Amirfattahi, and M. R. Ahmadzadeh, "New content-based image retrieval system based on optimised integration of DCD, wavelet and curvelet features," *IET Image Process.*, vol. 11, no. 2, pp. 89–98, 2017.

[3] A. Alzu'bi, A. Amira, and N. Ramzan, "Semantic content-based image retrieval: A comprehensive study," *J. Vis. Commun. Image Represent.*, vol. 32, pp. 20-54, 2015.

[4] R. Chaudhari and A. M. Patil, "Content Based Image Retrieval Using Color and Shape Features," *Int. J. Adv. Res. Electr. Electron. Instrum. Eng.*, vol. 1, no. 5, pp. 386-392, 2012.

[5] L. K. Pavithra and T. S. Sharmila, "An efficient framework for image retrieval using color, texture and edge features," *Comput. Electr. Eng.*, vol. 0, pp. 1–14, 2017.

[6] M. K. Alsmadi, "An efficient similarity measure for content based image retrieval using memetic algorithm," *Egyptian J. of Basic and Appl. Sci*., vol. 4, no. 2, pp. 1-11, 2017.

[7] E. Mehdi, E. Aroussi, and N. El Houssif, "Content-Based Image Retrieval Approach Using Color and Texture Applied to Two Databases (Coil-100 and Wang)". *Springer International Publishing, https://doi.org/10.1007/978-3-319-76357-6_5*, pp. 49-59, 2018.

[8] M. A. Ansari, M. Dixit, D. Kurchaniya, and P. K. Johari, "An Effective Approach to an Image Retrieval using SVM Classifier" *Int. J. of Computer Sciences and Engineering*, vol. 5, no. 6, pp. 64-72, March 2018.

[9] Y. Mistry, D. T. Ingole, and M. D. Ingole, "Content based image retrieval using hybrid features and various distance metric," *J. Electr. Syst. Inf. Technol.*, no. 2016, pp. 1–15, 2017.

[10] J. M. Guo, H. Prasetyo, and H. Su, "Image indexing using the color and bit pattern feature fusion," *J. Vis. Commun. Image Represent.*, vol. 24, no. 8, pp. 1360–1379, 2013.

[11] P. Srivastava and A. Khare, "Utilizing multiscale local binary pattern for content-based image retrieval," *Mutimedia Tools and App.,* DOI 10.1007/s11042-017-4894-4, pp. 1-27, 2017.

[12] C. Singh and K. Preet Kaur, "A fast and efficient image retrieval system based on color and texture features," *J. Vis. Commun. Image Represent.*, vol. 41, no. October, pp. 225–238, 2016.

[13] M. V. Lande, P. Bhanodiya, and P. Jain, "An Effective Content-Based Image Retrieval Using Color, Texture and Shape Feature," *Intelligent Computing, Networking, and Informatics, Advances in Intelligent Systems and Computing 243, DOI: 10.1007/978-81-322-1665-0_119,* pp. 1163–1170, 2014.

[14] M. Arevalillo-Herráez, F. J. Ferri, and S. Moreno-Picot, "A hybrid multi-objective optimization algorithm for content based image retrieval," *Appl. Soft Comput.*, vol. 13, no. 11, pp. 4358–4369, 2013.

[15] M. Mosbah and B. Boucheham, "Distance selection based on relevance feedback in the context of CBIR using the SFS meta-heuristic with one round," *Egypt. Informatics J.*, vol. 18, no. 1, pp. 1–9, 2017.

[16] E. Walia, S. Vesal, and A. Pal, "An Effective and Fast Hybrid Framework for Color Image Retrieval," *Sens Imaging*, vol. 15, no. 1, pp. 1-23, 2014.

[17] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," *Pattern Recognit.*, vol. 29, no. 8, pp. 1233-1244, 1996.

[18] C.-H. Lin, R.-T. Chen, and Y.-K. Chan, "A smart content-based image retrieval system based on color and texture feature," *Image Vis. Comput.*, vol. 27, no. 6, pp. 658–665, 2009.

[19] D. Pandey, "An Efficient Low Level Features of CBIR using Wavelet,

GLDM and SMOSVM Method," *Int. J. of Signal Processing, Image Processing and Pattern Recog.,* vol. 10, no. 3, pp. 13–22, 2017.

[20] C. S. Parkhi and A. Gupta, "A comparative implementation of Content Based Image Retrieval techniques in Wavelet and Cosine domains," *IOSR J. Electron. Commun. Eng.*, vol. 9, no. 2, pp. 25–30, 2014.

[21] N. Neelima, E. Sreenivasa Reddy, and N. Kalpitha, "An Efficient QBIR System Using Adaptive Segmentation and Multiple Features," *Procedia Comput. Sci.*, vol. 87, pp. 134–139, 2016.

[22] P. Sharma, "Improved shape matching and retrieval using robust histograms of spatially distributed points and angular radial transform," *Opt. - Int. J. Light Electron Opt.*, vol. 145, pp. 346–364, 2017.

[23] P. Vikhar, R. Scholar, and P. Karde, "Improved CBIR System using Edge Histogram Descriptor (EHD) and Support Vector Machine (SVM)." in *Proc. of IEEE Int. Conf. on ICT in Business, Industry and Government (ICTBIG),* Indore, India, pp. 1-5, 2016.

[24] R. Usha and K. Perumal, "Content Based Image Retrieval using Combined Features of Color and Texture Features with SVM Classification," *Int. J. of Comp. Science & Comm. Networks* vol. 4, no. 5, pp. 169–174.

[25] K. S. Arun, V. K. Govindan, "A Hybrid Deep Learning Architecture for Latent Topic-based Image Retrieval," *Data Sci. Eng.*, vol. 3, no. 2, pp. 166–195, 2018.

[26] R. R. Saritha, V. Paul, and P. G. Kumar, "Content based image retrieval using deep learning process," *Cluster Comput.*, no. February, pp. 1–14, 2018.

[27] M. Mansourvar, S. Shamshirband, R. G. Raj, and R. Gunalan, "An Automated System for Skeletal Maturity Assessment by Extreme Learning Machines," *PLOS One,* doi:10.1371/journal.pone.0138493, pp. 1–14, 2015.

[28] Y. Kaya, L. Kayci, R. Tekin, and Ö. Faruk, "Evaluation of texture features for automatic detecting butterfly species using extreme learning machine," *J. of Experimental & Theoretical Artificial Intelligence,* vol. 26, no. 2, pp. 37–41, 2015.

[29] R. Fu, B. Li, Y. Gao, and P. Wang, "Content-Based Image Retrieval Based on CNN and SVM," In *Proc. of IEEE Int. Conf. on Computer and Comm., Chengdu, 14-17*, pp. 1-5, 2016.

[30] K. Hemachandran, A. Paul, and M. Singha, "Content-based image retrieval using the combination of the fast wavelet transformation and the colour histogram," *IET Image Process.*, vol. 6, no. 9, pp. 1221–1226, 2012.

[31] S. M. Singh and K. Hemachandran, "Content- Content - Based Image Retrieval using Color Moment and Gabor Texture Feature," *Int. J. of Comp. Science Issues*, vol. 9, no. 5, pp. 299–309, 2012.

[32] W.D. Carlos, M.C.R. Renata and A. L. B. Candeias, "Texture classification based on co-occurance matrix and self-organizing map," In *Proc. of IEEE Int. Conf. on Systems, Man and Cybernetics*, Istanbul, Turkey, 10-13, October, pp. 2487–2491, 2010.

[33] M. P. Kushwaha and R. R. Welekar, "Feature Selection for Image Retrieval based on Genetic Algorithm," *Int. J. of Interactive Multimedia and Artificial Intelligence*, vol. 4, pp. 16–21, 2016.

[34] J. Annrose and C. S. Christopher, "An efficient image retrieval system with structured query based feature selection and filtering initial level relevant images using range query," *Optik*, vol. 157, pp. 1053–1064, 2018.

[35] J. D. Pujari, R. Yakkundimath, and A. S. Byadgi, "SVM and ANN Based Classification of Plant Diseases Using Feature Reduction Technique," *Int. J. of Interactive Multimedia and Artificial Intelligence*, vol. 3, pp. 6–14, 2016.

[36] S. Ding and X. Xu, "Extreme learning machine: algorithm, theory and applications," *Neural comuting and Applications,* vol. 25, no. (3-4), pp. 1-11, June, 2013.

[37] A. Suliman and B. S. Omarov, "Applying Bayesian Regularization for Acceleration of Levenberg-Marquardt based Neural Network Training," *Int. J. of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 1, pp. 68-72, 2018.

[38] S. Liu, H. Wang, J. Wu, and L. Feng, "Incorporate Extreme Learning Machine to content-based image retrieval with relevance feedback," In *Proc. of IEEE 11th World Congress on Intelligent Control and Automation,* Shenyang, China, 29 june-4 july, pp. 1010-1013, March 2015.

[39] G. Liu, "Content-Based Image Retrieval Based On Visual Attention And The Conditional Probability," In *proc. of Int. Conf. on Chemical, Material and Food Engg.*, pp. 838–842, 2015.

[40] S. Bhardwaj, G. Pandove, and P. K. Dahiya, "An Intelligent Multi-resolution and Co-occuring local pattern generator for Image Retrieval," *EAI Endorsed Transactions on Scalable Information Systems,* vol. 4, no. 1, e1, pp. 1–12, 2019.

Shikha Bhardwa

She received her B.Tech degree in ECE from Kurukshetra University, Haryana, India in 2006, Master's degree in ECE from M.M University, Ambala, Haryana, India in 2009. Now, pursuing Ph.D in Electronics and Communication Engineering Department from DCRUST University, Murthal, Haryana, India. Since 2011, she is currently working as an Assistant Professor in University Institute of Engineering & Technology, Kurukshetra University, Haryana, India. Her major research areas are Image processing, Artificial intelligence, Machine learning etc.

Gitanjali Pandove

She received her Ph.D degree in ECE. She is working as Professor in ECE Deptt., in DCRUST, Murthal, Haryana, India. Her research areas are optical communication, digital image processing, power systems, etc. She has more than 70 publications in various international reputed journals.

Pawan Kumar Dahiya

He received his Ph.D degree in ECE. He is working as a Professor in ECE Deptt., in DCRUST, Murthal, Haryana, India. His research areas are Digital and Embedded System Design, ANPR, VANETs, Soft Computing Techniques, Image Processing, Internet of Things (IoT), etc. He has more than 75 publications in various national and international reputed journals.

Adjectives Grouping in a Dimensionality Affective Clustering Model for Fuzzy Perceptual Evaluation

Wenlin Huang[1], Qun Wu[2, 3,4]*, Nilanjan Dey[5], Amira S. Ashour[6], Simon James Fong[7,8], Rubén González Crespo[9,10]

[1] School of Media & Art Design, Wenzhou Business College, Wenzhou (PR China)
[2] Institute of Universal Design, Zhejiang Sci-Tech University, Hanzhou (PR China)
[3] Collaborative Innovation Center of Culture, Creative Design and Manufacturing Industry of China Academy of Art, Hangzhou (PR China)
[4] Zhejiang Provincial Key Laboratory of Integration of Healthy Smart Kitchen System, Hangzhou (PR China)
[5] Department of Information Technology, Techno International New Town, West Bengal (India)
[6] Department of Electronics and Electrical Communications Engineering, Faculty of Engineering, Tanta University (Egypt)
[7] Department of Computer and Information Science, University of Macau, Macau SAR (China)
[8] DACC Laboratory, Zhuhai Institutes of Advanced Technology of the Chinese Academy of Sciences (China)
[9] Department of Computer Science and Technology, Universidad Internacional de La Rioja, Logroño (Spain)
[10] Department of Computer Science and Technology, Marconi International University / UNIR LLC, Miami, Florida (USA)

Received 23 March 2020 | Accepted 25 May 2020 | Published 27 May 2020

Abstract

More and more products are no longer limited to the satisfaction of the basic needs, but reflect the emotional interaction between people and environment. The characteristics of user emotions and their evaluation scales are relatively simple. This paper proposes a three-dimensional space model valence-arousal-dominance (VAD) based on the theory of psychological dimensional emotions. It studies the clustering and evaluation of emotional phrases, called VAdC (VAD-dimensional clustering), which is a kind of the affective computing technology. Firstly, a Gaussian Mixture Model (GMM) based information presentation system was introduced, including the type of the presentation, such as single point, plain, and sphere. Subsequently, the border of the presentation was defined. To increase the ability of the proposed algorithm to handle a high dimensional affective space, the distance and inference mechanics were addressed to avoid lacking of local measurement by using fuzzy perceptual evaluation. By comparing the performance of the proposed method with fuzzy c-mean (FCM), k-mean, hard -c-mean (HCM), extra fuzzy c-mean (EFCM), the proposed VADdC performs high effectiveness in fitness, inter-distance, intra-distance, and accuracy. The results were based on the dataset created from a questionnaire on products of the Ming style chairs online evaluation system.

Keywords

Affective Computing, Product Evaluation, Fuzzy Set, Clustering, Valence-Arousal-Dominance.

DOI: 10.9781/ijimai.2020.05.002

I. Introduction

PERCEPTUAL engineering is a new branch of engineering which combines the perceptual and engineering technologies. It mainly designs products by analyzing human sensibility and manufactures products according to human preference [1]-[2]. The sensibility of perceptual engineering is a dynamic process, which changes fashion, trend and individual timely. It is difficult to grasp and quantify the perceptual issues, but they can be measured, quantified, and analyzed by modern technology, and their rules still can be grasped [3]. Some uncertain inference with evaluations was widely applied in the industrial design field, especially, for the products preference evaluation [4]-[6]. Researchers from Hiroshima University were the first to introduce perceptual analysis into the field of engineering research. In 1970, with the beginning of the comprehensive consideration of the emotions and desires of occupants in residential design, the study of how to embody the sensibility of occupants into engineering technology in residential design was originally called "emotional engineering" [7].

The customer's emotional evaluation of the product exists in its natural language description, while the vocabulary in natural language is often inaccurate and vague [8]-[11]. The difficulty in dealing with natural language also complicates the study of implicit emotions in products. The traditional fuzzy set theory coarse-grained natural language and the formation of linguistic variables [12] reduce the computational complexity, which brings a feasible direction to language processing under weakening conditions [13]-[15]. Osgood's semantic difference work is composed of three-dimensional indicators [16]-[18]. One important dimension is the range of "Valence" from pleasant to unpleasant; the other dimension is "Arousal", which measures of calm

* Corresponding author.
E-mail address: wuq@zstu.edu.cn

to excited; and "dominance" means the perceived degree of control in a (social) situation, so-called VAD model. Mehrabian and Russell used 150 words for experiments, and Bellezza *et al*. used 450 words for experiments; Bradley & Lang launched the psychological quantity calibration experiment to launch an emotional rating data set of about 1000 words, and more than 1000 words were divided into 56 groups, each group consisted of 14 lines of 4 words per line for the subject to rate [19]-[20]. The corresponding vocabulary of the 56 vocabularies is subjected to the three-dimensional sentiment rating of VAD, which is set to a discrete scale [21].

By statistically calculating the data, the VAD emotional space mean and its standard deviation (SD) of each vocabulary are calculated separately. Since the words of the emotional vocabulary are not all suitable for product and perceptual evaluation, some of the vocabularies need to be selected to reflect the product. Emotional elements are carried; therefore, perceptual engineering classifies and merges certain words using grouping, and forms about 25 representative words (perceptual vocabulary groups) through group computing of perceptual vocabulary. This work conducted an emotional rating experiment based on 25 vocabularies and modified the Self-Assessment Manikin (SAM) and Affective Norms for English Words (ANEW) accordingly. Because the traditional SAM-based class experiments use pure manual methods, they face the problem of inefficient data collection, which also imposes constraints on the design of SAM sentiment rating experiments [8]. For example, a discrete scale design is mainly to consider statistical convenience. In this experiment, the expression scale of SAM is still used, but when setting the number of rating points, the continuous point method is adopted, which is also in line with Osgood's theory of "continuous psychological quantity".

In this study, a SAM continuous-scale sentiment rating system was proposed. The data points are implemented using scroll bars, so that the acquired VAD sentiment spatial data is no longer restricted to the discrete distribution of [0-8]. This will be beneficial to discover the microscopic mode of the VAD space; where, the point set distribution form in the VAD space. The purpose of our experiment is to find the relationship between the perceptual vocabulary and the corresponding VAD space, and the distribution of each perceptual vocabulary and VAD emotional space point set, and to find the consistency degree of the emotional vocabulary from the analysis of the distribution state of the point set [22]-[25]. Specifically, the feelings of the user are relatively consistent; these are all expected to be obtained in the experimental data and later analysis. Similarly, in the specific application, the feature set image of the product can be used to perform the same VAD rating experiment to obtain the VAD data of a certain product feature, so that the product's characteristics and emotional elements pass the VAD emotion.

Visually it reproduces the VAD sentiment data of 25 vocabularies; thus, the rating data of the single dimension is drawn separately. The results established that each dimension data has an aggregation effect, which means only from a single dimension (Valence, Arousal or Dominance) (K= Control), the rating data has certain stability; and from the two-dimensional data point distribution observation (K=Elegant), the data still shows a certain aggregation effect. As the number of data increases, people's ratings are not only stable, but also related to Valence, Arousal and Dominance. However, they cannot be generalized by linear regression. The traditional method is to use cluster analysis to select representative point sets for classification, but the results of clustering still make each emotional element become an isolated point [26]-[27].

In this research, the VAD emotional space was constructed by using the modeling proposed in Section II. Two clustering algorithms were applied for emotion clustering and presented the ANEW system words in VAD space, the 3-dimensional emotion space that was potentially applied in an industrial design product evaluation by using a fuzzy inference system. Section III deployed the results and discussions, while Section IV concluded the proposed methods effectively in some applications. The framework of the research is illustrated in Fig. 1.

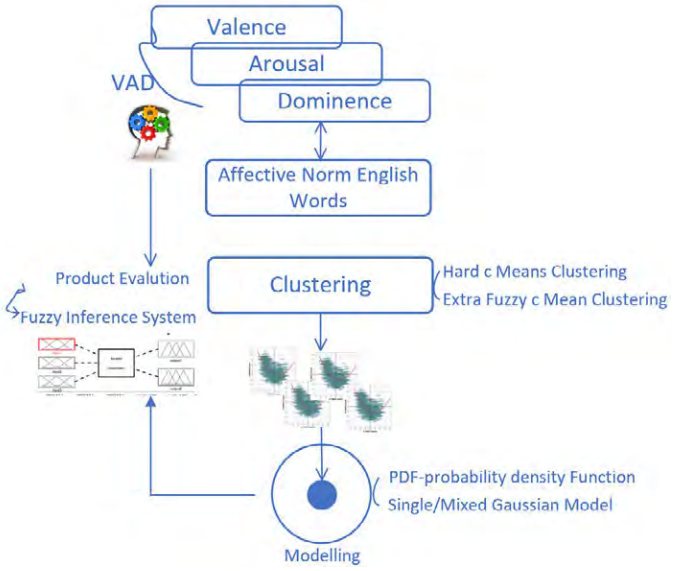


Fig. 1. The framework of VAD emotional model for evaluation by using dimensional clustering methods.

II. Modeling

The emotional evaluation by using a perceptual vocabulary is more concentrated and the emotional meta-core can be defined as a single -point set. The center is the average of the three-dimensional coordinates of the point set. The standard deviation is the domain radius. The point set distribution has a certain range, which can be calculated by its standard deviation. This kind of data distribution can set the kernel of the emotional element as an ellipsoid set indicating that the VAD identity of the emotional element is general. The VAD point set data is a distributed approximation on a plane indicating the VAD identity of the emotional element, such as a specific 25-word VAD point set topology. The content is to carry out the artificial psychological quantity labeling experiment based on the VAD emotional space, and the system design is carried out according to the method of ANEW experimental design. Due to the traditional manual labeling method, it is not conducive to large-scale data acquisition, and there are statistical difficulties. The continuous scale method (VAD emotion rating)does no longer adopt the 9-point rating system through the online survey system. VAD labeling experiments can be performed on a large scale to collect more data. The VAD labeling of the product and the VAD labeling of the relevant sensible vocabulary can establish a kind of mapping of the product feature set to the corresponding sensible vocabulary while it is not a one-to-one correspondence function.

The VAD sentimental spatial point set of each perceptual vocabulary shows that the distribution state of VAD spatial point sets of different perceptual vocabularies is different, which can be roughly divided into three categories: single point set (indicating that the emotional evaluation has a strong one), plane set (indicating that the sentiment evaluation is poorly consistent), and ellipsoid set (indicating that the sentiment evaluation has a medium consistency). Through the different spatial geometric topologies of the point set distribution, these emotions need to be defined separately. The emotional cell metamodel is a very special semantic cell model whose domain is a three-dimensional VAD emotional space. As a special semantic cell,

it is composed of two parts: the semantic kernel and the semantic shell. The semantic kernel is a set of VAD values, which represent the typical VAD values of all emotional cell elements. The emotional cell element shell represents the boundary of the field covered by the perceptual vocabulary, which is essentially uncertain, so a distance density function is used to represent this uncertainty. For the kernel of emotional cell elements, there are many forms, such as single point set, and sphere set. For the outer shell of the emotional cell elements, there are also many forms of the density function, such as the Gaussian Mixture Model (GMM).

A. Adjectives Space Construction

$\forall P \in \Omega$ is of $\Omega = \{(E_1, E_2, ..., E_n) : E_i \in P.\ i = 1, 2, ..., n$ where n is the number evaluations. Particularly, in the valence-arousal-dominent dimentinality space, P is described as $P = (Valence, Arousal, Dominance)$ and simplified as $P = (v, a, d)$, where $\Omega = \{(v, a, d) \mid v, a, d \in R\}$. A given metric $d = \|\cdot\|$ in VAD dimentionality can be defined as [26]:

$$d(P, P) = \|P\| = \sqrt{(v^2 + a^2 + d^2)} \tag{1}$$

$\forall P, Q \in \Omega$, we have that,

$$d(P \pm Q) = \sqrt{((v_P \pm v_Q)^2 + (a_P \pm a_Q)^2 + (d_P \pm d_Q)^2)} \tag{2}$$

Besides, $\forall \alpha, \beta \in R, P, Q \in \Omega$, d can be given by:

$$d(\alpha P \pm \beta Q) = \sqrt{(\alpha v_P \pm \beta v_Q)^2 + (\alpha a_P \pm \beta a_Q)^2 + (\alpha d_P \pm \beta d_Q)^2} \tag{3}$$

Thus, the following expression is applied:

$$d(\alpha P + \beta Q) \leq |\alpha| \cdot d(P) + |\beta| \cdot |d(Q) \tag{4}$$

Definition 1: $\forall P \in \Omega$, there exists a neighbor of P which is defined as:

$$N_P^{\varepsilon} = \{X \mid \|P - X\| < \varepsilon, X \in \Omega\} \tag{5}$$

Definition 2: For adjectives K, if the VAD values belong to a single point kernel, then the kernel is defined by:

$$\{P_K \mid P_K = \frac{1}{\|K\|} \sum_{P_i \in K} P_i(\rho(P_i))\} \tag{6}$$

where, $\rho(P_i)$ is the probability density function (PDF) of P_i.

Definition 3: The sphere kernel is defined as:

$$\{P_j \mid P_j \in N_{P_K}^{\varepsilon}\} \tag{7}$$

where, $P_K = \frac{1}{\|K'\|} \sum_{P_i \in K'} P_i(\rho(Pi)), K' \subset K$,

and $K' = \{P_i \mid \rho(P_j) \leq \rho_T\}$, ρ_T is a given constant to limit the size of the kernel.

Definition 4: The plain kernel is defined as a union of sphere kernels $\{\cup_i P_k\}$, where P_K is subject to Definition 2 and 3.

Definition 5: The border of the kernel is defined by upper and lower sets, which are respectively given by:

$$UP_B = \{P_l \mid P_l \in N_{P_K}^{\varepsilon_u}\} \tag{8}$$

$$LP_B = \{P_l \mid P_l \in N_{P_K}^{\varepsilon_l}\} \tag{9}$$

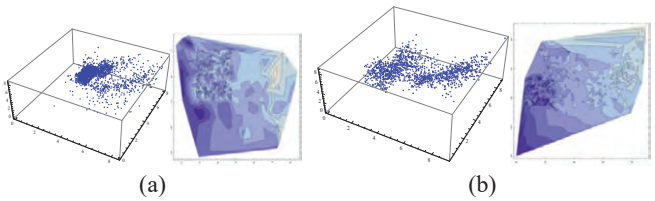
Then, the border is given by:

$$P_B = UP_B \setminus LP_B \tag{10}$$

B. Metric Function Acquisition

1. Linear Based Function

The density calculation of a point set is a relatively complicated process, and a linear function can be used to simplify the calculation. As the constant function $\rho_i(x) = a_i$, it is clear that when the density function is linear, it will reflect the uniform distribution state of the point set. Fig. 2 shows the grouping of the adjective, where (a) control, and (b) modern in the perceptual space of emotions in the VAD model.



(a) (b)

Fig. 2. The density distribution (middle) and contour map (right) of the linear density function of the VAD spatial point, where (a) K=control, and (b) K=modern.

2. Gaussian Based Metric

In the point set topology type of the VAD emotion space, both the single point set and the spheroid set are considered to be approximately spherical. In this case, the Simple Gaussian Model (SGM) is used to describe the probability density of these points, which is defined as follows:

$$\rho(P; \mu, \Delta) = \frac{1}{\sqrt{(2\pi)^3 |\Delta|}} e^{-\frac{1}{2}(P-\mu)^T \Delta^{-1} (P-\mu)} \tag{11}$$

where Δ is the covariance matrix, and μ is the center point of the density function. The characteristics of the density function are determined by (Δ, μ). Then, to achieve the best description of the point set to feature, the parameter p of (Δ, μ) should be estimated. For any $P_i \in \Omega$ of the VAD space, the density probability is $\rho(P_i, \mu, \Delta)$, for any adjective K, each point $P_i \in K$ is independent, then probability density of K can be calculated by:

$$\rho_K = \rho(K; \mu, \Delta) = \prod_i \rho(P_i; \mu, \Delta) \tag{12}$$

By using the maximum likelihood estimation (MLE), the estimating parameter pairs of (Δ, μ) to be applied for the maximization are calculated using the following formula, in which O is an estimation on (μ, Δ) [26] :

$$\begin{aligned} O(\mu, \Delta) &= \ln(\prod_i \rho(P_i; \mu, \Delta)) \\ &= \sum_i \ln(\rho(P_i; \mu, \Delta)) \\ &= \sum_i [-\frac{3}{2}\ln(2\pi) - \frac{1}{2}\ln|\Delta| + \frac{1}{2}(P_i - \mu)^T \Delta^{-1} (P_i - \mu) \\ &= -\frac{3n}{2}\ln(2\pi) - \frac{n}{2}\ln|\Delta| - \frac{1}{2}\sum_i [P_i - \mu)^T \Delta^{-1} (P_i - \mu)] \end{aligned} \tag{13}$$

To get μ, continuously, equate the differentiation of Eq(13) by the variable μ by 0 as follows:

$$\begin{aligned} \partial_\mu (O(\mu, \Delta)) &= -\frac{1}{2}\sum_i [-2\Delta^{-1}(P_i - \mu)] \\ &= \Delta^{-1}\sum_i [(P_i - \mu)] \\ &= \Delta^{-1}\sum_i P_i - n\mu = 0 \end{aligned} \tag{14}$$

Also, the estimated μ and Δ are given by:,

$$\hat{\mu} = \frac{1}{n}\sum_i P_i \tag{15}$$

$$\hat{\Delta} = \frac{1}{n-1}\sum_i (P_i - \hat{\mu})(P_i - \hat{\mu})^T \tag{16}$$

Any density of the point in the kernel can be presented formally as $\rho(P; \hat{\mu}, \hat{\Delta})$. The definition of VAD and the center point's parameter estimation μ can be rewritten as:

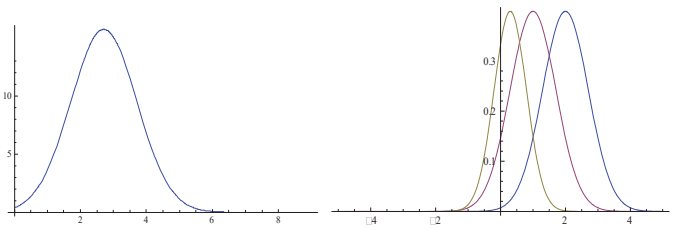
$$\hat{\mu} = (\frac{1}{n}\sum_i v_i, \frac{1}{n}\sum_i a_i, \frac{1}{n}\sum_i d_i) \tag{17}$$

The parameter estimation of covariance $\hat{\Delta}$ can be rewritten as:

$$\hat{\Delta} = \frac{1}{n-1}\sum_i [v_i - \hat{\mu}_1, a_i - \hat{\mu}_2, d_i - \hat{\mu}_3] \begin{bmatrix} v_i - \hat{\mu}_1 \\ a_i - \hat{\mu}_2 \\ d_i - \hat{\mu}_3 \end{bmatrix}$$

$$= \frac{1}{n-1}\sum_i [(v_i - \hat{\mu}_1)^2 + (a_i - \hat{\mu}_2)^2 + (d_i - \hat{\mu}_3)^2] \tag{18}$$

where $n = 2^P$, i.e. the number of elements of the set; where the discreet scale can be used if the VAD is in one-dimensional space, where the mode visualization is shown in Fig. 3.



(a) (b)

Fig. 3. One dimensional affective SGM model.

By using K=Elegant in the VAD dataset, after SGM, we have μ = (2.71, 4.79, 5.75), Δ = 2.07*I*, which is visualized in Fig. 4, where

$$\rho(P_{Elegant}) = (2\pi)^{-\frac{3}{2}} e^{-\frac{1}{2}((x-2.71)^2+(y-4.79)^2)} \tag{19}$$

$$\rho(P_{Elegant}) = (2\pi)^{-\frac{3}{2}} e^{-\frac{1}{2}((x-1.70)^2+(y-2.45)^2)} \tag{20}$$

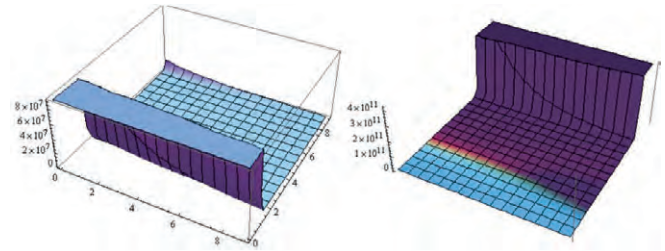


Fig. 4. SGM model in VAD space.

For "*P* is elegant" in VAD is visualized in Fig. 5, where

$$\rho(P_{Elegant}) = (2\pi)^{-\frac{3}{2}} e^{-\frac{1}{2}((x-2.71)^2+(y-4.79)^2+(z-5.75))^2} \tag{21}$$

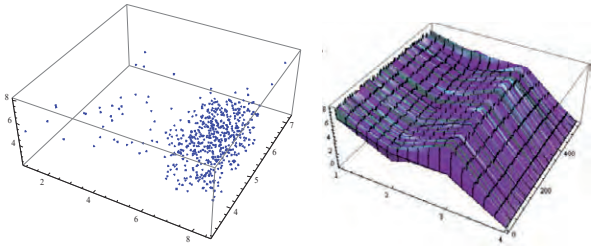
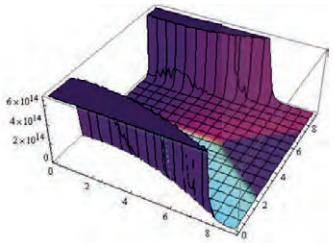


Fig. 5. VAD effective space by using the metric of θ = (2.71, 4.79, 5.75, *I*) and *n* =568.

The kernel of the plane set form is formed by the VAD emotional space, which considered that the recognition of the corresponding emotional vocabulary is low. Thus, the GMM describes the characteristics of the VAD emotional space efficiently. To describe the density value distribution corresponding to the equivalent dimension value, the plot of the Valence Arousal dimension is shown in Fig. 6.



$$\rho(P) = (2\pi)^{-\frac{3}{2}}(e^{-\frac{1}{2}((x-1.71)^2+(y-2.79)^2)} + e^{-\frac{1}{2}((x-3.05)^2+(y-5.25)^2)} + e^{-\frac{1}{2}((x-1.05)^2+(y-7.25)^2)}).$$

Fig. 6. The VAD of the three kernels combination in the SGM distribution.

3. The Density Function of the Single Dimension Affective

For the VAD emotional space, it is not yet possible to accurately obtain the correlation between the three-dimensional emotional data. However, from the medical observation of the brain magnetic resonance imaging data, there is a certain correlation between VAD. Traditional perceptual engineering uses single-dimensional emotional rating data. When the data point set is on the VAD axis, it means that a certain perceptual semantics presents a one-dimensional emotion. For theoretical completeness, this section gives a method for estimating the density of single-dimensional emotions in a single point set [28]-[32]. The point set of the coordinate axes (*v*, 0, 0), (0, a, 0), and (0, 0, *d*) can be used to estimate the three- dimensional variables separately. For the VAD point set (v_i, 0, 0) the Parzen-Borel kernel estimation is used, which is given by:

$$f_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K(\frac{x - v_i}{h_n}) \tag{22}$$

Kernel estimation is a type of the nonparametric statistical methods to determine *K*(*u*), which is a uniform density distribution function on [−1,1]. Thus, the kernel estimate in Eq.(22) is degenerated into an average, where *K*(*u*) is usually used with the forms of $K = \frac{1}{2}$, Epanechikov kernel of $\frac{3}{4}(1-t^2), t \in [-1,1]$, $\frac{5}{16}(1-t^2)^2, \ t \in [-1,1]$, and Gaussian of $\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}$. The best distribution of the function is Gaussian as shown in Fig. 7.

Through the definition of the kernel and the outer shell of the perceptual concept, the definition of the neighborhood-based method is used for the point set of different forms. At the same time, the outer layer of the emotional cell element is represented by the approximate set to express the characteristics of the soft film. Secondly, in the definition of the density function part, if the cell element of the VAD sentiment space has a single point set form, a single Gaussian density

function is used. Also, for the planar form, the Gaussian mixture model is used. Through the iterative method, the parameters to be determined can be estimated to describe the distribution of the point set of the VAD emotion space [33].

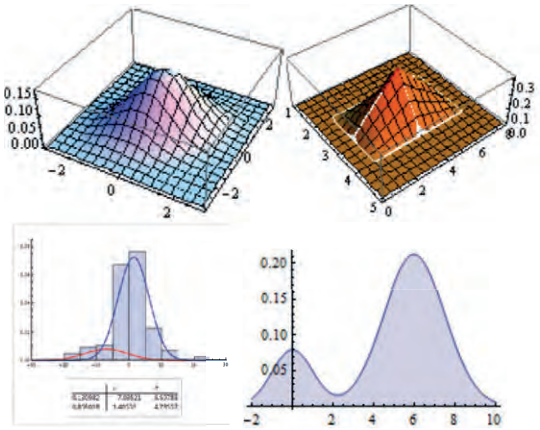


Fig. 7. An example of Gaussian distribution for the dimensional adjectives grouping model.

C. Fuzzy Perceptual Inference Using the VAD Model for Clustering Evaluation

1. Fuzzy Evaluation Process

When using fuzzy inference rules, the conditional statement "IF x is A, THEN y is B" is converted into fuzzy relations rules. The Linguistic Variable (LV) was subsequently introduced for translating the natural language to a membership of fuzzy set as shown in Fig. 8.

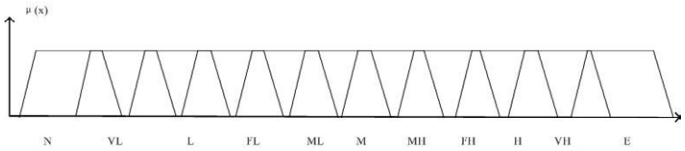


Fig. 8. Linguistic variables and their membership functions.

The fuzzy sets have been studied following fuzzy systems development rapidly. The selection of implication operators and fuzzy reasoning is closely related to the effect associated to the triangular norms. Besides, the implication operators study the fuzzy reasoning and fuzzy logic combined [34]-[35]. Therefore, the purpose is to accompany each other based on the triangular norms and implication operators to establish a new form of fuzzy propositional calculus system [36]. In the basic form of the production $P \rightarrow Q$ or "IF P THEN Q", P is a prerequisite for production(front piece), which gives the possibility to use a production prerequisite based on a logical combination to form; Q is a conclusion or operation (post-production pieces). By using the Gaussian fusion as the density functions, the Fuzzy Inference System (FIS) rules fusion operation is of two "IF-THEN" rules integration. Supposed that,

RULE: IF "x is A", THEN "y is B", the assertion "x is A" satisfies the Gaussian distribution, which is given by:

$$f(P,A,\sigma)=\frac{1}{\sqrt{2\pi}\sigma_A}e^{-\frac{1}{2}[\frac{(P-A)^2}{\sigma_A{}^2}]} \tag{23}$$

Such as by Mamndini (Fig. 9), it is found that:

$$Mamndini_{FIS}=Min(\frac{1}{\sqrt{2\pi}\sigma_A}e^{-\frac{1}{2}[\frac{(P-A)^2}{\sigma_A{}^2}]},\frac{1}{\sqrt{2\pi}\sigma_B}e^{-\frac{1}{2}[\frac{(P-B)^2}{\sigma_B{}^2}]}) \tag{24}$$

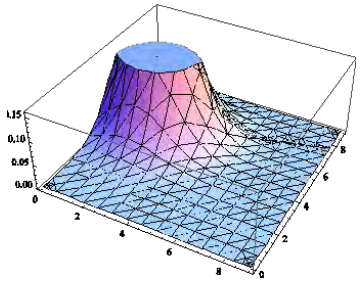


Fig. 9. IF-THEN rule by Mamndini Gaussian implication.

2. FIS for Fusion and Decision Making

Suppose that the m knowledge rules in v_k under Gaussian model (see formula (12)) conclude a particular assertion at the δ_k level, thus, we have that:

IF x_1 is v_1 and x_2 is v_2 and ..., and x_m is v_m THEN I_O is V_O, ϖ (25)

where ω is the result, this can be simplified to the following rules:

IF v_1 and v_2 and, ..., and v_m THEN ϖ (v_1, v_2, ..., v_m) (26)

IF x_1 and x_2 and, ..., and x_m THEN $x_1 \wedge x_2 \wedge ... x_m$ (27)

IF d_1 and d_2 and, ..., and d_m THEN $d_1 \wedge d_2 \wedge ... d_m$ (28)

IF ρ_1 and ρ_2 and, ..., and ρ_m THEN ϖ (ρ_1, ρ_2, ..., ρ_m) (29)

Consider the density function δ_k in FIS and by letting $\Delta=[\delta_1, \delta_2, ..., \delta_n]$, it is found that:

IF Δ THEN ϖ (Δ) (30)

Since ϖ (Δ) is a Gaussian density function, the rule is re-labeled as "IF X THEN $f(X)$", and for this rule set, we have that:

R_i: IF X THEN $f_i(X)$ (31)

However, as $f(X)$ is a nonlinear function, it is difficult to find its minimum point under the Mamdani model, so we need to linearize $f(X)$ and use the nonlinear conjugate gradient algorithm to optimize the parameters of $f(X)$. To conduct sub-dataset indexing by some indices using fuzzy transformation, it is necessary to use the fuzzy factor analyst on a matrix type dataset [37].

3. Proposed Model for Affective Clustering

Basic hard c-means: First, the initial cluster centers are given, and all elements are assigned to each cluster according to the closest assignment principle to the cluster center. Afterward, to solve the new cluster center (element centroid) for each cluster, these steps are repeated until the centroid is no longer significantly changed, then the clustering is completed. The distance used depends on the nature of the data or project requirements; and the classification of distance can refer to the A-star algorithm overview and the Manhattan distance, diagonal distance and Euclidean distance can be considered. It is equivalent to solve the minimum problem of a global state function, which is the sum of the distances of each element to the nearest cluster center. The characteristics of this algorithm are, firstly, it does not necessary obtain a global optimal solution; while the initial cluster center does not meet the requirements, only locally optimal solution may be obtained [38]. For globally optimal solution, the algorithm changes the method name to k-means; secondly, the influence of noise points on clustering cannot be ruled out. Thirdly, the cluster shape is required to be nearly circular. The algorithm can be described as follows [39]:

Algorithm 1: Hard c-means
Required: Number of clusters, epsilon [threshold]
Outputs: Clusters
a. Set cluster number's value k.
b. Choose the k cluster center randomly
c. **WHILE** center is not empty DO
    Compute the mean or center of each cluster
    Compute the distance of pixels and cluster's center
    IF |distance – center| < epsilon THEN
      Move points to the cluster.
    ELSE:
      Move points to the next cluster
    ENDIF
    Re-estimate the center
  ENDWHILE
e. Output clusters

Extra fuzzy-c-mean: EFCM is a clustering algorithm based on a fuzzy number evaluation system that is also an improvement of the fuzzy c- mean (FCM) algorithm. The EFCM clustering model includes initialization, loop and output; firstly, the method is used to determine the initial clustering center to ensure the optimal solution, while the second step is to determine the degree of membership $U(i, j)$ of each point to each cluster center. In which, $C^{(k)}$ is a weighted index. Thirdly, the system needs to determine the new cluster center and mark the change track of the cluster center to determine whether the changing amplitude of the cluster center is less than the given error limit. If not, it returns to the previous steps, otherwise it exits the loop. Finally, the system outputs the cluster center trajectory and clustering results. The characteristics of this algorithm are similar to ordinary k-means clustering. Full clustering is required, and noisy points cannot be distinguished. The center of clustering is more consistent, but the calculation efficiency is relatively low. The concepts of smoothing parameters and membership are adopted so that the points are not directly attached to a single cluster center. Algorithm 2 shows the process of the model [40]-[43].

Algorithm 2: Extra fuzzy c means clustering
Require: Epsilon [threshold], X [points]
Outputs: Centers, clusters
a. Initializing a segment matrix U
b. Compute $C^{(k)} = [c_j]$ with $U^{(k)}$

$$c(j) = \sum_{m=1}^{n} u(i,j)^m \bullet p(j) \Big/ \sum_{m=1}^{n} u(i,j)^m$$

c. Update $U^{(k)}$ and $U^{(k+1)}$

$$u(i,j) = \left(\sum_{k=1}^{m-1} | p(j) - c(k) | \Big/ | p(j) - c(i) | \right)^{-\frac{2}{m-1}}$$

d. **IF** $|U^{(k+1)} - U^{(k)}|$ < epsilon THEN
    STOP
  ELSE:
    GOTO Step b
  ENDIF
e. Output clusters

VAD-dimensional clustering: The Valence-Arousal-Dominance dimensional Clustering (VADdC) model was deployed based on hard c-mean and extra fuzzy c-mean. Each element in the initial state is a cluster, and the cluster with the smallest distance between clusters is merged each time until the number of clusters meets the requirements or merges more than 90%. Similar to the Huffman tree algorithm, the union set also needs to be checked. For normal c-mean algorithms (HCM or EFCM), there are also several classifications of the distance definitions: including the minimum distance between cluster elements, the maximum distance between cluster elements, and the centroid distance of clusters. The characteristics of this algorithm are as follows.

a) The storage space consumed by the agglomeration clustering is higher than several other methods. The interference of the noise points can be eliminated, noise points may be divided into a cluster. Suitable for cases with irregular shapes and when complete clustering is not required.
b) The merging operation must have a merging limit ratio, otherwise excessive merging may cause all classification centers to aggregate, causing clustering failure.

Similar to the decision trees, the advantage of hierarchical clustering is that the entire tree can be obtained at one time, and the control of certain conditions, whether depth or width, is controllable. Several problems may occur, such as calculation on the division is determined and cannot be changed; and the cohesion/divisions combined is not "optimal" every time. However, the proposed VADdC algorithm is easy for local optimization and can be performed by appropriate random operations. It uses balanced iterative reducing and clustering. It first divides neighboring sample points into micro-clusters and then uses the c-means algorithm for these micro-clusters (**Algorithm** 3).

Algorithm 3: VADdC-valence arousal dominance dimensional clustering
Required: POINTS, points with index group, cluster numbers
Outputs: Groups, POINTS (clusters, centers)
a. POINTS ← [v, a, d] ## for each point in required VAD space
b. Number of Groups ← index ## from 1 to index in range of length of POINTS
c. **FOR** index, point in POINTS:
    IF type of point is required THEN
      FOR index of the point in POINTS of group index
        ITERATE number of groups with point group
        DELETE number of groups with points of
            index of point of the group
        MERGE the groups with points and POINTS with
            index of point
      ENDFOR
    ELSE IF type of point is not required:
      CALCULATE core point of group index with type
          of POINTS
      COUNT the number of groups with core point
          group index
      DELETE the number of groups with point group
      The group of point ← core point with group index
    ENDIF
    COUNT number of groups with sorted iterations using
        key ← lambda
    FOR key in several groups:
      INCREASE c ## c=c+1
      FOR point in POINTS:
        IF group of the point is key*c THEN
          CONTINUE
        ENDIF
      ENDFOR
      IF c >= cluster number THEN
        BREAK ## exit the loop
      ENDIF
    ENDFOR
  EDNFOR
d. Outputs POINTS, group ## clusters and centers

III. Results and Discussion

A. Questionnaire System and Data Acquisition

We developed a questionnaire system for data acquisition to be applied for evaluation. Chair styles are selected from an online shopping website, and the style has a list with values: glam, farmhouse, traditional, eclectic, bohemian, modern, global-inspired, coastal, American traditional, ornate glam, beachy, posh luxe, modern farmhouse, rustic, French country, industrial, ornate traditional, modern rustic, Scandinavian, sleek chic modern, mid-central modern, Asian-inspired, bold eclectic modern, cabin lodge, cottage Americana, nautical, tropical, and Victorian. The questionnaire system for the Ming-style chairs for fuzzy perceptual evaluation was proposed by using (1-9) fuzzy numbers. An ancient furniture system in industrial design fields was evaluated. The relative dataset of results has been employed for clustering by using the VADdC model (shown in Table I).

TABLE I. Shape Classification and Their Sub-shape Code in the Questionnaire System

Shape Category	**Sub-shape**			
Top Rail (C1)	C11	C12	C13	C14
	Scroll	Luoguo	Bow	Round
Seat-Back (C2)	C21	C22	C23	C24
	Screen	Comb-Teeth	Relief	Slatted
Armrest (C3)	C31	C32	C33	C34
	Square	Round	Openwork	Relief
Foot (C4)	C41	C42	C43	C44
	Carved Foot	Horseshoe	Square	Circle-Center

B. Fuzzy Perceptual Evaluation using VADdC

The adjective evaluation data is illustrated in Fig. 10 and 11, which show the result after grouping by using the VADdC algorithm. Fuzzy perceptual evaluation by using fuzzy number [1,3,5,7,9] is applied on the shape of the Ming style chairs. Performance by using a fuzzy perceptual evaluation system is designed using 3 inputs, and 2 outputs basic Mamdani inference model based on the rule system as demonstrated in Fig. 12.

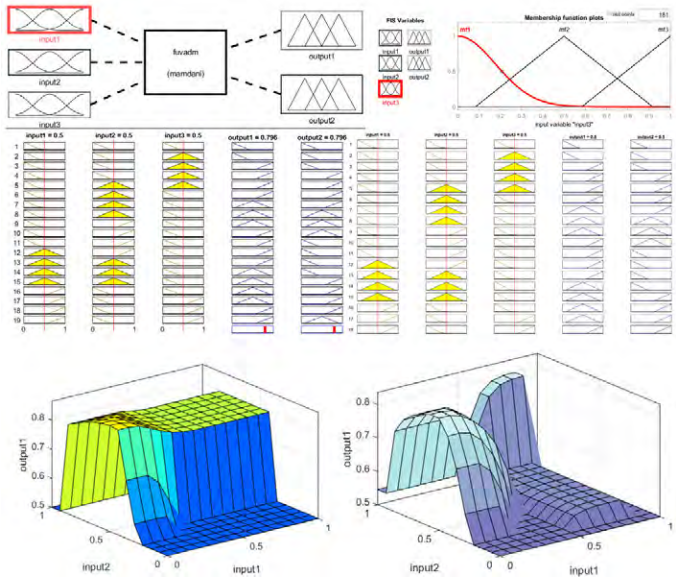


Fig. 12. 3-inputs 2-outputs basic fuzzy Mamdani inference model and rules.

Adjectives	functional	free	friendly	fresh	neat	individuality	artistic	traditional	gum texture	valuable	modern	classic	elegant	unique	technical	professional	arrogant	confident	fun	future	authoritative	strong	reliable
	k1	k2	k3	k4	k5	k6	k7	k8	k9	k10	k11	k12	k13	k14	k15	k16	k17	k18	k19	k20	k21	k22	k23
functional	k1																						
free	k2	3																					
friendly	k3	5	5																				
fresh	k4	3	5	1																			
neat	k5	3	1	1	3																		
individuality	k6	7	5	3	3	9																	
artistic	k7	1	5	3	3	3	3																
traditional	k8	5	1	3	2	3	7	1															
gum texture	k9	3	7	3	3	1	9	3	3														
valuable	k10	1	5	3	3	1	1	3	3	3													
modern	k11	9	5	3	1	3	3	5	1	3	3												
classic	k12	3	5	1	3	3	9	3	3	3	1	1											
elegant	k13	1	1	5	3	3	3	3	3	5	3	1	1										
unique	k14	5	5	0	1	3	1	3	5	3	5	3	3	5									
technical	k15	3	1	7	3	3	3	1	3	1	3	3	9	3	3								
professional	k16	1	5	7	1	3	9	3	5	1	3	1	7	5	5	3							
arrogant	k17	1	3	7	1	1	3	1	3	1	1	3	7	1	9	1	7						
confident	k18	1	3	3	1	3	5	1	1	1	3	5	3	3	5	1	1	0					
fun	k19	1	1	5	1	1	1	3	2	3	3	3	3	1	3	3	3	3	5				
future	k20	1	3	9	1	3	3	1	3	7	1	1	1	1	1	9	7	0	1	5			
authoritative	k21	1	1	3	1	5	3	1	3	3	3	5	3	5	3	3	1	0	3	5	5		
strong	k22	5	1	1	3	3	1	1	5	7	1	1	5	1	3	1	7	1	1	1	0	5	
reliable	k23	1	1	1	1	3	3	1	1	1	1	1	3	3	1	1	3	1	1	1	1	7	1

Fig. 10. Adjectives fuzzy perceptual evaluation before grouping.

Adjectives	functional	friendly	technical	reliable	valuable	confident	arrogant	traditional	individuality	professional	authoritative	free	fresh	neat	artistic	gum texture	modern	classic	elegant	unique	fun	future	strong
	k1	k3	k15	k23	k10	k18	k17	k8	k6	k16	k21	k2	k4	k5	k7	k9	k11	k12	k13	k14	k19	k20	k22
functional	k1																						
friendly	k3	0.694																					
technical	k15	0.688	0.576																				
reliable	k23	0.699	0.701	0.723																			
valuable	k10	0.678	0.665	0.675	0.596																		
confident	k18	0	0	0	0	0																	
arrogant	k17	0	0	0	0	0	0																
traditional	k8	0	0	0	0	0	0	0.588															
individuality	k6	0	0	0	0	0	0	0.678	0.665														
professional	k16	0	0	0	0	0	0	0.824	0.871	0.842													
authoritative	k21	0	0	0	0	0	0	0.794	0.814	0.796	0.788												
free	k2	0	0	0	0	0	0	0	0	0	0	0											
fresh	k4	0	0	0	0	0	0	0	0	0	0	0	0										
neat	k5	0	0	0	0	0	0	0	0	0	0	0	0	0									
artistic	k7	0	0	0	0	0	0	0	0	0	0	0	0	0	0.911								
gum texture	k9	0	0	0	0	0	0	0	0	0	0	0	0	0	0.651	0.861							
modern	k11	0	0	0	0	0	0	0	0	0	0	0	0	0	0.947	0.914	0.921						
classic	k12	0	0	0	0	0	0	0	0	0	0	0	0	0	0.896	0.852	0.451	0.963					
elegant	k13	0	0	0	0	0	0	0	0	0	0	0	0	0	0.796	0.863	0.952	0.971	0.992				
unique	k14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
fun	k19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.991		
future	k20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.985	0.841	
strong	k22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.963	1	0.982

Fig. 11. Grouping adjectives by using clustering operators.

TABLE II. SELECTED ADJECTIVES GROUPING AND THEIR VALUE STANDARD DEVIATION IN VAD DIMENSIONAL SPACE

Adjectives	V	V-SD	A	A-SD	D	D-SD	P
Functional	4.65	2.22	5.43	2.10	3.98	1.23	single
Free	5.22	2.54	3.78	3.02	7.32	2.43	plain
Easy To Use	8.42	1.12	5.33	2.50	5.32	2.13	plain
Fresh	7.24	1.09	5.35	2.13	4.97	1.43	sphere
Neat	3.56	3.02	7.23	2.21	3.29	1.67	single
Individuality	5.77	2.24	5.52	1.75	6.76	1.90	plain
Artistic	4.33	2.87	4.39	2.45	6.54	3.87	plain
Traditional	2.98	2.43	8.43	1.65	5.33	3.29	single
Gum Texture	4.39	2.56	7.32	2.08	8.12	1.98	sphere
Valuable	8.22	3.10	5.44	2.34	1.56	2.32	sphere
Modern	7.34	1.65	6.32	3.89	1.75	2.21	plain
Classic	5.55	2.04	5.52	1.65	6.76	3.01	sphere
Elegant	7.65	1.13	4.39	2.55	6.13	2.01	sphere
Unique	7.66	2.04	5.52	2.18	2.55	2.29	plain
Technical	4.56	2.04	5.52	2.09	1.65	2.29	plain
Professional	5.76	2.04	5.52	2.72	4.57	2.29	plain
Arrogant	5.15	1.68	5.83	2.33	5.59	2.40	sphere
Confident	6.68	1.29	6.22	2.41	7.68	1.94	single
Fun	8.12	1.11	7.22	2.01	6.80	1.85	single
Future	7.77	2.04	5.52	2.72	6.76	2.29	sphere
Authoritative	6.03	2.09	5.76	2.04	6.98	2.20	single
Strong	7.58	2.04	5.52	2.72	6.76	2.29	plain
Reliable	6.89	1.87	4.90	2.33	5.98	3.20	single

Table II shows the norms with VAD values from affective norm English Words (ANEW). Table III shows the comparing results of the VADdC model with FCM, K-mean, self-organization mapping network clustering (SOM), HCM, and EFCM. The VADdC performs 91.37% overall accuracy. Due to the superiority of the proposed method, different clustering and machine learning methods can be integrated with the proposed method to solve other applications as in [44]-[51].

TABLE III. COMPARING ANALYSIS OF HCM, EFCM AND VADDC

Clustering Algorithms	Fitness value	Inter-cluster Distance	Intra-cluster Distance	Elapse time	Accuracy (%)
FCM	57.60	8.36	85.36	0.2865	79.68
K-mean	47.23	7.56	76.37	0.3638	79.75
SOM	59.62	10.36	122.35	0.4212	89.56
HCM	60.89	8.566	89.32	0.3568	89.34
EFCM	68.65	8.698	158.37	0.4251	90.07
VADdC	17.53	9.210	126.36	0.4352	91.37

IV. CONCLUSION

In classical research, word vectors of features (terms, including words, words, phrases, etc.) are often used to build text vectors, and cluster analysis is performed based on the similarity between text vectors. To evaluate the clustering quality of the unsupervised clustering algorithm even in the case of the overlapping cluster centers, a VAD vector space model data was constructed for clustering analysis by studying the principle of clustering algorithms and applying clustering algorithms. For clustering a given set of objects, there can be multiple meaningful divisions due to the distance (or similarity) between objects, which has multiple implicit definitions. The present work constructed a VAD based model including distance, borders and type of centers. It also developed a rule-based inference system using fuzzy perceptual evaluation and introduced dimensional affective based VAD clustering called VADdC, taking successfully application on a dataset that has been acquired from an online questionnaire system.

The comparing analysis reported that the performance of the proposed method is better than others in the fitness of 17.53, elapse time of 0.4352 and 91.37% accuracy finally. In future work, high-dimensional data can be involved, where the data distribution in a high-dimensional space may be very sparse, and highly skewed. In practical applications, it may be necessary to perform clustering under various conditions. Data grouping with clustering characteristics is very challenging. The most difficult question here is how to identify the "specific constraints" implicit in the problem we are trying to solve, and what algorithms should be used to best "fit" these constraints.

ACKNOWLEDGMENT

This work is supported by Science Foundation of Ministry of Education of China (Grant No:18YJC760099) and Zhejiang Provincial Key Laboratory of Integration of Healthy Smart Kitchen System (Grant No: 19080049-N)

REFERENCES

[1] Dey, N., Babo, R., Ashour, A. S., Bhatnagar, V., & Bouhlel, M. S. (Eds.). "Social networks science: Design, implementation, security, and challenges: From social networks analysis to social networks intelligence", *Springer*. doi:10.1007/978-3-319-90059-9_1.

[2] Chiu, M.-C. and K.-Z. Lin. "Utilizing text mining and Kansei Engineering to support data-driven design automation at conceptual design stage", *Advanced Engineering Informatics* vol.38, pp.826-839, 2018, doi:10.1016/j.aei.2018.11.002.

[3] Ahmed, S. S., Dey, N., Ashour, A. S., et al. "Effect of fuzzy partitioning in Crohn's disease classification: a neuro-fuzzy-based approach", *Medical &*

Biological Engineering & Computing, vol.55, no 1, pp.101-115, 2017, doi: 10.1007/s11517-016-1508-7

[4] Han, Z., et al. "Clustering and retrieval of mechanical CAD assembly models based on multi-source attributes information", *Robotics and Computer-Integrated Manufacturing*, vol.58, pp.220-229, 2019, doi:10.1016/j.rcim.2019.01.003

[5] Lacko, D., et al. "Product sizing with 3D anthropometry and k-medoids clustering", *Computer-Aided Design*, vol. 91, pp. 60-74, 2017, doi: 10.1016/j.cad.2017.06.004

[6] Li, W., et al. "A Comprehensive Approach for the Clustering of Similar-Performance Cells for the Design of a Lithium-Ion Battery Module for Electric Vehicles", *Engineering* vol. 5, no. 4, pp. 795-802, 2019, doi: 10.1016/j.eng.2019.07.005

[7] Franco, M. A. "A system dynamics approach to product design and business model strategies for the circular economy", *Journal of Cleaner Production,* vol. 241, 118327, 2019, doi:10.1016/j.jclepro.2019.118327

[8] Chick, C. F. "Cooperative versus competitive influences of emotion and cognition on decision making: A primer for psychiatry research", *Psychiatry Research,* vol. 273, pp. 493-500, 2019, doi: 10.1016/j.psychres.2019.01.048.

[9] Mhetre, N. A., Deshpande, A. V., & Mahalle, P. N. "Trust management model based on fuzzy approach for ubiquitous computing", *International Journal of Ambient Computing and Intelligence*, vol. 7, no.2, pp. 33-46, 2016, doi: 10.4018/978-1-5225-9866-4.ch022.

[10] Pleyers, G. "Shape congruence in product design: Impacts on automatically activated attitudes", *Journal of Retailing and Consumer Services*, 101935, 2019, doi: 10.1016/j.jretconser.2019.101935.

[11] Balakrishnan, J., et al. "Product recommendation algorithms in the age of omnichannel retailing -An intuitive clustering approach", *Computers & Industrial Engineering,* vol. 115, pp. 459-470, 2017, doi: 10.1016/j.cie.2017.12.005.

[12] Büyüközkan, G. and M. Güler. "Smart watch evaluation with integrated hesitant fuzzy linguistic SAW-ARAS technique", *Measurement* vol. 153, 107353, 2019, doi: 10.1016/j.measurement.2019.107353.

[13] Lipor, J. and L. Balzano, "Clustering quality metrics for subspace clustering", *Pattern Recognition*, 107328, 2020, doi: 10.1016/j.patcog.2020.107328.

[14] Pandey, M. M. "Evaluating the strategic design parameters of airports in Thailand to meet service expectations of Low-Cost Airlines using the Fuzzy-based QFD method", *Journal of Air Transport Management,* vol. 82, 101738, 2019, doi: 10.1016/j.jairtraman.2019.101738.

[15] Qiao, W., et al. "A methodology to evaluate human factors contributed to maritime accident by mapping fuzzy FT into ANN based on HFACS", *Ocean Engineering,* vol. 197, 106892, 2019, doi: 10.1016/j.oceaneng.2019.106892.

[16] Hsu, C.-C., et al. "Relationship between eye fixation patterns and Kansei evaluation of 3D chair forms", *Displays,* vol. 50, pp. 21-34, 2017 doi: 10.1016/j.displa.2017.09.002.

[17] Cambria, E., Poria, S., Hussain, A., & Liu, B. "Computational Intelligence for Affective Computing and Sentiment Analysis", *IEEE Computational Intelligence Magazine,* vol. 14, no. 2, pp. 16-17, 2019, doi: 10.1109/MCI.2019.2901082.

[18] Gonzalez CB, García-Nieto J, Navas-Delgado I, Aldana-Montes JF, "A Fine Grain Sentiment Analysis with Semantics in Tweets", *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 3, no.6 pp. 22-28, 2016, doi: 10.9781/ijimai.2016.363.

[19] Kossaifi, J., et al. "AFEW-VA database for valence and arousal estimation in-the-wild", *Image and Vision Computing,* vol. 65, pp.23-36, 2017, doi: 10.1016/j.imavis.2017.02.001.

[20] Jaeger, S. R., et al. "Valence, arousal and sentiment meanings of 33 facial emoji: Insights for the use of emoji in consumer research", *Food Research International,* vol. 119, pp. 895-907, 2018, doi: 10.1016/j.foodres.2018.10.074.

[21] Nalepa, G. J., et al. "Affective computing in ambient intelligence systems", *Future Generation Computer Systems,* vol. 92, pp. 454-457, 2018, doi: 10.1016/j.future.2018.11.016.

[22] Trabelsi, I., Bouhlel, M. S., & Dey, N. "Discrete and continuous emotion recognition using sequence kernels", *International Journal of Intelligent Engineering Informatics,* vol. *5, no.* 3 pp.194-205, 2017, doi:10.1504/IJIEI.2017.086608.

[23] Sarkar, D., Debnath, S., Kole, D. K., & Jana, P. "Influential Nodes Identification Based on Activity Behaviors and Network Structure with Personality Analysis in Egocentric Online Social Networks", *International Journal of Ambient Computing and Intelligence,* vol.10 no.4, pp.1-24, 2019, doi:10.4018/IJACI.2019100101.

[24] Pass-Lanneau, A., et al. "Perfect graphs with polynomially computable kernels", *Discrete Applied Mathematics,* vol. 272, pp. 69-74, 2018, doi: 10.1016/j.dam.2018.09.027.

[25] Person, O., et al. "Should new products look similar or different? The influence of the market environment on strategic product styling", *Design Studies,* vol. 29, no. 1, pp. 30-48, 2007, doi: 10.1016/j.destud.2007.06.005

[26] He, T., et al. "Curvature manipulation of the spectrum of Valence-Arousal-related fMRI dataset using Gaussian-shaped Fast Fourier Transform and its application to fuzzy KANSEI adjectives modeling", *Neurocomputing,* vol. 174, pp. 1049-1059. 2015, doi: 10.1016/j.neucom.2015.10.025.

[27] Lee, W. and M. D. Norman. "Affective Computing as Complex Systems Science", *Procedia Computer Science,* vol. 95, pp. 18-23, 2016, doi: 10.1016/j.procs.2016.09.288.

[28] Salgado, S. and O. S. Kingo. "How is physiological arousal related to self-reported measures of emotional intensity and valence of events and their autobiographical memories?", *Consciousness and Cognition,* vol. 75, 102811, 2019, doi: 10.1016/j.concog.2019.102811.

[29] Scult, M. A. and A. R. Hariri. "A brief introduction to the neurogenetics of cognition-emotion interactions", *Current Opinion in Behavioral Sciences,* vol. 19, pp.50-54, 2017, doi: 10.1016/j.cobeha.2017.09.014.

[30] Wang, W. M., et al. "Multiple affective attribute classification of online customer product reviews: A heuristic deep learning method for supporting Kansei engineering", *Engineering Applications of Artificial Intelligence,* vol. 85, pp. 33-45, 2019, doi: 10.1016/j.engappai.2019.05.015.

[31] Zabotto, C. N., et al. "Automatic digital mood boards to connect users and designers with kansei engineering", *International Journal of Industrial Ergonomics,* vol. 74, 102829, 2019, doi: 10.1016/j.ergon.2019.102829.

[32] Hyun, K. H. and J.-H. Lee. "Balancing homogeneity and heterogeneity in design exploration by synthesizing novel design alternatives based on genetic algorithm and strategic styling decision", *Advanced Engineering Informatics,* vol. 38, pp. 113-128, 2018, doi: 10.1016/j.aei.2018.06.005.

[33] Kratzwald, B., et al. "Deep learning for affective computing: Text-based emotion recognition in decision support", *Decision Support Systems,* vol. 115, pp. 24-35, 2018, doi: 10.1016/j.dss.2018.09.002.

[34] Vučetić, M., et al. "Fuzzy functional dependencies and linguistic interpretations employed in knowledge discovery tasks from relational databases", *Engineering Applications of Artificial Intelligence,* vol. 88, 103395, 2019, doi: 10.1016/j.engappai.2019.103395.

[35] Bagherinia, A., et al. "Reliability-Based Fuzzy Clustering Ensemble", *Fuzzy Sets and Systems, 2020,* doi: 10.1016/j.fss.2020.03.008.

[36] Chen, Z., et al. "Explore and evaluate innovative value propositions for smart product service system: A novel graphics-based rough-fuzzy DEMATEL method", *Journal of Cleaner Production,* vol. 243, 118672, 2020, doi:10.1016/j.jclepro.2019.118672.

[37] Eustáquio, F. and T. Nogueira. "Evaluating the numerical instability in fuzzy clustering validation of high-dimensional data", *Theoretical Computer Science,* vol. 805, pp. 19-36, 2019, doi: 10.1016/j.tcs.2019.10.039.

[38] Hernández-Julio, Y. F., et al. "Fuzzy clustering and dynamic tables for knowledge discovery and decision-making: Analysis of the reproductive performance of the marine copepod Cyclopina sp", *Aquaculture,* 735183, 2020, doi: 10.1016/j.aquaculture.2020.735183.

[39] Li, W., et al. "Boosted K-nearest neighbor classifiers based on fuzzy granules", *Knowledge-Based Systems* 105606, 2020, doi: 10.1016/j.knosys.2020.105606.

[40] Rani, P., et al. "A novel approach to extended fuzzy TOPSIS based on new divergence measures for renewable energy sources selection", *Journal of Cleaner Production,* vol. 257, 120352, 2020, doi: 10.1016/j.jclepro.2020.120352.

[41] García-Díaz, Vicente, Jordán Pascual Espada, Rubén González Crespo, B. Cristina Pelayo G-Bustelo, and Juan Manuel Cueva Lovelle. "An approach to improve the accuracy of probabilistic classifiers for decision support systems in sentiment analysis", *Applied Soft Computing,* vol. 67, pp. 822-833, 2018, doi: 10.1016/j.asoc.2017.05.038.

[42] Naik, Anima, Suresh Chandra Satapathy, Amira S. Ashour, and Nilanjan Dey. "Social group optimization for global optimization of multimodal functions and data clustering problems", *Neural Computing and*

Applications, vol. 30, no. 1, pp. 271-287, 2018, doi: 10.1007/s00521-016-2686-9.

[43] Hamzaoui, Y., Amnai, M., Choukri, A., & Fakhri, Y. "Novel clustering method based on K-medoids and mobility metric", *International Journal of Interactive Multimedia and Artificial Intelligence,* vol.5, no.1, pp. 29-33, 2018. doi: 10.9781/ijimai.2017.11.001.

[44] Cao, Luying, Nilanjan Dey, Amira S. Ashour, Simon Fong, R. Simon Sherratt, Lijun Wu, and Fuqian Shi. "Diabetic plantar pressure analysis using image fusion", *Multimedia Tools and Applications,* vol. 79, pp.11213-11236, 2020, doi: 10.1007/s11042-018-6269-x.

[45] Mondragon, Victor M., Vicente García-Díaz, Carlos Porcel, and Rubén González Crespo. "Adaptive contents for interactive TV guided by machine learning based on predictive sentiment analysis of data", *Soft Computing,* vol. 22, no. 8, pp. 2731-2752, 2018, doi: 10.1007/s00500-017-2530-x.

[46] Tian, Zongmei, Nilanjan Dey, Amira S. Ashour, Pamela McCauley, and Fuqian Shi. "Morphological segmenting and neighborhood pixel-based locality preserving projection on brain fMRI dataset for semantic feature extraction: an affective computing study", *Neural Computing and Applications,* vol. 30, no. 12 pp. 3733-3748, 2018, doi: 10.1007/s00521-017-2955-2.

[47] Li, Zairan, Kai Shi, Nilanjan Dey, Amira S. Ashour, Dan Wang, Valentina E. Balas, Pamela McCauley, and Fuqian Shi. "Rule-based back propagation neural networks for various precision rough set presented KANSEI knowledge prediction: a case study on shoe product form features extraction", *Neural Computing and Applications,* vol. 28, no. 3, pp. 613-630, 2017, doi:10.1007/s00521-016-2707-8.

[48] Pourvali, Mohsen, Salvatore Orlando, and Hosna Omidvarborna. "Topic models and fusion methods: a union to improve text clustering and cluster labeling", *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 5, no. 4, pp. 28-34, 2019, doi:10.9781/ijimai.2018.12.007.

[49] Shinde, Gitanjali Rahul, and Henning Olesen. "Beacon-Based Cluster Framework for Internet of People, Things, and Services (IoPTS)", *International Journal of Ambient Computing and Intelligence,* vol. 9, no. 4, pp. 15-33, 2018, doi:10.4018/IJACI.2018100102.

[50] Ali, Md Nawab Yousuf, Md Golam Sarowar, Md Lizur Rahman, Jyotismita Chaki, Nilanjan Dey, and João Manuel RS Tavares. "Adam deep learning with SOM for human sentiment classification." *International Journal of Ambient Computing and Intelligence,* vol.10, no. 3, pp. 92-116, 2019, doi:10.4018/IJACI.2019070106.

[51] Sengupta, Diganta. "Taxonomy on Ambient Computing: A Research Methodology Perspective", *International Journal of Ambient Computing and Intelligence,* vol.11, no. 1, pp.1-33, 2020, doi:10.4018/IJACI.2020010101.

Wenlin Huang

Wenlin Huang, was born in Wenzhou City, Zhejiang Province, China in 1975. In 2000, he received a bachelor's degree in fine arts education from Wenzhou Normal University and a master's degree in fine arts from Tianjin Academy of Fine Arts in 2007. In 2010, he was employed as a lecturer in the Art and Design Department of City College of Wenzhou University, and in 2019, he was employed as an Associate Professor in the Arts and Crafts Department of Wenzhou Business College. His research direction is esthetic cognition and cultural design.

Qun Wu

Qun Wu, is an Associate Professor of Human Factor at the Institute of Universal Design, Zhejiang Sci-Tech University, China. He received his Ph.D. in College of Computer Science and Technology from Zhejiang University, China, in 2008. He holds a B.E. degree in Industrial Design from Nanchang University, China, in 2001, and a M.E. degree in Mechanical Engineering from Shaanxi University of Science and Technology, China, in 2004. His research interests include machine learning, human factor and product innovation design.

Nilanjan Dey

Nilanjan Dey, was born in Kolkata, India, in 1984. He received his B.Tech. degree in Information Technology from West Bengal University of Technology in 2005, M. Tech. in Information Technology in 2011 from the same University and Ph.D. in digital image processing in 2015 from Jadavpur University, India. In 2011, he was appointed as an Asst. Professor in the Department of Information Technology at JIS College of Engineering, Kalyani, India followed by Bengal College of Engineering College, Durgapur, India in 2014. He is now employed as an Asst. Professor in Department of Information Technology, Techno India College of Technology, India. His research topic is signal processing, machine learning, and information security. Dr. Dey is an Associate Editor of IEEE Access and is currently the Editor-in-Chief of the International Journal of Ambient Computing and Intelligence, and Series co-editor of Springer Tracts of Nature-Inspired Computing (STNIC).

Amira S. Ashour

He received her B.S. degree in Electrical Engineering (Electronics and Electrical Communications Engineering 'EEC') from Faculty of Engineering, Tanta University, Egypt in 1997, M.Sc. in Image Processing for Nondestructive Evaluation Performance, EEC Department, Faculty of Engineering, Egypt in 2000, and Ph.D. in Smart Antenna (Direction of arrival estimation using local polynomial approximation) from Faculty of Engineering, Tanta University, Egypt in 2005. Dr. Ashour is currently an Assistant Professor and Head of EEC Department, Faculty of Engineering, Tanta University, Egypt, since 2016. She is a member in the Research and Development Unit, Faculty of Engineering, Tanta University, Egypt. She is the ICT manager of Huawei-Tanta Academy, Tanta University. She was the Vice-chair of Computer Engineering Department, Computers and Information Technology College (CIT), Taif University, KSA for one year from 2015. She was the Vice-chair of Computer Science Department, CIT College, Taif University, KSA for 5 years till 2015. Her research interests include image processing and analysis, medical imaging, computer-aided diagnosis, signal/image/video processing, machine learning, smart antenna, direction of arrival estimation, targets tracking, inverse problems, optimization, and neutrosophic theory. She is series Co-Editor of Advances in Ubiquitous Sensing Applications for Healthcare series, Elsevier.

Simon James Fong

Simon Fong graduated from La Trobe University, Australia, with a 1st Class Honours BEng. Computer Systems degree and a PhD. Computer Science degree in 1993 and 1998 respectively. Simon is now working as an Associate Professor at the Computer and Information Science Department of the University of Macau, as an Adjunct Professor at Faculty of Informatics, Durban University of Technology, South Africa. He is a co-founder of the Data Analytics and Collaborative Computing Research Group in the Faculty of Science and Technology. Prior to his academic career, Simon took up various managerial and technical posts, such as systems engineer, IT consultant and e-commerce director in Australia and Asia. Dr. Fong has published over 450 international conference and peer-reviewed journal papers, mostly in the areas of data mining, data stream mining, big data analytics, meta-heuristics optimization algorithms, and their applications. He serves on the editorial boards of the Journal of Network and Computer Applications of Elsevier, IEEE IT Professional Magazine, and various special issues of SCIE-indexed journals. Simon is also an active researcher with leading positions such as Vice-chair of IEEE Computational Intelligence Society (CIS) Task Force on "Business Intelligence & Knowledge Management", TC Chair of IEEE ComSoc e-Health SIG and Vice-director of International Consortium for Optimization and Modelling in Science and Industry (iCOMSI).

Rubén González Crespo

Dr. Rubén González Crespo has a PhD in Computer Science Engineering. Currently he is Vice Chancellor of Academic Affairs and Faculty from UNIR and Global Director of Engineering Schools from PROEDUCA Group. He is advisory board member for the Ministry of Education at Colombia and evaluator from the National Agency for Quality Evaluation and Accreditation of Spain (ANECA). He is member from different committees at ISO Organization. Finally, He has published more than 200 papers in indexed journals and congresses.

An Experimental Study on Microarray Expression Data from Plants under Salt Stress by using Clustering Methods

Houda Fyad*, Fatiha Barigou, Karim Bouamrane

Laboratoire d'informatique d'Oran (LIO), Département d'informatique, Université Oran1 Ahmed Benbella (Algeria)

Received 9 June 2019 | Accepted 13 February 2020 | Published 27 May 2020

Abstract

Current Genome-wide advancements in Gene chips technology provide in the "Omics (genomics, proteomics and transcriptomics) research", an opportunity to analyze the expression levels of thousand of genes across multiple experiments. In this regard, many machine learning approaches were proposed to deal with this deluge of information. Clustering methods are one of these approaches. Their process consists of grouping data (gene profiles) into homogeneous clusters using distance measurements. Various clustering techniques are applied, but there is no consensus for the best one. In this context, a comparison of seven clustering algorithms was performed and tested against the gene expression datasets of three model plants under salt stress. These techniques are evaluated by internal and relative validity measures. It appears that the AGNES algorithm is the best one for internal validity measures for the three plant datasets. Also, K-Means profiles a trend for relative validity measures for these datasets.

Keywords

Clustering Methods, Clustering Validity Indices, Gene Chips Analysis, Gene Expression, Plant Datasets.

DOI: 10.9781/ijimai.2020.05.004

I. Introduction

Abiotic stresses significantly reduce agricultural productivity worldwide. Plant growth and crop productivity are affected by environmental factors, especially saline stress [1]. Therefore, it is important to know the genes implicated in tolerance to salinity [1]. Furthermore, technological advances in the field of genomics, such as DNA sequencing, have generated a wealth of genetic information [2]. Such information includes expression profile levels of thousands of genes under various experimental conditions [3]. Hence, a better biological view of the presumed gene functions can be obtained.

Therefore, a wide range of machine learning methods such as clustering have been developed [4], [5]. They are being used in a variety of applications, such as cancer diagnosis [6], pharmacovigilance [7] and plant breeding [8]. "Omics research" has thereafter relied on clustering techniques to group genes. The main objective of clustering techniques is exploring the results of DNA chips to classify and group identical expression profiles [4], identify co-expressed genes [4], find their biological functions [8], [9], and explain their regulatory mechanisms [9], [10].

In Gene chips data analysis, some classical clustering techniques were implemented. One of these is the Hierarchical algorithm commonly called UPGMA [11], [12]. It generates dendrograms and heat maps that display and intuitive visualization of genes and their relationships [12]-[15]. Other clustering methods, called Partitioning methods like K-Means, PAM and CLARA were developed. Their purpose is to partition the gene expression dataset into (k) coherent clusters with same biological characteristics [16]-[18]. Model based clustering methods, such as Self Organization Map (SOM) is also another clustering technique. Their aim is similar to K-Means and Hierarchical algorithms. The advantage of the SOM algorithm is its ability to visualize and optimize the high-dimensional data on an output map of neurons with similar gene functions [19].

Some of these methods have been combined. For instance, Hierarchical algorithm with SOM algorithm designated Self-Organizing Tree Algorithm (SOTA) [20] and Self-Dynamically Growing Self-Organizing Tree (DGSOT) [20] algorithms were developed for improving clustering performance when there is noisy data and determining a good quality of partition of the gene expression data [20], [21].

Another combination of the Hierarchical algorithm that was associated with the K-Means algorithm is called Hierarchical K-Means [22]. This combination took advantage of both algorithms. The hierarchical algorithm provided a tree structure of groups that was used by the K-Means algorithm to determine relevant and compact gene expression groups [22].

Many more algorithms were implemented in order to enhance the convergence and efficiency of the clustering result. There are for example, Fuzzy clustering [17, 23], Fuzzy clustering based on Local Approximation of MEmbership (FLAME) [23], Graph-based clustering method like MST [24], Grid-based clustering method (STING, CLIQUE) [24], Density-based clustering method (OPTCS, DBSCAN) [24], Gaussians and Spectral Clustering methods [24].

* Corresponding author.
E-mail address: houdafyad82@gmail.com

While all these algorithms have been compared in different studies, there was no clear agreement on the most appropriate clustering algorithm to be used for clustering genes with their associated expression profiles [25]-[27].

Mostly, each clustering method has its own parameters for calculating clusters. The decision to use a particular method for clustering will depend on the nature of the datasets being studied and what the researcher expects to achieve using that method [26]-[29].

Based on these considerations, we decided to conduct a comparative study of seven most commonly used clustering algorithms on gene expression datasets from three model plants under saline stress. These methods are evaluated based on both internal and relative validity measures. The main objective of this study is to address biologists' concerns about the most appropriate algorithm to be used for achieving the desired gene clustering.

The remaining of this paper is organized as follows: Section II presents an overview of clustering techniques used in gene expression. Section III, is dedicated to gene expression experiments, the choice of clustering techniques and the clustering Validity concepts. Section IV, provides the results and discussion of the performance of the respective algorithms. Finally, Section V concludes by summarizing findings and identifying possible future work.

II. Related Work

In the previous section, we have mentioned the importance of analyzing and studying gene expression data with clustering techniques. These techniques have helped to answer several biological questions.

Hierarchical algorithms (HC) are the earliest ones used in gene expression data. Eisen et al. [12] used HC for an empirical analysis to classifying and visualizing gene expression on yeast Saccharomyces cerevisiae datasets. Dendrogram tree and Heat maps are well-known HC graphic tools that illustrate the correlation of these genes.

Alizadeh et al. [13] applied the same method on Diffuse large B-cell Lymphoma (DLBCL), HC has permitted the discovery of new molecular subtypes with three different genetic signatures.

Bajsa et al. [14] focused on the determination of the transcription level of the cellular pathways on the model plant Arabidopsis. The result of the heat map with dendrogram revealed the up-regulated gene and down-regulated ones in different time courses under salt stress.

Hossen et al. [15] analyzed the clustering proximity effect on two types of gene expression datasets (Affimetrix and cDNA). The authors implemented seven Hierarchical algorithms (Single Linkage, Complete Linkage, Average Linkage, Ward, Centroid, Median, Mcquitty) according to five proximity measures (Euclidean, Manhattan, Pearson, Spearman and, Cosine). The Ward method with Cosine distance was outperforming on both types of datasets.

Takahashi et al. [16] studied gene expression of 4 varieties of Wheat, the analysis concerned different levels of salinity tolerance. K-Means Clustering algorithm optimized to 3 the number of clusters: Cluster I included genes expressed as an early response that occurred within 24 hours under control conditions. Cluster II assembled genes expressed during the second day under control conditions and Cluster III included the genes expressed in the late response that occurred on the third day. The Hierarchical clustering (with Pearson correlation and average linkage) method and Principal component analysis were used for visualization of results [16].

Gasch et al. [17], worked on the Yeast gene expression profile. K-Means and Fuzzy C-means (FCM) have established the expression profile during seven periods of the cell cycle in Yeast. They validate their results with the Davis Bouldin Index (DBI). FCM has achieved the DBI of 0.31452 for K=3 and 0.37822 for K=4 which is better than K-Means clustering.

The study conducted by Ge et al. [18], with Hierarchical clustering and K-Means methods allowed identifying eight (08) distinctive gene groups regulated by abiotic stress in Glycine soja. The authors successfully discovered the corresponding co-regulated genes and their functions.

FLAME is an extension of Fuzzy clustering based on the Local Approximation of Membership that was implemented for microarray data [23]. The advantage of the FLAME algorithm compared to the FCM algorithm is its ability to define various and homogeneous groups of genes, and to give a relevant subdivision of biological functions patterns [23].

SOM algorithm was applied to Yeast Sporulation, Human Fibroblasts Serum and Rat CNS datasets [19]. This method provides a better result for the recognition and classification of the features in complex and multidimensional datasets. Luo et al. [30] have used the SOTA algorithm to discover Transcription Factor (TF) gene families in *Medicago sativa* during ABA treatment. In that case, 82 TF genes families were distributed into four clusters with the number of genes equal respectively 15, 34, 18 and, 14.

The comparison study between the following Clustering algorithms (HC, K-Means, and SOM) was performed on Solanum tuberosum genes showing differential expression in abiotic stress [25]. The author in this study, obtains almost the same number of the clusters for these different algorithms.

López-Kleine et al. [26] applied AGNES, DIANA, K-Means (with Euclidean and Manhattan distances) and SOM for clustering the genes involved in pathogen resistance on Tomato. The results showed that AGNES, K-Means, and SOM grouped these genes into two clusters: genes implicated or no in plant resistance. DIANA was abandoned because almost all genes were assigned to one cluster.

In other comparison work [27], Hierarchical algorithms with single, complete and average linkage, K-Means, Gaussians Clustering methods (FMC), Spectral Clustering (SP) methods and a Nearest Neighbour-based methods were evaluated on 35 gene expression datasets of various cancerous tissue types. FMC and K-Means were the most appropriate methods to recover the true structure of this kind of datasets.

Singh et al. [28] assessed the efficiency of K-Means (KM), Density-based clustering (DBC) and expectation maximization (EM) methods by using the sum of squared error, log-likelihood measures. These methods were tested on SRBCT, Lymphoma and three different Leukemia datasets. The results showed that EM algorithm gives the best result with log-likelihood measurement. KM and DBC algorithms produced similar results with regards to the sum of squared error measurement.

Bihari et al. [29] compared the performance of KM, HC clustering, SOM and DBSCAN on Iris flower gene expression data. The comparison results of these methods were validated by using internal and external indices. According to the experimental analysis KM is more appropriate for gene clustering.

In Table I, we summarize some algorithms that we have mentioned in the state of the art. We will give their main characteristics with respect to the following parameters: (i) influence of noisy data, (ii) ability to work properly with large dataset and (iii) algorithm computation time.

III. Methods

This section describes the different experiences conduced to analyze and compare the clustering methods for plant genes expression

TABLE I. Characteristics of Various Clustering Algorithms

"n" is the number of points in the dataset, "k" is the number of clusters, "l" is the number of iterations, "m" is the number of initial sub-clusters produced by the graph partitioning algorithm.

Clustering method	Algorithm	Type of data	Sensitive to noisy data	Dealing with high dimensional data	Scale	Computational time
Hierarchical	**AGNES**: a bottom-up approach [30], [44].	Numerical	Not very sensitive	No	NA	$O(n^2)$
	DIANA: a top-down approach [31], [44].	Numerical	Not very Sensitive	Yes	NA	$O(2^n)$
	BIRCH: agglomerative hierarchical based clustering algorithm [32], [44].	Numerical	Very Insensitive	No	Yes	$O(n)$
	CURE: has been developed to handle a huge volume of data, insensitive to outliers and capable of working with clusters of different shapes and sizes [33], [44].	Numerical	Insensitive	Yes	Yes	$O\ (n^2+nm_m m_a+\ n^2 log\ n)$
	ROCK: uses the concept of the number of links between two records to assess the similarity of the categorical attributes of the dataset [34], [44].	Categorical	Not very Sensitive	No	Yes	$O\ \ (n^2 log\ n)$
	CAMELEON: based on a dynamic model for merging clusters. It calculates the interconnectivity and the proximity of two clusters in order to discover the similarity between them [35], [44].	Numerical⁄ Categorical	NA	No	Yes	NA
Partitioning	**K-MEANS**: is a method which aims to divide the dataset elements into groups that are well separated from each other [36], [44].	Numerical	Sensitive	No	Yes	$O(Ikmn)$
	PAM: algorithm aims to find a sequence of objects called medoids that are located in the center of clusters. It is a more robust partitioning algorithm against outliers than the k-means partitioning algorithm [37], [44].	Numerical	Not very Sensitive	No	Yes	$O(k(n-k)^2)$
	CLARA: was developed in order to deal with large datasets. It does not work with the whole set of data, but with a small portion of the data which is chosen randomly [38], [44]	Numerical	Not very Sensitive	No	Yes	$O\ (k(40+k)^2+\ k(n-k))$
	CLARANS: is an extension of the CLARA algorithm. It is a combination of sampling techniques with the PAM algorithm [39], [44].	Numerical	Sensitive	No	Yes	$O(kn^2)$
Model based	**SOM**: consists in projecting the large data space observed on a 2 or 3 dimensional space called a map. This map is composed of groups of neurons connected together according to the concept of neighborhood [40], [44].	Numerical	Not very sensitive	Yes	Yes	$O(I)$
Fuzzy based	**FCM**: allows assigning an element to one or more clusters [41], [44].	Numerical	Sensitive	No	Yes	$O(n)$
Grid based	**CLIQUE**: finds clusters in subspaces of high density data [42], [44].	Numerical	Not very Sensitive	No	Yes	$O(n+k^2)$
Density based	**DBSCAN**: groups in the neighborhood of a point having a given radius (ε) a minimum number of points (MinPts) [43], [44].	Numerical	Very insensitive	No	No	$O(m\ log\ m)$

(NA) information is not mentioned by authors.

data. The proposed workflow is described in Fig. 1

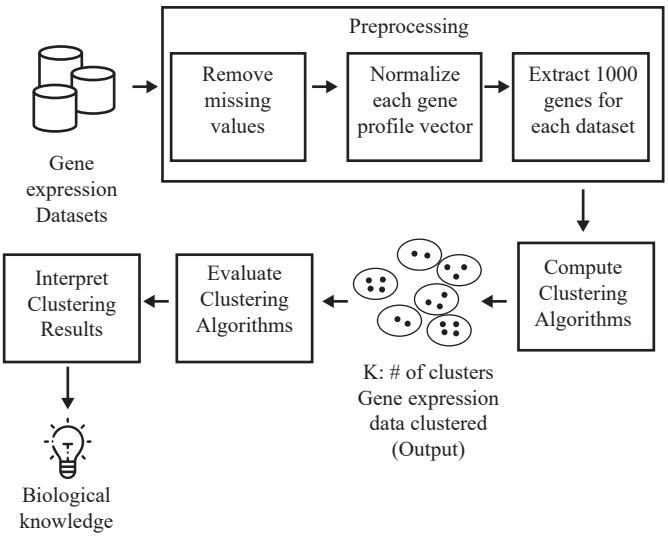


Fig. 1. Flowchart of the Experimental Process.

A. Datasets Selection

Three datasets of expression data relating to plants *Arabidopsis thaliana*, *Solanum lycopersicom* (Tomato) and *Medicago truncatula* under salt stress were considered. In this experimental study, we choose to work with these datasets because they are based on model plants. *Arabidopsis thaliana* is regarded to be the first most studied and investigated model plant. *Solanum lycopersicom* and *Medicago truncatula* are also model plants, each representing a family of plant species. Datasets for these model plants cover a broad spectrum of gene expressions.

1. Dataset 1: Arabidopsis Thaliana (A. Thaliana) Salt Stress

This dataset describes the salt stress experiment of model *Arabidopsis thaliana* leaves using Affymetrix Array, 2 samples of leaves from 3 genotypes of *A. thaliana* with and without 100 mM NaCl. This dataset shows the salt-stress influence on leaves from these 3 genotypes. The experiment results explain a global change on related genes and provide an insight into the molecular mechanisms underlying variation in salt stress responses [45]. The *Arabidopsis thaliana* dataset was downloaded from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16765.

2. Dataset 2: Solanum Lycopersicom (Tomato) Salt Stress

This dataset describes the salt stress experiment of an old Tomato leaves using Affymetrix Array, 6 samples of leaves-old with 200 mM NaCl for 5 h, 6 samples of leaves-old without 200 mM NaCl for 5 h. This dataset compares the salt-stress influence analysis on leaves-old from 2 genotypes of Tomato. The experiment results that the Wild tomato genotype is significantly more salt-tolerant than a Cultivar, *Solanum lycopersicom* [46]. The *Solanum lycopersicom* dataset was downloaded from http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16401.

3. Dataset 3: Medicago Truncatula (M. Truncatula) Salt Stress

This dataset describes the time-course salt stress experiment of model legume *Medicago truncatula* roots using Affymetrix Array, 6 samples of Medicago truncatula seedlings grew in two weeks in hydroponics media with 200mM NaCl salt stress at 0, 6, 24, 48 hours, 12 samples other of Medicago truncatula seedlings 3 days Petri dishes with 180mM NaCl salt stress at 0, 1, 2, 5, 10, 24 hours. This dataset reveals the salt stress effect on Medicago truncatula seedlings [47]. The *Medicago truncatula* dataset was downloaded from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14029.

These datasets are retrieved from the Gene Expression Omnibus database [48]. Table II gives all the information concerning these three datasets.

B. Preprocessing

As shown in Fig. 1, before gene clustering, it is necessary to pre-process datasets (removing missing values). And then, every gene vector is normalized according to whether its mean is equal to 0 and its standard deviation has a value of 1 [49]. For more homogeneity with the number of genes present in the Tomato dataset, only 1000 genes of the *Arabidopsis thaliana* and the *Medicago truncatula* datasets were randomly extracted for analysis. They were randomly selected to eliminate selection bias. This selection is due to the fact that the number of genes annotated on tomato is less important than for *Arabidopsis thaliana* and the *Medicago truncatula* in this type of dataset.

C. Clustering

For analyzing and evaluating the three datasets cited before, we used the open-source R environment which contains a variety of functions for data clustering. Among the clustering algorithms, we have chosen seven one: Hierarchical algorithms (AGNES, DIANA), Partitioning algorithms (K-MEANS, PAM and CLARA), Fuzzy Clustering (FANNY). These categories of methods are functions defined in R package named "cluster". Model-based Clustering (SOM) in this category of methods depends on R packages "kohonen" and "mclust". All these seven methods are contained in the R package named "clValid" that includes some validity measures that we used for testing our three datasets. These algorithms are the most commonly used, as the time complexity is low and they offer an easy interpretation of results by biologists. The code source link of each clustering method used and their validation is: http://github.com/Projet-82/New-Project/blob/master/Clustering_eva-lunation_codesource.R.

TABLE II. Dataset Description

Data set	#Genes	#Samples	Genotypes	Salt-Stress concentration	Time points
1	**A. thaliana_salt stress**				
	15 288	18	6 Ws	0 mM NaCl 100 mM NaCl	NA
			6 Col	0 mM NaCl 100 mM NaCl	NA
			6 Col(gl)	NaCl 100 mM NaCl	NA
2	**Tomato_salt stress**				
	1 000	12	6 Money maker	0 mM NaCl 200 mM NaCl	5 hours
			6 PI365967	0 mM NaCl 200 mM NaCl	5 hours
3	**M. truncatula_salt stress**				
	2 394	18	6 NA	180 mM NaCl	0, 1, 2, 5, 10, 24 hours
			12 NA	200 mM NaCl	0, 6, 24, 48 hours

D. Evaluation

To evaluate and compare the clustering algorithms, we consider two important concepts: Cohesion and separation. The Cluster cohesion measures how closely related are objects in a cluster. [50]. Cluster separation measures how distinct a cluster is from other clusters [50].

For this study, the following measures are used to assess the quality and consistency of the clusters on terms of the cohesion and separation of clusters resulting from different clustering algorithms: Connectivity index [50], [51], Dunn index [50], [51], and Silhouette coefficient [50], [51]. These 3 measurements are called internal measures.

In the other hand, the stability measures compare the results from clustering based on the full data to clustering based on removing each column, one at a time. These 4 measures work especially well if the data are highly correlated, which is often the case in high-throughput genomic data. They included Average proportion of non-overlap (APN) [50], [51], Average distance (AD) [50], [51], Average distance between means (ADM) [50], [51], and the figure of merit (FOM) [50], [51]. These lasts are called relative measures.

1. Connectivity Index

It measures how much neighbouring data points have been ranked in the same cluster [50], [51]. It is calculated by the following formula:

$$Conn(c)=\sum_{i=1}^{N}\sum_{j=1}^{L}x_{i,nn(j)} \tag{1}$$

With

$$x_{i,nn_{i(l)}}=\begin{cases}\frac{1}{j}, if \nexists c_k : i\in c_k \wedge nn_{i(j)}\in c_k\\ 0, otherwise\end{cases} \tag{2}$$

Where:

"K" is the total number of clusters.

"N" is the total number of rows (observations).

$nn(j)$ is the j^{th} nearest neighbour of the data point.

"L" is the parameter determining the number of neighbours that contribute to connectivity measure. Connectivity should be minimal.

2. Dunn Index

Dunn's goal is to identify dense and well-isolated clusters. It describes the proportion between the minimum and the maximum distances separating the clusters [50], [51]. It is computed by the following formula:

$$D=\frac{\min_{1\le i\le j\le n} d(i,j)}{\max_{1\le i\le j\le n} d'(k)} \tag{3}$$

Where:

$d(i\,j)$ describes the two cluster's distance i and j.

$d'(k)$ measures the intra-group distance of cluster k.

$d(i, j)$ is the inter-group distance. In this case the distance corresponds to the centroids distance.

3. Silhouette Coefficient

The silhouette width coefficient defines the compactness based on the paired distance between all items in the cluster, and the separation based on paired distance between all items on the cluster and all items in the nearest cluster [50], [51]. The Silhouette score is defined as:

$$S(i)=\frac{b(i)-a(i)}{\max\{a(i)-b(i)\}} \tag{4}$$

Where:

$a(i)$ is the average distance of gene i to other genes in the same cluster. $b(i)$ is the average distance of gene i to genes in its nearest neighbour cluster. The average of $S(i)$ across all genes reflects the overall quality of the clustering result.

4. Average Proportion of Non-overlap (APN)

The APN measure calculates the average proportion between observations that are not affected in their similar cluster by grouping together the complete data and grouping together the data with one column removed [50], [51]. The APN measure is denoted as follows:

$$APN(K)=\frac{1}{MN}\sum_{i=1}^{N}\sum_{l=1}^{M}\left(1-\frac{n(C^{i,l}\cap C^{i,0})}{n(C^{i,0})}\right) \tag{5}$$

Where:

"K" is the total number of clusters.

"M" is the total number of columns (attributes)

"N" is the total number of rows (observations).

"$n(C^{i,0})$" represents the cluster that contains observation i using the original clustering (based on all available data).

"$C^{i,l}$" represents the cluster that contains observation i where the clustering is based on the dataset with column removed.

5. Average Distance (AD)

The mean distance between observations that are not assigned in a similar cluster by grouping based on complete data and grouping based on data with one column deleted is estimated by the AD measure [50], [51] which is denoted as follows:

$$AD(K)=\frac{1}{NM}\sum_{i=1}^{N}\sum_{l=1}^{M}\frac{1}{n(C^{i,l}\cap C^{i,0})}\left[\sum_{i\in C^{i,0},j\in C^{i,l}}\left(dist(g_i,g_j)\right)\right] \tag{6}$$

Where:

$dist(g_i, g_j)$ is a distance (e.g. Euclidean, Manhattan, etc.) between two expression genes profiles i and j.

"K" is the total number of clusters.

"M" is the total number of columns (attributes).

"N" is the total number of rows (observations).

"$n(C^{i,0})$" represents the cluster that contains observation i using the original clustering (based on all available data).

"$C^{i,l}$" represents the cluster that contains observation i where the clustering is based on the dataset with column removed.

6. Average Distance between Means (ADM)

The ADM measure calculates the mean distance between cluster centers that are not assigned in a similar cluster by grouping based on complete data and grouping based on data with one column [50], [51]. The ADM measure is denoted as follows:

$$ADM(K)=\frac{1}{NM}\sum_{i=1}^{N}\sum_{l=1}^{M}dist(\bar{x}_{c^{i,l}},\bar{x}_{c^{i,0}}) \tag{7}$$

Where:

"M" is the total number of columns (a collection of samples, time points…).

"N" is the total number of rows (observations).

$\overline{x}_{C^{i,0}}$ is the mean of the observations in the cluster which contains observation i, when clustering is based on the full data.

$\overline{x}_{C^{i,l}}$ is the mean of the observations in the cluster which contains observation i, when clustering is based on the dataset with column removed. Currently, ADM only uses the Euclidean distance.

7. Figure of Merit (FOM)

The intra-cluster mean variance of the suppressed column observations is computed by the FOM measurement, the resulting classification is being based on the remaining samples (not cleared). This estimates the average error using predictions based on cluster averages [50], [51]. For a particular left-out column *l*, the FOM is:

$$FOM(l,K) = \sqrt{\frac{1}{N}\sum_{k=1}^{K}\sum_{i\in C_k(l)} dist(x_{i,l},\overline{x}_{C_k(l)})} \quad (8)$$

Where:

"K" is the total number of clusters.

"N" is the total number of rows (observations).

$x_{i,l}$ is the value of the i^{th} observation in the l^{th} column in the cluster.

$\overline{x}_{C_k(l)}$ is the average of the cluster $C_k(l)$. Currently, the only distance available for FOM is Euclidean.

IV. Results & Discussion

In this section, we comment on and discuss the results obtained. Tables III, IV and V describe the various performance measures and validity indices corresponding to the three best clustering algorithms applied on our three datasets. The rest of the clustering algorithms are not shown.

A. Dataset 1: A. Thaliana Salt Stress

It can be seen from the results of Table III that in the case of Hierarchical algorithms, AGNES gives, for an optimal number of clusters **K=4**, high performance with a lower Connectivity Index equal to **25.049**, best Silhouette index score equal to **0.488** and with best Dunn index value score equal to **0.228**. This algorithm performance is followed by K-Means and DIANA algorithms for the Dunn index, with a value of **0.0576** (resp. **0.0579**) for a cluster number equal to 10. SOM is less rated than DIANA by the index of Silhouette which is worth **0.2818** (resp. **0.5068**) for **K = 2**. Also, we found that FANNY algorithm does not provide any results because of its inability to generate measurable clusters.

On the other side, the cluster stability measures describes that the Model based Clustering algorithm, SOM presents high performance too with AD measure value equal to **3.6581** and FOM measure value equal to **0.6988** for **K = 10**.

AGNES, followed by DIANA gives good performance with an APN value equal to **0.0047** (respectively **0.0056**). And on the contrary, DIANA is followed by AGNES with an ADM value equal to **0.0518** (resp. **0.0677**) for **K = 2**. PAM is better than K-Means in AD measure with a value equal to **3.7822** (or **3.8037**). K-Means is better than PAM in FOM measurement with a value equal to **0.7083** (respectively **0.7196**) for **K=10**.

B. Dataset 2: Tomato Salt Stress

From the results presented in Table IV, we can observe that concerning Hierarchical Clustering algorithms, for the optimal numbers of clusters **K=4**, AGNES gives high performance with lower Connectivity index value equal to **25.128**, best Silhouette index score equal to **0.7229** and with best Dunn index value score equal to **0.124**. This performance is followed by the DIANA algorithm for the Dunn and Silhouette indices, whose value is **0.0648** (resp. **0.7161**) for a cluster number equal to 4 (resp.2). SOM is lower than DIANA, for the Silhouette index is worth **0.7161** (resp. **0.7122**) for **K = 2**.

TABLE III. Evaluation of the 3 Best Clustering Techniques on A. Thaliana Salt Stress Dataset

Algorithm rank		1	2	3
		Algorithm [parameter K] score		
Internal validation	Conn. index	AGNES[K=4] 25.048	DIANA[K=2] 37.637	K-Means[K=2] 147.272
	Dunn index	AGNES[K=4] 0.2281	K-Means[K=10] 0.0579	DIANA[K=10] 0.0576
	Silhouette index	AGNES[K=4] 0.4880	DIANA [K=2] 0.5068	SOM[K=2] 0.2818
Relative validation	APN measure	AGNES[K=2] 0.0047	DIANA[K=2] 0.0056	K-Means[K=2] 0.0236
	AD measure	SOM[K=10] 3.6581	PAM [K=10] 3.7822	K-Means[K=10] 3.8037
	ADM Measure	DIANA[K=2] 0.0518	AGNES[K=2] 0.0677	K-Means[K=2] 0.1045
	FOM measure	SOM[K=10] 0.6988	K-Means[K=10] 0.7083	K-Means[K=10] 0.7196

On another side, the relative measures show that the partitioning Clustering algorithm, PAM produces high performance too with AD measure equal to **0.992** and FOM measure equal to **0.324** for number of clusters **K=10**, DIANA as well performed with APN measure equal to **0.0076** followed by FANNY and K-Means values equal to **0.0111** (resp. **0.0116**) for an optimal number of **K = 2**. PAM done a good result with an AD measure score equal to 0.9924 and FOM with a score equal to **0.3241** for **K=10**.

CLARA obtains a good result too with AD and ADM measures value equal to **1.0751** (resp. **0.0817**) for **K=10**. FANNY presents a good performance with an APN and AD measures with values equal to 0.0111 (resp. **1.0690**) for **K=2** (resp. **K=10**) and K-Means presents the same behavior for ADM and FOM values equal to **0.0804** (resp. **0.3280**) for the same number of clusters.

TABLE IV. Evaluation of the 3 Best Clustering Techniques on Tomato Salt Stress Dataset

Algorithm rank		1	2	3
		Algorithm [parameter K] Score		
Internal validation	Conn. index	AGNES[K=4] 25.128	CLARA[K=2] 30.249	K-Means[K=2] 35.0825
	Dunn index	AGNES[K=4] 0.1239	DIANA[K=4] 0.0648	CLARA[K=10] 0.0353
	Silhouette index	AGNES[K=4] 0.7229	DIANA[K=2] 0.7161	SOM[K=10] 0.7122
Relative validation	APN measure	DIANA[K=2] 0.0076	FANNY [K=2] 0.0111	K-Means[K=2] 0.0116
	AD measure	PAM[K=2] 0.9924	FANNY[K=10] 1.0690	CLARA[K=10] 1.0751
	ADM measure	FANNY[K=2] 0.0538	K-Means[K=2] 0.0804	CLARA[K=10] 0.0817
	FOM measure	PAM[K=10] 0.3241	K-Means[K=10] 0.3280	SOM[K=10] 0.3290

C. Dataset 3: M. Truncatula Salt Stress

From the results reported in Table V, the Hierarchical Clustering algorithms, for the optimal numbers of clusters **K = 4**, AGNES presents high performance with lower Connectivity index value equal to **2.9290**, a best Dunn and with best Silhouette index of **0.8296** (resp. **0.9587**). This performance is followed by the DIANA method with Connectivity and Silhouette indices equal to **5.2869** and. **0.9406**, respectively, for a cluster number equal to **2**.

On the other side, the relative stability describes that DIANA as well performed with a value of APN measure equal to **0.0001**. This performance is followed by AGNES and CLARA with values equal to **0.0007** (resp. **0.0053**) and the same behavior with the inverse ordered algorithms is shown with ADM measure equal respectively for CLARA **0.0288** and AGNES **0.0307** with a cluster number of **K = 2**. K-Means performs in FOM measure with a value equal to **0.3487** followed by PAM and SOM with a value equal to **0.3497** (resp. **0.4059**) with a cluster number of **K = 10**.

TABLE V. Evaluation of the 3 Best Clustering Techniques on M. Truncatula Salt Stress Dataset

Algorithm rank		1	2	3
		Algorithm [parameter K] score		
Internal validation	Conn. index	AGNES[K=4] 2.9290	DIANA[K=2] 5.2869	K-Means[K=2] 24.5222
	Dunn index	AGNES[K=4] 0.8296	DIANA[K=2] 0.3359	SOM[K=6] 0.0044
	Silhouette index	AGNES[K=4] 0.9587	DIANA[K=2] 0.9406	K-Means[K=10] 0.8822
Relative validation	APN measure	DIANA[K=2] 0.0001	DIANA[K=2] 0.0007	K-Means[K=2] 0.0053
	AD measure	PAM[K=10] 0.9523	PAM[K=10] 0.9945	K-Means[K=10] 1.0745
	ADM measure	DIANA[K=2] 0.0001	AGNES[K=2] 0.0288	K-Means[K=2] 0.0307
	FOM measure	K-Means[K=10] 0.3487	K-Means[K=10] 0.3497	PAM[K=10] 0.4059

According to the three internal validity measures Connectivity, Silhouette and Dunn index value, Hierarchical Clustering (AGNES) appears to be the most efficient with **K = 4** clusters for the three datasets examined (Fig. 2, 3 and 4). However, the number of plant genes categories as found by authors who have submitted these different datasets is higher than 4 categories.

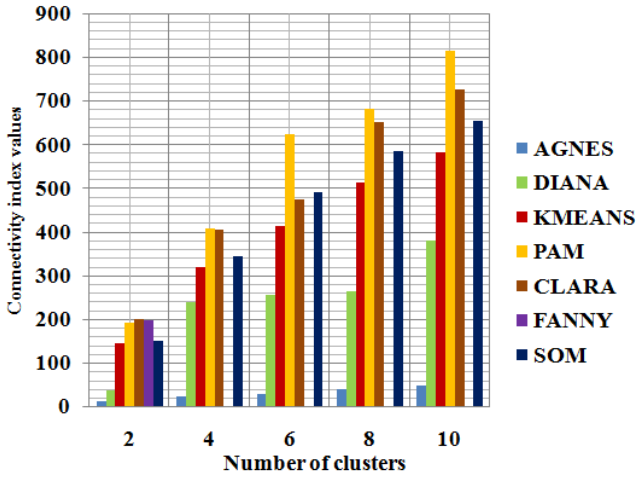


Fig. 2. Performance of Connectivity index using A. thaliana salt stress dataset.

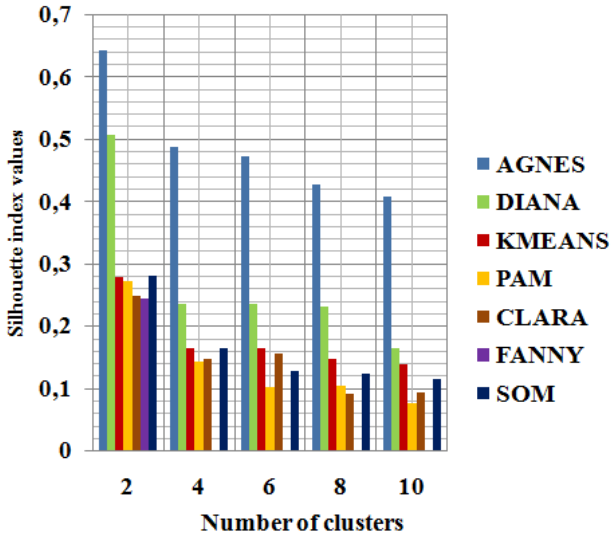


Fig. 3. Performance of Silhouette index using A. thaliana salt stress dataset.

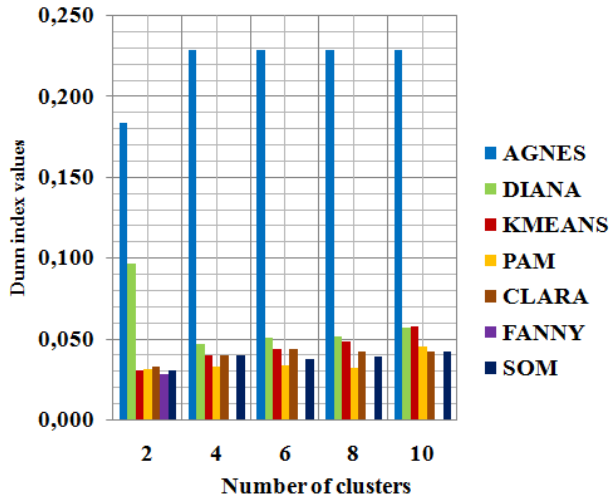


Fig. 4. Performance of Dunn index using A. thaliana salt stress dataset.

On the other side, the relative validity measures report that SOM algorithm performs well for dataset 1, with AD measure value equal to **3.6581** and FOM measure value equal to **0.6988** (Fig. 5 and 6). PAM and K-Mean provide good results for dataset 2 and 3 with the same number of clusters equal to **10** (Fig. 5 and 6). Also, this number of clusters would correspond biologically to the number of gene families found.

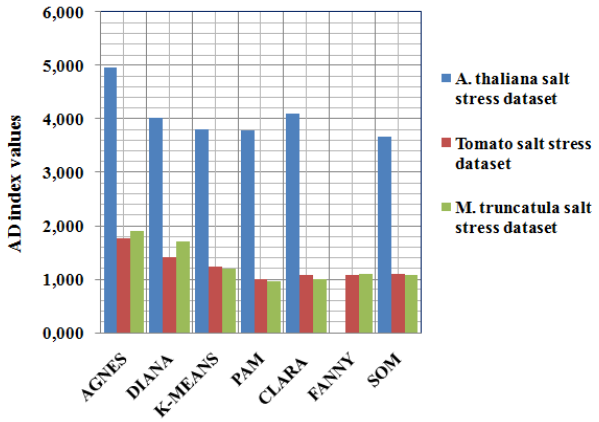


Fig 5. Performance of AD index with K=10.

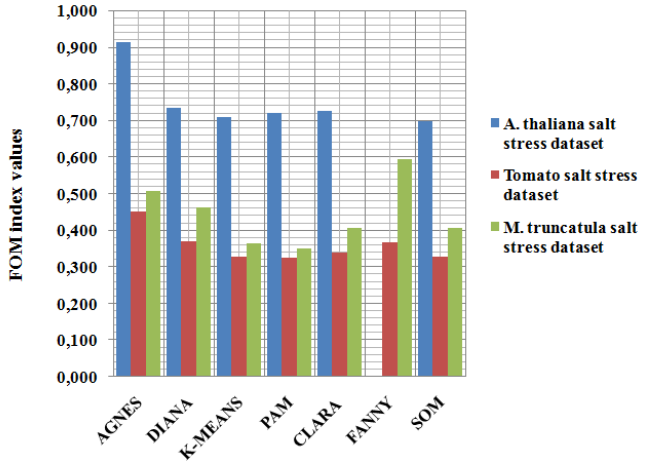


Fig 6. Performance of FOM index with K=10.

Finally, this study demonstrates that, according to the values of the validity indices (internal and relative) and the number of optimal clusters:

- For dataset 1, the SOM algorithm is the most efficient with the relative validation indices (ADM and FOM) for an optimal cluster number of 10. This cluster number is compatible with the biological reality of different gene families obtained by the submitters of this dataset. This algorithm has confirmed its performance in other datasets of complex organisms such as: Human Fibroblasts Serum and Rat CNS datasets [19].
- For datasets 2 and 3, PAM and K-Means algorithms are also distinguished by their performance for the same relative validation indices (ADM and FOM) and for the same cluster number equal to 10 compatible with the biological reality of different gene families obtained by the experimenters. These algorithms revealed interesting results in different kind of datasets: cancerous tissue types [52] and on the plant functions [53].

When we consider only the values of validity indices (internal and relative) without taking into count the cluster number expected by biologist:

- For all datasets, the AGNES algorithm presents the best internal indices values (Connectivity, Dunn and Silhouette) with an optimal number of clusters K=4. We also note that according to the relative validity indices (APN, ADM and FOM), the K-Means algorithm seems to be suitable for the three datasets with an APN index value of **0.0236** for dataset 1, which decreases to **0.0116** for dataset 2 and 0.0053 for dataset 3 when the number of clusters is set to 2. The ADM index also shows a decreasing trend with values of **0.1045** for dataset 1, and **0.0804** (respectively **0.0307**) for datasets 2 and 3 respectively with the same number of clusters K=2. For the last index, which is FOM, it is equal to **0.7083** for dataset 1 and decreases to **0.3280** for dataset 2 and **0.3497** for dataset 3 for K=10. And here the number of clusters is in adequacy with the biological reality of the families of genes.

However, firstly it should be kept in mind that these datasets are not reference datasets and therefore they are not necessarily "potentially groupable" which may explain the mismatch between the number of optimum clusters obtained and the number of expected clusters by biologists. Secondly, we have had to retain only a number of 1000 genes for each datasets. This reduced size of the gene sample is due to the number of genes annotated on tomato which is less than that of the other two plants. This fact may have contributed to this mismatch or competed to make the datasets less groupable.

V. Conclusion

In this paper, seven clustering algorithms were compared and evaluated on three sets of gene expression data from plants subjected to salt stress. The purpose was to determine the best performing algorithm that produces the optimal number of clusters reflecting the biological reality.

The results showed that the SOM algorithm allows a good distribution of genes for dataset 1. The partitioning algorithms PAM and K-Means for datasets 2 and 3 lead to the same results but with slightly lower validity index values. When we take into account only the internal validity indices, we see that the AGNES algorithm presents for the three data sets, the best values (Connectivity, Dunn and Silhouette) with a number of clusters equal to 4. In this case, we also note that the values of the relative validity indices allow the emergence of a trend indicating an acceptable performance of K-Means for the three sets of data.

This work has certain limitations: (i) The number of genes studied: Only 1000 genes are selected. (ii) Noise and outliers are inherent in the expression data. Clustering methods can be affected by this phenomenon. But, although K-Means is generally deemed as a sensitive method to outliers, it appears in this study that it is not the case. Because we obtained for this later a result with acceptable indices values and with an optimal cluster number identical to the one expected by biologists.

These results provide guidance for future work. The use of AGNES and K-Means clustering methods may be recommended for the analysis of this type of datasets. The additional orientation would be to associate the expression profiles (numerical aspect) with the corresponding annotations described by the ontologies (semantic aspect) in order to provide enrichment in the gene clustering.

Acknowledgment

The authors would like to thank the Directorate General of Science Research and Technological Development (DGRSDT), Ministry of Higher Education and Scientific Research of Algeria for their support in this work.

References

[1] Sharma. (2016). "Computational gene expression profiling under salt stress reveals patterns of co-expression", Genomics data, Vol. 7, pp. 214-221. DOI: https: //doi.org/10.1016/j.gdata.2016.01.009.

[2] F. M. Afendi, N. Ono, Y. Nakamura, K. Nakamura, L. K. Darusman, N. Kibinge, A. H. Morita, K. Tanaka, H. Horai, and M. Altaf-Ul-Amin. (2013), "Data mining methods for omics and knowledge of crude medicinal plants toward big data biology", Computational and Structural Biotechnology Journal, Vol. 4, No. 5, pp. e201301010. DOI: https: //dx.doi.org/10.5936/csbj.201301010.

[3] L. M. O. Mesa, L. F. N. Vasquez, and L. Lopez-Kleine. (2012). "Identification and analysis of gene clusters in biological data". In 2012 IEEE International Conference on Bioinformatics and Biomedicine Workshops, pp. 551-557.

[4] N. Pasquier, C. Pasquier, L. Brisson, and M. Collard. (2008). "Mining gene expression data using domain knowledge", International Journal of Software and Informatics (IJSI), Vol. 2, No. 2, pp. 215-231.

[5] K. Raza. (2012), "Application of data mining in bioinformatics", arXiv preprint arXiv: 1205.1125.

[6] R. Govindarajan, J. Duraiyan, K. Kaliyappan, and M. Palanisamy. (2012). "Microarray and its applications", Journal of pharmacy & bioallied sciences, Vol. 4, No. (Supp2), pp.S310. DOI: https: //dx.doi.org /10.4103/0975-7406.100283.

[7] W. Shannon, R. Culverhouse, and J. Duncan. (2003). "Analyzing microarray data using cluster analysis". Pharmacogenomics, Vol. 4, No. 1, pp. 41-52. DOI: https: //doi.org/10.1517/phgs.4.1.41.22581.

[8] W. A. Rensink and C. R. Buell. (2005). "Microarray expression profiling resources for plant genomics", Trends in plant science, Vol. 10, No. 12, pp. 603-609. DOI: https: //dx.doi.org/10.1016/j.tplants.2005.10.003.

[9] S. Y. Rhee and M. Mutwil. (2014). "Towards revealing the functions of all genes in plants". Trends in plant science, Vol. 19, No. 4, pp. 212-221. DOI: https: //dx.doi.org/10.1016/j.tplants.2013.10.006.

[10] K. Byron and J. T. Wang. (2018). "A comparative review of recent bioinformatics tools for inferring gene regulatory networks using time-series expression data". International journal of data mining and bioinformatics, Vol. 20, No. 4, pp. 320-340. DOI: https: //doi.org/10.1504/IJDMB.2018.094889.

[11] Y. Loewenstein, E. Portugaly, M. Fromer, and M. Linial. (2008), "Efficient algorithms for accurate hierarchical clustering of huge datasets: tackling the entire protein space", Bioinformatics, Vol. 24, No. 13, pp. i41-i49. DOI: https: //doi.org/10.1093/bioinformatics/btn174.

[12] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein. (1998). "Cluster analysis and display of genome-wide expression patterns" Proceedings of the National Academy of Sciences, Vol. 95, No. 25, pp. 14863-14868.

[13] A.Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, and X. Yu. (2000). "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling", Nature, No. 6769, Vol. 403, pp. 503. DOI: https: //doi.org/10.1038/35000501.

[14] J. Bajsa, Z. Pan, and S. O. Duke. (2011). "Transcriptional responses to cantharidin, a protein phosphatase inhibitor, in Arabidopsis thaliana reveal the involvement of multiple signal transduction pathways" Physiologia plantarum, Vol. 143, No. 2, pp. 188-205. DOI: https: //doi.org/10.1111/j.1399-3054.2011.01494.x.

[15] A. Hossen, H. A. Siraj-Ud-Doulah, and A. Hoque. (2015). "Methods for evaluating agglomerative hierarchical clustering for gene expression data: a comparative study", Computational Biology and Bioinformatics, Vol. 3, No. 6, pp. 88-94. DOI: https: //doi.org/10.11648/j.cbb.20150306.12.

[16] F. Takahashi, J. Tilbrook, C. Trittermann, B. Berger, S. J. Roy, M. Seki, K. Shinozaki, and M. Tester. (2015). "Comparison of leaf sheath transcriptome profiles with physiological traits of bread wheat cultivars under salinity stress", PLoS One, Vol. 10, No. 8, pp. e0133322. DOI: https: //doi.org/10.1371/journal.pone.0133322.

[17] P. Gasch and M. B. Eisen. (2002). "Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering", Genome biology, Vol. 3, No. 11, pp. research0059. 1, 2002. DOI: https: //doi.org/10.1186/gb-2002-3-11-research0059.

[18] Y. Ge, Y. Li, Y.-M. Zhu, X. Bai, D.-K. Lv, D. Guo, W. Ji, and H. Cai. (2010). "Global transcriptome profiling of wild soybean (Glycine soja) roots under NaHCO 3 treatment", BMC plant biology, Vol. 10, No. 1, pp. 153. DOI: https: //doi.org/10.1186/1471-2229-10-153.

[19] P. Tamayo, D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E. S. Lander, and T. R. Golub. (1999), "Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation", Proceedings of the National Academy of Sciences, Vol. 96, No. 6, pp. 2907-2912, 1999. DOI: https: //doi.org/10.1073/pnas.96.6.2907.

[20] S. Babichev, V. Lytvynenko, M. A. Taif, and A. Sharko. (2016). "Hybrid model of inductive clustering system of high-dimensional data based on the sota algorithm". No. 843, pp. 173-179.

[21] T. Deepika, and R. Porkodi. (2015). "A survey on microarray gene expression data sets in clustering and visualization plots". Int J Emerg Res Manag Technol, Vol. 4, No. 3, pp. 56-66.

[22] M. S. Hasan, and Z. H. Duan. "Hierarchical k-Means: A Hybrid Clustering Algorithm and Its Application to Study Gene Expression in Lung Adenocarcinoma". In Emerging Trends in Computer Science and Applied Computing, chap 4. Quoc Nam Tran and H. Arabnia, Eds. Boston: Morgan Kaufmann, 2015, pp. 51-67. DOI: https: //doi.org/10.1016/B978-0-12-802508-6.00004-1.

[23] C. Muruganantha, and D. Ramyachitra. (December 2014). "An Empirical Analysis of Flame and Fuzzy C-Means Clustering for Protein Sequences". International Journal of Computational Intelligence and Informatics Vol. 4, No. 3, pp. 214-220.

[24] J. Oyelade, I. Isewon, F. Oladipupo, O. Aromolaran, E. Uwoghiren, F. Ameh, M. Achas, and E. Adebiyi. (2016). "Clustering algorithms: Their application to gene expression data". Bioinformatics and Biology insights, Vol. 10, pp. BBI. S38316. DOI: https: //doi.org/10.4137/BBI.S38316.

[25] A. Sharma. (2016). "Computational gene expression profiling under salt stress reveals patterns of co-expression", Genomics data, Vol. 7, pp. 214-221. DOI: https: //doi.org/10.1016/j.gdata.2016.01.009.

[26] L. López-Kleine, J. Romeo, and F. Torres-Avilés. (2013). "Gene functional prediction using clustering methods for the analysis of tomato microarray data". In 7th International Conference on Practical Applications of Computational Biology & Bioinformatics. pp. 1-6. Springer, Heidelberg. DOI: https: //doi.org/10.1007/978-3-319-00578-2_1.

[27] N. Belacel, Q. Wang, and M. Cuperlovic-Culf. (2006), "Clustering methods for microarray gene expression data", Omics: a journal of integrative biology, Vol. 10, No. 4, pp. 507-531. DOI: https: //doi.org/10.1089/omi.2006.10.507.

[28] A. Bihari, S. Tripathi, and A. Deepak. (2019). "Gene Expression Analysis Using Clustering Techniques and Evaluation Indices". Available at SSRN 3350332.

[29] A. A. Singh, A. E. Fernando, and E. J. Leavline. (2016). "Performance Analysis on Clustering Approaches for Gene Expression Data". International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, No. 2, pp. 196-200. DOI: https: //doi.org/10.17148/IJARCCE.2016.5242.

[30] D. Luo, Y. Wu, J. Liu, Q. Zhou, W. Liu, Y. Wang ... & Z. Liu. (2019). "Comparative transcriptomic and physiological analyses of Medicago sativa l. indicates that multiple regulatory networks are activated during continuous ABA treatment". International journal of molecular sciences, Vol. 20, No. 1, pp.47. DOI: https: //doi.org/10.3390/ijms20010047.

[31] Rousseeuw, P. J., & Kaufman, L. (1990). Finding groups in Hoboken: Wiley Online Library.

[32] Zhang, T., Ramakrishnan, R., & Livny, M. (1996). BIRCH: an efficient data clustering method for very large databases. Paper presented at the ACM Sigmod Record.

[33] Guha, S., Rastogi, R., & Shim, K. (1998). CURE: an efficient clustering algorithm for large databases. Paper presented at the ACM Sigmod Record.

[34] Guha, S., Rastogi, R., & Shim, K. (2000). ROCK: A robust clustering algorithm for categorical attributes. Information systems, Vol. 25, No. 5, pp. 345-366. DOI: https://doi.org/10.1016/S0306-4379(00)00022-3.

[35] Karypis, G., Han, E.-H. S., & Kumar, V. (1999). Chameleon: Hierarchical clustering using dynamic modeling. Computer, Vol. 8, pp. 68-75. DOI: https: //doi.ieeecomputersociety.org/10.1109/MC.2005.258

[36] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Paper presented at the Proceedings of the fifth Berkeley symposium on mathematical statistics and probability.

[37] Kaufman, L., Rousseeuw, P., & Dodge, Y. (1987). Clustering by Means of Medoids in Statistical Data Analysis Based on the: L1 Norm,~ orth-Holland, Amsterdam.

[38] Deepa, M. S., & Sujatha, N. (2014). Comparative Studies of Various Clustering Techniques and Its Characteristics. International Journal of Advanced Networking and Applications, Vol. 5, No.6, pp. 2104.

[39] Ng, R. T., & Han, J. (2002). CLARANS: A method for clustering objects for spatial data mining. IEEE Transactions on Knowledge & Data Engineering, Vol. 5, pp. 1003-1016. DOI:https://doi.ieeecomputersociety.org/10.1109/TKDE.2002.1033770.

[40] Kohonen, T. (2013). Essentials of the self-organizing map. Neural networks, Vol. 37, pp. 52-65. DOI: https://doi.org/10.1016/j.neunet.2012.09.018

[41] Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. DOI: https:/doi: 10.1080/01969727308546046

[42] Agrawal, R., Gehrke, J., Gunopulos, D., & Raghavan, P. (2005). Automatic subspace clustering of high dimensional data. Data Mining and Knowledge Discovery, Vol. 11, No. 1, pp. 5-33. DOI: https://10.1007/s10618-005-1396-1

[43] Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Paper presented at the Kdd.

[44] Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., Lin, C.-T. (2017). A review of clustering techniques and developments. Neurocomputing, Vol. 267, pp. 664-681. DOI: https://doi.org/10.1016/j.neucom.2017.06.053.

[45] Z. Chan, R. Grumet, and W. Loescher. (2011). "Global gene expression analysis of transgenic, mannitol-producing, and salt-tolerant Arabidopsis thaliana indicates widespread changes in abiotic and biotic stress-related

genes", Journal of Experimental Botany, Vol. 62, No. 14, pp. 4787-4803. DOI: https://doi.org//10.1093/jxb/err130.

[46] W. Sun, X. Xu, H. Zhu, A. Liu, L. Liu, J. Li, and X. Hua. (2010). "Comparative transcriptomic profiling of a salt-tolerant wild tomato species and a salt-sensitive tomato cultivar", Plant and Cell Physiology, Vol. 51, No. 6, pp. 997-1006, 2010. DOI: https://doi.org/10.1093/pcp/pcq056.

[47] D. Li, Y. Zhang, X. Hu, X. Shen, L. Ma, Z. Su, T. Wang, and J. Dong. (2011). "Transcriptional profiling of Medicago truncatula under salt stress identified a novel CBF transcription factor MtCBF4 that plays an important role in abiotic stress responses". BMC plant biology, Vol. 11, No. 1, pp. 109. DOI: https://doi.org/10.1186/1471-2229-11-109.

[48] T. Barrett, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, and M. Holko. (2012). "NCBI GEO: archive for functional genomics data sets—update". Nucleic acids research, Vol. 41, No. D1, pp. D991-D995. DOI: https://doi.org/10.1093/nar/gks1193.

[49] A. Thalamuthu, I. Mukhopadhyay, X. Zheng, and G. C. Tseng. (2006). "Evaluation and comparison of gene clustering methods in microarray analysis", Bioinformatics, Vol. 22, No. 19, pp. 2405-2412. DOI: https://doi.org/10.1093/bioinformatics/btl406

[50] G. Brock, V. Pihur, S. Datta, and S. Datta. (2011). "clValid, an R package for cluster validation", Journal of Statistical Software (Brock et al., March 2008).

[51] Punitha, K. (2019). Extraction of Co-Expressed Degs From Parkinson Disease Microarray Dataset Using Partition Based Clustering Techniques. Paper presented at the 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT). DOI: https://ieeexplore.ieee.org/document/8869140

[52] M. C. de Souto, I. G. Costa, D. S. de Araujo, T. B. Ludermir, and A. Schliep. (2008). "Clustering cancer gene expression data: a comparative study", BMC bioinformatics, Vol. 9, No. 1, pp. 497. DOI: https://doi.org/10.1186/1471-2105-9-497.

[53] X. Yu, G. Yu, and J. Wang. (2017). "Clustering cancer gene expression data by projective clustering ensemble", PLoS One, Vol. 12, No. 2, pp. e0171429. DOI: https://doi.org/10.1371/journal.pone.0171429.

Houda Fyad

Houda Fyad is an Assistant Professor in computer science at University of Oran 2, Algeria. She received her engineering degree in computer science department (2006) from University of Oran1. She also received her master degree (2011) with specialization Informatique & Automatique from the same university. Currently she is preparing her PhD thesis within the Computer Science Department in the University of Oran1 with specialization Diagnostic and Decision-Making Assistance and Human Interaction Machine. In research field, she works on Machine Learning, Data mining, Bioinformatics and Ontologies.

Fatiha Barigou

Fatiha Barigou is a university lecturer at Computer Science Department at Université Oran 1. She is a research member of the AIR team in the LIO laboratory. She does research in Text Data Mining, Big data and Artificial Intelligence. Her current projects are Sentiment Analysis, AI, Fog and Cloud Computing in healthcare.

Karim Bouamrane

Karim Bouamrane received the PhD Degree in computer science from the Oran University in 2006. He is Professor of computer Science at the same university. He is the head of computer science laboratory (LIO) and Decision and piloting system team. His current research interests deal with decision support system in maritime transportation, urban transportation system, production system, and application of bio-inspired based optimization metaheuristic. He participates in several scientific committees' international/national conferences in Algeria and others countries in the same domain and collaborates in Algerian-French scientific projects. He is co-author of more than 40 scientific publications.

A Holistic Methodology for Improved RFID Network Lifetime by Advanced Cluster Head Selection using Dragonfly Algorithm

Pramod Singh Rathore[1]*, Abhishek Kumar[2], Vicente García-Díaz[3]

[1] Department of CSE, ACERC Ajmer (India)
[2] Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab (India)
[3] Department of Computer Science, Universidad de Oviedo (Spain)

Received 4 November 2019 | Accepted 9 May 2020 | Published 27 May 2020

Abstract

Radio Frequency Identification (RFID) networks usually require many tags along with readers and computation facilities. Those networks have limitations with respect to computing power and energy consumption. Thus, for saving energy and to make the best use of the resources, networks should operate and be able to recover in an efficient way. This will also reduce the energy expenditure of RFID readers. In this work, the RFID network life span will be enlarged through an energy-efficient cluster-based protocol used together with the Dragonfly algorithm. There are two stages in the processing of the clustering system: the cluster formation from the whole structure and the election of a cluster leader. After completing those procedures, the cluster leader controls the other nodes that are not leaders. The system works with a large energy node that provides an amount of energy while transmitting aggregated data near a base station.

Keywords

Wireless Sensor Networks (WSN), Sensor Nodes (SN), CH Node, Cluster, Cluster Head Selection.

DOI: 10.9781/ijimai.2020.05.003

I. Introduction

RFID is a well-known technology used to identify all kinds of objects. It is a fast-emerging technology that is likely to create massive financial gains in industries and in the digital world. The explanation of the "Internet of Things" concept is usually considered together with the RFID technology. Some of the fields of application include transportation, security, product tracking as well as access control. For the primary purposes, RFID systems can be used to register items, also supporting actions such as counting and tracking objects in motion [1].

The RFID Network protocol is tremendously complex. There is a limitation to design the scheme of any RFID network, being the main constraint the fact of providing power. In addition, diminishing power utilization with the energy consumption process is a significant issue in the intended use of the RFID network protocol. Moreover, RFID systems also have to focus on some aspects such as reliability, scalability or acknowledgment. Therefore, low power RFID readers are used to provide secure and reliable communication. For example, the communication mechanisms can create traffic congestion and they are mostly consuming the resources by transmitting information from the node to the base station. There is a requirement of understanding the process of reducing the burden on the communication mechanism using the straight transmission protocol which is being used for communicating with base station directly [2].

* Corresponding author.
E-mail address: pramodrathore88@gmail.com

In these circumstances, readers are usually separated from the base station and, due to the separation; energy cannot be reflected towards the reader. This separation is the basic reason for high consumption of energy in this particular process. Thus, low power batteries face difficulties in reflecting back the signals to readers. Some authors proposed methods on the MAC and network layers for creating some improvements. However, an important problem is still present when there are more than two nodes that want to act as leaders and confront each other to lead other nodes in the network. Clustering methods are the most important techniques to avoid all difficulties acting in the network [3], [4].

Within the existing system, every reader interprets the data and propel straight near the base station, where it reduces the competence of the readers. In favor of improving good organization of readers in the RFID, the network is based on the cluster method (Fig. 1). In every cluster, we consider the most accurate client and the group of clients which are performing similar operations. Thus, before sending the connection request, the cluster will combine all the related data of the client in advance.

In the proposed work, the RFID network life spans extended by an energy-efficient cluster-based protocol. In addition, with the help of the Dragonfly algorithm, the network life is also extended. This creates a large energy node, like a cluster head, that provides a smaller amount of energy while transmitting aggregated data near the base station. To reduce the loss of energy and to increase the efficiency of the network we need to use the clustering mechanism that can manage complex networks with reduced energy consumption.

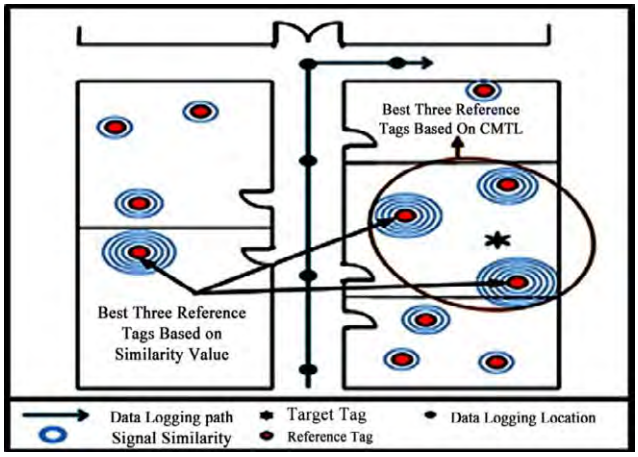


Fig. 1. Cluster-Based RFID Network.

II. Related Work

R. Koh et al. [1] gave a detailed explanation about the establishment of the network and the tracking of the network to identify the packet loss or any kind of miscommunications in between the sender and receiver. The network structure must be maintained nonvolatile and without maintaining any kind of distractions in establishment and the communication. There is a concept mentioned by this author about the Auto -ID which is to track the network including packets which are flowing in the current network. Tracking the system includes the information of the physical device which is a source or destination. In this network, communication source and destination communication devices must be straight forward without any interruption they have to communicate.

J. Kim et al. [2] used the methodologies of the RFID tags to the communication between source and destination which was further used for the reestablishment of the network and also the packet loss. In this scenario we can identify the devices which are connecting in the same network panel. For an example consider the army mission and if the soldiers are connected to the same channel using RFID tags of their device that will be very easy to track the person in any situation. RFID tags are further used for the information processing from source to destination and identifying the receivers based on their tags. Because all the people in the network need not to get the information.

Swimpy Pahuja et al. [5] produced a survey on the RFID tags and the pattern they are communicating among the people and the servers which are connected with the tags will monitor and backup of the communication and the operations are stored as logs. If there is any further same process identified in the same group of network then the same operations and the same log can be referred for the further operations in the network. If we are using the same protocol for the different purpose on the same log the duplication will occur and store the same in the log.

WaleedAlsalih et al. [6] proposed distance-based clustering, which is a technique that is anticipated to reduce the complexity of the network as it takes care of every part of the movements and masking it to make it unaffected from congestion of inactive tags. Shailesh M. et al. [7] suggested a solution for the problem life identifying the collision of the people in the network. If same network was shared between the people, then there will be biggest problem in the network and reconnecting the network or establishing new kind of it. Author suggested new protocols with the new RFID data transmission to recognize the duplication.

The security measures have to be taken very seriously and all the RFID tags must be protected with the latest firewall mechanisms and the information was mentioned in [8]-[10].With low cost we need to mention the security mechanisms and all the mechanisms must be processed according to the willingness of the communicators in the network. Because without any of their intervention nothing must be leaked out of the channel [11], [12].

Routing calculations are being done in this new RFID tags scenario and those are clearly mentioned in [13]. We need to calculate the security levels and proceed further will all kind of operations. Whether it may be secured or unsecured first we need to calculate the channel security and must establish all the related measures. Instead of purchasing the commercial tools it is better to use low cost commodity software or tools for network establishment. That was mentioned on [14]-[16].

Mann et al. presented Bee Swarm, an SI that is based on the energy-efficient hierarchical routing protocol for WSNs [17]. Mirzaie et al. has provided a multi-clustering algorithm that is based on fuzzy logic (MCFL) using a unique methodology that is exhibited to do node clustering in WSN [18].

Elshrkaweyetal., in 2018, proposed an advanced method to minimize energy usage and maximize network lifetime [19]. It needs to support information secrecy, safeguard the WSN and enhance the security. Lalwani et al. [20] presented the biogeography-based energy-saving routing architecture (BERA) for CH selection and routing. Consideration ought to be taken for basic protocols while choosing the CH to improve the life of the system. CSSDA moves further and this movement will require two parts: the first part is the cluster set-up and the second part is the regular cluster. Inside the cluster set up, the reader, who has more enduring power will choose for the cluster heads [21],[22]. The others, which are not chosen for cluster heads, will merge the cluster to their relevant cluster head. Another part is a regular group part, in that each datum established through the cluster heads are combined cumulative and sent toward the base station [23],[24]. In any case, studying the key factors that can be important for the design or routing techniques of wireless networks is a topic of interest in the research community [25].

III. Problem With the Conventional Approach

RFID plays an integral role within the different domains and their requirements are solved with tracking of the person or an object with the RFID tags. If there is a need of tracking animals in the forest, we need to attach a chip to track them. In cases of shopping areas in metro cities, there is no chance of maintaining an inventory of the products and scan them one by one to add in the inventory. In such cases we use these tags to add the bulk of similar products instead of barcodes or QR codes.

The communication channel security at the back end server is one of the most important factors and at the same time tagging the reader messages is necessary for mobile readers. Although, the security issues occur worldwide, and the safety becomes the major issue. This paper means to say that the verification protocol is on the way to ensure that a prohibited tag or reader should break the structure. Hash-based security measure, also unrestricted essential incorruption methods are executed on top of authentication protocol.

We proposed new scalable, anti-counterfeit and undetectable confirmation protocol entrenched in terms of hash method, even public key encryption knowledge for offering the safety and security in RFID tag-reader message procedure. The conventional approach is composed of the following steps:

- De-synchronization. The proposed work eliminates the de-synchronization issue since the server preserves two files of ID (ID also last). Thus, if the message is deemed fraudulent; we know how

to regain it and reverse as of the last ID.

- Anti-counterfeit issue. The Mannequin chip, amid a similar number, is impossible to make when the RFID chip is exclusive. Therefore, a reader might not print identically in order, and our logic is hence explained.
- Forward secrecy. After the modernization of IDi toward IDi+1, there was confusion regarding the one-way encoding method, and there was no other way to produce the original instruction that was ambiguous. That is, IDi+1 cannot create IDi.
- Un-traceability. To track the statement between the entities, using this method, it is hard to compare the two sides, that is the public-key encryption and the time stamp.
- Spoofing. RFID spoofing involves covertly reading and recording a data transmission from a RFID tag. When the data is retransmitted, it would contain the original tag's TID, making it appear to be valid.
- Item privacy. Public, as well as the crucial private encryption decoding, and the encoding of data building is nearly unfeasible to bother the confidentiality of the network.
- Replay attack. Replay attacks build on eavesdropping and specifically occur when one part of communication in an RFID system is recorded and then 'replayed' later to the receiving device in order to steal information or gain access.

IV. Proposed Approach

Through the usage of an energy-efficient cluster-based protocol, the RFID networks may extensively utilize the Dragonfly algorithm during their lifetime (Fig. 2). For every cluster member, the cluster head reader (CH) has a receiver tag charge in order. After, it conveys towards the base station (BS) and executes the aggregation development that received the data. For all RFID Networks, readers obtain the energy stage details from the base station. Under the circumstances, the optimized cluster head is chosen, the base station estimates the average energy levels for every reader presented within the network.

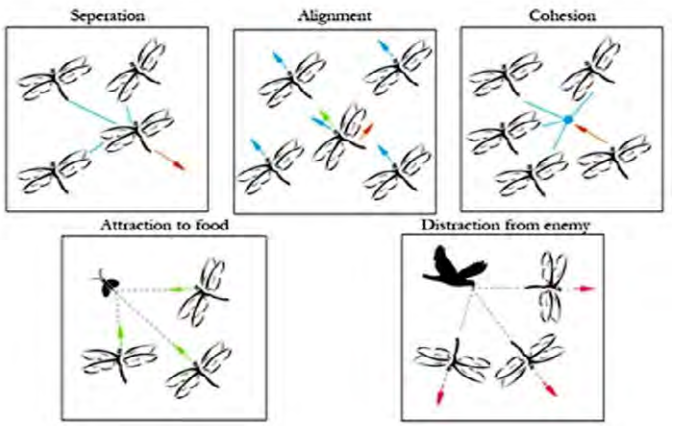


Fig. 2. Structure of the Dragonfly Algorithm.

The cluster head is chosen by the mobile RFID environment where the readers have high power and equivalent mobility, while effect of moving back a few parting readers also increases the network lifetime. On behalf of the dropping movement among the readers into the RFID, the network uses an energy-efficient clustering system. There are two parts of this method. Within the first part, cluster heads are selected on the basis of energy level and mobility of the reader. In the present work the cluster head will be connected to the most ambiguous node which can be solved with connectivity issues of the cluster and the reader can connect to the cluster head easily. Fig. 3 and Fig. 4 illustrate the Cluster Head Selection and the Formation.

The RFID network is enlarging its lifespan through the clustering method. With the aim of energy utilization for every node, the cluster heads are preferred within the clusters. All the readers transmit hello messages in route to their neighbors. In the communication range, every reader reclines and receives the hello message, and recognizes the neighbor and transfer as the reader. Even though inside the network, the entire set of readers may have found their neighbors. With a high probability of attaining, for example, mobility and energy, the cluster heads are chosen. All the readers' energy levels are evaluated among the threshold charge. A threshold value is defined the remaining necessary power get information on or after the entire set of readers aggregates the information and conveys it toward the base station.

Another way to identify a suitable cluster head is by the reader during soaring residual energy that is evaluated to be within the threshold value. Along with the available cluster head, applying Dragonfly Clustering, we choose the optimal cluster heads in the network. After the relevant cluster head is selected, and algorithm of a dragonfly is used. There are three old ethics in the Dragonfly algorithm: i) collision avoidance ii) Segregation iii) nearby Reader's distance.

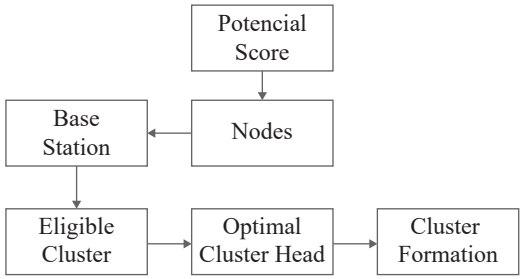


Fig. 3. Dragon Fly Algorithm block diagram.

Distance metrics are used to map the related variables to the specific cluster based on the standard distance metric algorithms. These algorithms identify the similar group elements. Administrators are chosen based on the direction and the speed of the readers that are used for the explanation. The neighbor reader count for every suitable cluster head through a network is known as cohesion. Within the mobile web, there are tags and readers. Given the system, the dynamic behaviors of the readers are three significant features that are worn within the network, the location updates of the nodes are cohesion, alignment and separation. Every element is taken as the best cluster head selection. The paragraph below illustrates that the behaviors are precisely modeled.

Separation: the ambiguity among the cluster head and their neighbor is intended after the cluster head is appropriately elected. To establish a node, it is closely considered by their detachment towards the data transmission in the RFID network. The procedure for separation is given by Eq.(1). The present node location is considered as (x1, y1), whichever reader otherwise tag; the neighboring node location is described as (x2, y2) also neighboring readers, and their count is noted as N between the present reader.

$$S_i = \sqrt{\{(x^2 - x^1)^2 + (y^2 - y^1)^2\}/N} \quad (1)$$

Alignment: after separation, as the moment of the cluster head, the RFID tag can recognize location of the object with respect the specific cluster head. The association of the cluster head should exist in parallel toward neighbor readers in the direction of the cluster headed for evading the RFID network rupture the cluster. Through the alignment, every mobility node is determined. The method of alignment is given by Eq.(2). Everywhere, V_j illustrates what is within the network

mobility of the n-*th* adjacent neighbor node. Later then the speed is choosing the same direction as cluster head direction.

$$A_i = (\sum_{j=1}^{N} V_j)/N \quad (2)$$

Cohesion: when the procedure of alignment is finished, the adjacent neighbor node is intended to favor the appropriate cluster heads. The neighbor nodes is said to be the number of nodes around the cluster head, for which the calculation is not as much of the cluster, which is synchronized, and becomes exaggerated. Therefore, anode must always remain separated to avoid collision; this is useful for the network and their competence. The prescription for cohesion is described in Eq.(3).

$$C_j = (\sum_{j=1}^{N} V_j)/N - X \quad (3)$$

Where, X is the position of the current individual, N is the number of neighborhoods and X_j is the position of the jth neighboring node in the RFID network.

The cohesion, alignment and separation ranges are added for every cluster head. Here, the cluster head will be chosen as a head. Based on low mobility, distance, and neighbor count, the cluster head will be selected.

After the optimal cluster head is chosen (based on the cluster) the cluster head mobility is formed due to the presence of readers and tags. In the network, to avoid the collision and termination of communication amongst the nodes, the cluster head must be selected with majority.

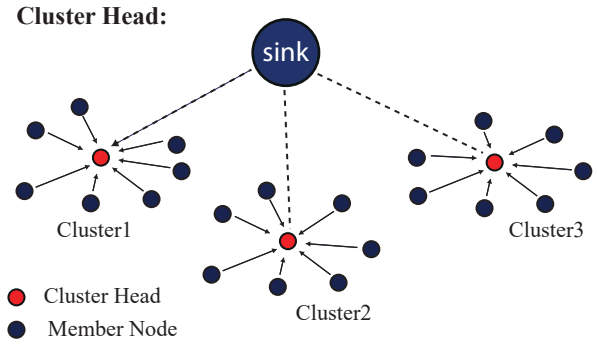


Fig. 4. Dragon fly algorithm representation.

In addition, the algorithm for cluster head selection and cluster formation is as follows:

Step 1: Initialize the nodes and labels in the system Ri

(i=1, 2, 3...n).

Step 2: Set the potential score (vitality and portability) to every node and label hubs.

Step 3: Send the vitality level of the Reader to Base Station and discover optimum nodes.

Step 4: If (Residual Energy (Ri)> Threshold esteem).

Then: Update the node as an Eligible Cluster Head.

else: Update the nodes as Remaining nodes in the system.

Step 5: Separation is determined for a qualified head in the system utilizing (Eq. (1)).

Step 6: Alignment is determined for the head in the order using Eq. (2).

Step 7: Cohesion is defined as a qualified leader in the system utilizing Eq. (3).

Step 8: Add the estimations of partition, arrangement, and union.

Step 9: If the worth is high, select as "Ideal head."

Step 10: else "Become customary nodes" in the system.

Step 11: End the Cluster development.

In the RFID network, taking into consideration the potential value of the readers, we can determine the optimal cluster head by employing a cluster formation algorithm. To obtain an adequate cluster configuration in the network, some procedures like separation or cohesion are applied.

The optimal set of cluster heads that belong with their associated cluster members are predicted by the base station. In the RFID network, the cluster head plays an important role to send the data from one cluster ID and vice versa. Cluster head also plays a role of local control center to organize this event and cluster base station acted as a Reader arranger.

Regarding data transmission, every reader begins to pass a signal for sensing the information after the cluster formation and cluster head selection. Within the clustering schedule, cluster head works as readers to detect data. The TDMA schedule is utilized in the cluster for arranging the readers. The reader throws the signal to tags with their corresponding range pedestal of the program. The reader sends the information to the cluster head once it senses the information. Subsequently, the next reader begins to pass the data. With the help of a cluster head, each reader finishes the data transmission and data aggregation.

V. Experiment

An object-driven network simulator, network simulator version 2 (NS-2), was developed at the University of California-Berkeley. Such a simulator utilizes two programming languages: C++ and Tcl. NS-2 is useful for the simulation of the wide-area and local networks. These programming languages are used for numerous reasons, most importantly, because of their internal characteristics. While C++ provides efficiency in its implementation of a specific design, it encounters some difficulties in graphic representation. Without a visual language that is easy-to-use and descriptive, it may be challenging to perform a modification and assembly of different components and alter distinct parameters.

The system test system is typically called an NS2; it is a capable test system for concentrating dynamic behavior of portable remote sensor organizers. NS2 boosts the re-enactment of an order from a physical radio transmission channel to the application layer. The NS2.35 test system is utilized for re-enactment and is directed under the Linux mint environment. This testing was conveyed using the standard network test system, NS-2.34, which has 100 nodes. These nodes are spread by an arbitrary request in a 100 x 100 area.

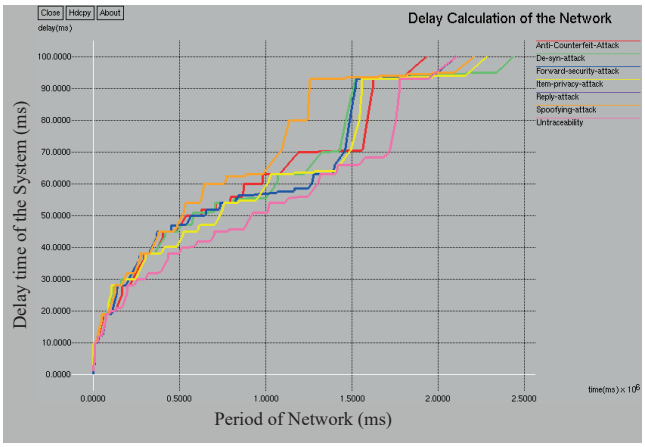


Fig. 5. Delay Calculation of the network.

While using any kind of implementation in the network the throughput must be identified and the latency of the connectivity

should be very less with respect to the implementation done in the present existing methodology. Fig. 5 explains the implementation of the delay calculation in the network. This delay can be identified while connecting the RFID tags to the base station to transform information from one location to another location. The existing system deals with the highest delay problem in connecting the server and with the problem with reconnection when there is any network drop.

In the existing system, several attacks are analyzed, and the results are calculated. Fig. 6 shows the calculation of the throughput of the network. The network throughput is defined as the number of packets transmitted according to the period. Hence, it is a two-dimensional figure that consists of two axes.

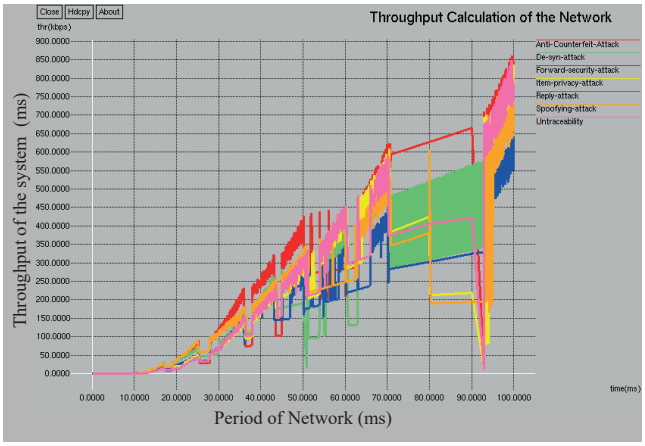


Fig. 6. Throughput Calculation of the network.

The through put can be exhibited with the network connecting time and the data acquisition from different repositories. In this data acquisition we gather data from different knowledge bases and connect the client to the base station.

In the existing system, several attacks are analyzed, and the results are calculated. Fig. 7 shows the calculation of the forwarding security of the network. The forward network security is defined as that the amount of count of security packets in the forward direction according to the time. Hence it is the two-dimensional figure. It consists of two axes.

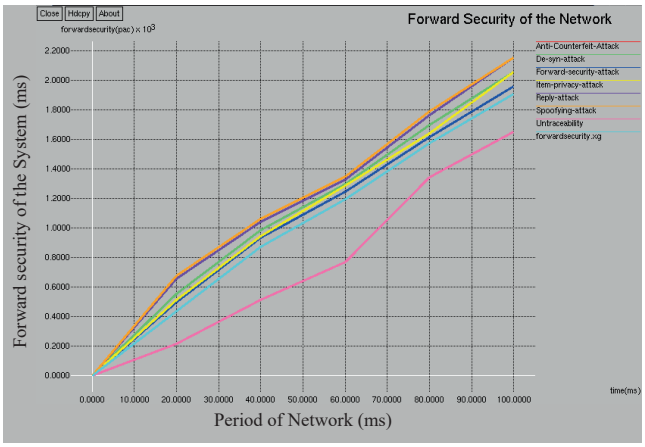


Fig. 7. Forward Security Calculation of the network.

There are different approaches to calculate the threat in the network and the best way to analyses the network security is by maintaining forward network security.

In the existing system, several attacks are analyzed, and the results are calculated. Fig. 8 shows the calculation of the received packets in the network. The network received packages is defined as that the number of packets received by the destination during the transmission of packets from the source to the destination, according to the period. Hence, it is a two-dimensional figure that consists of two axes.

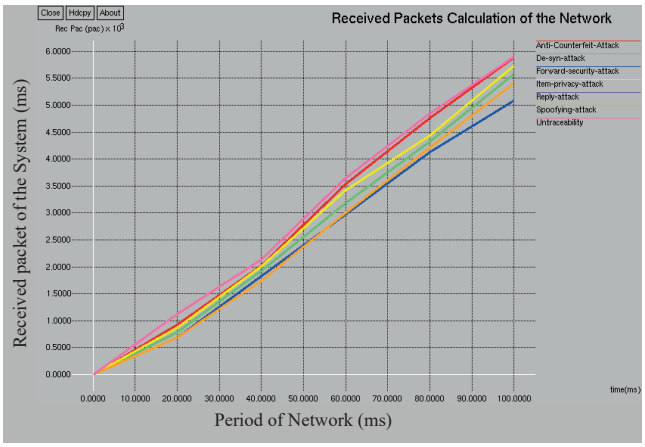


Fig. 8. Received Packets Calculation of the network.

Before completing the transmission we need to calculate the number of packets we are transmitting from one node to another and after completing the transmission we need to check the whether all the packets is received or not.

In the existing system, several attacks are analyzed, and the results are calculated. Fig. 9 shows the calculation of the sent packets in the network. The network sent packages is defined as the number of packets transmitted by the source (sender node) during the transmission of packets from the source and to the destination, according to the period. Hence, it is a two-dimensional figure that consists of two axes.

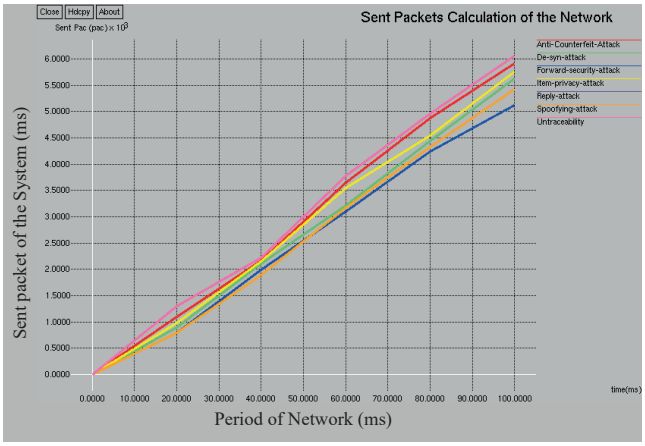


Fig. 9. Sent Packets Calculation of the network.

Fig. 10 shows the calculation of energy consumption, and it is a comparative analysis between the existing and the proposed method. The network energy consumption is defined as the amount of energy utilized or spent during the data transmission between the source and the destination, according to the period, until the end of the communication. Hence, it is a two-dimensional figure that consists of two axes.

Fig. 11 shows the calculation of energy efficiency, and it is a comparative analysis between the existing and the proposed method. The network energy efficiency is defined as the amount of energy saved during the data transmission, between the source and the destination, according to the time period and until the end of the transfer. Hence, it is a two-dimensional figure that consists of two axes.

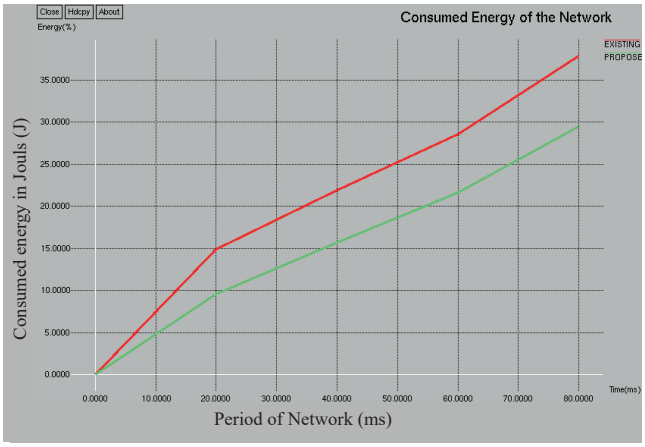


Fig. 10. Consumed Energy Calculation of the network.

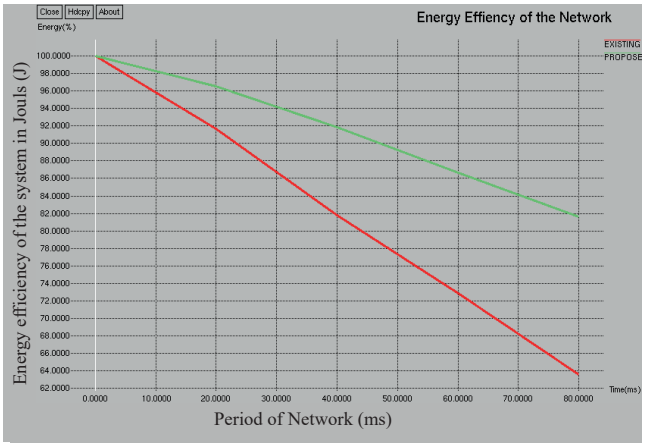


Fig. 11. Energy Efficiency Calculation of the network.

The energy of the network must be identified before performing any kind of operation. In this energy efficiency the network must use the minimum energy and execute the large task.

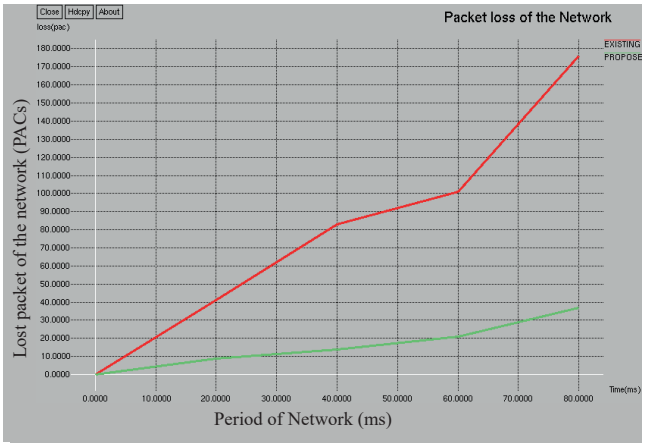


Fig. 12. Packet Loss Calculation of the network.

Fig. 12 shows the calculation of packet loss and is a comparative analysis between the existing and the proposed method. The network packet loss is defined as the number of packets lost during the data transmission between the source and the destination according to the time until the end of the communication. Hence, it is a two-dimensional figure and consists of two axes.

While transmitting the information the task we need to consider is the avoiding packet loss. While transmitting information sometimes because of problem in network establishment packets may loss. We need to avoid and calculate those and recover those.

Fig. 13 shows the calculation of generated packets and is a comparative analysis between the existing and the proposed methods. The system made packets are defined as that the number of packages sent during the data transmission between the source and the destination, according to the time period until the end of the transfer. Hence, it is a two-dimensional figure and consists of two axes.

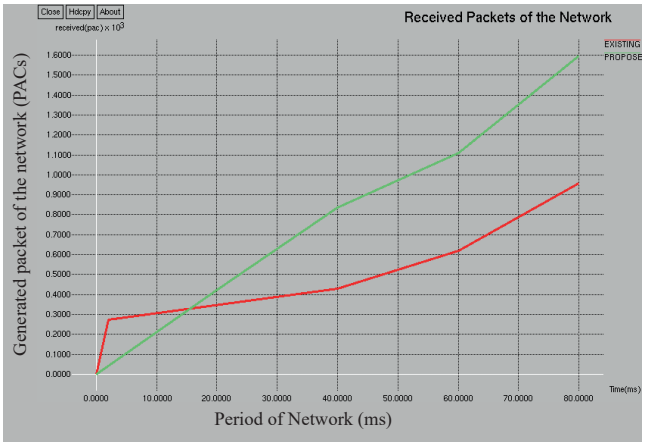


Fig. 13. Generated packets Calculation of the network.

The existing and proposed systems are compared with the service of different network admins while transmitting the information. We need to calculate the packets which are being generated.

Fig. 14 shows the calculation of received packets. This is a comparative analysis between the existing and the proposed methods. The system received packages, defined as the number of packets received during the data transmission between the source and the destination according to the time period and until the end of the transfer. Hence, it is a two-dimensional figure that consists of two axes.

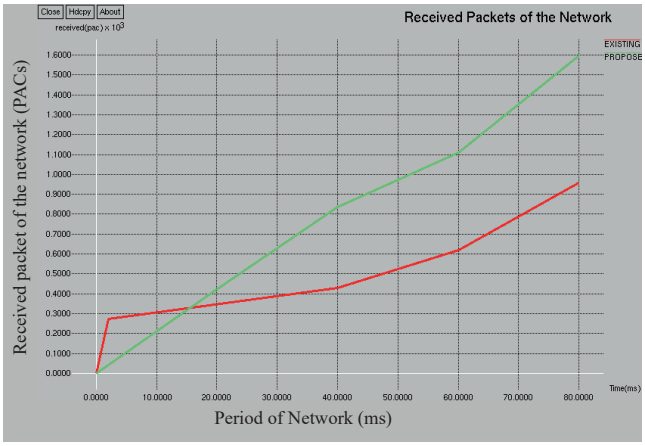


Fig. 14. Received packets Calculation of the network.

Fig.15 shows the calculation of the packet's delivery ratio, and it is a comparative analysis between the existing and the proposed method. The network packets delivery ratio is defined as that the percentage of packets received vs. the number of packages sent during the data transmission between the source and the destination according to the time period till the end of the transfer. Hence it is the two-dimensional figure. It consists of two axes.

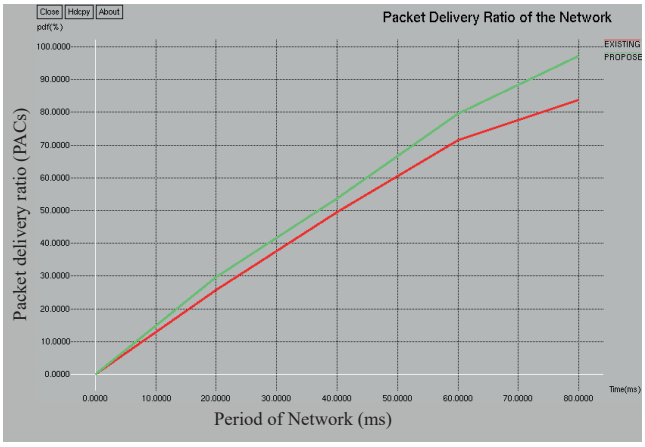


Fig. 15. Packets Delivery Ratio Calculation of the network.

Fig. 16 shows the calculation of throughput, and it is a comparative analysis between the existing and the proposed methods. The network throughput is defined as that the overall calculation of the number of packets sent during the data transmission between the source and the destination, according to the time period and until the end of the transfer. Hence, it is a two-dimensional figure that consists of two axes.

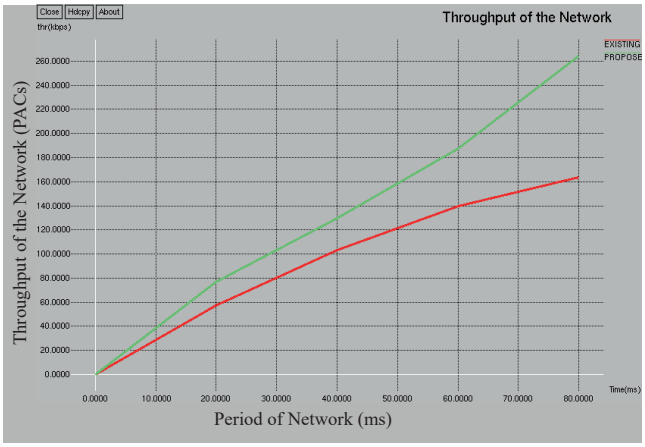


Fig. 16. Throughput Calculation of the network.

VI. Conclusions and Future Work

The network environment is designed and implemented using Network Simulator. The simulator executes the proposed dragonfly clustering protocol for cluster head selection and cluster formation to reduce the cluster breakage and to improve the reading efficiency in the network. In the considered 100 nodes in the network, there are three types of sensors such Readers, tags and cluster heads. All the readers are homogeneous in the cluster but perform different tasks. This type of distribution balances the operational load within each cluster and also results in improved network lifetime. The cluster head schedules data collection time in the network. Readers sense the data from tags and send to cluster head within the cluster. Cluster head performs aggregation of the gathered data before transmitting them to the base station. Our simulation result is evaluated in terms of number of parameters such as network lifetime (number of Active nodes) and cluster head selection rounds.

Earlier work in this domain had tried to provide a genuine solution through malicious attack. Our proposed work is an effort in the same direction. In this work the hash function and public-key encryption with timestamping on either side of a two-way communication are used to remove minimal chances of an attack on a wireless connection.

In the context of future work, it is necessary to emphasize the *IEEE* issue of time complexity; the next step will be to evaluate the time function, and to minimize it to make the system more efficient in a real-time scenario.

References

[1] R. Koh, E. W. Schuster, I. Chackrabarti, and A. Bellman, "Securing the Pharmaceutical Supply Chain," *White Paper MIT-AUTO ID-WH- 021, Auto-Id Center MIT,* Cambridge, MA 02139-4307, USA, 2003, Available at http://www.mitdatacenter.org

[2] J. Kim, D. Choi, I. Kim, and Kim H., "Product authentication service of consumer's mobile RFID device," *IEEE 10th International Symposium on Consumer Electronics*, pp. 1-6, 2006.

[3] Kim, J. and Kim, H., "A wireless service for product authentication in mobile RFID environment", *1st International Symposium on Wireless Pervasive Computing,* pp. 1-5, 2006.

[4] C. –L. Chen, Y. –Y. Chen, T. –F. Shih, T. –M.Kuo, "An RFID Authentication and Anti-counterfeit Transaction Protocol, " *International Symposium on Computer, Consumer and Control*, 2012, pp. 419-422.

[5] S. Pahuja, S. Negi, A. Verma, P. Rathi, N. Narang, R. Chawla, "An Authentication Protocol for secure tag-reader communication," IEEE Students' *Conference on Electrical, Electronics and Computer Science,* 2012.

[6] W. Alsalih, K. Ali, and H. Hassanein, "Optimal distance-based clustering for tag anti-collision in RFID systems," 2008 *33rd IEEE Conference on Local Computer Networks (LCN), Montreal, Que, 2008*, pp. 266-273, doi: 10.1109/LCN.2008.4664179.

[7] Birari S.M., Iyer S., "PULSE: A MAC Protocol for RFID Networks" in Enokido T., Yan L., Xiao B., Kim D., Dai Y., Yang L.T. (eds) *Embedded and Ubiquitous Computing – EUC 2005 Workshops. EUC 2005.* Lecture Notes in Computer Science, vol 3823. Springer, Berlin, Heidelberg.

[8] P.D'Arco and A. De Santis, "On ultralightweight RFID Authentication Protocols," *IEEE Transactions on Dependable and Secure Computing*, vol. 8, no. 4, July/August 2011.

[9] J. Yang, K. Ren and K. Kim, "Security and Privacy on Authentication Protocol for Low-Cost Radio", *Proc. 2005 Symp. Cryptography and Information Security,* 2005.

[10] J. Wang, D.Wang, Y.Zhao, and T. Korhonen, "A Novel Anti-Collision Protocol with Collision based Dynamic Clustering in Multiple-Reader RFID Systems," *International Conference on Applied Informatics and Communications*, 2008, pp. 417-422.

[11] M.V. Bueno Delgado, J. Vales Alonso, F.J. Gonzalez Castaño, "Analysis of DFSA Anti-collision Protocols in passive RFID environment" *2009 35th Annual Conference of IEEE Industrial Electronics,* Porto, 2009, pp. 2610-2617, doi: 10.1109/IECON.2009.5415261.

[12] M.Shuang, Y. Xiao-long, "An Efficient Authentication Protocol for Low-Cost RFID System in the Presence of Malicious Readers" 2012 9th *International Conference on Fuzzy Systems and Knowledge Discovery(FSKD 2012)*, pp. 2111-2114.

[13] S. Pahuja, S. Negi, A. Verma, P. Rathi, N. Narang, R. Chawla, "An Authentication Protocol for secure tag-reader communication," *2012 IEEE Students' Conference on Electrical, Electronics and Computer Science*, Bhopal, 2012, pp. 1-4, doi: 10.1109/SCEECS.2012.6184757.

[14] C. –L. Chen, Y. –Y. Chen, T. –F. Shih, T. –M.Kuo, "An RFID Authentication and Anti-counterfeit Transaction Protocol" *2012 International Symposium on Computer, Consumer and Control*, 2012, pp. 419-422.

[15] M.Shuang, Y. Xiao-long, "An Efficient Authentication Protocol for Low-Cost RFID System in the Presence of Malicious Readers" 2012 9 *International Conference on Fuzzy System and Knowledge Discovery (FSKD 2012)*, pp. 2111-2114.

[16] L. F.Bittencourt, E. R. Madeira, F.Cierre, and L.Buzato, "A path clustering heuristic for scheduling task graphs onto a grid," in 3rd *International Workshop on Middleware for Grid Computing (MGC05)*, 2005.

[17] P. S. Mann, and S. Singh, "Energy-efficient hierarchical routing for wireless sensor networks: a swarm intelligence approach." *Wireless Personal Communications*, vol. 92, no. 2, 2017, pp. 785-805.

[18] M. Mirzaie, and S. M. Mazinani, "MCFL: An energy-efficient multi-

clustering algorithm using fuzzy logic in the wireless sensor networks." *Wireless Networks*, vol. 24, no. 6, pp. 2251-2266, 2018.

[19] M. Elshrkawey, S. M. Elsherif, and M. E. Wahed, "An enhancement approach for reducing energy consumption in wireless sensor networks." *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 2, pp. 259-267, 2018.

[20] P. Lalwani, H. Banka, and C. Kumar, "BERA: a biogeography-based energy-saving routing architecture for wireless sensor networks," *Soft Computing*, vol. *22*, no. 5, pp. 1651-1667, 2018.

[21] S.-C. Wang, "Artificial Neural Network," in *Interdisciplinary Computing in Java Programming, Boston, MA: Springer* US, 2003, pp. 81–100.

[22] V. Estivill-Castro, "Why so many clustering algorithms," *ACM SIGKDD Explore. News*., vol. 4, no. 1, pp. 65–75, Jun. 2002.

[23] B. S. Harish, B. S. Kumar, "Anomaly-based Intrusion Detection using Modified Fuzzy Clustering," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 6, pp. 54-59, 2017.

[24] García CG, Núñez-Valdez ER, García-Díaz V, Pelayo G-Bustelo C, Cueva-Lovelle JM. "A Review of Artificial Intelligence in the Internet of Things." *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 5, no. 4, pp. 9-20, 2019.

[25] Bahuguna, Y., D. Punetha, and P. Verma. "An analytic Study of the Key Factors Influencing the Design and Routing Techniques of a Wireless Sensor Network." *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 4, no. 3, pp. 11-15, 2017.

Pramod Singh Rathore

Pramod Singh Rathore is pursuing his Doctorate in computer science & Engineering from Bundelkhand University and research is going on Networking and done M. Tech in Computer Sci. & Engineering from Government engineering college Ajmer, Rajasthan Technical University, Kota India. He has been working as an Assistant professor of Computer Science & Engineering Department at Aryabhatt Engineering College and Research centre, Ajmer, Rajasthan and also visiting faculty in Government University MDS Ajmer. He has total Academic teaching experience of more than 8 years with more than 45 publications in reputed, peer reviewed National and International Journals, books & Conferences like Wiley, IGI GLOBAL, Taylor & Francis Springer, Elsevier Science Direct, Annals of Computer Science, Poland, and IEEE. He has co-authored & edited many books with many reputed publisher like Wiley, CRC Press, USA. .His research area includes NS2, Computer Network, Mining, and DBMS.

Abhishek Kumar

Abhishek Kumar is Doctorate in computer science from University of Madras done M.Tech in Computer Sci. & Engineering from Government engineering college Ajmer, Rajasthan Technical University, Kota India. He has total Academic teaching experience of more than 8 years with more than 60 publications in reputed, peer reviewed National and International Journals. His research area includes- Artificial intelligence, Image processing, Computer Vision, Data Mining, Machine Learning. He has authored 6 books published internationally and edited 16 book with Wiley, IGI GLOBAL Springer, Apple Academic Press and CRC etc. He is also member of various National and International professional societies in the field of engineering & research like Senior Member of IEEE. He has got Sir CV Raman life time achievement national award for 2018 in young researcher and faculty Category.

Vicente García-Díaz

Vicente García-Díaz is an Associate Professor in the Department of Computer Science at the University of Oviedo. He has a PhD in Computer Science from the University of Oviedo and a Diploma in Advanced Studies, as well as Degrees in Computer Engineering and Technical Systems Computer Engineering. In addition, he possesses a Degree in Occupational Risk Prevention. He is part of the editorial and advisory board of several international journals. He has supervised 90+ academic projects and published 90+ research papers in journals, conferences, and books. His teaching areas are algorithm design techniques, and design and development of Domain-Specific languages. His research interests also include decision support systems and the use of technologies in teaching and learning.

Incremental Hierarchical Clustering driven Automatic Annotations for Unifying IoT Streaming Data

Sivadi Balakrishna[1]*, M.Thirumaran[1], Vijender Kumar Solanki[2], Edward Rolando Núñez-Valdez[3]

[1] Department of Computer Science and Engineering, Pondicherry Engineering College, Pondicherry (India)
[2] Department of Computer Science and Engineering, CMR Institute of Technology, Hyderabad, TS (India)
[3] Department of Computer Science, University of Oviedo (Spain)

Received 27 September 2019 | Accepted 23 January 2020| Published 21 March 2020

Abstract

In the Internet of Things (IoT), Cyber-Physical Systems (CPS), and sensor technologies huge and variety of streaming sensor data is generated. The unification of streaming sensor data is a challenging problem. Moreover, the huge amount of raw data has implied the insufficiency of manual and semi-automatic annotation and leads to an increase of the research of automatic semantic annotation. However, many of the existing semantic annotation mechanisms require many joint conditions that could generate redundant processing of transitional results for annotating the sensor data using SPARQL queries. In this paper, we present an Incremental Clustering Driven Automatic Annotation for IoT Streaming Data (IHC-AA-IoTSD) using SPARQL to improve the annotation efficiency. The processes and corresponding algorithms of the incremental hierarchical clustering driven automatic annotation mechanism are presented in detail, including data classification, incremental hierarchical clustering, querying the extracted data, semantic data annotation, and semantic data integration. The IHC-AA-IoTSD has been implemented and experimented on three healthcare datasets and compared with leading approaches namely- Agent-based Text Labelling and Automatic Selection (ATLAS), Fuzzy-based Automatic Semantic Annotation Method (FBASAM), and an Ontology-based Semantic Annotation Approach (OBSAA), yielding encouraging results with Accuracy of 86.67%, Precision of 87.36%, Recall of 85.48%, and F-score of 85.92% at 100k triple data.

Keywords

IoT Sensor Data, Semantics, Automatic Annotation, Incremental Hierarchical Clustering, Healthcare, Agent, SPARQL.

DOI: 10.9781/ijimai.2020.03.001

I. Introduction

THE semantic technologies address the problem of various heterogeneous devices, communication protocols, and data formats of the generated data in the Internet of Things. Annotation of IoT sensor data is the substance of IoT semantics [1]. The future generation of IoT not only deals with the physical sensor devices but also the meanings they carry with virtual representation of smart data. On an average, every day around 3.2 quintillion bytes of data are generated on the Internet. The CISCO predictions state that more than 60 billion devices will be connected to the internet by 2025, as a result zetta bytes of sensor data will be generated continuously and exponentially. The IoT sensors generated raw data is stored in the data repositories and it supports to heterogeneous smart city applications. Therefore, applying the raw data into applications may result in structural data with pre-notified format, date, source, affiliation, unit, and encryption. The next level of data is perception data that contains the multi abstraction from low-level to high-level applications to perform actionable and predictive data for the final evaluation. For understanding the perception data more concisely, the structural information is needed. Without structural information, the data may mislead to false results and may fail to integrate the real-time application data [2]. The perception data is extracted from the structured data that is more compressive and occupies less space than the raw data. Machine Learning (ML) clustering techniques are used for performing analysis on the perception data and automatic generation of semantic annotations. Moreover, in IoT, the real-time streaming data plays a major role to perform cluster analysis. The streaming data is flowing continuously as data stream from the IoT device to the peer network. The stream processing has been effectively analyzes the cluster data, improve the cluster efficiency, and able to make quicker decisions on clustered data [3]-[5].

The hierarchical clustering techniques are used for representing logical, temporal, and spatial relations on the IoT streaming data. The most important aspect of clustering IoT streaming data is its dynamic and heterogeneous nature. Therefore, a novel clustering mechanism is needed to represent the hierarchical relationships-based annotations for the IoT streaming data [6]. In this paper, incremental hierarchical clustering is deployed for unifying the streaming data in a hierarchical manner. SPARQL queries are used for extracting semantic annotations between the hierarchical clustered data. The agents will receive the raw data streams as input data from the IoT sensor devices and then perform the classification between the data streams for generating the RDF data patterns for the hierarchical clustering. The RDF data patterns are combined with the pre-notified metadata of the IoT sensors for the incremental hierarchical clustering process. At last, the hierarchical streaming data is annotated with the automatic semantic annotations using SPARQL queries.

* Corresponding author.
E-mail address: balakrishna.sivadi@pec.edu

Semantic annotation has mainly taken from the field of text annotation. It provides machine-readable descriptions along with labels for URIs. Dealing with IoT semantic data is a difficult and challenging task for researchers and developers with technical issues. To solve this problem on providing the manual annotation and semi-automatic annotation, one approach for providing a semantic annotation to IoT semantic data is proposed [7]-[8]. Using manual annotation and semi-automatic annotation cannot be applicable if the IoT sensor data is huge in volume. It consumes more time to annotate the huge data and unable to capture the IoT devices generated data [9]. Therefore, a new and innovative automatic semantic annotation with more efficient mechanisms are needed.

The main contributions of this work are listed as follows: Firstly, build an architectural model using hierarchical clustering driven automatic annotation for unifying IoT streaming data. Thereafter, add semantic annotations using SPARQL queries. Then extract and visualize the streaming data using the proposed IHC-AA-IoTSD mechanism and SPARQL queries. Afterwards, find the performance evaluation of the proposed model. Finally, comparison has been made of the proposed architectural model with existing approaches.

The remainder of this paper is described as follows: section II discuss background of the related work and the state of the art schemes. In section III, the authors discuss the proposed mechanism Incremental Hierarchical Clustering based Automatic Annotation for IoT Streaming data (IHC-AA-IoTSD). Experimental Methodology and Evaluation are described in section IV and section V respectively. Finally, section VI concludes this work along with the future scope.

II. Background and Related Work

In this section, the related work of semantic annotations in IoT platforms for unifying streaming data in efficient way is discussed. Majority of the researchers has put their efforts on how to deal with big volume and variety of data generated by IoT devices. As a result, ontologies and standards, mapping technologies and exchange systems, semantic annotations, data integration, interoperability, scalability, cluster efficiency and energy-efficient issues are identified. In semantic annotations, manual and semi-automatic annotations are time consuming and perform the annotation process with labels and manually. In addition, these all are dealing with web documents, text documents, and sensor networks. While thinking of Cyber-Physical Systems (CPS) and IoT dynamic data, it generates the big volume of data, therefore, it requires automatic annotation for handling large dynamic data.

Annotation is the process of adding additional information to the existing data, which is enriched with labels, keywords, things, etc. Semantic annotation is the term of enriching data with meanings and descriptions. Annotation plays a major role by providing semantics between humans and machines. These are categorized as three ways-Manual annotation, Semi-automatic annotation, and Automatic annotation. In manual annotation, the data is annotated manually. Here keywords are used for annotating the additional information with existing data. Humans with their self-imagination annotate the keywords. Therefore, it yields the highest accuracy, but it consumes more time to complete the entire triple data. In semi-automatic type of annotation, some part is carried out with keywords and the rest of the part is finished with trained pre-defined set automatically [21]. Two steps complete this process. In the first step, the annotator can annotate the data with keywords. In the second step, the semantic annotation tools are used to toggle the data. Both accuracy and efficiency are improved in this type of annotation system. Automatic annotation is the advanced and recently used annotation system by developers and researchers. In this, the whole process is measured by the annotation system. Annotation tools like Gruff (https://franz.com/agraph/gruff/), Jena (http://jena.apache.org/), and Protégé (https://protege.stanford.edu/) are used for this approach. Based on the instructions given by a user, the annotation tool will place corresponding predicates among the subject and object. At last, a meaningful label and property are assigned to it.

The existing research work on semantic annotation, majority of researcher's intention have been attentive on the semantic based Web documents, and a few researches pay attention to the IoT streaming data based automatic semantic annotation. As shown in Table I, the authors has been associated the former semantic annotation methods in seven aspects, such as "Automatic Annotation", "Semi-Automatic Annotation", "Manual Annotation", "Training Data Set", "Application Specific Domain", "Data Type based on" and "Model/Framework/Technology used". In Table I, the authors has been deliberate based seven aspect and it indicates the following:

- Supreme of the annotation methods focus on the Internet field and are applied for Web documents.
- The existing research of semantics for Web documents primarily pay attention towards Ontology based annotation methods.
- Majority of the existing works on semantic annotation methods in the IoT data are manual. Furthermore, they primarily focus on architectural models and deployable frameworks.

Nowadays, the methods compared in Table I are the most powerful and popular mechanisms to achieve semantic data integration in IoT platforms. The existing data models are updates with semantic annotations on providing semantic labels to become model elements. Kolozali et al. [23] proposed SensorSAX and SAX (Symbolic Aggregate Approximation) methods for adaptive and non-adaptive window size segmentation of data streams real-time processing. Their algorithms are efficient in improving data aggregation in streaming data. However, these are unfair while annotating the IoT dynamic data. Mazayev et al. [24] proposed a CoRE framework for data integration and profiling of objects, as a result, it facilitates semantic data annotation, validation of results, and reasoning of annotated data. This framework adopted the RESTful resources for validating the user profiling of objects with the COAP server. However, the proposed framework is limited for validating and annotating IoT dynamic data efficiently. Mayer et al. [25] developed an Open Semantic Framework (OSF) for industrial IoT applications to make the web of things into semantic web of things. This framework is widely designed to enable the industrial things with semantics to the IoT domains. However, the OSF is not implemented under consideration of various industrial applications.

Shi et al. [26] concentrated on data semantization in IoT applications. They reviewed and overviewed all architectural elements and applications supported for IoT domain. In addition, they surveyed on how to add semantics to the IoT dynamic data, discussed on current research issues and challenges faced by semantic scholars. However, they limited to perform analysis on IoT data integration techniques. Zamil et al. [27] have proposed automatic data annotation techniques for smart home environments by adopting temporal relations. In addition, they incorporated HMM and Random Field models for integrating temporal and spatial relations enhanced by detection accuracy rate. The produced results are moderate and there is a space for enhancement with other incremental clustering techniques. Moutinho et al. [28] have extended the semantic annotations for integrating XML-messages using generating translators under the domain of arrowhead framework. These annotations are not automatic and only domain specific. Therefore, it consumes more time and space for annotating the IoT dynamic data.

An exhaustive and optimistic survey has been conducted under the literature survey. Nevertheless, these all do not light the prerequisites

TABLE I. Comparison of Semantic Annotation Methods

Approaches/ Methods	Automatic Annotation (Yes/No)	Manual Annotation (Yes/No)	Semi-Automatic Annotation (Yes/No)	Training data set	Application Specific Domain	Data Type based on	Model/Framework Technology used
SRSM and MTCRF [9]	No	Yes	Yes	No	Internet	Web documents	Rule, CRFs
Chen et al., SSMIMCR [10]	No	Yes	Yes	No	Internet	Web documents	Conceptual relationships
De Maio et al., FBASAM [11]	Yes	Yes	Yes	Yes	Internet	Web documents	Relational concept analysis
Barnaghi et al., SM2SS [12]	No	Yes	Yes	Yes	IoT	Sensor networks	Sensor streams model
Kolozali et al., KBA4IoTDS [13]	No	Yes	Yes	Yes	IoT	IoT data streams	IoT data model
Wei and Barnaghi et al., SAM4SD [14]	No	Yes	Yes	No	IoT sensor network	Sensor networks	Sensors streams model
Chenyi et al., SOESAF [15]	No	Yes	Yes	No	IoT	IoT entity information	Entity semantic annotation framework
Bing, et al., SAM4IoTD [16]	No	Yes	Yes	No	IoT	Documents	Rule
Ming et al., SAM4WSDL [17]	No	Yes	Yes	Yes	IoT	WSDL documents	Rule, Machine learning
Charton et al., ASAM4NE [18]	Yes	Yes	Yes	No	Internet	Web documents	Semantic similarity, linked data
Diallo et al., OBSAA [19]	Yes	Yes	Yes	Yes	Biomedicine	Biomedical Texts	NLP, TF-IDF
Ahmed E. Khaled and Sumi Helal, ATLAS[22]	Yes	Yes	Yes	No	IoT	Text Labelling	Topic, REST
IHC-AA-IoTSD (Proposed)	Yes	Yes	Yes	Yes	IoT	Streaming sensor data	Hierarchical clustering, automatic annotations, SPARQL Queries

of the semantic scholars and users for adding automatic semantic annotations in IoT streaming data. Therefore, in this paper, we present IHC-AA-IoTSD mechanism using SPARQL queries to improve the clustering based annotating process in IoT sensor streaming data. Through a unification of machine learning and semantic technologies, the proposed approach gives better results in terms of efficiency, reliability, scalability and security compared to the state of the art schemes.

III. Proposed Mechanism

In this proposed research work, to achieve semantic annotations among data samples, a Resource Description Framework (RDF) is used to annotate the data objects in meaningful way. The authors have analysed an incremental hierarchical clustering driven automatic annotation architectural model based on IoT for unifying the streaming data. For this reason, in this paper, a new and novel IHC-AA-IoTSD mechanism is proposed for annotating the streaming data semantically. The Fig. 1 shows an overview of the simplified architectural model of the proposed work. At first, the data generated from IoT sensors are collected from IoT sensor data world. On identification of the sensor data, then the agents will classify and analyze the data. The IoT streaming data generated from the data repository section; firstly, to interpret the objects in the streaming data, the RDF framework is used. Secondly, to abstract the data from the triple store, SPARQL queries is required. The SPARQL Query Engine mainly consists of three subcomponents. Those are Query Parser (QP), Query Optimizer (QO), and Query Processor (QP). The Query Parser (QP) is used for generating the triple patterns in a sequential manner. With the use of the Query Optimizer (QO), the SPARQL queries are optimized and processed. This task is accomplished before it goes to the next component called Query Processor (QP). The SPARQL Query Engine depicts the overall picture and model of the proposed approach. Each component workflow descriptions discussed as follows:

A. Query Parser

This is the first subcomponent of the Query Execution Engine. This subcomponent finds the input healthcare related SPARQL queries from users, abstracts subsequent resources for the consequent subcomponent named as Query Optimizer (QO) and produces a node list for the Query Processor (QP). In this work, we used only basic SPARQL queries with simple SELECT and WHERE clauses. The proposed approach also supports other clauses, such as ORDERBY, GROUP BY, and FILTER.

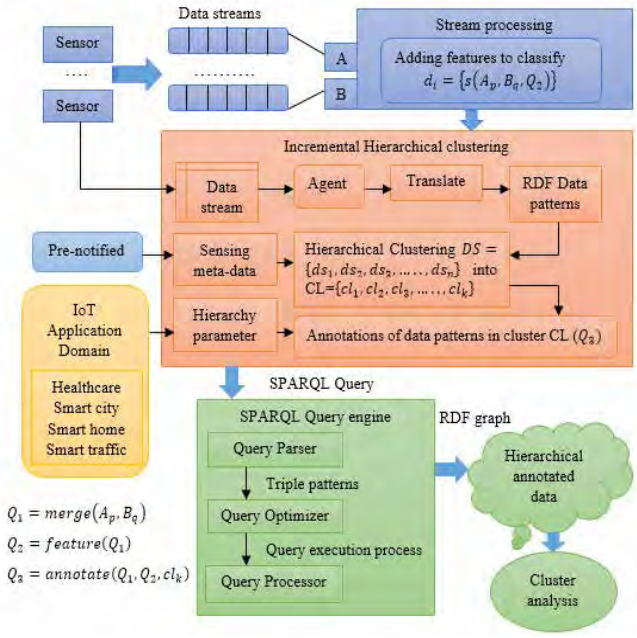


Fig. 1 Architectural model of the proposed mechanism.

B. Query Optimizer

This is the second subcomponent and generates a Query Execution Plan (QEP) for the SPARQL query. The processing of queries is optimized by assessing the input query patterns in a meaningful way. The query triple patterns are arranged in a hierarchical manner for finding matching value result of a triple pattern function for the subsequent triple pattern in the query execution plan.

C. Query Processor

In query processor subcomponent, the matching value results are found, verified with triple patterns and finally combined for answering the full query result. The validity of the triple patterns and input queries are arranged in a hierarchical and topological order. Then the intermediate mismatched patterns are reduced. Table II shows all the symbols or notations used in this paper.

TABLE II. SYMBOLS AND ITS DESCRIPTIONS

Symbol	Description	Symbol	Description
sN	starting Node	cML	common Matching List
Tp	Triple pattern	MV	Matching Value
nextN	next Node	nextNML	next Node Matching List
QEP	Query Execution Plan	nextcN	next common Node
TL	Triple List	tmpP	temporary Plan
Ts	Triple store	subP	sub Plan
TpL	Triple pattern List	nextT	next Triple
GenTriple()	Generate Triple	cN	common Node

The Algorithm 1 is used to translate the given label of information into a RDF label using *TranslateLabel()* function. The input is taken as *n* number of triples and correspondingly specify each type of label and the processing triple time *t* is measured. The annotated <*label*> of RDF data is the output. Firstly, it collects the various types of data from IoT devices, to store the triples data starting from 1 to n as decision iterator. If the condition $p\,(t, x_c) \leq 1$ is satisfied then extract every row and label whichever is matched. At last, the list L has to be returned.

Algorithm 1: *TranslateLabel ()* translates the given label into RDF label

Input: Number of triples denoted as n; type of each data item (type); t is the processing time.

Output: List out the annotated *(<label>)* from RDF.

1: First collect the various types of data generated from IoT sensors
2: **for** i: =1 to n // i is the decision iterator
3: **if** $p\,(t, x_c) \leq 1$ as per the Eq. 3
4: then extract every row and label
5: add the matched label
6: close the each completed label
7: List L: = add element
8: **end if**
9: **end for**
10. Finally, return the list L.

The Algorithm 2 is used to generate the given RDF label of information into triples using *GenTriple()* function. The input is taken as n number of triples value and correspondingly specify each type of label and subsequent time t is measured. The annotated <label> of RDF data is the output. Firstly, it collects the various types of data from IoT devices, to store the triples data starting from 1 to n as decision iterator. If the condition *label[i].isElement()* = 1 is satisfied then it extracts each label and annotates it as a triple, then it allocates the unique id for the namely added resource whichever is matched. Finally, the list L with RDF triples is returned.

Algorithm 2: For the generation of triples *GenTriple()* from a given RDF label

Input: Number of triples named as n; value measured; type of each data item (type); *t* is the measured time.

Output: List out the annotated *(<label>)* from RDF.

1. First, expand each label
2. **for** i: = 1 to n do // i is the decision iterator
3. **if** *label[i].isElement() = 1* // return 1 when the current label is matched element
4. extract every label and annotate it as triple
5. then allocate the unique id for namely added resource
6. List T: = add triples
7. close the completed triple list
8. **end if**
9. **end for**
10. return list L

Algorithm 3: IoT sensor data into annotated RDF data transformation

Input: Dataset to annotate; type of each data item (type)

Output: The annotated *(<label>)* is transformed into a reduced triple format from source data.

1: First, collect the various types of data generated from IoT sensor S
2: Repeat this collection process up to the end of the last triple
3: Then annotate the *List L: GenTriple(TranslateLabel())*
4: There after extract every label and annotate it as triple
5: Allocate the unique id for newly added resource
6: End process loop
7: Return list L

In Algorithm 3, the IoT sensor data into annotated RDF data transformation is shown. The input is taken as a dataset to annotate and specifies each data item type. The annotated *(<label>)* is transformed into a reduced triple format from source data. Firstly, it collects various type of sensor data. It repeats this process until the last triple item is matched. Then it annotates the *List L: GenTriple (TranslateLabel()).* Thereafter it extracts every label and annotates it as a triple. It allocates the unique id for the newly added resource.

D. Incremental Hierarchical Clustering driven Automatic Annotation Process

The agents will play a key role to place the classification of data in time basis by using the matching mechanism for grouping each instance resources occurrence.

The matching objects are denoted as m of the RDF data and current capture objects as C. the matching m ϵ RDF instances as shown in Eq. 3.1 and current capture C as shown in Eq. 3.2, at $t_n \epsilon$ time interval. X_m and X_{cc} form the instances ($X_{cc}^{t_1}$, $X_{cc}^{t_2}$, $X_{cc}^{t_3}$, …… , $X_{cc}^{t_n}$) with the corresponding time interval range $t_1 \leq t_2 \leq t_3$ .....$\leq t_n$, for each individual i.

$$X_m[i]\epsilon[min_i^m, max_i^m] \tag{3.1}$$

$$X_c^j[i]\epsilon[min_i^c, max_i^c] \tag{3.2}$$

Here j starts from 1 to n.

For pattern recognition of data, let us take a resource r that should be any category of data d. The scoring function S^d is used to calculate the matching pattern data d at a particular time T. The individual match value is x at time period t < T and is defined using Eq. (3.3).

$$Pr(T, x_p) = \begin{cases} Add\ data & whenever, T > t_r \text{ or } S(x^T) < 1 \\ Reject\ data & whenever, S(x_c^T) \approx S_r(x_m) \\ x & otherwise \end{cases} \tag{3.3}$$

In order to generate the hierarchical clustering driven tree of the IoT streaming data, the problem is formulated as follows: the input of the sensor raw data is classified with agents and represented as data streams $DS = \{ds_1, ds_2, ds_3, ..., ds_n\}$ in the D dimensional space, the pre-notified meta data as the k dimensions $\{x_1, …., x_k\}$, the pre-clustered streaming data values as $\{x_{k+1}, …., x_l\}$, and to measure the cluster distance among the data patterns as $dist(cl_1, cl_2)$. At starting, each classified data is assigned to its own cluster. Each data pattern in DS and the cluster $cl_i = \{ds_i\}$, $CL = \{cl_1, ..., cl_n\}$ are selected for measuring the minimum distance between data object. Then the merge operation is performed until the none of the cluster can be left blank or empty.

while $CL.size > 1$ do
  if $(cl_{min1}, cl_{min2}) = min\ dist\ (cl_i, cl_j)$ then
    for all (cl_i, cl_j) in cluster CL
    Remove cl_{min1}, cl_{min2} from cluster CL
    Add $\{cl_{min1}, cl_{min2}\}$ to cluster CL
  end if
end while

The $dist\ (cl_1, cl_2)$ is measured as, for example $cl_1 = \{ds_{11}, ds_{12}, ds_{13}, ..., ds_{1n}\}$ and $cl_2 = \{ds_{21}, ds_{22}, ds_{23}, ..., ds_{2n}\}$, then $dist(cl_1, cl_2) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} dist(ds_{1i}, ds_{2j})$. Whereas $dist\ (ds_{1i}, ds_{2j})$ may be calculated using any of the mahalanobis distance, Euclidean distance, or Minkowski distance function in the D dimensional space. The same procedure is performed until the semantic annotations are extracted from the hierarchical tree by cutting into horizontally or vertically and adding the data streams in incremental manner.

The following list of steps are required to design an incremental hierarchical clustering driven automatic annotations for unifying IoT streaming data.

The input data streams $DS = \{ds_1, ds_2, ds_3, ..., ds_n\}$ are obtained from the IoT sensor data repositories in the D dimensional space.

The incremental hierarchical clustering based nearest neighbor chain is used for clustering streaming data.

It starts with any node S in the hierarchical tree, elaborates it until a RNN (Reciprocal Nearest Neighbor) pair of data samples, and then agglomerates these data samples.

Continue the same process with the hierarchical tree of the previously annotated objects using RNN.

The RNN of object p and q, where object q must satisfy the condition

$$if dist(p, q) \leq min\{dist(p, r), dist(q, r)(r \neq p, q)\}$$

Thereafter, the clustering distance $dist(p, q)$ using the Euclidean similarity distance measuring function is measured.

The establishment of the linkage or distance between clusters of the hierarchical tree is done using wards method

$\Delta(a, b) = \frac{w_a w_b}{w + w_b} \|\overrightarrow{n_a} + \overrightarrow{n_b}\|$ where $\overrightarrow{n_a}$ is the center of the cluster e and n_e is the number of data samples involved in it.

Finally, the semantic representations between the clustered hierarchical trees are annotated with SPARQL queries.

In the Resource Description Framework (RDF), the data is generally warehoused as a combination of statements in triples format as {Subject Sb, Predicate Pr, Object Obj}, which is similar to an entity representation in DBMS as {entity e, property p, value v}. Subjects and predicates stored in triples are URIs when objects can be either Uniform Resource Identifiers (URIs) or literal values. SPARQL is a Simple Protocol and RDF Query Language is used for retrieving data stored in RDF repositories. Its syntax is similar to SQL; thus it contains two main clauses, e.g., SELECT and WHERE. The SELECT clause identifies the statements as triples that will appear in the query results. The WHERE clause provides the basic graph pattern to match against the data graph. We consider four disjoint sets Var (variables), Uri (URIs), Blnk (blank nodes) and Ltr (literals).

Almost every SPARQL query contains a set of triple patterns called a basic graph pattern. A basic graph pattern, BGP, is a finite set of triple patterns {tp1, tp2, tp3….,tpn}, in which each tp is a triple as shown in Eq. (3.4).

$$(Sb, Pr, Obj) \in (Var \cup Uri \cup Blnk)\ X(Var \cup Uri)\ X\ (Var \cup Uri \cup Blnk \cup Ltr) \tag{3.4}$$

The sequence of the patterns is framed with different combinations of the triples as shown in Eq. (3.5)

$$SId: [node] = Value: [resource] \cap Sub\_Pre\_Obj: \{S_1 : [p_1, p_2, ; o_1, o_2], S_2 : [p_1, p_2, ; o_1, o_2]\} \tag{3.5}$$

The Query Execution Plan (QEP) is measured based on the sequence of patterns generated by triples (tp_1, tp_2, tp_3, … tp_n) as long as the long sequence patterns are generated. Such that there is at least one common medium of sequence patterns among tp_1 and tp_{i+1} from (Subject as S, Object as O, and Predicate as P) being selected, and it follows any one of the patterns as shown in Eq. (3.6-3.8).

$$S(tp(i)) = S(tp(i+1))\ and\ O(tp(i+1)) \tag{3.6}$$

$$P(tp(i)) = S(tp(i+1))\ and\ P(tp(i+1)) \tag{3.7}$$

$$O(tp(i)) = S(tp(i+1))\ and\ O(tp(i+1)) \tag{3.8}$$

Here S(tp), P(tp), and O(tp) are the Subject, Predicate, and Object respectively. If anyone of the triple patterns is satisfied with the required query then the query execution plan is assigned to the Query

Optimizer (QO) subcomponent.

In order to develop the query execution plan, the query in Algorithm 4 is processed and stored in the triple store (ts). The loaded query is mapped with each triple pattern to subsequent nodes. Sometimes, it refers to other triple patterns that are matched with the stored triple values i.e. (node, [adjacent_TL]). The subject, predicate, and object manner are matched by applying the SPARQL query; finally, the corresponding RDF graph is generated. In Algorithm 4, the query execution plan is shown. The motivation behind this query execution plan is engaging the ordered triple patterns to indexed RDF data and improved version of ordered triple patterns to process the queries in an efficient manner. After generating an ordered triple pattern list for the execution plan, the residual sequential triple patterns that are in the triple list are not attached to the current execution plan. The Tp is considered as the triple pattern in QEP. The nextN <= get_nextN (Tp, sN) is placed subject as first node and object as the second node and vice versa generated. The subP is the sub plan used for storing the remaining triple pattern part for QEP. The appended data triple pattern is a subset of adjacent triple list and TpL is the not visited list then create an intermediator plan for annotating current pattern objects. Such that, consider this one as the current query evaluation plan for executing queries. Finally, the next triple and triple pattern are merged with adjacent triple list, evaluated with QEP.

Algorithm 4: GET_ Query_Execution_PLAN (sN, Tp)

Input: sN - starting Node; tp- triple pattern; ts- triple store (for storing every query generated triple).

Output: QEP, - generates longest triple patterns in RDF format

1: QEP **<=** Tp // triple pattern is considered in QEP

2: *nextN***<=** get_nextN (Tp, sN)

// if N is the Node and it is in subject place, nextN is its object.

// if N is the Node and it is in object place, nextN is its subject

3: *adjacentTL***<=** getTL (Ts, nextN) *// TL- Triple List; Ts- Triple store*

4: *subP***<=** Ø *// load the remaining triple pattern part for QEP; subP- sub Plan*

5: for each data triple pattern TpL ϵ *adjacentTL* **do** *// TpL- Triple pattern List*

6: **if** (TpL ≠ visited node) **then**

7: tmpP **<=** GET_Query_PLAN (nextN, TpL)

8: **if** len *(subP)*< len *(tmpP)***then**

9: *subP* **<=***tmpP*

10: *nextT***<=** *TpL*

11: **end**

12: **end**

13: end

14: QEP<= *adjacentTL* \ {*nextT*, Tp} *// nextT is included in subP*

15: QEP<=*subP*

16: return QEP

The following are the list of steps required to execute the query plan.

1. Firstly, go to the file menu, in that select new triple store- appropriate path name that has been given for storing work in the triple store.
2. Once the path is identified by triple store, a maximum number of estimated triples are selected. (E.g. 100000).
3. Then load the triples of any format (E.g. N-triples, RDF/XML, N-Quads).
4. Go to the Query view in View menu bar. The required query is applied for annotating the data in RDF format.
5. Then run the SPARQL query, it shows the result as ?s ?p ?o in tabular form.
6. Finally, click on the create visual graph icon, then it generates the annotated RDF graph. Make changes on the graph as per the neediness of the user.

In the Query Execution Plan (QEP), the subject, predicate, and object are placed in triple patterns format. Therefore, at any point, only the vertices and edges can be placed. The time complexity for generating the query execution plan is $O(|S|).|P|)$. Here, $|S|$ is the number of Subjects placed in the healthcare dataset, and $|P|$ is the number of Predicates placed in the healthcare dataset. Therefore, the proposed algorithm 4 and algorithm 5 take total computational time $O(|S|) + |P|)$ as the time complexity, because this work uses every Subject and Predicate or node only once.

IV. Experimental Methodology

In this section, the proposed mechanism is described with automatic annotations for unifying the IoT streaming data. In addition, the implementation results of the proposed mechanism with SPARQL queries processing is discussed.

A. Adding Semantic Annotations to the IoT Streaming Data

In this proposed research work, to achieve semantic annotations, the query processing mechanism and triple patterns are used. The authors have analyzed and tested on three different healthcare datasets using incremental hierarchical clustering driven automatic annotations based on IoT streaming data.

B. Query Processing

In this section, the queries are processed and executed based on the query execution plan so Algorithm 4 is used. To perform that operation we need to observe the correct triple patterns from the triple store or find the invalid annotation results. The following common steps used in Algorithm 5 using a triple store are followed:

1. Firstly, the input cN is considered as the common node or node pattern to retrieve the subsequent triple pattern (tp) from query execution plan and generate the annotation results from triple store to the common node cN.
2. The matching common node cN is attached to the common matching list cML.
3. For each common annotated data is resulted to merge the annotator list of final matching list.
4. Then, each matching value is a subset of the final matching list and contains the annotated attributes for matching value identification.
5. The next value is placed on the basis to get the next node and add the mapping value to the triple store.
6. If any node is mapped with the next node then the mapping annotations consisting of next node, matching value, and next node matching list are added. If any node is not matched with the next node then all corresponding matching values and associated annotations are removed.
7. This entire process is repeated until the all-existing triples are reached and mapping of the common node cN exists that was taken from triple pattern tp.

The SPARQL queries are processed for annotating the matching healthcare data and its associated values. However, the queries are different triple patterns (tp_1, tp_2, ..., tp_n) and the matching subject of any common node is retrieved as well as its corresponding predicates.

Algorithm 5: GET_ PROCESS_TRIPLE_PATTERN (cN)

Input: cN is the common Node (node pattern).

Data: QEP- Query Execution Plan for processing triple patterns using SPARQL queries. Tp- Triple pattern;

Ts- Triple store (for storing every intermediate generated data). Ant-annotation

Output: SPARQL query and RDF triple data mapping

1: Tp<= QEP.getNext () // *next triple pattern is considered in QEP and is placed*

2: cML <=M (*cN*) // *common node match list placed*

3: for each ant Є getAnnotatorList (*cML*) **do**

4: **for each** MV Є cML.getMV (*ant*) **do**

5: nextN <=get_nextN (*Tp, cN*)

6: nextNML <=findMatches (*nextN, MV*)

7: **if** any node is mapping with *nextN* **then**

8: M. addMap (*nextN, MV, nextNML*) // *now MV is the annotator of nextN*

9: **else**

10: remove (*MV*) // *remove the matching values and associated annotators*

11: **end**

12: **end**

13: end

14: if any node is matched with result of *Tp* **then**

15: *nextcN*<=findNextcN (*Tp*)

16: PROCESS_TRIPLE_PATTERN (*ncN*)

17: end

```
select ? s ?  ? p ? o  where
{<https://Resource url> ? s ? p ? o
limit 100
}
```

Fig.2. SPARQL Query.

The Fig. 2 is a sample SPARQL query for annotation of triples such as subject, predicate, and object manner. The SPARQL queries are widely used for annotating the RDF data for machine-readable and semantically describable data. There is another option in SPARQL queries to extract the full dataset attribute information with a limit basis like 10k, 20k, 30k, 100k, and so on triples.

V. Performance Evaluation

This section employs the experimental datasets used for the proposed IHC-AA-IoTSD mechanism. In addition, the performance evaluation metrics are discussed for evaluating the performance of the IHC-AA-IoTSD in detail. In the final analysis, the time complexity of the proposed algorithms are measured.

A. Data Setup

For evaluation of the proposed mechanism IHC-AA-IoTSD, three different kinds of healthcare datasets, namely Heart diseases, Heart attack, and Diabetes are taken. These are openly available datasets from the UCI Machine learning repository**.** Table III shows the dataset details including names of datasets, the number of triples in the datasets, and downloadable resources information.

TABLE III. Dataset Details

S.No	Dataset Name	No. of Triples	Source
1	Heart diseases	212154	https://archive.ics.uci.edu/ml/datasets/heart+Disease
2	Heart Attack	112896	https://www.kaggle.com/imnikhilanand/heart-attack-prediction
3	Diabetes	142547	https://archive.ics.uci.edu/ml/datasets/diabetes

B. Experimental Environment

To evaluate the performance proposed mechanism, a conventional and regular laptop was used with the configuration of Windows 10 Home 64-bit, 8 GB RAM, 1 TB HDD, 2 cores, 2.2 GHz CPU clock speed, and Intel® Core™ i7-8th Gen-8750H CPU type. The Gruff tool with Java 1.8.0 platform was used to experiment the healthcare data. The Tableau and Allegro Graph tools support to visualize the data in a good manner for users. The SPARQL query language was used for annotating the healthcare data to communicate patient and doctors in a meaningful way.

C. Performance Metrics

To evaluate the performance of the proposed framework, the following metrics are considered for measuring the framework. These metrics are generated from the confusion matrix as shown in Table. IV.

TABLE. IV. Confusion Matrix

	Predicted as "YES"	**Predicted as "NO"**
Actually as "YES"	True Positive $[Cl_{ant} \rightarrow Cl_{ant}]$	False Negative $[Cl_{ant} \rightarrow NCl_{ant}]$
Actually as "NO"	False Positive $[NCl_{ant} \rightarrow Cl_{ant}]$	True Negative $[NCl_{ant} \rightarrow NCl_{ant}]$

- *True Positive* $Cl_{ant} \rightarrow Cl_{ant}$: This is an assessment of correctly clustered annotations considered correctly as clustered annotations.
- *True Negative* $NCl_{ant} \rightarrow NCl_{ant}$: This is an assessment of non-clustered annotations considered correctly as non-clustered annotations.

$$TPR = \frac{Cl_{ant} \rightarrow Cl_{ant}}{[Cl_{ant} \rightarrow Cl_{ant} + Cl_{ant} \rightarrow NCl_{ant}]} \quad (4.1)$$

- *False Positive* $NCl_{ant} \rightarrow Cl_{ant}$: This is an assessment of non-clustered annotations considered incorrectly as clustered annotations.
- *False Negative* $Cl_{ant} \rightarrow NCl_{ant}$: This is an assessment of clustered annotations considered incorrectly as non-clustered annotations.

1. True Positive Rate (TPR)

TPR states the sensitivity value and measures correctly clustered annotations from the dataset as shown Eq. (4.1). Eq. (4.2) corresponds to the true negative rate (TNR).

$$TNR = \frac{NCl_{ant} \rightarrow NCl_{ant}}{[NCl_{ant} \rightarrow NCl_{ant} + NCl_{ant} \rightarrow Cl_{ant}]} \quad (4.2)$$

2. False Positive Rate (FPR)

FPR measures the significance level, which scales the proportion of non-clustered annotations that are interpreted as clustered annotations in the automatic annotation process, and generated as input dataset sequence as shown Eq. (4.3).

$$FPR = \frac{NCl_{ant} \rightarrow Cl_{ant}}{[NCl_{ant} \rightarrow Cl_{ant} + NCl_{ant} \rightarrow NCl_{ant}]} \quad (4.3)$$

3. False Negative Rate (FNR)

FNR scales the proportion of clustered annotations that are interpreted as non-clustered annotations in the clustered data annotation process as shown Eq. (4.4).

$$FNR = \frac{Cl_{ant} \rightarrow NCl_{ant}}{[Cl_{ant} \rightarrow NCl_{ant} + Cl_{ant} \rightarrow Cl_{ant}]} \quad (4.4)$$

4. Accuracy

Accuracy is the first step towards performance measure where it defines the ratio between the total counts of correct clustered annotations made to a total count of clustered annotations made as shown Eq. (4.5).

$$Accuracy \frac{(Cl_{ant} \rightarrow Cl_{ant} + NCl_{ant} \rightarrow NCl_{ant})}{[Cl_{ant} \rightarrow Cl_{ant} + NCl_{ant} \rightarrow NCl_{ant} + NCl_{ant} \rightarrow Cl_{ant} + Cl_{ant} \rightarrow NCl_{ant}]} \quad (4.5)$$

5. Precision, Recall & F-measure

Precision discourses about the exactness of the clustered data, and the Recall voices about completeness of the data. The Precision and Recall discuss more about the detected accuracy of the data, and the accuracy should not deal much about false results. The F-measure is the mean of precision and recall. The equations depicted from (4.6) to (4.8) is Precision, Recall, and F-measure respectively.

$$Precision = \frac{Cl_{ant} \rightarrow Cl_{ant}}{[Cl_{ant} \rightarrow Cl_{ant} + NCl_{ant} \rightarrow Cl_{ant}]} \quad (4.6)$$

$$Recall = \frac{Cl_{ant} \rightarrow NCl_{ant}}{[Cl_{ant} \rightarrow Cl_{ant} + Cl_{ant} \rightarrow NCl_{ant}]} \quad (4.7)$$

$$F - measure = 2 * \left(\frac{Precision * Recall}{Precision + Recall}\right) \quad (4.8)$$

These ML metrics are used on the proposed mechanism for improving cluster efficiency and unifying the IoT streaming data.

D. Experimental Results and Discussions

This experiment is conducted under the stimulus of three healthcare datasets namely- Heart Diseases, Heart Attack, and Diabetes by applying various triple sizes with 10k, 20k, 30k, 40k, 50k, and 100k respectively. Annotating the objects of streaming data, six SPARQL queries are used to evaluate the proposed IHC-AA-IoTSD mechanism as represented in Fig.3 to Fig.10.

```
select ? drug ? type ? value where
{<https://www.drugs.tms/drugs/resource/MeDRAconceptType> ? drug ? type ? value
limit 100
}
```

Fig. 3. SPARQL Query 1.

The SPARQL query 1 shown in Fig. 3, queries for the drug types and values annotated with hierarchical clustered data. The SPARQL query 2 shown in Fig. 4 is used to extract unique heart attack attributes and their values from heart attack dataset.

```
select ? heart attack ? value where
{<https://data.medicare.gov/resource/ygty-mm5a3> ? heart attack ? value
limit 100
}
```

Fig. 4. SPARQL Query 2.

The role of SPARQL queries is highly enrich to all attributes for annotation. In addition, the queries are effectively annotated various attributes in lower execution time.

```
select ? heart attack ? type where
{<https://data.medicare.gov/resource/ygty-mm5a3> ? heart attack ? type
limit 100
}
```

Fig. 5. SPARQL Query 3.

The SPARQL query 3 is as shown in Fig. 5 and its resultant RDF graph as shown in Fig. 7. The hierarchical tree based predicates are annotated over the various triple data objects.

```
select ? Row ID ? Member ? Year ? Number of diabetes deaths where
{<https://sensormeasurement.appspot.com/m3#diabetesdeaths>
? Row ID ? Member ? Year ? Number of diabetes deaths
limit 100
}
```

Fig. 6. SPARQL Query 4.

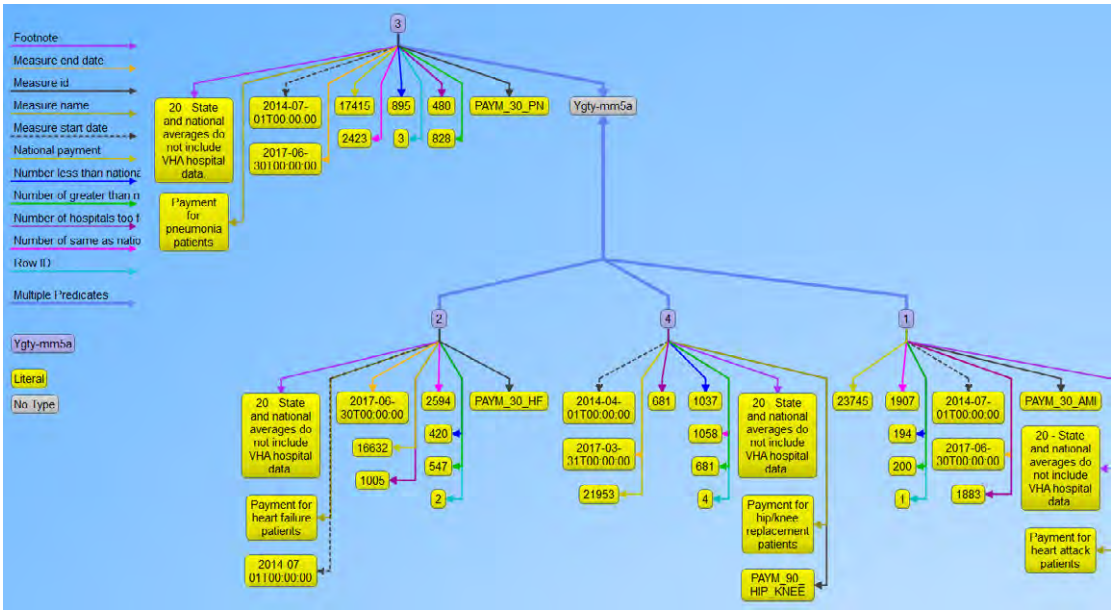


Fig. 7. Annotated heart attack diagnostic measurement values in a hierarchical tree.

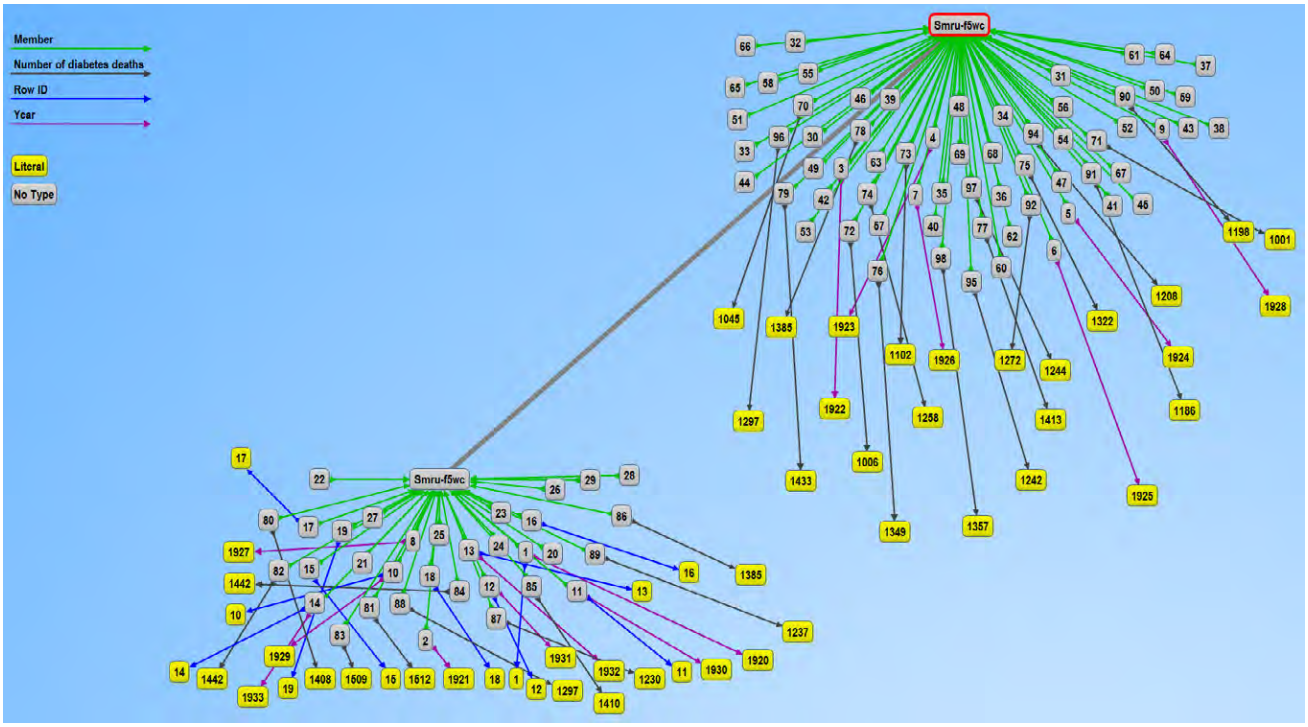


Fig. 8. Annotated diabetes diagnostic measurement values.

The SPARQL query 4 as shown in Fig. 6. The Fig. 8 shows the resulted output of query 4. Moreover, the diabetes data set contains of predicates as row id, value, and number of the deaths, etc. Using annotation process, the representation of the year wise death rates have been enriched as well as extracted.

```
select ? subject ? predicate ? object where
{<https://data.cdc.gov/resource/bi63-dtpu> ? subject ? predicate ? object
limit 100
}
```

Fig. 9. SPARQL Query 5.

The SPARQL query 5 shown in Fig. 9 has been performed on heart diseases dataset for annotating the healthcare records by means of subject, predicate, and object manner. It indicates that the annotations performed on the whole dataset with accurate annotations. The SPARQL query 6 shown in Fig. 10 is widely used for annotating the heart diseases data on value and predicate basis annotations. In this, the corresponding predicate as the number of national payments on year wise, payment for heart diseases, measure id, measure name, measure start date, measure end date, type and corresponding values are annotated. The SPARQL query 5 and query 6 are used in this paper to annotate the healthcare data by varying triple data size up to 100k triples. These results have not been presented because these annotations make the things complex and not visible to the users.

```
select ? value ? predicate where
{<https://data.maryland.gov/resource/smru-ftwc> ? value ? predicate
limit 100
}
```

Fig. 10. SPARQL query 6.

However, the results of SPARQL query 1 to query 6 clearly indicate that automatic annotations are more concisely preferable than the manual and semi-automatic annotations. Because in automatic semantic annotations, the trained and classified data are labelled using an automated annotation system. The average execution time of the various queries are measured, and it achieves the lowest compared with ATLAS [22], FBASAM [11], and OBSAA [19] approaches.

The first experimental investigation of IHC-AA-IoTSD is validated through TPR by applying various triples with respect to a stable FPR 10, 20, 30 and 40% over the benchmark mechanisms such as ATLAS, FBASAM, and OBSAA is observed in Figs. 11, 12, and 13 respectively.

Fig.11 (a-d) shows the leading TPR value on Heart Diseases dataset of proposed IHC-AA-IoTSD over ATLAS, FBASAM, and OBSAA with respect to the stable FPR of 10%, 20%, 30%, and 40% respectively. Fig. 11 (a) result proves that IHC-AA-IoTSD is capable to preserve the TPR around 0.95 at dynamically allocated triples and this TPR value infers 12% success rate than the benchmark mechanisms ATLAS, FBASAM, and OBSAA respectively under 10% FPR. Fig. 11 (b) shows the dominant TPR value of IHC-AA-IoTSD over ATLAS, FBASAM, and OBSAA respectively under 20% FPR and is capable to maintain its TPR value around 0.92 at dynamically allocated triples even the FPR is increased. In addition, the proposed IHC-AA-IoTSD proves a greater TPR around 13% than the benchmark mechanisms ATLAS, FBASAM, and OBSAA respectively. Likewise, Fig. 11 (c) represent the TPR value of IHC-AA-IoTSD over ATLAS, FBASAM, and OBSAA under 30% FPR and is capable to withstand its TPR value around 0.9 at various dynamically allocated triples and proves a greater TPR around 11% than the benchmark mechanisms ATLAS, FBASAM, and OBSAA respectively. Similarly, Fig. 11 (d) represent the TPR value of IHC-AA-IoTSD over ATLAS, FBASAM, and OBSAA respectively under 40% FPR. Besides, proposed IHC-AA-IoTSD is achieved a marginable TPR around 0.88 at dynamically allocated triples and proves this TPR value infers 8% higher accurate than the benchmark mechanisms ATLAS, FBASAM, and OBSAA respectively.

Fig.12 (a-d) shows the dominant TPR value on Heart Attack dataset of proposed IHC-AA-IoTSD over ATLAS, FBASAM, and OBSAA with respect to the stable FPR of 10%, 20%, 30%, and 40% respectively. Fig. 12 (a) result proves that IHC-AA-IoTSD is capable to preserve the TPR around 0.92 at dynamically allocated triples and this TPR value infers 12% success rate than the benchmark

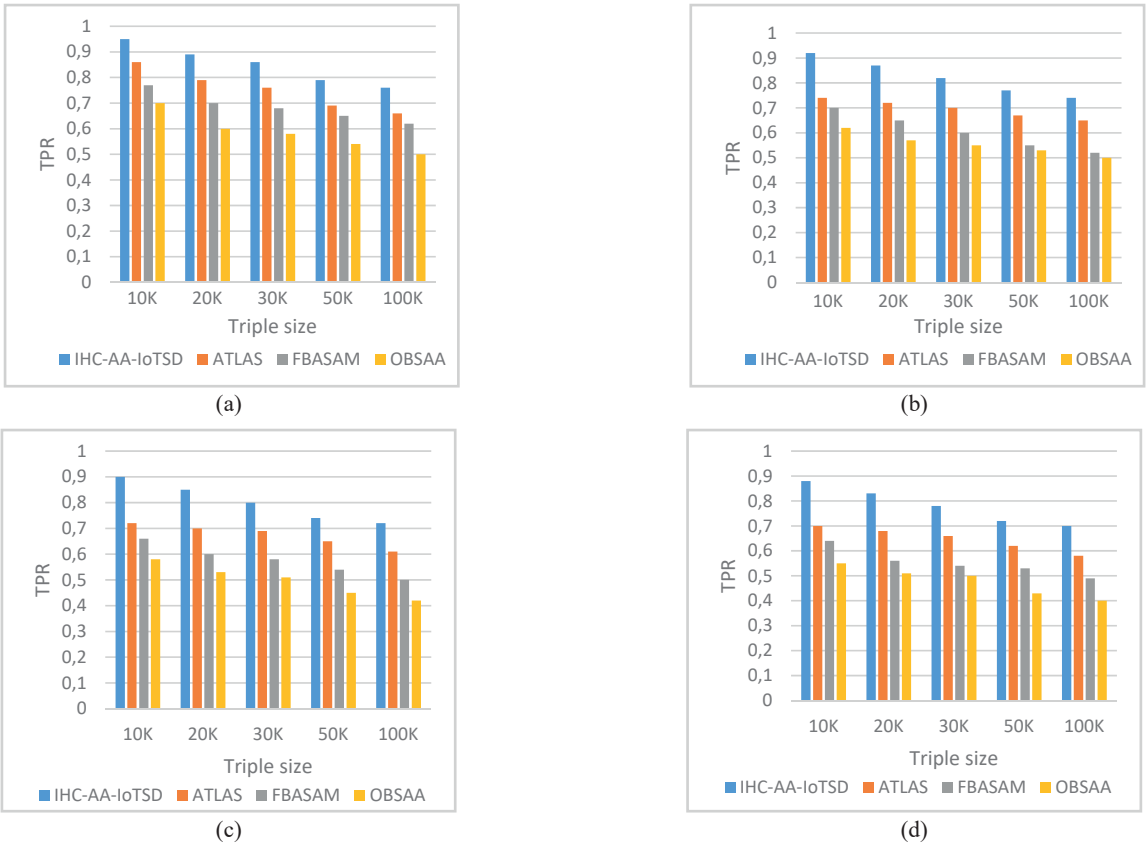


Fig. 11. True Positive Rate (TPR) on Heart Diseases dataset by varying triple size (a) false positive rate =10%, (b) false positive rate =20%, (c) false positive rate =30%, (d) false positive rate =40%.

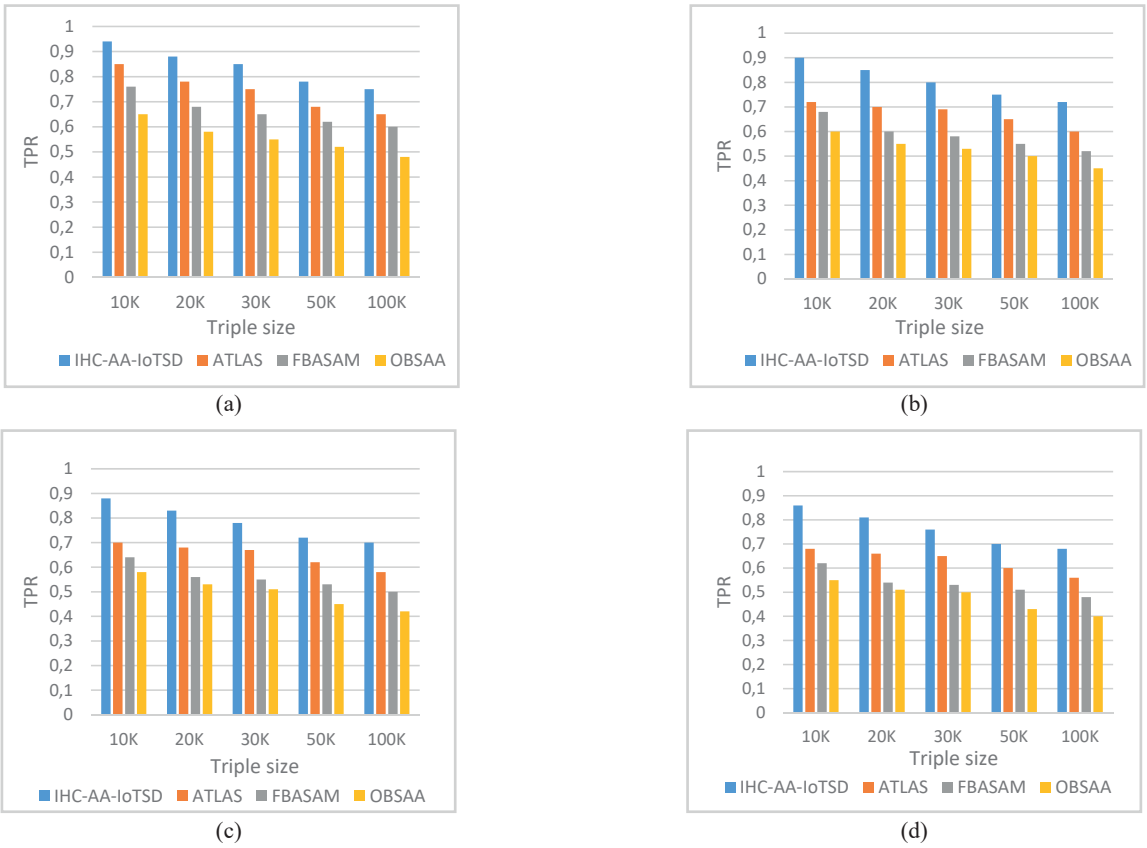


Fig. 12. True Positive Rate (TPR) on Heart Attack dataset by varying triple size (a) false positive rate =10%, (b) false positive rate =20%, (c) false positive rate =30%, (d) false positive rate =40%.

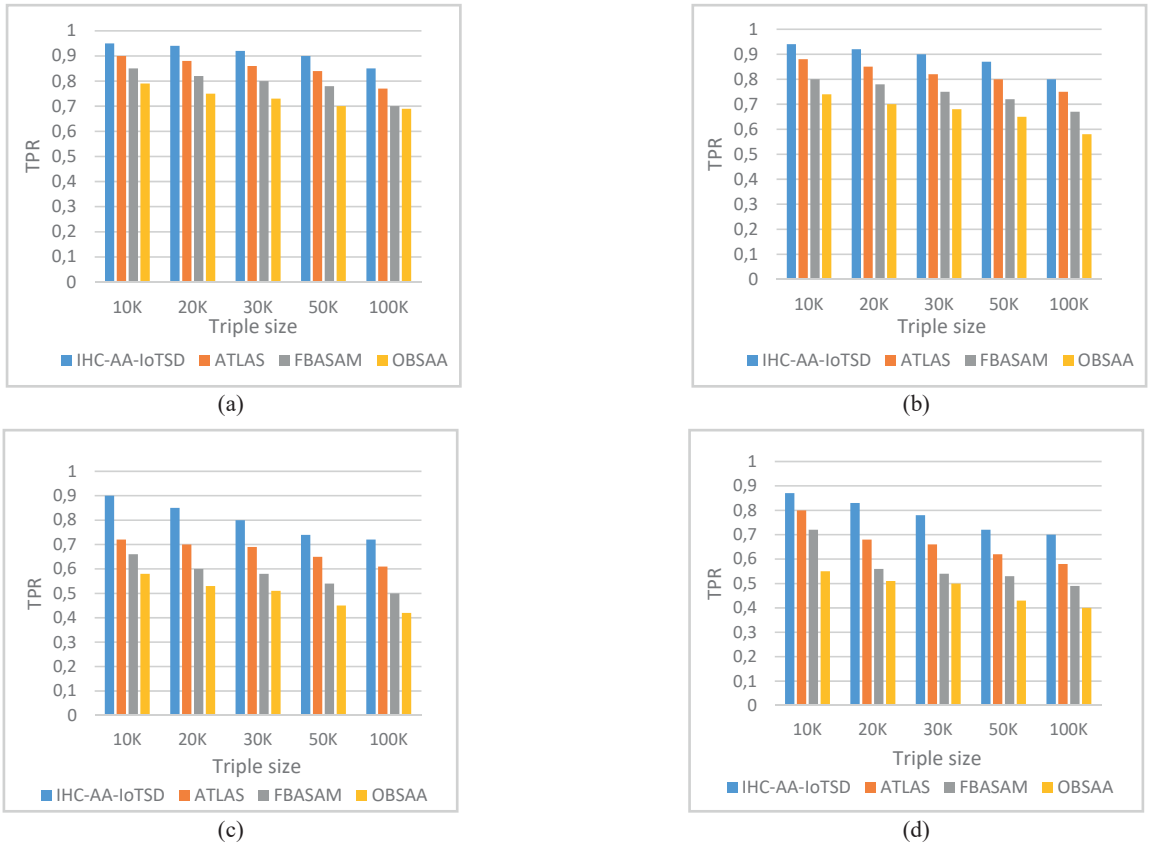


Fig. 13. True Positive Rate (TPR) on Diabetes dataset by varying triple size (a) false positive rate =10%, (b) false positive rate =20%, (c) false positive rate =30%, (d) false positive rate =40%.

mechanisms ATLAS, FBASAM, and OBSAA respectively under 10% FPR. Fig. 12 (b) shows the dominant TPR value of IHC-AA-IoTSD over ATLAS, FBASAM, and OBSAA respectively under 20% FPR and is capable to maintain its TPR value around 0.9 at dynamically allocated triples even the FPR is increased. In addition, the proposed IHC-AA-IoTSD proves a greater TPR around 13% than the benchmark mechanisms ATLAS, FBASAM, and OBSAA respectively. Likewise, Fig. 12 (c) represent the TPR value of IHC-AA-IoTSD over ATLAS, FBASAM, and OBSAA under 30% FPR and is capable to withstand its TPR value around 0.88 at various dynamically allocated triples and proves a greater TPR around 11% than the benchmark mechanisms ATLAS, FBASAM, and OBSAA respectively. Similarly, Fig. 12 (d) represent the TPR value of IHC-AA-IoTSD over ATLAS, FBASAM, and OBSAA respectively under 40% FPR. Besides, proposed IHC-AA-IoTSD is achieved a marginable TPR around 0.86 at dynamically allocated triples and proves this TPR value infers 7% higher accurate than the benchmark mechanisms ATLAS, FBASAM, and OBSAA respectively.

Fig.13 (a-d) shows the dominant TPR value on Heart Diseases dataset of proposed IHC-AA-IoTSD over ATLAS, FBASAM, and OBSAA with respect to the stable FPR of 10%, 20%, 30%, and 40% respectively. Fig. 13 (a) result proves that IHC-AA-IoTSD is capable to preserve the TPR around 0.94 at dynamically allocated triples and this TPR value infers 13% success rate than the benchmark mechanisms ATLAS, FBASAM, and OBSAA respectively under 10% FPR. Fig. 13 (b) shows the dominant TPR value of IHC-AA-IoTSD over ATLAS, FBASAM, and OBSAA respectively under 20% FPR and is capable to maintain its TPR value around 0.92 at dynamically allocated triples even the FPR is increased. In addition, the proposed IHC-AA-IoTSD proves a greater TPR around 12% than the benchmark mechanisms ATLAS, FBASAM, and OBSAA respectively. Likewise, Fig. 13 (c) represent the TPR value of IHC-AA-IoTSD over ATLAS, FBASAM, and OBSAA under 30% FPR and is capable to withstand its TPR value around 0.9 at various dynamically allocated triples and proves a greater TPR around 11% than the benchmark mechanisms ATLAS, FBASAM, and OBSAA respectively. Similarly, Fig. 13 (d) represent the TPR value of IHC-AA-IoTSD over ATLAS, FBASAM, and OBSAA respectively under 40% FPR. Besides, proposed IHC-AA-IoTSD is achieved a marginable TPR around 0.87 at dynamically allocated triples and proves this TPR value infers 9% higher accurate than the benchmark mechanisms ATLAS, FBASAM, and OBSAA respectively.

In the second experimental investigation of IHC-AA-IoTSD validated through the detection accuracy, TNR, FNR, TPR, Precision, and FPR over the benchmark mechanisms such as ATLAS, FBASAM, and OBSAA techniques respectively.

Fig. 14 (a) represents the average detection accuracy of IHC-AA-IoTSD on three healthcare datasets with various triple sizes. The results confirm that IHC-AA-IoTSD is capable to accomplish superior detection accuracy in heart dataset from the UCI data repository, and it acquired detection accuracy of 9–94% from 10k triples to 100k triples respectively. Nevertheless, ATLAS facilitates a detection accuracy of 7–90% from 10k triples to 100k triples respectively, FBASAM achieves a detection accuracy of 5–81% from 10k triples to 100k triples respectively and OBSAA ensures a detection rate of 2–75% from 10k triples to 100k triples respectively. Performing tests on heart attack dataset from the kaggle data repository, it got a detection accuracy of 11–97% from 10k triples to 100k triples respectively. Nevertheless, ATLAS facilitates a detection accuracy of 9–93% from 10k triples to 100k triples respectively, FBASAM achieves a detection accuracy of 7–89% from 10k triples to 100k triples respectively and OBSAA ensures a detection rate of 4–84% from 10k triples to 100k triples respectively. Performing tests on diabetes dataset from the UCI data repository, it

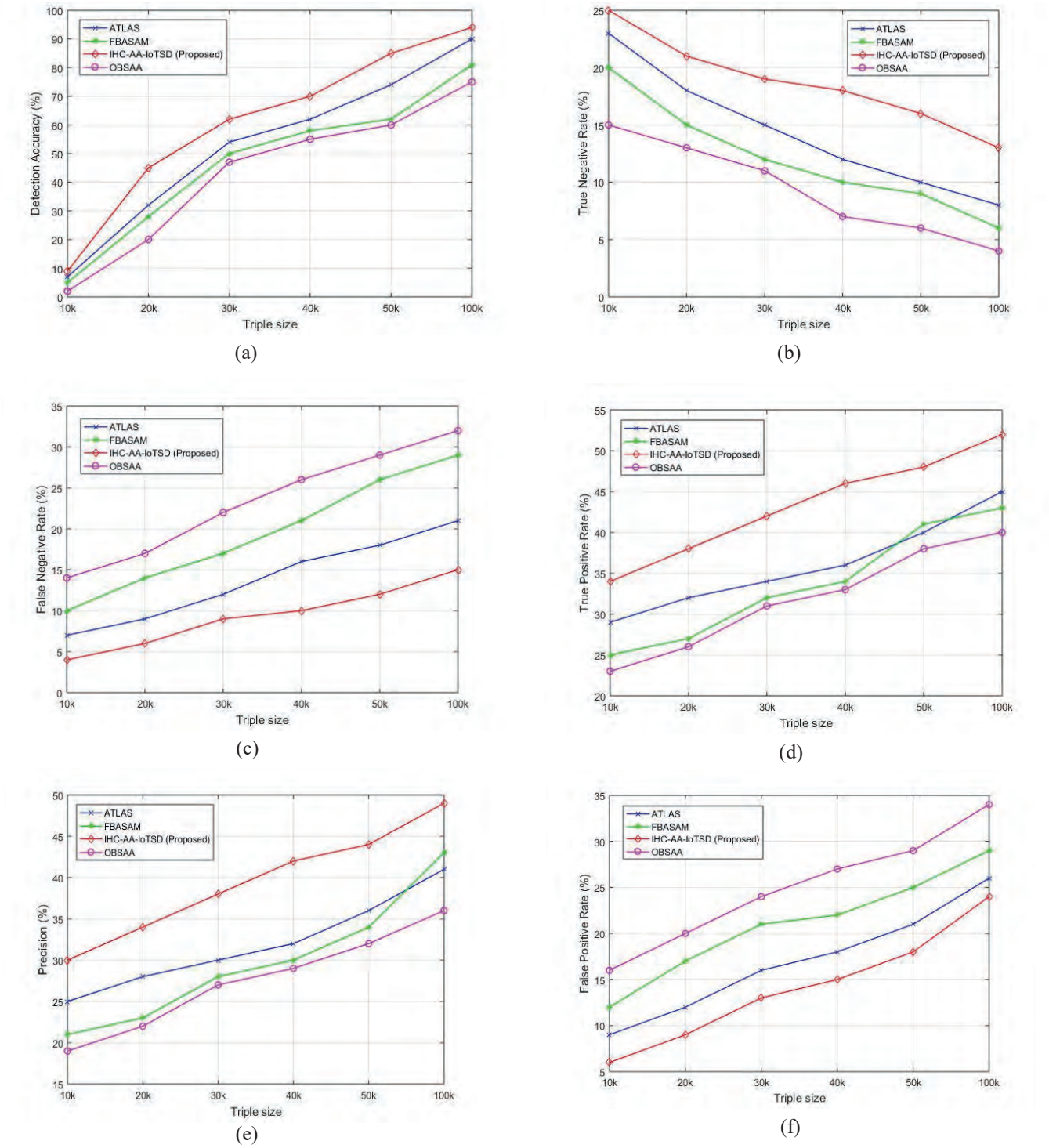


Fig. 14. (a) Detection Accuracy of IHC-AA-IoTSD under various triple sizes, (b) True Negative Rate of IHC-AA-IoTSD under various triple sizes (c) False Negative Rate of IHC-AA-IoTSD under various triple sizes (d) True Positive Rate of IHC-AA-IoTSD under various triple sizes (e) Precision Rate of IHC-AA-IoTSD under various triple sizes (f) False Positive Rate of IHC-AA-IoTSD under various triple sizes.

got a detection accuracy of 10–96% from 10k triples to 100k triples respectively. Nevertheless, ATLAS facilitates a detection accuracy of 7–90% from 10k triples to 100k triples respectively, FBASAM achieves a detection accuracy of 5–84% from 10k triples to 100k triples respectively and OBSAA ensures a detection rate of 3–80% from 10k triples to 100k triples respectively. On an average IHC-AA-IoTSD got the 4% detection accuracy increases from ATLAS mechanism at 10k triples whereas at 100k triples got the same improvement. After combining the three healthcare dataset results with increased detection accuracy, the results indicating that 2%, 4%, and 7% decrease than the ATLAS, FBASAM, and OBSAA techniques respectively. This effectiveness of IHC-AA-IoTSD by means of detection accuracy is primarily payable to the enhanced process of multi-agent based semantic annotation used for classifying and testing. This detection accuracy is also because of the agent-based automatic semantic process stimulated in the IHC-AA-IoTSD annotation mechanism.

Fig.14 (b) shows the TNR value of IHC-AA-IoTSD under varying triple data size and the result endorses that it is effective in enlightening

the TNR value by 15-23% differing to ATLAS, FBASAM, and OBSAA, which enable an improvement of 2%, 5%, and 10% from 10k triples to 100k triples. The results about the enhancement of TNR value prove that the IHC-AA-IoTSD performs better because of the patient and doctor annotating the healthcare data enabled in the detection process.

Fig.14 (c) depicts the reduced FNR of IHC-AA-IoTSD under changing triple rate and ensures that, it can minimize the FNR of about 16–30%, which is hardly 8% decrease at 10k triple size and 12% decrease at 100k triple size than ATLAS framework testing on heart diseases dataset. Testing on heart attack dataset, it can minimize the false negative rate of about 17–29%, which is nearly 8% decrease at 10k triple size and 8% decrease at 100k triple size than ATLAS framework. Similarly, by testing on diabetes dataset, it can minimize the false negative rate of about 18–29%, which is nearly 8% decrease at 10k triple size and 12% decrease at 100k triple size than ATLAS approach. The results depict that, the decrease in false positive rate at 10k triple size is nearly 8, 14, and 16% testing on Heart diseases dataset, nearly 8, 12, and 22% testing on Heart Attack dataset, and nearly 6, 14, 22% testing on Diabetes dataset than the ATLAS, FBASAM, and OBSAA techniques respectively. After combining the three healthcare dataset results with reduced False Negative Rate (FNR), the results indicate a 7%, 13%, and 18% decrease compared to the ATLAS, FBASAM, and OBSAA techniques respectively.

Fig. 14 (d) represents the TPR value of IHC-AA-IoTSD under changing triple rate and the result evidences its capacity of enhancing the TNR value by 34–23%, which is nearly 5, 9 and 11% higher than the TNR obtained by ATLAS, FBASAM, and OBSAA tested on three healthcare datasets at 10k triples. The results are considered on an average and it achieves the 5, 9, and 11% more than the TPR value achieved by ATLAS, FBASAM, and OBSAA. The importance of IHC-AA-IoTSD is based on agent preprocessing mechanism used for annotating the triple data and SPARQL queries that could be optimally applicable for healthcare data annotations.

Fig. 14 (e) represents the Precision rate of IHC-AA-IoTSD under varying data triple sizes at 10k, 20k, 30k, 40k, 50k, and 100k on 3 different datasets. The result evidences by enhancing the Precision value around 5–21%, which is nearly 5, 10 and 17% higher than the Precision rate simplified by ATLAS, FBASAM, and OBSAA testing on Heart Diseases dataset using 10k triples. Similarly, nearly 6, 10 and 15% higher than the Precision rate simplified by ATLAS, FBASAM, and OBSAA testing on Heart Attack dataset at 10k triples, and nearly 7, 12 and 21% higher than the Precision rate obtained by ATLAS, FBASAM, and OBSAA testing on Diabetes dataset using 10k triples. After combining the three healthcare dataset results with increased Precision rate, the results indicate that around 6%, 11%, and 17% increase than the ATLAS, FBASAM, and OBSAA techniques respectively.

Fig.14 (f) depicts the reduced FPR of IHC-AA-IoTSD under varying data triples and ensures that, it can minimize the FPR around 6–24%, which is nearly 3% decrease at 10k triple size and 6% decrease at 100k triple size compared to the ATLAS framework, testing on heart diseases dataset. Testing on heart attack dataset, it can reduces the FNR about 8–29%, which is nearly 5% decrease at 10k triple size and 8% decrease at 100k triple size compared to the ATLAS framework.

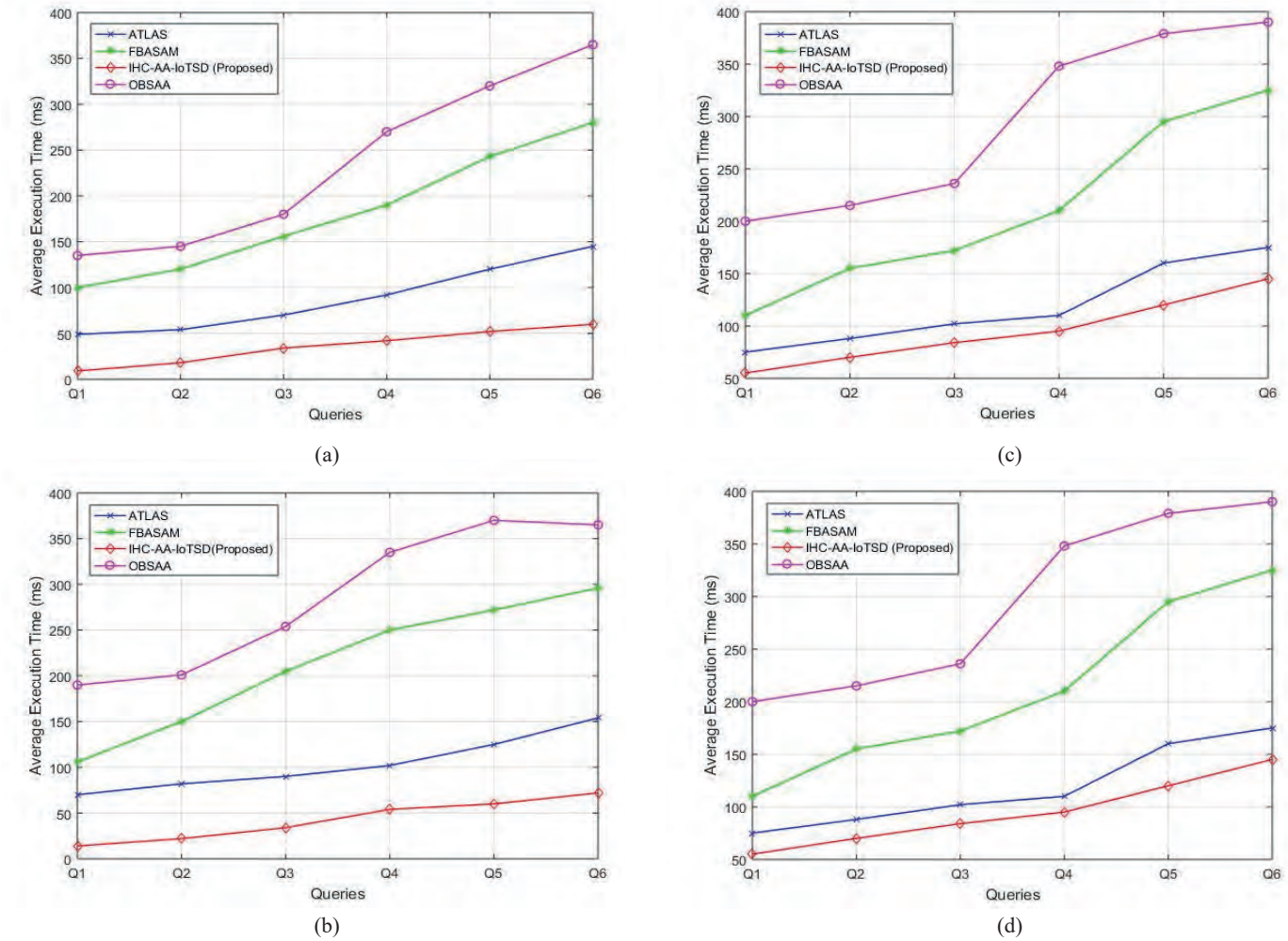


Fig.15. Average Execution Time (ms) by various queries (a) at 10k triples (b) at 20k triples (c) at 30k triples (d) at 50k triples

Similarly, by testing on diabetes dataset, it can minimize the false negative rate of about 5–25%, which is nearly 7% decrease at 10k triple size and 11% decrease at 100k triple size than ATLAS approach. The results depict that, the decrease in false positive rate at 10k triple size is nearly 3%, 6%, and 10% testing on Heart diseases dataset, nearly 5%, 8%, and 14% testing on Heart Attack dataset, and nearly 7%, 11%, and 16% testing on Diabetes dataset compared to the ATLAS, FBASAM, and OBSAA techniques respectively. After combining the three healthcare dataset results with reduced False Positive Rate (FPR), the results indicate 5%, 8%, and 13% decrease compared to the ATLAS, FBASAM, and OBSAA techniques, respectively.

In the third experimental investigation of IHC-AA-IoTSD validated through the Average Execution Time of various queries over the benchmark mechanisms such as ATLAS, FBASAM, and OBSAA techniques respectively.

Fig. 15 (a-d) shows measured average execution time by various queries from Q1 to Q6 at 10k triples, 20k triples, 30k triples, and 50k triples, respectively. The result proves that IHC-AA-IoTSD is able to maintain the Average Execution Time of 27 ms at various queries and this Average Execution Time infers 12% success rate higher than ATLAS, FBASAM, and OBSAA. Figs. 15 (a-d) highlights the predominance Average Execution Time of IHC-AA-IoTSD over ATLAS, FBASAM, and OBSAA under the 10k triples, 20k triples, 30k triples, and 50k triples respectively. The result confirms that IHC-AA-IoTSD is able to endure its Average Execution Time of 86 ms at various queries even when the triple size is increased. IHC-AA-IoTSD enables a superior Average Execution Time of 16% when compared to ATLAS, FBASAM, and OBSAA with all the queries.

E. Complexity Analysis

Moreover, the time complexity of IHC-AA-IoTSD scheme, which used algorithms from 1 to 3, is determined to be $T(n)$ for algorithm perceived instances staring from j is 1 to n and i value between 1 to 9. The time complexity of algorithm 1 is calculated by $T_1(n)$, algorithm 2 is by $T_2(n)$, and algorithm 3 is by $T_3(n)$. At last, these three times complexities will be combined to get the overall time complexity $T(n)$. Let us see how to find the time complexity of $T_1(n)$, it is as follows in Eq. (5.1).

$$T_1(n) = t_1 + t_2 + \cdots + t_9,$$
$$T_1(n) = 1 + (1 + 1 + 1 + 1 + 1 + 1 + 1 + 1)n,$$
$$T_1(n) = 1 + 9n = \theta(n) \tag{5.1}$$

Similarly, the time complexity is generated for algorithm 2 as follows in Eq. (5.2).

$$T_2(n) = t_1 + t_2 + \cdots + t_9,$$
$$T_2(n) = 1 + (1 + 1 + 1 + 1 + 1 + 1 + 1 + 1)n,$$
$$T_2(n) = 1 + 9n = \theta(n) \tag{5.2}$$

Similarly, the time complexity is generated for algorithm 3 as follows in Eq. (5.3).

$$T_3(n) = t_1 + t_2 + \cdots + t_9,$$
$$T_3(n) = 1 + n + (n * n) + (n * n * n) + 1,$$
$$T_3(n) = 1 + n + n^2 + n^3 + 1 = \theta(n^3) \tag{5.3}$$

Hence the total time complexity of IHC-AA-IoTSD is $T(n)=\theta(n^3)$.

VI. Conclusion and Future Work

In the IoT streaming data era, the sensor devices are generating dynamic data continuously, which is heterogeneous. The IoT data also consists of the real-time streaming data. To perform analysis and annotating the streaming data is a current research problem faced by researchers. Therefore, in this paper, the authors proposed IHC-AA-IoTSD mechanism for unifying the hierarchical clustered data using SPARQL queries. The experimental investigation of IHC-AA-IoTSD has been conducted on three popular healthcare datasets by varying triple data and measuring detection accuracy, precision, TPR, TNR, FPR, and FNR. In the first experimental investigation, the TPR value has been measured under the streaming of triples with stable FPR diverse with 10, 20, 30 and 40%, respectively. In the second experimental investigation, the average results have been taken for an account and proves that the IHC-AA-IoTSD outperforms compared to benchmark mechanisms such as ATLAS, FBASAM, and OBSAA. In the third experimental investigation, the query average execution time has been calculated by taking six different queries under 10k, 20k, 30k, and 50k triples. Considering that IoT streaming data is dynamic and heterogeneous, the proposed mechanism overwhelmed by efficiently annotating the hierarchical clustered data. Moreover, the proposed IHC-AA-IoTSD mechanism outperforms compared to the existing state of the art schemes. In future, the proposed mechanism can be optimized by considering the hash table (key, value pair) for storing SPARQL queries. In addition, artificial intelligent systems need quicker decisions on streaming data. In this scenario, the proposed mechanism may be useful and can achieve efficient results. Besides, it can be considered applying advanced deep learning techniques like Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN), for annotating IoT sensor data with optimum results.

References

[1] G. Xiao, J. Guo, L. D. Xu, and Z. Gong, "User interoperability with heterogeneous IoT devices through transformation," IEEE Transactions on Industrial Informatics, vol. 10, no. 2, pp. 1486– 1496, 2014.

[2] Rohit Dhall & Vijender Kumar Solanki, "An IoT Based Predictive Connected Car Maintenance Approach," International Journal of Interactive Multimedia and Artificial Intelligence, ISSN 1989-1660,Vol 4, no 3, pp 1-13, 2017.

[3] Sivadi Balakrishna, M Thirumaran, R. Padmanaban, and Vijender Kumar Solanki "An Efficient Incremental Clustering based Improved K-Medoids for IoT Multivariate Data Cluster Analysis", Peer-to-Peer Networking and Applications, Springer, Vol 13, no 3, pp 1-23, 2019.

[4] Sivadi Balakrishna, M Thirumaran, and Vijender Kumar Solanki "Machine Learning based Improved GMM Mechanism for IoT Real-Time Dynamic Data Analysis", Journal of Revista Ingeniería Solidaria, Vol 16, No 30, e-ISSN 2357-6014, pp 1-29, 2020.

[5] H. T. Lin, "Implementing Smart Homes with Open Source Solutions", International Journal of Smart Home Vol.7 Issue. 4, pp 289–295, 2013.

[6] Antunes, Mário, Diogo Gomes, and Rui L. Aguiar. "Towards IoT data classification through semantic features." Future Generation Computer Systems, Vol. 8 no 6, pp 792-798, 2018.

[7] M. Junling, J. Xueqin, and L. Hongqi, "Research on Semantic Architecture and Semantic Technology of IoT," Research and Development, vol. 8, no. 5, pp. 26–31, 2014.

[8] Q. Xu, P. Ren, H. Song, and Q. Du, "Security enhancement for IoT communications exposed to eavesdroppers with uncertain locations," IEEE Access, vol. 4, pp. 2840–2853, 2016.

[9] D. Rong, "The Research on Automatic Semantic Annotation Methods", Lanzhou University of Technology, Lanzhou, China, 2012.

[10] F. Chen, C. Lu, H.Wu. Wu, and M. Li, "A semantic similarity measure integrating multiple conceptual relationships for web service discovery," Expert Systems with Applications, vol. 6 Issue.7, pp. 19–31, 2017.

[11] C. De Maio, G. Fenza, M. Gallo, V. Loia, and S. Senatore, "Formal and relational concept analysis for fuzzy-based automatic semantic annotation," Applied Intelligence, vol. 40, no. 1, pp. 154– 177, 2014.

[12] P. Barnaghi, W. Wang, L. Dong, and C. Wang, "A linked-data model for semantic sensor streams," IEEE International Conference on and IEEE Cyber, Physical and Social Computing, Green Computing and Communications (GreenCom '13), Beijing, China, pp. 468–475August

2013.
[13] S. Kolozali, M. Bermudez-Edo, D. Puschmann, F. Ganz, and P. Barnaghi, "A knowledge-based approach for real-time IoT data stream annotation and processing," in Proc: International Conference on Internet of Things, IEEE, pp. 215–222, 2014.
[14] W. Wei and P. Barnaghi, "Semantic annotation and reasoning for sensor data," in Smart Sensing and Context, vol. 5741 of Lecture Notes in Computer Science, pp. 66–76, Springer, Berlin, Germany, 2009.
[15] P. Chenyi, Service-oriented entity semantic annotation in internet of things [M.S. thesis], South China University of Technology, Guangzhou, China, 2015.
[16] J. Bing, "Research on semantic-based service architecture and key algorithms for the internet of things", [Ph.D. thesis], Jilin University, Changchun, China, 2013.
[17] Z. Ming, "Research on several key issues in internet of things applications", [Ph.D. thesis], Beijing University of Posts and Telecommunications, Beijing, China, 2014.
[18] E. Charton, M. Gagnon, and B. Ozell, "Automatic semantic web annotation of named entities," in Advances in Artificial Intelligence, vol. 6657 of Lecture Notes in Comput. Sci., Springer, Berlin, Germany, pp. 74–85, 2011.
[19] G. Diallo, M. Simonet, and A. Simonet, "An approach to automatic ontology-based annotation of biomedical texts," Lecture Notes in Computer Science, vol. 40 no. 31, pp. 1024–1033, 2006.
[20] M. Jacoby, A. Antonic, K. Kreiner, R. Lapacz, J. Pielorz. "Semantic interoperability as key to IoT platform federation," in LNCS 10218: Interoperability and Open- Source for the Internet of Things, pp. 3-19, 2017.
[21] A.P. Plageras, K.E. Psannis, C. Stergiou, H. Wang, B.B. Gupta, " Efficient IoT- based sensor BIG Data collection- processing and analysis in Smart Buildings", Future Generation Computer Systems, 82, pp 349-357, 2018.
[22] A. E. Khaled, S. Helal, "Interoperable communication framework for bridging RESTful and topic-based communication in IoT", Future Generation Computer Systems, Elsevier, 92, pp 628-643, 2019.
[23] Kolozali, S. Puschmann, D.; Bermudez-Edo, M.; Barnaghi, P. "On the Effect of Adaptive and Non adaptive Analysis of Time-Series Sensory Data", IEEE Internet Things J., 3, pp 1084–1098, 2016.
[24] Mazayev, Andriy, Jaime A. Martins, and Noélia Correia. "Interoperability in IoT through the Semantic Profiling of Objects." IEEE Access 6, pp 19379-19385, 2017.
[25] Mayer, Simon, Jack Hodges, Dan Yu, Mareike Kritzler, and Florian Michahelles. "An open semantic framework for the industrial Internet of Things." IEEE Intelligent Systems 32, no. 1, pp 96-101, 2017.
[26] Shi, Feifei, Qingjuan Li, Tao Zhu, and Huansheng Ning. "A survey of data semantization in internet of things." Sensors 18, no. 1, 313, 2018.
[27] Al Zamil, Mohammed Gh, Majdi Rawashdeh, Samer Samarah, M. Shamim Hossain, Awny Alnusair, and Sk Md Mizanur Rahman. "An annotation technique for in-home smart monitoring environments." IEEE Access 6, pp 1471-1479, 2018.
[28] Moutinho, Filipe, Luís Paiva, Julius Köpke, and Pedro Maló. "Extended Semantic Annotations for Generating Translators in the Arrowhead Framework." IEEE Transactions on Industrial Informatics 14, no. 6, pp 2760-2769. 2018.

Sivadi Balakrishna

He received his Bachler of Technology (B.Tech) in the Department of Computer Science and Engineering from Jawaharlal Nehru Technological University (JNTU) in 2010 and Master of Technology (M.Tech) in the Department of Computer Science and Engineering from Jawaharlal Nehru Technological University (JNTU) in 2013, Kakinada, AP, India. He is currently a Full-time Ph.D research scholar from Pondicherry Engineering College in the Department of Computer Science and Engineering, Pondicherry University (A Central University), Pondicherry, India. He has qualified NET (National Eligibility Test) in Dec-2018, which was conducted by UGC. He has more than 4 years of teaching experience in various reputed institutions in computer science and engineering department. He has published more than 15 research articles in various reputed International Journals, International Conferences and Book Chapters. His current research interests are Internet of Things (IoT), Machine Learning, and Semantic Technologies.

M.Thirumaran

He is currently working as an Assistant Professor in Department of Computer Science and Engineering in Pondicherry Engineering College. He has completed his B.Tech in Pondicherry Engineering College in the year 2000 and completed his Post Graduation in Pondicherry University in 2002. He has qualified NET examination for three consecutive years from 2004 to 2006 which was conducted by UGC. He completed his Ph.D in Pondicherry University in the year 2014. He has interested in the domains of Service Oriented Architecture, Web Technology, Web Application Security, Principles of Compiler Design and Automata Theory and Computation. Also he has teaching experience around 15 years in the field of Computer Science Engineering. He has published more than 80 research papers in various reputed International Conferences and International Journals.

Vijender Kumar Solanki

Vijender Kumar Solanki, Ph.D., is an Associate Professor in Department of Computer Science & Engineering, CMR Institute of Technology (Autonomous), Hyderabad, TS, India. He has more than 11 years of academic experience in network security, IoT, Big Data, Smart City and IT. Prior to his current role, he was associated with Apeejay Institute of Technology, Greater Noida, UP, KSRCE (Autonomous) Institution, Tamilnadu, India & Institute of Technology & Science, Ghaziabad, UP, India. He has attended an orientation program at UGC-Academic Staff College, University of Kerala, Thiruvananthapuram, Kerala &Refresher course at Indian Institute of Information Technology, Allahabad, UP, India. He has authored or co-authored more than 50 research articles that are published in journals, books and conference proceedings. He has edited or co-edited 10 books in the area of Information Technology. He teaches graduate & post graduate level courses in IT at ITS. He received Ph.D in Computer Science and Engineering from Anna University, Chennai, India in 2017 and ME, MCA from Maharishi Dayanand University, Rohtak, Haryana, India in 2007 and 2004, respectively and a bachelor's degree in Science from JLN Government College, Faridabad Haryana, India in 2001.He is Editor in International Journal of Machine Learning and Networked Collaborative Engineering (IJMLNCE) ISSN 2581-3242, Associate Editor in International Journal of Information Retrieval Research (IJIRR), IGI-GLOBAL, USA,ISSN: 2155-6377 | E-ISSN: 2155-6385 also serving editorial board members with many reputed journals. He has guest edited many volumes, with IGI-Global, USA, InderScience & Many more reputed publishers.

Edward Rolando Núñez-Valdez

Ph.D. from the University of Oviedo in Computer Engineering. Master's in software engineering from the Pontifical University of Salamanca and B.S. in Computer Science from Autonomous University of Santo Domingo. He has participated in several research projects; He has taught computer science at various schools and universities and has worked in software development companies and IT Consulting for many years. He has published several articles in international journals and conferences. Currently working as Assistant Professor at the University of Oviedo in Spain. His research interests include Software Engineering, Object-Oriented technology, Web Engineering, Recommendation Systems, Artificial Intelligence, Distributed Systems and DSL.

NFC and VLC based Mobile Business Information System for Registering Class Attendance

Sergio Rios-Aguilar[1,2]*, Iñigo Sarría[3], Marta Beltrán Pardo[4]

[1] Universidad Rey Juan Carlos (URJC), Madrid (Spain)
[2] Universidad Politécnica de Madrid (UPM), Madrid (Spain)
[3] Universidad Internacional de La Rioja (UNIR), Logroño (Spain)
[4] Universidad Rey Juan Carlos (URJC), Madrid (Spain)

Received 20 March 2020 | Accepted 29 April 2020 | Published 4 May 2020

Abstract

This work proposes a Mobile Information System for class attendance control using Visible Light Communications (VLC), and the students' own mobile devices for automatic clocking in and clocking out. The proposed information system includes (a) VLC physical infrastructure, (b) native Android and iOS apps for the students, and (c) a web application for classroom attendance management. A proof of concept has been developed, setting up a testbed representing a real-world classroom environment for experimentation, using two VLC-enabled LED lighting sources. After three rounds of testing (n=225) under different conditions, it has been concluded that the system is viable and shows consistent positive detections when the smartphones are on the classroom desk within non-overlapped areas of the light circles generated by the LED lighting sources on the table surface. The performed tests also show that if mobile devices are placed within those overlapping areas, the likelihood of a detection error could increase up to nearly 10%, due to multipath effects, and actions can be taken should it happen. Finally, it has to be highlighted that the proposed autonomous class attendance system allows lecturers to focus on making the most of their time in class, transferring knowledge instead of spending time in attendance management task.

Keywords

Class Attendance, Visible Light Communications, NFC, Presence Information Systems.

DOI: 10.9781/ijimai.2020.05.001

I. Introduction

In the traditional on-site class learning model, student attendance rates have been positively linked in a consistent manner to an increase in academic performance. This attendance is also encouraged by lecturers, which routinely maintain daily records of students attending classes. The attendance recording process, by means of roll call, can be easily unfeasible in large classes, wasting valuable teaching time.

Also, transferring the responsibility of logging their own attendance to the students, passing out a sign-in sheet and collecting it at the end of class may open the possibility of fraud, as absent students could ask other classmates to mark them as "present", and this impersonation can only be detected by means of doing a headcount of the students, which, once again, can take significant time and is prone to error, especially in large classrooms [1],[2].

Therefore, it would be very advisable to devise a highly autonomous class attendance system so that faculties could focus on making the most of their time in class transferring knowledge instead of spending time in attendance management tasks. Also, as the proposed system should be available in every single classroom of the involved teaching institution, the overall cost is an important metric to be considered.

This paper proposes an innovative information system for class attendance management, using Visible Light Communications (VLC) and a mobile app that meet the aforementioned requirements, minimizing costs and also providing every stakeholder involved with a seamless experience.

The rest of the paper is organized as follows: Section II provides an overview of the related work on class attendance systems and the background on enabling technologies, specifically VLC and NFC, discussing their adequacy for the proposed platform. Section III presents the proposed system. Section IV presents the testbed and methods used for experimentation, the results obtained for validation and evaluation of the proposed system, and discussion of these results. Finally, Section V summarizes conclusions, as well as opportunities and future work.

* Corresponding author.
E-mail address: sj.rios@alumnos.urjc.es

II. Related Work And Technology Background

A. On Class Attendance Systems

Several research works have been published related to class attendance recording employing biometric systems, such as fingerprint verification, speech recognition and face recognition [3]-[7].

Systems based in fingerprint verification usually involve high costs, as it is necessary to install a reader in every classroom where attendance control is needed. Also, a single reader does not allow multiple simultaneous verifications so it could lead to potential jams, queues and/or delays in the process.

On the other hand, current non-fingerprint based biometric methods for registering attendance are considered too intrusive and sometimes stressful. For example, [8] acknowledges this problem and proposes methods to make the attendance recording process stress-free

using face biometrics. The system just had 70.83% accuracy during verification when facial expressions were varied along with variations in lighting conditions during enrollment, and thus the process is still stressful for participants. [2] performs several experiments showing that the deflection angles of faces have severe impacts in the process of face identification. Almost no one of the algorithms evaluated could identify the faces with big pitches, roll or yaws.

In a broad sense, low-cost technology is a general requirement for this kind of systems in the academic world, in order to allow the educational institution to provide all students with it, or to take advantage of technology the students may already be in possession of. It should be a secure technology, additionally, which would reduce insofar as possible, any attempt of fraud (i.e. impersonation) [9],[10]

Also, instead of forcing the student to interact directly with a control device in the classroom, the use of wireless technologies is a usual requirement. Several wireless technologies are considered in research works:

- **Bluetooth**: It is a rather widespread wireless technology based on a radio frequency that uses the ISM (Industrial, Scientific & Medical) free radio band on the 2.4 GHz. One of the main advantages of this technology is its widespread in current mobile devices. However, a successful exchange of data does not require that emitting and receiving devices are in the same room, so it is not possible with this technology to prove presence at a classroom [9].
- **Radio-Frequency Identification tags (RFID)**: the student would be issued a personal RFID card, and a card reader would be placed in the classroom, which would send data to the server. Thus there is a need to install a device in the classroom with expensive deployment costs, and, as in the case of fingerprint readers, the time needed for clocking-in could lead to queues. Not to mention the difficulty of preventing impersonation frauds.
- **Near Field Communication (NFC)**: This short-range wireless technology is a promising alternative to the abovementioned technologies in class attendance systems, mainly due to the low cost of NFC tags and the increasing availability of the technology in smartphones. One successful approach is described in [9], but the described clock-in process does require the students to tap an NFC card at the beginning of the class, creating a bottleneck and crowds in large classrooms, and also this approach does not prevent early dropouts.

B. On NFC Technology

Near-field communication (NFC) is a radio-frequency identification system (RFID) that enables devices to communicate in close proximity of around 4cm, and operates at the High Frequency (HF) band of 13.56 MHz, using electromagnetic induction between two loop antennas. The information exchanged is stored in tiny microchips, called Tags in this domain. These tags can be used over a wide range of materials, such as walls in smart posters, glass, ceramic coffee mugs, plastic and, using a special design, over metals [35],[36].

NFC technology is very convenient and easy to use, enabling users to access digital content or perform contactless transactions with a single tap. Most current smartphones are NFC-enabled and support three operation modes:

- NFC reader/writer, in which an NFC-enabled smartphone acts as an initiator device, generating an RF field, and the target device -inexpensive NFC tags- uses that electromagnetic field to power itself for communication, and then transmits some information back to the smartphone.
- NFC card emulation, in which smartphones behave as contactless smart cards, allowing the users to make payments.
- NFC peer-to-peer, in which a smartphone can exchange information directly with another smartphone, creating an ad-hoc link via WiFi or Bluetooth pairing.

To ensure interoperability when different smartphones retrieve information from a passive tag, the NFC Forum defined a standard data format: NDEF (NFC Data Exchange Format), allowing the use of different types of information: plain text, uniform resource identifiers (URIs), location data, vCard, etc. [30].

C. On VLC Technology

Over the last 15 years, the overwhelming development of light-emitting diodes (LEDs) and LED lighting has been one of the major revolutions in the energy sector, mainly thanks to (i) lighting technology development and (ii) global policies put into force worldwide in favour of renewable energy and energy efficiency (LEDs are very environmental-friendly, with reduced emission of CO2 and with no toxic elements, like mercury) [11],[12].

In fact, LEDs have not only very high efficiency (in terms of energy consumption LEDs can save up to 85% compared to classic incandescent lamps, and up to 60% in the case of fluorescents lamps, on an equal basis of brightness and illumination power) but also a long life span (they are expected to last several years), making them ideal for lighting everywhere [13],[14].

One crucial characteristic of LEDs is that visible light (wavelengths from about 380 to 740 nanometers) is harmless to humans (there are no health risks identified distinctly from those of current lighting systems), unlike Infrared (IR) and lasers.

Now, LEDs are also the key pieces that make possible a completely new kind of indoor wireless communications. The light emitted from white lighting LEDs not only can provide illumination, but also can carry binary information at the same time, by means of switching ON and OFF at high frequencies (this is the basis of Visible Light Communications, VLC).

In addition, since LEDs modulation is quite easy and straightforward, it is possible to integrate VLC systems into previous LED lighting-only infrastructure. Finally, in terms of spectrum usage, in VLC it is fully unrestricted, unlike RF, whose spectrum is highly regulated for both military and civil use.

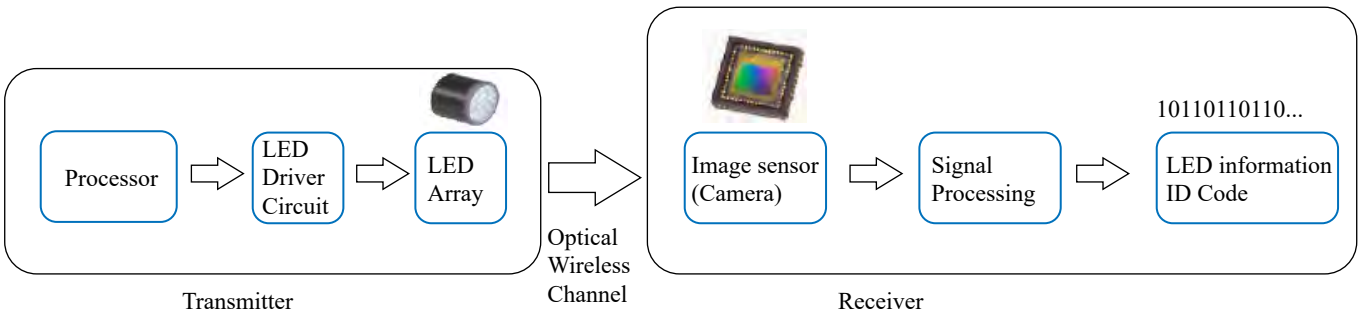


Fig. 1. VLC system components for Optical Camera Communications (OCC).

In the context of VLC as optical wireless communications, we can find two main system approaches: image sensor based VLC and photodiode-based systems. In the first case, the light is emitted by a LED and is detected using an image sensor, for example, a camera available in a consumer electronic device, and this case is also known as Optical Camera Communications (OCC) [15],[16]. In the second case, the detection is made using a photodiode, a very low-cost device with high sensitivity, but due to its own nature as a discrete semiconductor device, it has to be part of custom-made hardware designs (i.e. not easily available to users like other ubiquitous and commonplace consumer electronic devices) [17],[18].

Today, most of the use cases of VLC are based in two key elements. First, the interference-free nature of the light frequency range, and second, the optical communication takes place exclusively within the cone of light.

As shown in Fig. 1, when a LED transmits the binary information of its own ID Code, using On-Off Keying modulation (OOK), the image sensor captures a series of images for further processing, using a wide range of available algorithms (pixel detection, edge detection, centre of colour balance, etc.) in order to evaluate the ON or OFF status to extract the ID Code frame [19].

In the past few years, many research works on VLC have been published; most of them focused on the study of system design, medium access and channel characteristics [20]-[22]. Other research works focus on VLC applications, with a high proportion of them related to indoor positioning (knows as Visible Light Positioning, VLP) [11][19][33] as an alternative to other indoor positioning systems based on WiFi, sound or ultrasound or Bluetooth which show lack of accuracy due to the multipath propagation and radio frequency (RF) signal interference. These VLP-related proposals have a common issue: they usually involve high costs and unnecessary complexity, just for the burden of knowing the precise location of the user [34]. In this case, the need is less complex and limited to provide a system with proof of presence of a sensor under a given VLC LED lamp, and thus those costs are not justified. The work [23] surveyed smart lighting and optical wireless communication based on smartphones. As far as the authors know, at the moment of writing this proposal no research work facing the problem of class attendance recording using VLC has been published.

On the basis of the above background, this research work proposes the usage of VLC for class attendance recording, in the form of Optical Camera Communications, using image sensors -cameras- available in smartphones and a native app, avoiding the costs and complexity of a VLP platform.

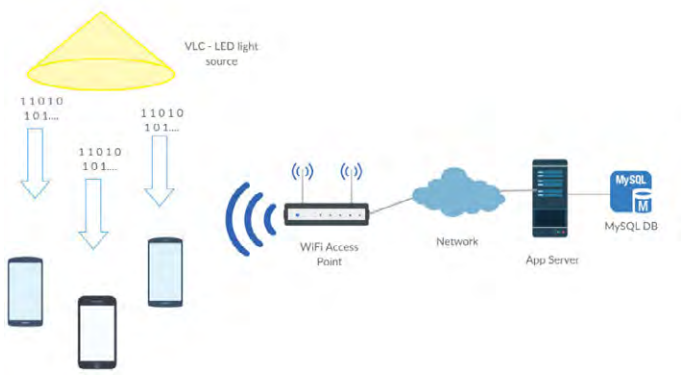


Fig. 2. Proposed System's architecture, including LED lighting sources, Server Application, and mobile phones running a native App.

III. Proposed System

Before describing the relevant components of the proposed system, it is necessary to point out several issues and assumptions made:

- In order to comply with health safety regulations, in particular, to avoid potential epileptic seizures, it has been verified that the modulation of LED lights used is set at a minimum of 1MHz, rendering the visible light generated as flicker-free (over 4.000 times higher than the top-notch 240Hz refresh rate found in current professional high-end monitors). So, it is assumed that the LED switching is too fast and not perceivable by humans or animals [24].
- It was decided that the end-users of the project should be the students themselves. Thus, the workload and responsibility of the lecturer are reduced as he would not have to know each and every student from their first day in class, nor keep an eye on who is missing, and can focus completely on teaching the syllabus [9].

The proposed system is comprised of (a) VLC physical infrastructure, (b) native Android and iOS apps for the students, and (c) a server-based web application for classroom attendance management, as shown in Fig. 2.

The workflow in the proposed information systems is as follows:

1. The mobile phones of the students involved must be registered in the class attendance management web application, with proof of identity. This is a one-time in-person onboarding process, common for all of the courses in which the student has enrolled, and it is carried out using NFC.

   A central service at the University (or Department depending on the system scale) checks the real identity of the students in-situ, showing them a pre-programmed NFC card so that the students can tap it using their mobile phones, after starting the classroom attendance app. The NFC Card shown to the students can be replaced with a different one -containing a different NDEF message, known to the class attendance management web application- daily (See Fig. 3).

2. The classroom attendance app then reads an encoded text string found in the NFC card and allows the students to complete the registration process. The app sends out both the encoded text received and a unique device identifier (ANDROID_ID or Apple's identifierForVendor), as well as other personal data (Name, Surname and Student_ID).

   Before completing this registration process, the students must agree with terms of service, agreeing to abide by the University rules against academic dishonesty specifically related to this classroom attendance system (i.e. impersonation, usage of multi-sim technology in dual-sim mobile phone, etc.).

3. The management web application keeps relational information about classes, classrooms, students, LED infrastructure in the classroom (VLC IDs), and WiFi Access Points reachable at every classroom (BSSIDs).
4. The students start the classroom attendance app and get into the classroom, placing their mobile phones upwards over the table/desk, with the front camera facing the ceiling. A connection to the local WiFi access point is required (the BSSID).
5. The app grabs from the server information about lecture times and random-generated times for automatic clocking-in within time margins specified by faculty members. There will be at least two checks for attendance during each class session, one at the beginning (to prevent lateness) and other at the end (to prevent dropouts). So, let's say the initial margin is defined between [+5min, +15min] after the scheduled beginning of the class. The server randomly forces a check for attendance within the initial

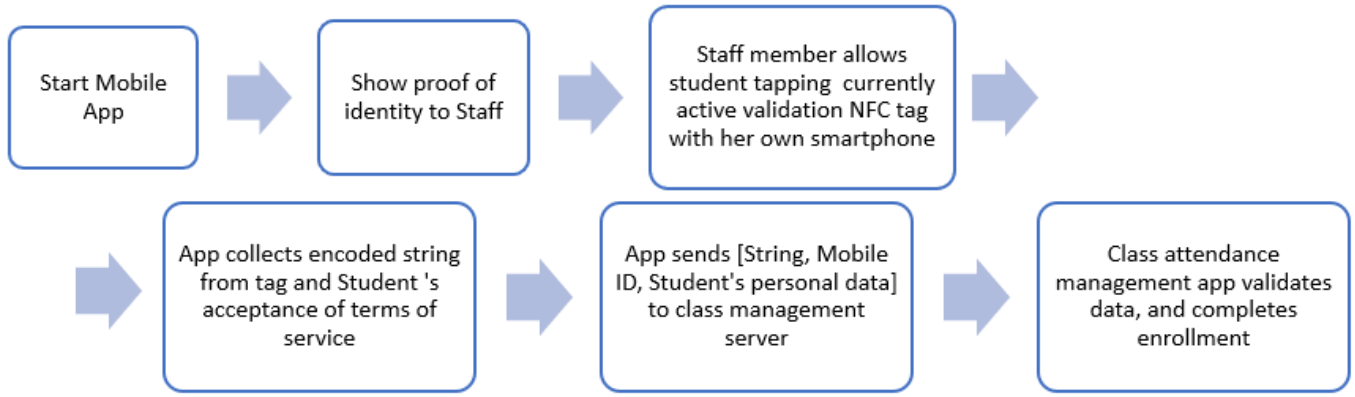


Fig. 3. Flow of the enrollment process (Steps 1-2 in the text), in which the physical validation of the student's identity is carried out offline, followed by an online logical validation.

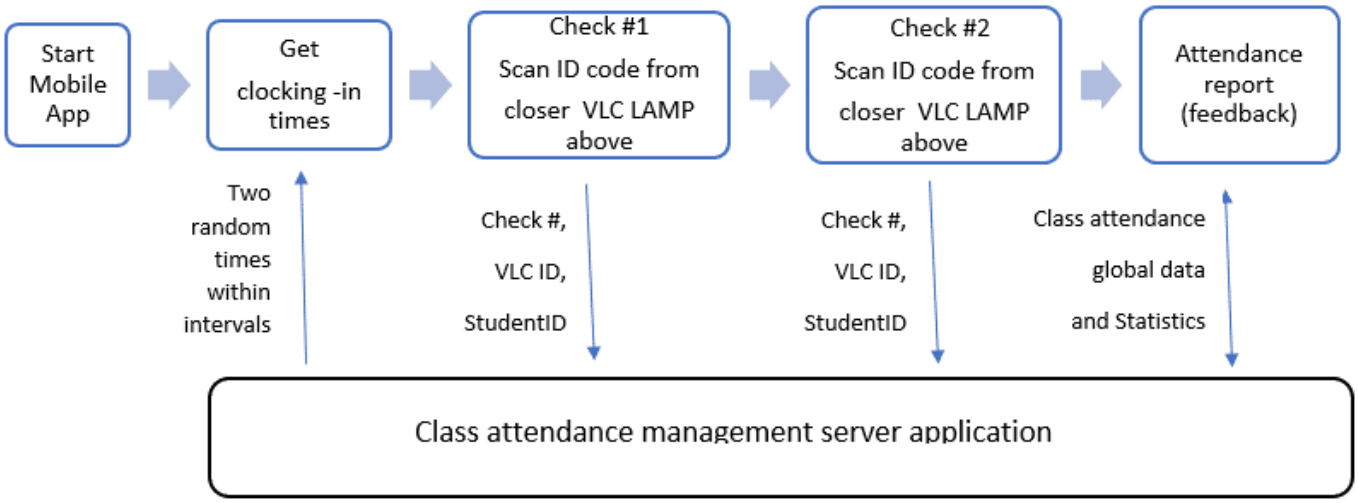


Fig. 4. Flow of operations for registering class attendance in the proposed system, showing the main interactions between the student's mobile app and the class attendance management Web application (Steps 3 to 9 in the text).

margin, for example at +11min (See Fig. 4).

6. The app, at this specific time, activates the front camera sensor of the mobile phone and scans for the ID code of the VLC lamp more closely located above. The transmitted IDs are decoded through the extraction of the bright and dark stripes of modulated LEDs, using the barcode detection algorithm [25].
7. The app sends the decoded ID, the check number, and a timestamp to the server over an encrypted channel using SSL/TLS, and then the server app validates those data and logs the results (PresenceCheckOK, PresenceCheckNotOK) in the database, associating them with the student ID.
8. The app will provide the student with feedback at the scheduled end of class, be it the time spent attending class, or other results (being late but with correct clocking-out, or clocking-in but without a corresponding clock-out).
9. The faculty member can get attendance statistics by using the management class attendance web application.

Regarding security, some of the aforementioned procedures have been included in the proposed system in order to prevent different forms of fraud of impersonation: checking the BSSID of WiFi's access points, using random times for clocking-in and clocking-out and relying on mobile phones' OS-level Unique IDs, mainly.

Also, VLC systems benefit from immunity to radiofrequency (RF) interference, and by the own nature of VLC technology, all types of possible sources of interferences must be in the visible spectrum, and thus they can be easily spotted by humans. This makes Denial-of-Service (DoS) attacks to the LEDs infrastructure easily detectable.

As for the mobile app, there exists the possibility of a DoS attack to the WiFi infrastructure trying to render this kind of required data communications unusable. In fact, DoS attacks based on jamming are really easy to perform on wireless networks, even without sophisticated hardware appliances [32].

But, once again, this kind of attack can be easily detected as it affects a wide range of communication services at the University campus other than the required by the mobile app, so lecturers and other staff can be aware of the problem very early, and act accordingly.

IV. Validation and Evaluation

A. Testbed and Methods

In order to evaluate the proposed system, a testbed has been set up representing a typical classroom environment for two students, which includes a desk table for two students, and two VLC light fixtures (see Fig. 5). Table I shows the detailed arrangements for these elements.

Uniform lighting coverage has been assumed in this work, with purposely generated overlapping light cones over the table, and thus this proposed system will use wide Field-of-View LEDs (with beam angles ≥ 45°). This mode of operation has been chosen due to its closer

resemblance to the real lighting conditions found in classrooms, and also due to the greater complexity of this mode compared to simple spotlighting (discrete and non-overlapping light beams).

Given the beam angle and height of the lamps over the desk table, each light cone intersects the surface of the table resulting in a light circle with a radius r= 0.86m. Given the dimensions of the table, this resulting radius makes sure the lights radiate over the full surface of the table. The distance between 100% light intensity points on the table radiated from both LEDs is d=1.5m.

Fig. 5. Representation of the testbed used for the experiments, with one large classroom table and two VLC-enabled LED lighting sources.

TABLE I. Elements Used in the Testbed, with Placement Information and other Descriptive Items

Element	Data
Room size (m)	4.00 (W) x 5.00 (D) x 2.80 (H)
Number of lamps	2
Individual LED optical power	3W
LED beam angle	60 degrees
LED light color	Warm white – 3000K
Lamp locations (m)	(1.25, 1.50, 1.70) and (2.75, 1.50, 1.70)
Mobile phone locations (m)	([1.10-2.00], [1.10-1.60], 0.70)
Desk table dimensions (m)	2.00 (W) x 0.70 (D) x 0.70(H)

Both light circles overlap -on purpose for testing- in the centre of the table, and the intersection area can be computed as follows:

$$A_{intersection} = 2r^2 \cos^{-1}\left(\frac{d}{2r}\right) - \frac{1}{2}\sqrt{d^2(2r-d)(2r+d)} \quad (1)$$

Using the provided values of r and d in eq. (1) yields:

$$A_{intersection} = 1.10\ m^2$$

This intersection area was marked on the desk table, in order to facilitate the tests in this experiment.

In order to evaluate the proposed system, several tests were made:

- In the first round of tests, the transmitter LEDs were set up to send the same ID Code
- In the second round of tests, the transmitter LEDs were set up to send different ID Codes
- In both rounds of tests, 50% of them were carried out placing the smartphone in several positions within the non-overlapping area of light circles, and the other 50% were done correspondingly within that overlapping area.

For every test, what was measured is the detection of the ID code binary pattern using the mobile app. The possible outcomes registered were: Positive Detection (detection with one single scan) and Negative Detection (no detection at all or wrong ID Code detection). Please note that the tests with the mobile phones placed outside the overlapping areas with the same and different IDs are equivalent and therefore have been subsumed into just one case.

The tools used for the measurements were two high-end smartphones running the test app with the following specifications:

- Samsung Galaxy S10, Android version: 9.0 (Android Pie), front camera: 10MP, f/1.9, 26mm, capable of 1080p at 30 frames per second.
- Apple iPhone XR, iOS version: 12, front camera: 7MP f/2,2, 32mm, capable of 1080p at 60 frames per second.

In both cases, the native test app was developed using Oledcomm GEOLiFi API for VLC Communications following the IEEE 802.15.7 standard, using a 32-bits LED lamp ID Code.

B. Results

The results obtained from the several rounds of tests are summarized in Table II.

TABLE II. Results of the Evaluation of the Proposed System

Test	N	Positive Detection	Negative Detection
Samsung/Android smartphone out of the overlapping area	75	100 % (75)	0%
Samsung/Android smartphone within the overlapping area and Transmitter LEDs sending same ID Code	75	90.7% (68)	9.3%
Samsung/Android smartphone within the overlapping area and Transmitter LEDs sending different ID Code	75	92% (69)	8%
Apple iPhone/iOS out of the overlapping area	75	100 % (75)	0 %
Apple iPhone/iOS within the overlapping area and Transmitter LEDs sending same ID Code	75	93.3% (70)	6.7%
Apple iPhone/iOS within the overlapping area and Transmitter LEDs sending different ID Code	75	90.7% (68)	9.3 %

C. Discussion

After an initial analysis of the obtained results, it seems clear that the barcode OCC detection algorithm is working seamlessly when both types of smartphones are placed within the non-overlapping and interference-free areas of the light circles on the table, (as if the mode of operation had been pure spotlighting from the start).

The problems arise when placing the smartphones within the overlapping area of the light circles on the table. This was somehow expected as in Single Frequency Networks (SFN) structures there are usually severe multipath effects, resulting in inter-symbol interferences [28], [29].

In effect, the rate of negative detections is a bit high, ranging from 6.7% to 9.3%. In any case, such multipath effects will only appear in the overlapping areas, due to the usually good spatial directivity of light in the visible spectrum [27].

Anyway, the consistent behaviour of the results when using both smartphones within the overlapping area is noteworthy, regardless the use of equal or different LED ID codes (5 to 7 false positives in absolute terms in all cases).

As shown in Table II, the results are slightly better when the LED lamps transmit the same binary ID code, and the iPhone device is used. One possible explanation for this may be related to the image sensor size in the iPhone's front camera, with 23% more surface available and doubling the sampling rate than those available in the Samsung Galaxy, and both elements contribute to better behaviour of the OCC detection algorithm when inter-symbol interferences are present.

There are several options to compensate the behaviour of the proposed system in real conditions in a classroom: (i) provide the students with visual feedback of the clock-in process, suggesting them to move the smartphone to a more suitable place on the table in case of error (this could be done when they start the class attendance app), (ii) surveying the uniform lighting conditions of the classroom before using the system and adjusting the usually dimmable LED light fixtures trying to minimize the overlapping areas of the light cones at the tables' height.

V. Conclusions And Future Work

This work proposes an information system for class attendance control using Visible Light Communications (VLC), using the students' own mobile phone devices for automatic clocking in and clocking out.

A complete flow of operations has been described, and a proof of concept has been developed, setting up a testbed representing real-world classroom environment for experimentation. After three rounds of testing (n=225), it has been concluded that the system is viable and works seamlessly when the smartphones are on the desk table within non-overlapped areas of the light circles generated by VLC-enabled LED lighting sources on the table surface. The tests performed also show that if the mobile devices are placed within the aforementioned overlapping areas, the likelihood of a detection error could increase up to nearly 10% (for the specific smartphone makes and models used in the tests), due to multipath effects.

Therefore, it is essential to minimize those multipath effects by (a) reducing the intersections of the light cones at the height of the classroom desks via local dimming of the light sources without affecting the uniform lighting and/or (b) providing the students with an initial setup of the classroom attendance app to check the adequacy of the current placement of their smartphones on the table.

The proposed information system for class attendance control is practical and brings benefits to both faculty staff and students, resulting in greater control for the educational institution, with fewer management burdens and obligations for lecturers, and additionally providing the students with higher quality teaching with a clear increase in actual teaching time.

One limitation of this work refers to the use of just two smartphones for testing, and that the chosen makes and models are very high-end devices at present. It is needed to perform more intensive testing using a wider range of mobile devices, representing the variety and diversity of student-owned smartphones.

Regarding future works, one interesting idea to develop could be an integration of the modulation circuits at the LED sources with the application server, so that the class attendance application could change on a per-class basis the ID Codes being transmitted, thus making more difficult potential impersonations based in reply-attacks.

Another future line of work will involve the study of the impact in positive detections at overlapping areas if LED sources are organized in a cell system where adjacent LED sources used slightly different frequencies (i.e. colour temperatures) not distinguishable by the human eye.

Author Contributions

Sergio Rios-Aguilar: Conceptualization, Methodology, Writing-original draft, Software, Design of Experiment, Data acquisition and analysis, Validation.

Iñigo Sarría: Data Analysis (Review of numerical calculations).

Marta Beltran Pardo: Conceptualization, Writing - Review & Editing, Supervision.

References

[1] E. Garcia, H. Rivera, N. Ponder, R. Kuo, and J. Zheng, "Efficient and cost-effective class attendance management with a smartphone-based system," presented at the Society for Information Technology & Teacher Education International Conference, 2017, pp. 965–972.).

[2] J. He, Y. Zhao, B. Sun, and L. Yu, "Research on video capture scheme and face recognition algorithms in a class attendance system," presented at the Proceedings of the International Conference on Watermarking and Image Processing, 2017, pp. 6–10.

[3] S. Rao and K. J. Satoa, "An attendance monitoring system using biometrics authentication," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, no. 4, pp. 379–383, 2017.

[4] T. Nawas, S. Pervaiz, A. Korrani, and Azhar-ud-din, "Development of academic attendance monitoring system using fingerprint identification," International Journal of Computer Science and Network Security, vol. 9, no. 5, 2009.

[5] Y. Kawaguchi, T. Shoji, W. Lin, K. Kakusho, and M. Minoh, "Face recognition-based lecture attendance system", Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University, 2009.

[6] H. Zaman, J. Hossain, T. Anika and D. Choudhury. "RFID based attendance system", 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2017.

[7] S. Dey, S. Barman, R. Bhukya, R. Das, B. Haris, S. Prasanna and R. Sinha, "Speech biometric based attendance system", 2014 Twentieth National Conference on Communications (NCC), 2014.

[8] K. Okokpujie, E. Noma-Osaghae, S. John, K.-A. Grace, and I. Okokpujie, "A face recognition attendance system with GSM notification," presented at the 2017 IEEE 3rd International Conference on Electro-Technology for National Development (NIGERCON), 2017, pp. 239–244.

[9] M.J.L. Fernández, J. Fernández, S. Rios-Aguilar, B. Selvi, and R.G. Crespo. "Control of attendance applied in higher education through mobile NFC technologies". Expert systems with applications, vol. 40, no. 11, pp. 4478-4489, 2013.

[10] S. Rios-Aguilar, J. Pascual-Espada, and R. González-Crespo. "NFC and Cloud-Based Lightweight Anonymous Assessment Mobile Intelligent Information System for Higher Education and Recruitment Competitions". Mobile Networks and Applications, vol. 21, no. 2, pp. 327-336, 2016.

[11] J. Lian, Z. Vatansever, M. Noshad, and M. Brandt-Pearce. "Indoor visible light communications, networking, and applications". Journal of Physics: Photonics, vol. 1, no. 1, p. 012001, 2019.

[12] C. Jurczak. "Review of LiFi visible light communications: research and use cases". Lucibel White Paper. arXiv preprint arXiv:1802.01471, 2017

[13] M. Figueiredo, C. Ribeiro, A. Dobesch, L. N. Alves, and O. Wilfert, "Consumer LED lamp with ODAC technology for high-speed visible light communications," IEEE Trans. Consumer Electron., vol. 63, no. 3, pp. 285-290, August 2017.

[14] A. Jovicic, J. Li, and T. Richardson, "Visible light communication: opportunities, challenges and the path to market". IEEE Commun. Mag., vol. 51, no. 12, pp. 26-32, December 2013.

[15] D. Ganti, W. Zhang and M. Kavehrad. "VLC-based indoor positioning system with tracking capability using Kalman and particle filters." In 2014 IEEE International Conference on Consumer Electronics (ICCE), IEEE, pp. 476-477, January 2014.

[16] M. Yoshino, S. Haruyama and M. Nakagawa. "High-accuracy positioning system using visible LED lights and image sensor." In Radio and Wireless Symposium, IEEE, pp. 439- 442, January 2008.

[17] T.H., Do, M.Yoo. "TDOA-based indoor positioning using visible light".

Photonic Netw. Commun. no. 27, 2014
[18] Sari Yamaguchi, Vuong V. Mai, Truong C. Thang and Anh T. Pham. "Design and Performance Evaluation of VLC Indoor Positioning System using Optical Orthogonal codes. In 2014 IEEE Fifth International Conference on Communications and Electronics (ICCE), pp. 54-59, IEEE, 2014
[19] N. U. Hassan, A. Naeem, M.A. Pasha, T. Jadoon and C. Yuen. "Indoor positioning using visible led lights: A survey". ACM Computing Surveys (CSUR), vol. 48, no.2, p. 20, 2015
[20] D. Karunatilaka, F. Zafar, V. Kalavally, and R. Parthiban, "LED Based Indoor Visible Light Communications: State of the Art," IEEE communications surveys and tutorials, vol. 17, pp. 1649-1678, 2015
[21] P. H. Pathak, X. Feng, P. Hu, and P. Mohapatra, "Visible Light Communication, Networking, and Sensing: A Survey, Potential and Challenges," *IEEE Communications Surveys & Tutorials,* vol. 17, pp. 2047-2077, 2015.
[22] M. A. Khalighi and M. Uysal, "Survey on free space optical communication: A communication theory perspective," IEEE Communications Surveys & Tutorials, vol. 16, pp. 2231-2258, 2014.
[23] A. Sevincer, M. Bhattarai, M. Bilgi, M. Yuksel, and N. Pala, "LIGHTNETs: Smart LIGHTing and mobile optical wireless NETworks—A survey," IEEE Communications Surveys & Tutorials, vol. 15, pp. 1620-1641, 2013.
[24] H. Haas. "A light-connected world". Physics World, vol. 29, no. 8, p. 30, 2016.
[25] C. Fu, C. W. Cheng, W. H. Shen, Y. L. Wei and H. M. Tsai, "Lightbib: marathoner recognition system with visible light communications," IEEE Int. Conf. Data Science and Data Intensive Sys., pp. 572-578, 2015.
[26] Y. Li, Z. Ghassemlooy, X. Tang, B. Lin, and Y. Zhang, "A VLC smartphone camera based indoor positioning system". IEEE Photonics Technology Letters, vol. 30, no.13, pp.1171-1174, 2018.
[27] J. Song, W. Ding, F. Yang, H. Yang, B. Yu, B. and H. Zhang. "An indoor broadband broadcasting system based on PLC and VLC". IEEE Transactions on Broadcasting, vol. 61, no. 2, pp.299-308, 2015.
[28] R. Kaur and S. Arora. "Nature Inspired Range Based Wireless Sensor Node Localization Algorithms". International Journal of Interactive Multimedia & Artificial Intelligence, vol. 4, no. 6, pp. 7-17, 2017.
[29] M. Nighot, A. Ghatol and V. Thakare. "Self-Organized Hybrid Wireless Sensor Network for Finding Randomly Moving Target in Unknown Environment". International Journal of Interactive Multimedia & Artificial Intelligence, vol. 5, no. 1, pp.16-28, 2018.
[30] A. Lesas and S. Miranda. "The art and science of NFC Programming". Hoboken, NJ: Jonh Wiley & Sons, Inc, 2017.
[31] C. Rohner, S. Raza, D. Puccinelli and T. Voigt. "Security in visible light communication: Novel challenges and opportunities". Sensors & Transducers Journal, vol. 192, no.9, pp. 9-15, 2015
[32] A.D. Wood, J.A. Stankovic and G. Zhou. "DEEJAM: Defeating energy-efficient jamming in IEEE 802.15. 4-based wireless networks". In 2017 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, pp. 60-69. IEEE. June, 2017
[33] D. H. Mai and A. T. Pham. Implementation and Evaluation of VLC-Based Indoor Positioning Systems for Smart Supermarkets. In 2018 9th International Conference on Awareness Science and Technology (iCAST), pp. 273-278. IEEE, 2018.
[34] Y. Zhuang, L. Hua, L. Qi, J. Yang, P. Cao, Y. Cao, Y. Wu, J. Thompson and H. Haas. "A Survey of Positioning Systems Using Visible LED Lights". IEEE Communications Surveys & Tutorials, 2018. 10.1109/COMST.2018.2806558.
[35] D. Saeed, R. Iqbal, h.H.R. Sherazi and U.G. Khan." Evaluating Near-Field Communication tag security for identity theft prevention". Internet Technology Letters, vol. 2, no. 5, p. e123, 2019
[36] A. Lazaro, R.Villarino, and D. Girbau. "A survey of NFC sensors based on energy harvesting for IoT applications". Sensors, vol. 18, no.11, 3746, 2018.

Sergio Ríos Aguilar

Sergio Rios received the Master's degree in Telematics and Computer Systems from Universidad Rey Juan Carlos (Spain). He also holds a PhD degree (Cum Laude) in Economics (Organization Engineering) from the Department of Management, School of Business and Economics, Universidad de Granada (Spain) in 2013. Dr. Ríos has worked as a consultant in projects for several international companies like Samsung Electronics, Nokia, INTU Group, and others. He is a member of the European Global Navigation Satellite Systems Agency's (GSA) Galileo Raw Measurement Taskforce. His current research interests include Mobile Business Information Systems, Mobile Presence and attendance Systems, Location based Services, Mobile Edge Computing and Internet of Things.

Iñigo Sarría Martínez de Mendivil

Iñigo Sarria holds a Doctorate (Cum Laude and special prize) from the International University of La Rioja, a Bachelors degree in Mathematics from the University of the Basque Country, a University Expert qualification in Analysis of the Knowledge Society from The International University of La Rioja, and a Certificate in Teaching from the Complutense University of Madrid. He is Associate Vice-Chancellor of Academic Affairs and PDI of the Computer Science and Technology Area of the School of Engineering and Technology at the International University of La Rioja. ORCID code: https://orcid.org/0000-0002-2584-9671

Marta Beltrán Pardo

Marta Beltrán received the master's degree in Electrical Engineering from Universidad Complutense of Madrid (Spain) in 2001, the master's degree in Industrial Physics from UNED (Spain) in 2003 and the PhD degree from the Department of Computing, Universidad Rey Juan Carlos, Madrid (Spain) in 2005. She is currently working with this department as an Associate Professor. She is the leader of the GAAP research group, co-founder of the Cybersecurity Cluster and she has published extensively in high-quality national and international journals and conference proceedings in the areas of security&privacy, and parallel and distributed systems. Her current research interests are Cloud computing, Edge/Fog Computing and Internet of Things, specifically, Identity and Access Management and privacy-preserving mechanisms for these paradigms.

On Improvement of Speech Intelligibility and Quality: A Survey of Unsupervised Single Channel Speech Enhancement Algorithms

Nasir Saleem[1]*, Muhammad Irfan Khattak[1], Elena Verdú[2]

[1] Department of Electrical Engineering, University of Engineering & Technology, Peshawar (Pakistan)
[2] Universidad Internacional de La Rioja, Logroño (Spain)

Received 20 September 2019 | Accepted 27 November 2019 | Published 18 December 2019

Abstract

Many forms of human communication exist; for instance, text and nonverbal based. Speech is, however, the most powerful and dexterous form for the humans. Speech signals enable humans to communicate and this usefulness of the speech signals has led to a variety of speech processing applications. Successful use of these applications is, however, significantly aggravated in presence of the background noise distortions. These noise signals overlap and mask the target speech signals. To deal with these overlapping background noise distortions, a speech enhancement algorithm at front end is crucial in order to make noisy speech intelligible and pleasant. Speech enhancement has become a very important research and engineering problem for the last couple of decades. In this paper, we present an all-inclusive survey on unsupervised single-channel speech enhancement (U-SCSE) algorithms. A taxonomy based review of the U-SCSE algorithms is presented and the associated studies regarding improving the intelligibility and quality are outlined. The studies on the speech enhancement algorithms in unsupervised perspective are presented. Objective experiments have been performed to evaluate the potential of the U-SCSE algorithms in terms of improving the speech intelligibility and quality. It is found that unsupervised speech enhancement improves the speech quality but the speech intelligibility improvement is deprived. To finish, several research problems are identified that require further research.

Keywords

Unsupervised Speech Enhancement, Speech Quality, Speech Intelligibility, Noise.

DOI: 10.9781/ijimai.2019.12.001

I. Introduction

SINGLE channel speech enhancement (SCSE) [1]-[16] is one of the significant researched problems in many speech related applications; such as, Automatic Speech Recognition (ASR) [17], Speaker Identification (SI) [18], Human-Machine interaction [19], etc. The problem occurs whenever an interfering noise signal degrades the target speech signal. The interfering noise signals could be convolutive [20] or additive. The convolutive noise signal is produced because of the reverberation. However, additive noise signals are usually supposed since this supposition expresses the uncomplicated solutions and practically more adequate results have been attained with the algorithms structured on such theory [21] [22]. The additive noise distortions significantly aggravate the quality and intelligibility of the speech signals. For this reason, the objective of the speech enhancement algorithms is to quantify the estimate of the underlying clean speech from the noisy speech to increase the intelligibility and quality of the noisy speech signal [21] [22]. A fundamental structure of the SCSE is shown in Fig. 1. A variety of speech related applications exists in our everyday situations where speech enhancement is required as for example: (i) the humans are present in the noisy environments and communicating on the mobile phones, (ii) listening to a call in the noisy street or in the factory, (iii) sitting in subway or travel in a car. In these situations, a speech enhancement could be used to ease the communication by reducing the noise signals. A number of speech enhancement methods have been designed at the front-end to create robust ASR systems by decreasing the discrepancies between the training and testing stages. In ASR, a speech enhancement method is applied to minimize the noise prior to the feature extraction phase. An additional imperative application of the speech enhancement system is for those individuals using hearing aid devices. The speech signals show extremely redundancy and normal hearing listeners can comprehend the target speech signals even in adverse signal-to-noise ratios (SNRs) [23]-[26]. For instance, a normal hearing individual can comprehend approximately 50% of the words spoken in a multitalker corrupted speech at signal to noise ratio equal to 0 dB [27]. However, for individuals with hearing problem (hearing loss), various speech parts could totally be inaudible or significantly distorted. Therefore,

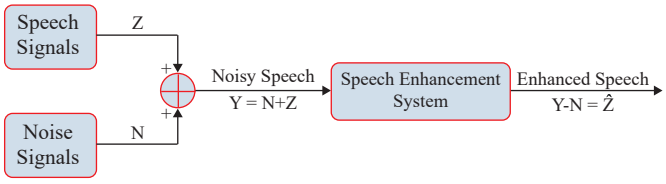


Fig. 1. Single Channel Speech Enhancement.

* Corresponding author.
E-mail address: nasirsaleem@gu.edu.pk

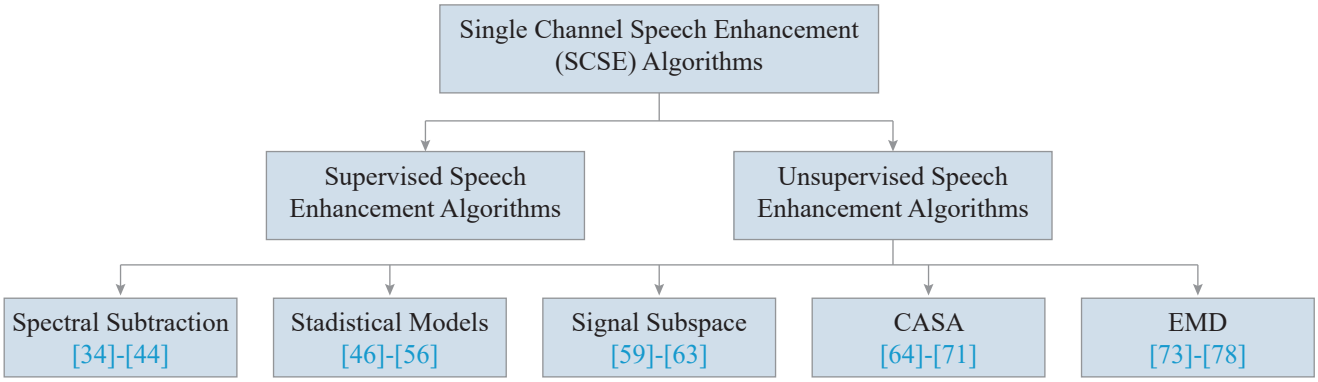


Fig. 2. General Classification of the U-SCSE Algorithms.

the perceived speech signals have small redundancy. Consequently, the individuals with hearing loss feel problem in the noisy environments [28]-[30]. Large attention towards designing the robust speech enhancement algorithms is given to decrease the listening effort and improve the speech intelligibility [31] [32]. The combinations of such algorithms with contemporary digital signal processing systems are implemented in a number of speech related devices.

Single-channel speech enhancement algorithms are divided into two major categories: Supervised SCSE (S-SCSE) algorithms and Unsupervised SCSE (U-SCSE) algorithms. In U-SCSE algorithms, a statistical model is used for speech/noise and the estimate of the underlying clean speech is quantified from the noisy speech devoid of prior facts about speaker identity and noise. Thus, no supervision and classification of the signals is required. Alternatively, the S-SCSE algorithms use models for speech and noise. The model parameters are learned through training of the speech and noise samples and models are defined by mixing the separate models for the speech and noise and the speech enhancement task is performed. In this category, therefore, prior supervision and classification of the speech or noise type is a requisite. The emphasis of this paper is to present a survey on the U-SCSE algorithms.

The remaining paper is organized as follows: Section II shows an extensive review of U-SCSE algorithms in terms of the speech intelligibility and quality. Section III presents experiments performed to evaluate the speech intelligibility and quality potentials of U-SCSE algorithms. Section IV presents the concluding remarks of the survey. Finally, section V presents important research problems which require further study.

II. Classification of U-SCSE Algorithms

This category includes a wide range of U-SCSE algorithms; however, general classification is not limited to the presented algorithms. In U-SCSE algorithms, a statistical model is used. The estimated underlying clean speech is quantified from the input noisy speech utterances devoid of previous facts about speaker identity and noise. A general classification and fundamental framework of the U-SCSE algorithms is shown in Fig. 2-3. In subsequent sub-sections, we provide a taxonomy based review of the U-SCSE algorithms.

A. Spectral Subtraction-based Speech Enhancement Algorithms

Spectral subtraction (SS) based speech enhancement is simple, effective and traditionally one of the pioneer methods proposed for reducing noise distortion. Noise signals are assumed to be additive. Spectral subtraction based speech enhancement algorithms were initially proposed by Boll [33]. In SS, the estimate of the underlying clean speech spectrum could be obtained by subtracting the estimate of noise spectrum from the noisy spectrum. The noise spectrum

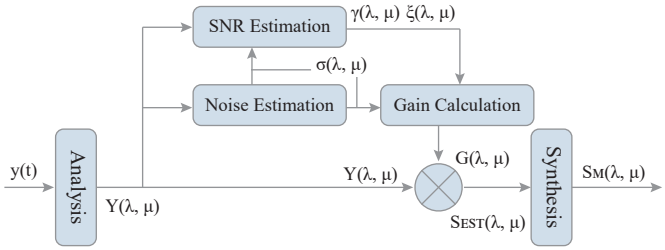


Fig. 3. U-SCSE System.

is estimated and updated during pause periods i.e., absence of the speech signals. The hypotheses for designing such algorithms are: **(i)** the stationary or slowly varying process and, **(ii)** the noise spectra do not vary drastically during updating periods. The enhanced speech is acquired by using inverse transform of the estimated spectrum using noisy phase. According to the basic principle of SS, let us assume that a noisy signal $z(n)$ is composed of the clean speech $s(n)$ and the additive noise signal, $e(n)$

$$z(n) = s(n)+e(n) \tag{1}$$

Computing the STFT of (1), we obtain:

$$Z(\omega,k)=S(\omega,k)+E(\omega,k) \tag{2}$$

Subtract noise magnitude spectrum $|D(\omega,k)|$ from the noisy speech magnitude spectrum $|Y(\omega,k)|$ and finally take the inverse Fourier transform of the difference spectra using the noisy phase to produce the enhanced speech signal, given by equation as:

$$\hat{S}(\omega,k)=[Z(\omega,k)-E(\omega,k)]e^{j\phi_Z(\omega,k)} \tag{3}$$

Since, noise signals are non-stationary and time-variant in the real-world environments; the SS-based enhancement approaches produce negative values for the estimated magnitude spectrum of the clean speech and result in musical noise artifact in enhanced speech. The research is done in near past to reduce the musical noise artifact. Some highly ranked researches on the SS for the speech enhancement are reviewed.

Lu and Loizou [34] proposed a spectral subtraction algorithm based on the geometric approach for the speech enhancement which addressed the inadequacy of the traditional SS algorithm. An efficient scheme to estimate the cross-terms is proposed which is involved in the phase differences between the speech and noise signals. After analyzing the suppression function of the proposed algorithm, it is examined that the algorithm holds the properties of the conventional minimum mean square error (MMSE) algorithm. The evaluation confirmed that geometric approach for the speech enhancement performed considerably better than the conventional spectral subtractive algorithm. A similar approach is also presented in [35].

Paliwal *et al.* [36] examined SS in the modulation domain, an

unconventional acoustic domain for the speech enhancement task, and showed capability of SS in the new domain. Analysis-modification-synthesis (AMS) framework is included and reduced musical noise artifact by applying the modulation-domain based SS algorithm. Moreover, consequences of the frame duration on speech quality have been examined. The outcomes of research indicated that frames with duration with 180-280 msec provided optimized results in terms of the spectral distortions and temporal slurring. For further improvements in the speech quality, a fusion with the MMSE principle has been presented in the short-time spectral domain by joining the magnitude spectrum of the proposed speech enhancement algorithm. Consistent improvements in speech quality have been achieved for different SNRs.

Zhang and Zhao [37] proposed an approach, and performed subtraction on the real and imaginary spectrum independently in modulation-domain. An enhanced magnitude and phase is achieved through the SS approach. Inoue *et al*. [38] provided a theoretical investigation of the musical noise artifact created by the SS on higher order statistics. It is assumed that power SS approach is a common used form. Generalization of SS for the unpredictable exponent parameters has been provided and the quantity of the musical noise artifact has been compared between several exponent-domains. A less musical noise artifact has been observed for a lower exponent spectral-domain and offered good quality and intelligible speech.

Miyazaki *et al*. [39] provided a theoretical examination of the musical noise artifact with an iterative SS method. Iteratively weak-nonlinear signal processing technique has been used to obtain a high quality speech with low musical noise artifact. The generation of musical noise artifact has been formulated by marking changes in kurtosis of the noise spectrum. Optimal internal parameters have been derived theoretically in order to produce no musical noise and explained that with a fixed point in kurtosis yield no musical artifacts.

Antonio *et al*. [40] proposed an improved algorithm based on the SS for real-time noise cancellation and applied the algorithm to the gunshot acoustical signals. A pre-processing approach based on spectral suppression algorithm is applied instead of post-filtering, which requires *a priori* information concerning the direction of arrival of desired signals. Ban and Kim [41] proposed an algorithm for reducing the reverberant noise to the application of remote-talking speech recognition. The SS has been used and the spectra of late reverberant signals are estimated by considering the delayed and attenuated versions of reverberant signals. The unpredictable weight sequences have been estimated via a Viterbi-decoding method based on the reverberation model. The weight sequences are then replaced with fixed weights in SS without estimating the reverberation time.

Hu and Wang [42] proposed a novel algorithm to separate the unvoiced speech signals from the non-speech interfering signals. The voiced speech and periodic parts of interfering signals have been firstly removed. The interference became stationary and the noise energy has been estimated in unvoiced intervals utilizing the separated speech in adjacent voiced intervals. The SS is applied to create time-frequency segments in unvoiced intervals and the unvoiced segments are then grouped. The grouping of segments is based on the frequency characteristics of unvoiced segments by considering thresholding and Bayesian classification.

Kokkinakis *et al*. [43] described and evaluated the capabilities of SS to suppress the late reflections and compared to ideal reverberant masking (IRM) approach. Speech intelligibility outcomes indicated that SS approach can suppress additive reverberant energy to a degree similar to that attained by the IRM. Hu and Yu [44] proposed an adaptive noise spectral estimator to deal with subtraction-based techniques for speech enhancement. The proposed method derived the noise spectrum from a primary estimate of noise spectrum together with the current noisy speech spectrum in an adaptive style. The fundamental framework of SS remained uninterrupted even in case of the gain for all spectral components is altered. The listening tests confirmed the superiority of the noise adaptation technique in suppressing the musical noise artifact and quality improvements.

B. Statistical Model-based Speech Enhancement Algorithms

In the statistical model based speech enhancement algorithms, speech and noise signals are assumed stationary and the resultant filter coefficients remain unchanged. The suppression of noise signals could effortlessly be realized utilizing Finite Impulse Response (FIR) or Infinite Impulse Response (IIR) filters. However, noise sources and particularly the speech signals are highly non-stationary. The speech generation trails a time-varying process. By using the noisy spectrum $Z(\omega,k)$, the short-time noise power spectral density (PSD) and the frequency-domain signal-to-noise ratio (SNR) are quantified to determine the weighting gains. The actual spectral weighting is achieved by multiplying the noisy spectrum $Z(\omega,k)$ by weighting gains $G(\omega,k)$ resulting in quantifying DFT coefficients of underlying clean speech according to the following equation:

$$\hat{S}(\omega,k)=G(\omega,k)Z(\omega,k) \tag{4}$$

The computation of the weighting gains rely on the particular speech enhancement algorithms and is usually a function of short-term noise PSD estimate $P_D^2(\omega,k)$ and the SNR estimates $\gamma(\omega,k)$ and $\xi(\omega,k)$ as:

$$\gamma(\omega,k)=\frac{|Z(\omega,k)|^2}{P_E^2(\omega,k)}, \quad \xi(\omega,k)=\frac{P_S^2(\omega,k)}{P_E^2(\omega,k)} \tag{5}$$

Where $\gamma(\omega,k)$ and $\xi(\omega,k)$ indicate *a posteriori* and *a priori* SNR estimate, $P_S^2(\omega,k)=E\{S(\omega,k)\}$ and $P_E^2(\omega,k)=E\{E(\omega,k)\}$ show variance of the clean speech and noise signals. $P_E^2(\omega,k)$ is calculated during the non-speech/ pauses-periods by using standard recursive equation, given as:

$$\hat{P}_E^2(\omega,k)=\beta\hat{P}_E^2(\omega,k-1)+(1-\beta)\hat{P}_Z^2(\omega,k-1) \tag{6}$$

Where, β is the smoothing factor and $\hat{P}_Z^2(\omega,k-1)$ is the noise estimate in the previous frame. The *a priori* SNR can be estimated by using Decision Direct (DD) [1] approach, given as:

$$\xi(\omega,k)=\alpha.\frac{\hat{X}^2(\omega,k-1)}{\hat{P}_D^2(\omega,k-1)}+(1-\alpha).\max\left[\frac{Y^2(\omega,k)}{\hat{P}_D^2(\omega,k)}-1,0\right] \tag{7}$$

Where, α is weighting parameter, $\hat{X}^2(\omega,k-1)$ and $\hat{P}_D^2(\omega,k-1)$ represent the power spectrum estimation of the clean speech and noise at *k-1* frame, respectively. In the following subsections distinguished and latest statistical speech enhancement algorithms based on the Wiener filtering (WF), minimum means square error (MMSE), Gaussian and super-Gaussian models are surveyed.

1. Wiener Filtering

Wiener filtering based speech enhancement minimizes the mean square error (MSE) between the estimated speech magnitude spectrum and the original signal magnitude spectrum. The formulation of the optimal wiener filter gain is as follows: [45]

$$G(\omega,k)=\frac{\xi(\omega,k)}{\xi(\omega,k)+1} \tag{8}$$

Over the years, Wiener filtering and its variants are used for the speech enhancement task. We discuss and review some of the highly ranked research studies on WF algorithms.

Huijun *et al* [46] proposed a SCSE algorithm which exploited connections between various time-frames to minimize residual noise. Contrasting to the traditional speech enhancement methods that apply a post-processor after standard algorithms like spectral subtraction, the proposed method applied a hybrid Wiener spectrogram filter (HWSF)

to reduce noise, trailed by a multi-blade post-processor that exploited two-dimension features of the spectrograms to retain the speech quality and to further reduced the residual noise. Spectrograms comparison showed that the proposed method significantly reduced the musical noise distortions. The usefulness of the proposed method is additionally confirmed by the use of objective assessments and unceremonious subjective listening tests.

Jahangir and Douglas [47] proposed a frequency-domain optimal linear estimator with perceptual post-filtering. The proposed method incorporated the masking properties of human hearing system to make the residual noise inaudible. A modified way is presented to quantify the tonality coefficients and relative threshold offsets for the best possible estimation of noise masking threshold. The proposed speech enhancement method has been evaluated for noise reduction and speech quality under many noisy conditions and yielded better results than [1].

Almajai and Milner [48] examined the visual speech information to enhance the noisy speech. The visual and audio speech features are analyzed which identified a pair with the highest audio-visual connection. The research revealed that high audio-visual connections exist inside individual phoneme rather entire speech. This connection is used in the application of a visually-driven Wiener filtering, which achieved clean speech and noise power spectrum statistics from the visual features. Clean speech statistics are quantified from the visual features using a maximum *a posteriori* structure and is incorporated inside the states of hidden Markov network to afford phoneme localization. Noise statistics are achieved by using a novel audio-visual voice activity detector, which used visual speech features to formulate the robust speech/nonspeech classifications. The efficiency of the proposed method is evaluated subjectively and objectively which confirmed the superiority.

Marwa *et al.* [49] presented adaptive Wiener filtering approach for speech enhancement. The proposed approach depended on the adaptation of the filter transfer function from sample-to-sample based speech signal statistics (the local mean and variance). The method is implemented in the time-domain to contain time-varying nature of the speech. The approach is evaluated against conventional frequency domain spectral subtraction, wavelet denoising methods and Wiener filtering using different speech quality metrics. The results showed superiority of the proposed Wiener filtering method.

Xia and Bao [50] proposed a Weighted Denoising Auto-encoder (WDA) and noise classification based speech enhancement approach. Weighted reconstruction loss function is established into standard Denoising Auto-encoder (DAE) and link between the power spectrums of underlying clean and noisy speech is expressed by WDA structure. The sub-band power spectrums of underlying clean speech are quantified using the WDA structure from the noisy speech. The *a priori* SNR is quantified using *a Posteriori* SNR Controlled Recursive Averaging (PCRA) approach. The enhanced speech is achieved by the Wiener filter in the frequency-domain. Moreover, GMM-based noise classification method is engaged to make the proposed method appropriate for various conditions. The experimental results demonstrated that the proposed method achieved improved objective speech quality. Effective noise reduction and SNR improvements are attained with less speech distortion.

Kristian and Marc [51] investigated speech-distortion weighted inter-frame Wiener filters for the SCSE in a filterbank configuration. The filterbank configuration utilized a regularization parameter as a tradeoff between speech distortion and noise reduction. The method depends on the quantification of inter-frame correlation coefficients, and it is shown that these coefficients could be robustly estimated using a secondary higher resolution filterbank. It is then demonstrated that real-valued scalar gains can be applied directly in higher resolution filterbank rather than inter-frame filtering in the primary filterbank, which leads to a robust noise reduction performance for any value of regularization parameter.

2. MMSE Estimators

The minimum means square error (MMSE) estimator [1] inheres to vital class of the estimators and quantifies the spectral magnitudes. The MMSE estimator reduces the quadratic error of the spectral speech amplitudes according to the following equation:

$$E\left\{(S(\omega,k)-\hat{S}(\omega,k))^2\right\} \rightarrow Min \tag{9}$$

Considering the Gaussian model of the speech and noise, the final weighting rule is given according to [1] as:

$$\hat{S}(\omega,k) = E\left\{S(\omega,k)/Z(\omega,k)\right\} \tag{10}$$

$$\hat{S}(\omega,k) = \frac{\sqrt{\nu(\omega,k)}}{\lambda(\omega,k)}\Gamma(1.5)\,F_1(-0.5,1,\nu).Z(\omega,k)$$
$$\nu(\omega,k) = \frac{\xi(\omega,k)}{\xi(\omega,k)+1}.\gamma(\omega,k) \tag{11}$$

Γ(.) and F1(.) shows Gamma function and Hypergeometric function, respectively. We discuss and review some of the highly ranked research studies on MMSE algorithms.

Basheera *et al.,* [52] proposed novel optimum linear and nonlinear estimators. They are derived based on the MMSE sense to reduce the distortion in original speech. Linear and nonlinear bilateral Laplacian gain estimators are proposed. The observed signal is first decorrelated through a real transform to achieve its moment coefficients and then applied to the estimated speech signal in the decorrelated domain. The mathematical aspect of MSE of estimators is evaluated suggesting significant improvement. Kandagatla and Subbaiah [53] derived joint MMSE estimation of speech coefficients provided phase uncertainty by assuming the speech coefficients. Uncertain phase is used for amplitude estimation. Furthermore new Phase-blind estimators are designed utilizing the Nagakami power spectral density function and the generalized Gamma for speech and noise priors.

Hamid *et al*. [54] addressed the problem of speech enhancement using β-order MMSE-STSA. The advantages of the Laplacian speech modeling and β-order cost function are taken in MMSE estimation. An investigative solution is presented for the β-order MMSE-STSA estimator deeming Laplacian priors for DFT coefficients of the clean speech. A Gaussian distribution for the real and imaginary parts of the DFT coefficients of the noise is presupposed. Using estimates for the joint PDF and the Bessel function, a better closed-form adaptation of the estimator is also presented.

Gerkmann and Krawczyk [55] derived a MMSE optimal estimator for underlying clean speech spectral amplitude. It is shown that the phase contains extra information which can be used to differentiate outliers in the noise from the target signals. *Matthew and Bernard* in [56] proposed a Bayesian STSA stochastic deterministic speech model, which included *a priori* information by utilizing a non-zero mean. For the speech STFT magnitude, investigative expressions are derived in the MMSE principle whereas phase in maximum-likelihood principle. An approach for quantifying *a priori* stochastic deterministic speech model parameters is explained based on the harmonically related sinusoidal parts in the STFT frames and deviations in magnitude and phase of components between succeeding STFT frames.

C. Signal Subspace-based Speech Enhancement Algorithms

Signal subspace [57] [58] based SE approaches use KLT, SVD and EVD to disintegrate noisy speech signals into the noise plus

signal subspace known as the signal-subspace, whereas eliminates the noise signal that falls within orthogonal noise-subspace. The signal-subspaces are processed separately to remove noise components utilizing a diagonal gain matrix based on uncorrelated components in subspace. The components of the gain matrix are quantified by time-domain or spectral-domain estimators. The covariance matrix R_Z of the noisy speech can be written as:

$$R_Z = R_s + R_E \tag{12}$$

R_s and R_E are the covariance matrices of the clean speech and noise signals. R_Z is supposed to have a higher rank than R_s. The EVD of the covariance matrices is given as:

$$R_S = V \Lambda V^T \tag{13}$$

$$R_D = V(\sigma_W^2 I)V^T \tag{14}$$

$$R_Y = V(\Lambda + \sigma_W^2 I)V^T \tag{15}$$

Λ indicates a diagonal matrix that contains the Eigen-values, V indicates an orthonormal matrix containing eigenvectors; σ shows variance of noise whereas I indicate identity matrix. Speech enhancement process is represented by a filtering operation input speech vector as:

$$z = \{z(1), z(2), y(3), \ldots . , z(n)\}^T \tag{16}$$

$$\hat{S} = \Psi Z \tag{17}$$

The term Ψ is the filtering matrix, given by equation (18) as:

$$\Psi = V_P G_P V_P^T \tag{18}$$

Where G_p holds weighted Eigen values of R_Z, and V_p and V_p^T shows KLT and its inverse matrices, respectively. We discuss and review highly ranked research studies on SigSub algorithms.

Borowicz and Petrovsky [59] examined speech enhancement methods based on the perceptually motivated signal subspace. Lagrange multipliers are used to modify the spectral-domain-constrained (SDC) estimator. The residual noise power spectrums are shaped with an algorithm for accurate computing the Lagrange multipliers. The proposed approach uses masking phenomena for residual noise shaping and is optimal for the case of colored noise. Results show that the proposed method outperformed the competing methods and provided high noise reduction and improved speech quality.

Mohammad *et al.* [60] proposed a non-unitary spectral transformation of the residual noise based on diagonalization of covariance matrices associated to the clean speech and noise signals. Through this transformation, the optimization problem is solvable devoid of any constraints on the structure of contributed matrices.

Vera [61] pointed out that estimation of the dimension of signal subspace is critical and depends on the noise variance as well as SNR. Both fluctuate along temporal segments of speech and frequency bands. It is anticipated to work over frames in all critical bands utilizing the threshold noise variance. Belhedi *et al* [62] used soft mask as a core in the proposed approach. The method produces two separate signals of dissimilar qualities and made them available in two separate channels. The classification of the channels is made via Fuzzy logic that needs two separate parameters. One parameter determines quality and intelligibility whereas the second parameter determines the gender of the speaker via F_0 tracking method. The proposed approach achieved an average 59.5% improvement in SIR, 67.9% progress in PESQ, and 10.5% improvement in TPS.

Sudeep and Kishore [63] proposed a perceptual subspace approach via masking properties of the human auditory system with variance normalization to decide the gain parameters. An estimator is used to determine the filter coefficients. The noise is handled by substituting the noise variance by Rayleigh quotient. Normalization of variance is made by removing the spikes to evade rapid increase or decrease in power of the output samples making the output more intelligible.

D. Computational Auditory Scene Analysis-based Speech Enhancement Algorithms

The field of computational study intends to achieve human performance in the Auditory Scene Analysis (ASA) by using single microphone recordings of the acoustic prospect. This definition describes the biological relevance of the field by limiting the microphone number to two and its functional goal of Computational Auditory Scene Analysis (CASA). The CASA uses perceptually motivated mechanisms. Over the years, CASA based methods are used for the speech enhancement; here we are reviewing some of the work in recent years.

A new ideal ratio mask (IRM) depiction is proposed by Bao and Abdulla in [64] by utilizing inter-channel correlation. The power ratio of the speech and noise during the structuring of ratio mask is adaptively reallocated; therefore more speech components are held and noise components are masked simultaneously. Channel-weight contour is assumed to modify the mask in all Gammatone filterbank channels.

Wang *et al.* [65] proposed IRM estimation that relies on the spectral dependency into the speech cochleagram to enhance noisy speech. A data field representation is established to design time-frequency connection of the cochleagram with adjacent spectral information to estimate IRM. Firstly, a pre-processed section is used to achieve initial time-frequency values of noise and speech. Then the data field model is used to obtain the forms of speech and noise potentials. Subsequently, the optimal potentials that reveal their respective optimal distribution are achieved by the optimal influence factors. Lastly, masking values are obtained via the potentials of the speech and noise for reinstating the clean speech signals.

Wang *et al.* [66] considered a novel approach of speech and noise models, and presented two model-based soft decision methods. A ratio mask is computed by the exact Bayesian estimators of speech and noise. Additionally, a probabilistic mask is estimated with a variable local criterion. Liang *et al.* [67] considered local correlation knowledge from two aspects for improved performance. The time-frequency segmentation-based potential function is derived to represent the local correlation between mask labels of neighboring units directly. It is demonstrated that time-frequency unit that belongs to one segment is mostly dominated by one source. Alternatively, a local noise level tracking phase is integrated. The local level is attained by averaging many neighboring time-frequency units and is considered as a method for accurate noise energy. It is utilized as an intermediary auxiliary variable to signify the correlation. A high dimensional posterior distribution is simulated by a Markov Chain Monte Carlo (MCMC) approach. During iterations, the correlation is fully utilized to quantify the acceptance ratio. The estimated ideal binary mask (IBM) is achieved using the expectation operator. The proposed approach is compared and evaluated with a Bayesian approach and the approach yielded considerably large performance gain in terms of SNR gain and HIT-FA rates.

Narayanan and Wang [68] presented a system for robust SNR estimation based on CASA. The proposed method used an estimate of the IBM to separate a time–frequency illustration of the noisy speech signal into speech and noise dominated sections. Energy inside each region was totaled to gain the filtered global SNR. SNR transformation was established to translate the estimated SNR to the true global SNR of the noisy speech signal.

Hu and Wang [69] proposed a tandem algorithm to estimate the pitch of a target speech utterance and separated the voiced regions

of the target speech. First, a coarse estimate of the target pitch was obtained and then the estimate is used to segregate target speech using harmonicity and temporal continuity. Lee and Kwon [70] proposed a CASA-based speech separation system and matched the missing speech parts by using the shape analysis method.

May and Dau [71] presented a method based on the estimate of the ideal binary mask from noisy speech in supervised learning of AMS features and auditory inspired modulation filterbanks with logarithmically scaled filters were used. Spectro-temporal integration stage was incorporated to obtain speech activity information in neighboring time-frequency units.

E. Empirical Mode Decommission-based Speech Enhancement Algorithms

Empirical Mode Decomposition (EMD) [72] directly extracts the energy related to different intrinsic time scales. EMD is an adaptive approach and follows some necessary steps to decompose nonlinear and nonstationary data. (**i**) First, the EMD obtains the local maxima and minima. (**ii**) Secondly, the EMD finds the local maximum and local minimum envelopes. (**iii**) Third, the EMD finds the mean of the obtained local extrema envelopes and finally subtracts this mean envelope from the input data to attain the residual intrinsic mode function (IMF).

Upadhyay and Pachori [73] proposed a novel speech enhancement method for suppressing stationary and non-stationary noise sources. The variational mode decomposition (VMD) and EMD approaches are combined to develop the new idea for speech enhancement. Firstly, the EMD decomposes the input noisy speech into the IMFs. The VMD is then applied on the summation of preferred IMFs. The Hurst exponent was used to select the IMFs. The proposed speech enhancement method reduced low and high-frequency noise sources and showed enhanced speech quality.

Khaldi *et al.* [74] presented a speech enhancement method that exploited the combined effects of EMD and the local statistics of the speech signal by utilizing the adaptive centre weighted averaging filter. The speech signals were segmented into frames and all frames were segmented down by EMD into IMFs. The filtered IMFs depend on the voiced or unvoiced frame. An energy norm was utilized to classify the voiced frames and a stationarity index was used between unvoiced and transient chain. Zao *et al.*, [75] proposed a speech enhancement scheme based on the adoption of Hurst exponent during the selection of IMFs to reconstruct the target speech.

Hamid *et al.*, [76] proposed a novel data adaptive thresholding approach. The noisy speech signals and fractional Gaussian noises were mixed to generate the complex noisy signal. Bivariate EMD was used to decompose the complex noisy signal into complex-valued IMFs and all IMFs were segmented into short-time frames for processing. The variances of the IMFs of fractional Gaussian noise computed inside the frames were used as the reference to categorize subsequent frames of noisy speech into signal-dominant and noise-dominant frames, respectively. A soft thresholding method is used at noise-dominant frames to decrease the effects of noise. Every frame and IMF of the speech signals were combined to yield the enhanced speech signal.

Chatlani and Soraghan [77] used the EMD as a post-processing stage for filtering low frequency noise. An adaptive approach was designed to choose IMF index for sorting out the noise component from speech components. This separation was carried out by using a second-order IMF statistics. The low-frequency noise components were removed by the biased reconstruction from the IMFs. Khaldi *et al.*, [78] used EMD for fully data-driven based approaches for noise reduction. Noisy speech signal was decomposed adaptively into IMFs using sifting process. The signal reconstruction with IMFs was done using the MMSE filter and thresholded using a shrinkage function.

The U-SCSE algorithms provide acceptable speech quality and noise reduction in many real-world noise sources. The U-SCSE algorithms along with several advantages also came with some limitations. The TableI and Table II provides advantages and limitations of various U-SCSE algorithms. These limitations will point out several

TABLE I. Problem Statements, Methodologies, Contributions and Limitations of U-SCSE Algorithms

Method	Problem Statement	Methodology	Contribution	Limitation
GA-SS [34]	Speech enhancement to improve speech quality and to reduce the musical noise distortion.	Compute the magnitude spectrum of the noisy signal using the FFT. The noise spectrum is updated using noise estimators. The gain is estimated using modified gain and multiplied with noisy spectrum to enhance speech.	Performed significantly better than the traditional spectral subtraction algorithm in terms of speech quality and musical noise artifact.	Speech intelligibility is not evaluated. Additionally informal tests were conducted for evaluations. Noise reduction impact on speech intelligibility research is required.
MOD-SS [36]	Speech enhancement to improve speech quality and intelligibility in Modulation domain.	The SE method used AMS-based modulation domain. Each frequency component of the acoustic magnitude spectra is processed frame-wise across time using a modulation AMS framework, and the enhanced modulation spectrum is computed.	New Speech enhancement domain in terms of SS is explored. Better speech quality and speech intelligibility is obtained. Better noise reduction is offered.	Although the proposed method offered better results, the combination with other domains produces complexity in the proposed method. The complexity of the method is not discussed.
MOD-SS [37]	Speech enhancement to improve speech quality in Modulation domain.	The magnitude subtraction is adopted and extended into the modulation frequency domain for the separate enhancements of the real and imaginary spectra. The noise is estimated in real and imaginary spectra and the estimated speech is recreated.	Perform subtraction on the real and imaginary spectra separately in the modulation frequency domain. Better noise reduction and speech quality is achieved.	The speech intelligibility potential of the proposed method is not discussed. The method estimated the phase, thus the complexity of the method is not discussed.
VAD-SS [39]	The Speech enhancement for better results and musical noise reduction in the Kurtosis of noise spectra.	Iteratively weak-nonlinear method is used to obtain quality speech with less musical artifact. The generation of musical artifact is formulated by marking changes in kurtosis of the noise spectrum. Optimal internal parameters are derived theoretically to produce no musical artifact in kurtosis.	The proposed method provided better results and generation of musical noise artifact is formulated in the Kurtosis of noise spectra.	No theoretical explanation is given, only experimental results are presented. Speech quality and intelligibility is not discussed.

TABLE II. Problem Statements, Methodologies, Contributions and Limitations of U-SCSE Algorithms

Method	Problem Statement	Methodology	Contribution	Limitation
SDW-IFWF [51]	Speech enhancement for better quality and to reduce the musical noise distortion	Speech-distortion weighted inter frame Wiener filters for noise reduction is implemented in a filter bank structure. The filters utilized a regularization parameter as a tradeoff between speech distortion and noise reduction. The method depends on the estimation of inter frame correlation coefficients and these coefficients are more robustly estimated using a secondary higher resolution filter bank.	The contribution of the paper is the implementation of the scalar SDW-IFWF gain in a HRFB, matching a principle in the crucial lower-resolution filter bank to improve the speech quality and noise reduction with less musical artifact	The algorithm provided improved results in terms of the speech quality However, speech intelligibility potential of the proposed algorithm is not discussed and evaluated.
LBLG-NBLG [52]	Speech enhancement for better quality speech and low speech Distortion	The estimators are derived on the basis of MMSE to reduce the distortion of the fundamental speech. The musical artifact is reduced without affecting the noise reduction. LBLG and NBLG estimator are proposed. The input signal is decorrelated to obtain moment coefficients. The estimators are applied to estimate the clean signal in the decorrelated domain. The original signal is obtained in time domain.	The proposed method obtained better speech quality and noise reduction. Non-linear and linear bilateral Laplacian estimators are derived to improve the speech quality.	Although method produced better speech quality as compare to traditional methods; however, the speech intelligibility and complexity potentials are not fully explored.
EPW-Sub [60]	Speech Separation in optimized subspace for improved quality and intelligibility.	The separation is achieved by optimizing the subspace via decomposing the mixture signal into three subspaces: sparse, sub-sparse and low-rank subspaces. Soft masking is used for the final verdict. Two signals of different qualities are provided in two separate channels. The channel classification is made by using Fuzzy logics with two parameters. F0 tracking algorithm is proposed to classify gender.	Embedded pre-whitening subspace method is proposed based on controlled spectral-domain for better speech quality and noise reduction in colored noises.	Although the proposed method offered better results but the speech intelligibility in non-stationary noise sources is not discussed.
CASA-SE [65]	The Speech enhancement for improved quality and intelligibility in the data driven field of cochleagram.	Iteratively weak-nonlinear method is used to obtain quality speech with less musical artifact. The generation of musical artifact is formulated by marking changes in kurtosis of the noise spectrum. Optimal internal parameters are derived theoretically to produce no musical artifact in kurtosis	Ideal Ratio Mask is estimated in the data driven field of cochleagram to enhance the noisy speech. The proposed method obtained considerable gain in speech quality. Better results in terms of energy loss and residue noise are contributed.	The proposed algorithm has not incorporated the DF model into the STFT domain. The complexity of the algorithm is not discussed.

research areas which need further research.

III. Speech Intelligibility and Quality Potential of Various U-Scse Algorithms

The Table I-II illustrates the problem statements, methodologies, contributions and limitations of U-SCSE algorithms. It is clear from Table I-II that the U-SCSE algorithms addressed the problem of the speech enhancement effectively for noise reduction, musical noise artifact and speech quality. Speech enhancement is usually used as the front-end to Automatic speech recognition systems where speech intelligibility is the more important attribute. It is observed from the survey of the above different classes that speech intelligibility attributes is not fully explored in most of the U-SCSE algorithms. This section provides an intense experimental evaluation to observe the quality and intelligibility potentials of the U-SCSE approaches.

A. Methods

The experiments represent the measures used to evaluate and validate the performance of speech enhancement algorithms. In experiments, the U-SCSE algorithms are evaluated by using a set of 60 noisy speech sentences belonging to female and male speakers in terms of the speech intelligibility and quality. The noisy stimuli are generated by adding four real-time background noises to the clean speech utterances at several signal-to-noise ratios (SNR). The clean speech sentences are selected from the standard IEEE database [85] randomly. Four nonstationary noise sources (street, exhibition hall, airport, and multitalker babble noise) are chosen from the Aurora database [86]. The speech utterances are mixed at four SNR from 0dB to 15dB, spacing 5dB applying the ITU-T P.51. The sampling rate is fixed at 8 kHz. Five classes of U-SCSE algorithms are included in the experiments performed for speech quality and intelligibility. The U-SCSE classes include Spectral Subtraction (SS), Wiener Filtering (WF), Minimum Mean Square Error (MMSE) estimators, Signal Subspace (SigSub) and EMD type. Table III provides the details of speech enhancement algorithms used in the experiments. Two evaluation measures are quantified in order to access the U-SCSE algorithms. The PESQ [87] is preferred for the speech quality; an ITU-T P.862 standard that substituted the obsolete ITU-T P.861 standard because of inadequate performance to evaluate the speech enhancement. The PESQ score follows the range of −0.5 and 4.5, but, during experiments the score follows the mean opinion score (MOS), that is, a range of 1.0 to 4.5. The PESQ scores are calculated using the following equation:

$$PESQ = \eta 0 + \eta 1.DSYM + \eta 2.DASYM \quad (18)$$

Where $\eta 0 = 4.5$, $\eta 1 = -0.1$ and $\eta 2 = -0.039$.

TABLE III. List of U-SCSE Algorithms in Experiments

S. No	Speech Enhancement Class	Speech Enhancement Algorithm
1	Spectral Subtractive (SS)	SS [79] SS-RDC [80] MBSS [81]
2	Wiener Filtering (WF)	WF [45] WWF [82]
3	Minimum Mean Square Estimation (MMSE)	MMSE-SPU[1] LMMSE [2]
4	Signal Subspace (SigSub)	KLT [83] PKLT [84]

TABLE IV. PESQ Analysis of U-SCSE Algorithms

Noise Type	SNR (dB)	Spectral Subtractive			Wiener Type		Statistical-Model		Signal Subspace		EMD
		SS	RDC	MBSS	WF	WWF	MMSE	LMMSE	KLT	PKLT	H-EMD
Airport	0dB	1.59	1.69	1.81	1.92	1.18	1.23	1.95	1.78	1.51	1.84
	5dB	2.03	2.16	2.20	2.12	2.03	1.43	2.12	2.13	2.02	2.01
	10dB	2.39	2.35	2.54	2.43	2.27	1.54	2.45	2.29	2.08	2.63
	15dB	2.95	2.74	3.12	3.05	2.62	1.65	3.03	2.79	2.42	2.93
Babble	0dB	1.45	1.68	1.98	1.78	1.16	1.26	1.92	1.34	1.34	1.91
	5dB	2.07	2.16	2.28	2.12	2.13	1.53	2.12	2.11	1.98	2.19
	10dB	2.42	2.36	2.59	2.46	2.34	1.67	2.53	2.37	2.25	2.72
	15dB	2.60	2.61	2.75	2.67	2.55	1.85	2.71	2.61	2.51	2.88
Exhibition Hall	0dB	1.25	1.49	1.43	1.69	1.33	1.33	1.72	1.37	1.63	1.77
	5dB	1.87	1.91	2.01	2.01	1.81	1.61	1.95	1.89	1.50	1.93
	10dB	2.47	2.14	2.44	2.40	2.39	1.79	2.46	2.44	2.28	2.57
	15dB	2.82	2.46	2.82	2.78	2.65	1.95	2.79	2.86	2.50	2.89
Street	0dB	1.49	1.51	1.54	1.60	1.53	1.53	1.72	1.59	1.55	1.79
	5dB	2.05	1.98	2.14	2.06	2.08	1.88	2.04	2.12	2.12	2.14
	10dB	2.49	2.36	2.61	2.60	2.33	2.03	2.52	2.32	2.14	2.63
	15dB	2.92	2.53	2.89	2.74	2.65	2.25	2.77	2.84	2.55	2.92

TABLE V. Across-Class Comparative Analysis of U-SCSE Algorithms in terms of PESQ

Noise Type	SNR (dB)	Spectral Subtraction			Wiener Filtering		MMSE Estimation		Signal Subspace		EMD
		SS	RDC	MSS	WF	WWF	MMSE	LMMSE	KLT	PKLT	H-EMD
Airport	15dB			*			*	*			
	10dB			*			*	*			*
Babble	15dB			*	*		*	*			*
	10dB			*	*		*	*			*
Exhibition Hall	15dB			*	*		*	*			*
	10dB			*	*		*	*			*
Street	15dB			*	*		*	*			*
	10dB			*			*	*			

TABLE VI. Across-Class Comparative Analysis of U-SCSE Algorithms in terms of STOI

Noise Type	SNR (dB)	Spectral Subtraction			Wiener Filtering		MMSE Estimation		Signal Subspace		EMD
		SS	RDC	MSS	WF	WWF	MMSE	LMMSE	KLT	PKLT	H-EMD
Airport	15dB	*		*			*	*		*	
	10dB			*			*	*			*
Babble	15dB		*	*	*	*	*	*	*	*	*
	10dB			*	*		*	*			*
Exhibition Hall	15dB	*	*	*	*		*	*	*		*
	10dB			*	*		*	*			*
Street	15dB	*		*	*	*	*	*	*		*
	10dB			*			*	*			

Note: Algorithms specified by asterisks sign executed equally well whereas algorithms without asterisks sign executed poorly.

5	Empirical Mode Decomposition (EMD)	H-EMD [75]

A separate evaluation metric is used to access the intelligibility of the enhanced speech. The short-time speech intelligibility (STOI) [88] is considered for this purpose. The STOI sores are calculated by the equation given as:

$$\mathrm{f(STOI)} = \frac{100}{1+\exp(a\mathrm{STOI}+b)} \tag{19}$$

The parameters a, b are set according to [8], $a = -17.4906$ and b = 9.6921.

B. Results and Discussion

A performance comparison analysis at two levels is presented in this section. First, within-class performance comparison of the U-SCSE algorithms is established. The five classes are Spectral Subtractive, Statistical-models, Wiener-Filtering type, Subspace and EMD-type. This performance comparison was conducted to observe the significant performance differences within-class algorithms. Secondly, across-classes performance comparison is conducted to evaluate and find the algorithm(s) that performed better in all noisy situations.

1. Within-Class Algorithm Comparison

Table IV provides the results for PESQ (speech quality) whereas average speech intelligibility results are demonstrated in Fig. 4. Of three tested spectral-subtractive algorithms, the multi-band spectral subtraction (MBSS) [81] performed constantly the best across all noisy situations in terms of the speech quality. The MBSS and SS-RDC [80] methods performed equivalently well excluding 0dB exhibition hall

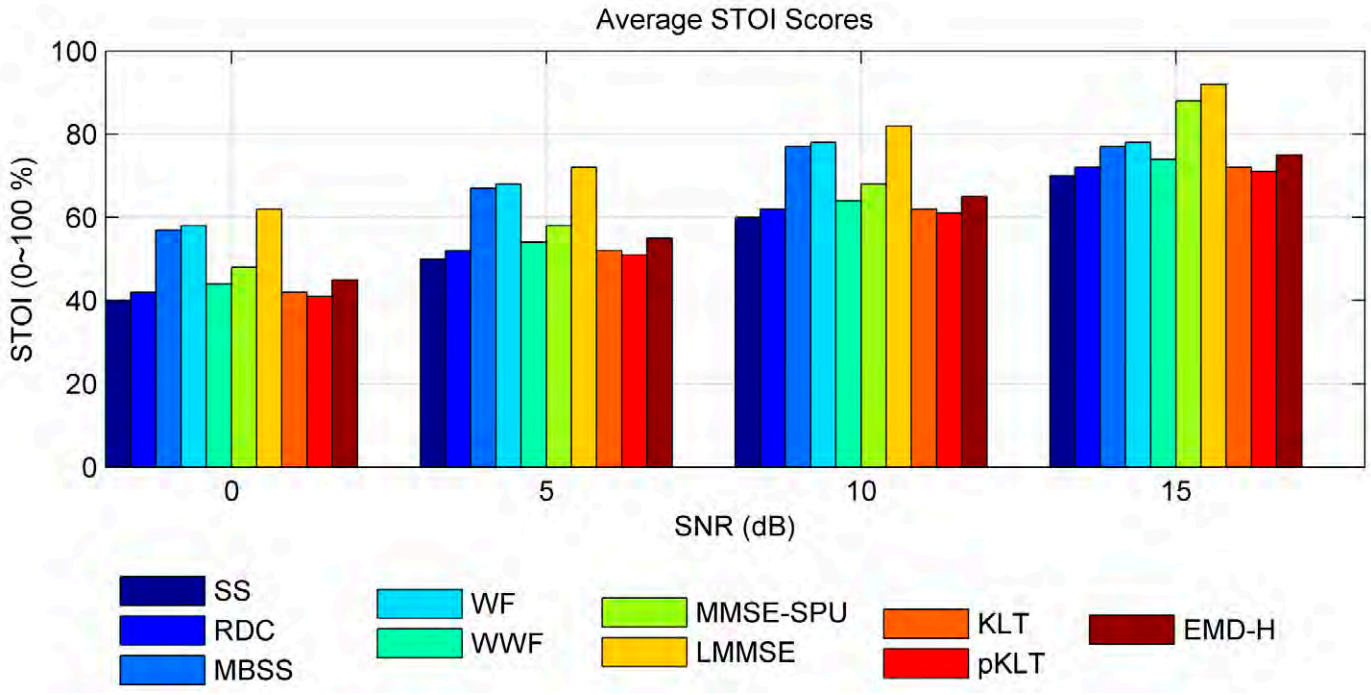


Fig. 4. Average Speech Intelligibility prediction for U-SCSE algorithms in terms of STOI.

noise and 0dB street noise conditions. Noise distortion of SS-RDC algorithm was considerably less than the MBSS and SS [79] approaches in all noisy situations. In terms of speech intelligibility, the MBSS and SS-RDC approaches equally performed in most of the noisy situations excluding 0dB exhibition hall noise and 0dB street noisy situations, where MBSS algorithm performed notably superior and presented less speech distortion. In brief, MBSS performed better than SS-RDC and SS, providing better overall speech intelligibility and quality. For speech quality, the two subspace approaches performed equally for the most of SNRs and noise types, excluding 0 dB babble noise.

The two Wiener-type algorithms performed well for most SNR conditions and four types of noise except 0dB airport noise and 0dB babble noise. For speech quality, the WF [45] performed significantly better than the WWF [82] approach at all SNRs and noise sources. WWF performed poorly in all noise sources at almost all SNRs and significant residual noise is experienced in the enhanced speech. On the other hand WF-as offered better speech quality and the noise reduction capabilities were significant. For speech intelligibility, the WF-as performed well at all SNRs and noise sources as compared to WWF method. There is significant speech distortion observed in the output speech utterance of the WWF approach.

The two statistical-model based approaches performed good for most of SNRs and noise types. The log-MMSE (LMMSE) [2] performed significantly better than the MMSE-SPU [1] approach at all SNRs and noise sources. MMSE performed poorly in all noise sources at almost all SNRs, and significant residual noise observed in the enhanced speech. On the other hand LMMSE offered better speech quality and noise reduction capabilities were significant. For speech intelligibility, the MMSE-SPU performed very poorly at all SNRs and noise sources. The small speech intelligibility signifies the higher speech distortion offered by MMSE-SPU. LMMSE offered better speech intelligibility and comparatively less speech distortion is experienced in the output speech.

The generalized subspace approach, KLT [83] performed significantly better than the pKLT [84] approach at all SNRs and noise sources except 0dB exhibition hall noise. The KLT approach was more successful in suppressing the background noise and perceptual speech quality. In terms of speech intelligibility, KLT and pKLT approaches performed equally well at all SNRs and noise sources except 0dB exhibition hall noise. There is no significant improvement in speech intelligibility observed for pKLT approach. On the other hand, KLT improved speech intelligibility marginally.

In terms of the speech quality, the EMD-H [75] algorithm performed well for all SNRs and noise types, except at 0dB exhibition hall noise and 0dB street noise. The EMD-H was successful in suppressing the background noise and improving the perceptual quality and speech intelligibility at all SNRs and the noise sources.

2. *Across-Class Algorithm Comparison*

Table V-VI indicates the results achieved by using ANOVA statistical analysis for the speech quality and intelligibility. Asterisk sign in Table V-VI show lack of statistical significant difference between algorithms with the utmost scores and the denoted algorithms. The U-SCSE algorithms marked by the Asterisk sign in Table V performed similarly. Table V indicates no single algorithm is categorized as the best, and several speech enhancement algorithms performed equally well across SNRs situations and noise types. In terms of the speech quality, MMSE-SPU, LMMSE, WF, EMD-H and MBSS performed equally well across all SNRs situations. Table VI indicates the results achieved from the ANOVA statistical analysis for speech intelligibility. The MMSE-SPU, LMMSE, MBSS and WF performed well. All algorithms produced low speech distortion (high intelligibility) across all SNRs situations and noise sources. KLT, SS-RDC and WWF algorithms also performed well in isolated SNR situations.

IV. Conclusion

This paper presented a comprehensive review of the different classes of the single-channel speech enhancement algorithms in unsupervised perspective in order to improve the intelligibility and quality of the contaminated speech. Various classes of the unsupervised speech enhancement approaches for enhancing the noisy speech have been discussed. We have summarized possible algorithms of the Spectral Subtraction (SS), Wiener Filtering (WF), Minimum Mean

Square Error (MMSE) estimators, Signal Subspace (SigSub) and EMD type, explained state-of-the-art approaches and a many related studies have been reviewed. The review suggested that unsupervised speech enhancement methods show an acceptable speech quality but speech intelligibility potential remains medium. The algorithms of unsupervised class show better noise reduction however; decrease of the residual noise artifact and speech distortion requires further research. Different unsupervised speech enhancement approaches have distinctive advantages that make these algorithms appropriate for speech enhancement; in contrast, these algorithms have some serious limitations as well. Table I-II summarized the problem statements, methodologies, contributions and the limitations of many speech enhancement algorithms. On the basis of the limitations extracted from the reviewed papers and also from the experimental results, it is concluded that unsupervised speech enhancement improves the speech quality but the speech intelligibility improvement potential requires further research. The algorithm can use the noise estimators, but accurate estimate is also a difficult task. A too aggressive estimation may lose important speech contents which in turn affect the speech intelligibility whereas too low noise estimation may lead to the residual noise. We have outlined various problems that need research to design robust single-channel speech enhancement algorithms. This rapid progress in the unsupervised speech enhancement algorithms will possibly persist in the future. To conclude, some following open research problems are outlined that are extracted from research studies:

1. Generalization to the Nonstationary Noise Sources: Although U-SCSE algorithms provide promising speech quality results in stationary noise sources, however, their performance in nonstationary noise sources is not high. Effective noise estimation must be integrated with U-SCSE algorithms for better speech quality and noise reduction results.

2. Speech Intelligibility in Nonstationary Noise Sources: U-SCSE provides enhanced speech with very low speech intelligibility. More effective algorithms are required that can improve speech intelligibility in nonstationary noise sources.

3. Musical Noise Artifact and Speech Distortion: Unsupervised speech enhancement algorithms provide acceptable noise reduction, however reduction of the residual noise artifact and speech distortion requires further research.

References

[1] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 32, no. 6, pp. 1109-1121, 1984.

[2] E. Yariv and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE transactions on acoustics, speech, and signal processing*, vol. 33, no. 2, pp. 443-445, 1985.

[3] E. Yariv and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Transactions on speech and audio processing,* vol. 3, no. 4, pp. 251-266, 1995.

[4] I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE signal processing letters*, vol. 9, no.1, pp. 12-15, 2002.

[5] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal processing*, vol. 81, no. 11, pp. 2403-2418, 2001.

[6] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Transactions on speech and audio processing*, vol. 7, no. 2, pp. 126-137, 1999.

[7] N. Saleem, M. I. Khattak, G. Witjaksono, and G. Ahmad, "Variance based time-frequency mask estimation for unsupervised speech enhancement," *Multimedia Tools and Applications*, 1-25, 2019.

[8] M. Bahoura and J. Rouat, "Wavelet speech enhancement based on the teager energy operator," *IEEE signal processing letters*, vol. 8, no. 1, pp. 10-12, 2001.

[9] T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model," *EURASIP Journal on Advances in Signal Processing*, vol. 2005, no. 7, pp. 354850, 2005.

[10] A. Rezayee and S. Gazor, "An adaptive KLT approach for speech enhancement," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 2, pp. 87-95, 2001.

[11] D. E. Tsoukalas, J. N. Mourjopoulos, and G. Kokkinakis, "Speech enhancement based on audible noise suppression," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 6, pp. 497-514, 1997.

[12] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 845-856, 2005.

[13] C. Plapous, C. Marro, and P. Scalart, "Improved signal-to-noise ratio estimation for speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 2098-2108, 2006.

[14] I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," *IEEE Signal Processing Letters,* vol. 9, no. 4, pp. 113-116, 2002.

[15] Y. Hu and P.C. Loizou, "Speech enhancement based on wavelet thresholding the multitaper spectrum," *IEEE Transactions on Speech and Audio processing*, vol. 12, no. 1, pp. 59-67, 2004.

[16] J. H. Hansen and M. A. Clements, "Constrained iterative speech enhancement with application to speech recognition," *IEEE Transactions on Signal Processing*, vol. 39, no. 4, pp. 795-805, 1991.

[17] S. Watanabe, M. Delcroix, F. Metze, and J.R. Hershey, Eds., *New era for robust speech recognition: exploiting deep learning*, Springer, 2017.

[18] N. Saleem and T. G. Tareen, "Spectral Restoration based speech enhancement for robust speaker identification," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 1, pp. 34-39, 2018.

[19] J. Lazar, J. H. Feng, and H. Hochheiser, *Research methods in human-computer interaction*, Morgan Kaufmann, 2017.

[20] N. Saleem, E. Mustafa, A. Nawaz, and A. Khan, "Ideal binary masking for reducing convolutive noise," *International Journal of Speech Technology*, vol. 18, no. 4, pp. 547-554, 2015.

[21] P. Vary and R. Martin, *Digital speech transmission: Enhancement, coding and error concealment*, John Wiley & Sons, 2006.

[22] P. C. Loizou, *Speech enhancement: theory and practice*, CRC press, 2007.

[23] D. G. Jamieson, R. L. Brennan, and L. E. Cornelisse, "Evaluation of a speech enhancement strategy with normal-hearing and hearing-impaired listeners," *Ear and hearing*, vol. 16, no. 3, pp. 274-286, 1995.

[24] B. C. Moore, "Speech processing for the hearing-impaired: successes, failures, and implications for speech mechanisms," *Speech communication*, vol. 41, no. 1, pp. 81-91, 2003.

[25] K. H. Arehart, J. H. Hansen, S, Gallant, and L. Kalstein, "Evaluation of an auditory masked threshold noise suppression algorithm in normal-hearing and hearing-impaired listeners," *Speech Communication*, vol. 40, no. 4, pp. 575-592, 2003.

[26] Y. Hu and P. C. Loizou, "A comparative intelligibility study of speech enhancement algorithms," In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07,* vol. 4, IEEE, 2007, pp. IV-561.

[27] Y. Hu and P. C. Loizou, "A comparative intelligibility study of single-microphone noise reduction algorithms," *The Journal of the Acoustical Society of America*, vol. 122, no. 3, pp. 1777-1786, 2007.

[28] S. Gordon-Salant, "Effects of acoustic modification on consonant recognition by elderly hearing-impaired subjects," *The Journal of the Acoustical Society of America*, vol. 81, no. 4, pp. 1199-1202, 1987.

[29] J. B. Allen, "How do humans process and recognize speech?," *IEEE Transactions on speech and audio processing*, vol. 2, no. 4, pp. 567-577, 1994.

[30] H. Levitt, "Noise reduction in hearing aids: A review," *Journal of rehabilitation research and development*," vol. 38, no. 1, pp. 111-122, 2001.

[31] G. Kim, Y. Lu, Y. Hu, and P. C. Loizou, "An algorithm that improves speech intelligibility in noise for normal-hearing listeners," *The Journal of the Acoustical Society of America*, vol. 126, no. 3, pp. 1486-1494, 2009.

[32] G. Kim and P.C. Loizou, "Improving speech intelligibility in noise using environment-optimized algorithms," *IEEE transactions on audio, speech, and language processing*, vol. 18, no. 8, pp. 2080-2090, 2010.

[33] S. Boll "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 27, no. 2, pp. 113-120, 1979.

[34] Y. Lu and P. C. Loizou, "A geometric approach to spectral subtraction," *Speech communication*, vol. 50, no. 6, pp.453-466, 2008.

[35] S. Nasir, A. Sher, K. Usman, and U. Farman, "Speech enhancement with geometric advent of spectral subtraction using connected time-frequency regions noise estimation," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 6, no. 6, pp. 1081-1087, 2013.

[36] K. Paliwal, K. Wójcicki, and B. Schwerin, "Single-channel speech enhancement using spectral subtraction in the short-time modulation domain," *Speech Communication*, vol. 52, no. 5, pp. 450-475, 2010.

[37] T. Inoue, H. Saruwatari, Y. Takahashi, K. Shikano, and K. Kondo, "Theoretical analysis of musical noise in generalized spectral subtraction based on higher order statistics," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 19, no. 6, pp. 1770-1779, 2010.

[38] Y. Zhang and Y. Zhao, "Real and imaginary modulation spectral subtraction for speech enhancement," *Speech Communication*, vol. 55, no. 4, pp. 509-522, 2013.

[39] R. Miyazaki, H. Saruwatari, T. Inoue, Y. Takahashi, K. Shikano, and K. Kondo, "Musical-noise-free speech enhancement based on optimized iterative spectral subtraction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 7, pp. 2080-2094, 2012.

[40] A. L. Ramos, S. Holm, S. Gudvangen, and R. Otterlei, "A spectral subtraction based algorithm for real-time noise cancellation with application to gunshot acoustics," *International Journal of Electronics and Telecommunications*, vol. 59, no. 1, pp. 93-98, 2013.

[41] S. M. Ban and H. S. Kim, "Weight-Space Viterbi Decoding Based Spectral Subtraction for Reverberant Speech Recognition," *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1424-1428, 2015.

[42] K. Hu and D. Wang, "Unvoiced speech segregation from nonspeech interference via CASA and spectral subtraction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1600-1609, 2010.

[43] K. Kokkinakis, C. Runge, Q. Tahmina, and Y. Hu, "Evaluation of a spectral subtraction strategy to suppress reverberant energy in cochlear implant devices," *The Journal of the Acoustical Society of America*, vol. 138, no. 1, pp. 115-124, 2015.

[44] H. T. Hu and C. Yu, "Adaptive noise spectral estimation for spectral subtraction speech enhancement," *IET Signal Processing*, vol. 1, no. 3, pp. 156-163, 2007.

[45] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proceedings of the IEEE*, vol. 67, no. 12, pp. 1586-1604, 1979.

[46] H. Ding, Y. Soon, S. N. Koh, and C. K. Yeo, "A spectral filtering method based on hybrid wiener filters for speech enhancement," *Speech Communication*, vol. 51, no. 3, pp. 259-267, 2009.

[47] M. J. Alam and D. O'Shaughnessy, "Perceptual improvement of Wiener filtering employing a post-filter," *Digital Signal Processing*, vol. 21, no. 1, pp. 54-65, 2011.

[48] I. Almajai and B. Milner, "Visually derived wiener filters for speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1642-1651, 2010.

[49] M. A. A. El-Fattah, M. I. Dessouky, A. M. Abbas, S. M. Diab, E. S. M. El-Rabaie, W. Al-Nuaimy, ... and F. E. A. El-Samie, "Speech enhancement with an adaptive Wiener filter," *International Journal of Speech Technology*, vol. 17, no. 1, pp. 53-64, 2014.

[50] B. Xia and C. Bao, "Wiener filtering based speech enhancement with weighted denoising auto-encoder and noise classification," *Speech Communication*, vol. 60, pp. 13-29, 2014.

[51] K. T. Andersen and M. Moonen, "Robust speech-distortion weighted interframe Wiener filters for single-channel noise reduction," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 1, pp. 97-107, 2017.

[52] B. M. Mahmmod, A. R. Ramli, S. H. Abdulhussian, S. A. R. Al-Haddad, and W. A. Jassim, "Low-distortion MMSE speech enhancement estimator based on Laplacian prior," *IEEE Access*, vol. 5, pp. 9866-9881, 2017.

[53] R. K. Kandagatla and P. V. Subbaiah, "Speech enhancement using MMSE estimation of amplitude and complex speech spectral coefficients under phase-uncertainty," *Speech Communication*, vol. 96, pp. 10-27, 2018.

[54] H. R. Abutalebi and M. Rashidinejad, "Speech enhancement based on β-order MMSE estimation of Short Time Spectral Amplitude and Laplacian speech modeling," *Speech Communication*, vol. 67, pp. 92-101, 2015.

[55] T. Gerkmann and M. Krawczyk, "MMSE-optimal spectral amplitude estimation given the STFT-phase,"·*IEEE Signal Processing Letters*, vol. 20, no. 2, pp. 129-132, 2012.

[56] M. McCallum and B. Guillemin, "Stochastic-deterministic MMSE STFT speech enhancement with general a priori information," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 7, pp. 1445-1457, 2013.

[57] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 4, pp. 251-266, 1995.

[58] K. Hermus and P. Wambacq, "A review of signal subspace speech enhancement and its application to noise robust speech recognition," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 045821, 2006.

[59] A. Borowicz and A. Petrovsky, "Signal subspace approach for psychoacoustically motivated speech enhancement," *Speech Communication*, vol. 53, no. 2, pp. 210-219, 2011.

[60] M. Kalantari, S. R. Gooran, and H. R. Kanan, "Improved embedded pre-whitening subspace approach for enhancing speech contaminated by colored noise," *Speech Communication*, vol. 99, pp. 12-26, 2018.

[61] E. V. de Payer, "The subspace approach as a first stage in speech enhancement," *IEEE Latin America Transactions*, vol. 9, no. 5, pp. 721-725, 2011.

[62] B. Wiem, P. Mowlaee, and B. Aicha, "Unsupervised single channel speech separation based on optimized subspace separation," *Speech Communication*, vol. 96, pp. 93-101, 2018.

[63] P. Sun, A. Mahdi, J. Xu, and J. Qin, "Speech enhancement in spectral envelop and details subspaces," *Speech Communication*, vol. 101, pp. 57-69, 2018.

[64] F. Bao, W. H. Abdulla, F. Bao, and W. H. Abdulla, "A New Ratio Mask Representation for CASA-Based Speech Enhancement," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP),* vol. 27, no. 1, pp. 7-19, 2019.

[65] X. Wang, F. Bao, and C. Bao, "IRM estimation based on data field of cochleagram for speech enhancement," *Speech Communication*, vol. 97, pp. 19-31, 2018.

[66] X. Wang, C. Bao, and F. Bao, "A model-based soft decision approach for speech enhancement," *China Communications*, vol. 14, no. 9, pp. 11-22, 2017.

[67] S. Liang, W. Liu, and W. Jiang, "A new Bayesian method incorporating with local correlation for IBM estimation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 3, pp. 476-487, 2012.

[68] A. Narayanan and D. Wang, "A CASA-based system for long-term SNR estimation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 9, pp. 2518-2527, 2012.

[69] G. Hu, and D. Wang, "A tandem algorithm for pitch estimation and voiced speech segregation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 2067-2079, 2010.

[70] Y. K. Lee and O. W. Kwon, "Application of shape analysis techniques for improved CASA-based speech separation," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 1, pp. 146-149, 2009.

[71] T. May and T. Dau, "Computational speech segregation based on an auditory-inspired modulation analysis," *The Journal of the Acoustical Society of America*, vol. 136, no. 6, 3350-3359, 2014.

[72] N. Rehman, C. Park, N. E. Huang, and D. P. Mandic, "EMD via MEMD: multivariate noise-aided computation of standard EMD," *Advances in Adaptive Data Analysis*, vol. 5, no. 02, pp. 1350007, 2013.

[73] A. Upadhyay and R. B. Pachori, "Speech enhancement based on mEMD-VMD method," *Electronics Letters*, vol. 53, no. 7, pp. 502-504, 2017.

[74] K. Khaldi, A. O. Boudraa, and M. Turki, "Voiced/unvoiced speech classification-based adaptive filtering of decomposed empirical modes for speech enhancement," *IET Signal Processing*, vol. 10, no. 1, pp. 69-80, 2016.

[75] L. Zao, R. Coelho and P. Flandrin, "Speech enhancement with emd and hurst-based mode selection," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 5, pp. 899-911, 2014.

[76] M. E. Hamid, M. K. I. Molla, X. Dang, and T. Nakai, "Single channel

speech enhancement using adaptive soft-thresholding with bivariate EMD," *ISRN signal processing*, 2013.
[77] N. Chatlani and J. J. Soraghan, "EMD-based filtering (EMDF) of low-frequency noise for speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 4, pp. 1158-1166, 2011.
[78] K. Khaldi, A. O. Boudraa, A. Bouchikhi, and M. T. H. Alouane, "Speech enhancement via EMD," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, pp. 873204, 2008.
[79] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *ICASSP'79, IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, IEEE, 1979, pp. 208-211.
[80] H. Gustafsson, S. E. Nordholm, and I. Claesson, "Spectral subtraction using reduced delay convolution and adaptive averaging," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, 799-807, 2001.
[81] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," in *ICASSP*, vol. 4, 2002, pp. 44164-44164.
[82] Y. Hu and P. C. Loizou, "Speech enhancement based on wavelet thresholding the multitaper spectrum," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 1, pp. 59-67, 2004.
[83] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 4, pp. 334-341, 2003.
[84] F. Jabloun and B. Champagne, "Incorporating the human hearing properties in the signal subspace approach for speech enhancement,"·*IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 700-708, 2003.
[85] E. H. Rothauser, "IEEE recommended practice for speech quality measurements," *IEEE Trans. on Audio and Electroacoustics*, vol. 17, pp. 225-246, 1969.
[86] H. G. Hirsch and D. Pearce, "The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions," In *ASR2000-Automatic Speech Recognition: Challenges for the new Millenium ISCA Tutorial and Research Workshop (ITRW)*, 2000.
[87] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *Proceedings of* IEEE *International Conference on Acoustics, Speech, and Signal Processing 2001 (ICASSP'01)*, vol. 2, 2001, pp. 749-752.
[88] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125-2136, 2011.

Nasir Saleem

Engr. Nasir Saleem received the B.S degree in Telecommunication Engineering from University of Engineering and Technology, Peshawar-25000, Pakistan in 2008 and M.S degree in Electrical Engineering from CECOS University, Peshawar, Pakistan in 2012. He was a senior Lecturer at the Institute of Engineering and Technology, Gomal University, D.I.Khan-29050, Pakistan. He is now Assistant Professor in Department of Electrical Engineering, Gomal University, Pakistan. His research interests are in the area of digital signal processing, speech processing and enhancement.

Muhammad Irfan Khattak

Muhammad Irfan Khattak is working as an Associate Professor in the Department of Electrical Engineering in the University of Engineering and Technology Peshawar. He did his B.Sc Electrical Engineering from the same University in 2004 and did his PhD from Loughborough University UK in 2010. His research interest involves Antenna Design, On-Body Communications, Speech processing and Speech Enhancement.

Elena Verdú Pérez

Elena Verdú received her master's and Ph.D. degrees in telecommunications engineering from the University of Valladolid, Spain, in 1999 and 2010, respectively. She is currently an Associate Professor at Universidad Internacional de La Rioja (UNIR) and member of the Research Group "Data Driven Science" of UNIR. For more than 15 years, she has worked on research projects at both national and European levels. Her research has focused on e-learning technologies, intelligent tutoring systems, competitive learning systems, accessibility, speech and image processing, data mining and expert systems.

A Collaborative Filtering Probabilistic Approach for Recommendation to Large Homogeneous and Automatically Detected Groups

Remigio Hurtado[1,2]*, Jesús Bobadilla[2], Abraham Gutiérrez[2], Santiago Alonso[2]

[1] Universidad Politécnica Salesiana, Calle Vieja 12-30, 010102 Cuenca (Ecuador)
[2] Universidad Politécnica de Madrid, Carretera de Valencia Km 7, 28031 Madrid (Spain)

Received 29 November 2019 | Accepted 17 February 2020 | Published 21 March 2020

Abstract

In the collaborative filtering recommender systems (CFRS) field, recommendation to group of users is mainly focused on stablished, occasional or random groups. These groups have a little number of users: relatives, friends, colleagues, etc. Our proposal deals with large numbers of automatically detected groups. Marketing and electronic commerce are typical targets of large homogenous groups. Large groups present a major difficulty in terms of automatically achieving homogeneity, equilibrated size and accurate recommendations. We provide a method that combines diverse machine learning algorithms in an original way: homogeneous groups are detected by means of a clustering based on hidden factors instead of ratings. Predictions are made using a virtual user model, and virtual users are obtained by performing a hidden factors aggregation. Additionally, this paper selects the most appropriate dimensionality reduction for the explained RS aim. We conduct a set of experiments to catch the maximum cumulative deviation of the ratings information. Results show an improvement on recommendations made to large homogeneous groups. It is also shown the desirability of designing specific methods and algorithms to deal with automatically detected groups.

Keywords

Collaborative Filtering Clustering, Dimensionality Reduction, Group Recommendation, Homogenous Groups, Recommender Systems.

DOI: 10.9781/ijimai.2020.03.002

I. Introduction

THIS section is divided into four subsections: 1) Fundamental concepts of Recommender Systems (RS), 2) Clustering to improve collaborative filtering, 3) Matrix decomposition-based clustering, and 4) Recommendation to groups of users and the proposed approach.

A. Recommendations to Individual Users

RS [1] field is relevant in the Artificial Intelligence scenario since it significantly reduces the information overload problem on the Internet. RS suggest to the users about the items they probably will like. Depending on the item nature, a variety of RS can be implemented: e-learning [2], tourism [3], [4], films [5], restaurants [6], networks [7], healthcare [8], industrial operators [9], etc. The most accurate type of RS is the Collaborative Filtering (CF) one [10],[11]. CF RS are based on the preferences of users about items; preferences can be explicit (votes) or implicit (listened songs, purchased items, watched movies, etc.). CF RS have been traditionally implemented by using the K-Nearest Neighbors (KNN) machine learning method [12], although current CF RS kernels are usually based on the Matrix Factorization (MF) algorithm [13], [14]. MF converts the sparse matrix of ratings (users x items) to two dense matrices: 'users x factors' and 'factors x items'. Since the number of factors is very little (usually 20 to 80), MF makes a matrix reduction [15] and it extracts the most important information from the original sparse rating matrix. Factors are called 'hidden factors' and they condense the relevant information from the CF dataset. Predictions to each user can be computed by making the dot product of the user and the items factors. Well known MF variations are Positive Matrix Factorization (PMF) [16], [17], Bayesian Nonnegative Matrix Factorization (BNMF) [18], and Elementwise Alternating Least Squares (eALS) [19]. Finally, state of the art CF RS implementations make use of neural networks [76], some of them joining MF and the Multilayer Perceptron [20]. These state of the art neural network approaches cover many different scopes, such as music [21] and videos [19]. Beyond CF, it is usual to improve accuracy results by merging RS types (hybrid RS) [22]. CF RS can be reinforced by means of demographic [6], content-based [23], context-aware [24], and social [7] information.

* Corresponding author.
E-mail address: rhurtadoo@ups.edu.ec

B. Clustering to Improve Collaborative Filtering

Clustering is a recurrent resource to improve CF RS [25]. Clustering can be performed on several types of RS information: content-based [26], demographic [27], hybrid [28], etc. CF clustering has traditionally performed on items [29] or users [30]; CF clustering based on the ratings information is the current most published approach [31], [25]. Fuzzy C-Means has proved to be accurate to CF RS purposes [26] and to improve coverage. MF can also be improved by means of a user clustering model [32]. To establish the number of clusters parameter (K) is a trial and error process that requires knowledge of the data and experience; in [33] the number of clusters can be dynamically set

to arrange the RS data size variations, and results show an accuracy improvement. The co-clustering method has also been used to improve MF results [34]. Making use of the user changes on their preferences, an evolutionary clustering algorithm has been proposed [35]; it takes into account the temporal evolution of features. Clustering similar items is a simple and useful way to improve CF RS accuracy; this is done with e-commerce products on [36]. Learning analytics with clustering is a potential benefit that grows with the size of the users. This has improved digital education processes on [37]. Association rules mining has been used to reduce the size of the data, and then to make the clustering process more efficient. Specifically, [38] has reduced the item space. Centroid selection is a key process to tackle most of the CF clustering approaches; it improves performance and reduces the processing time [39].

C. Matrix Decomposition-Based Clustering

In the context of this paper, it is important to address the matrix decomposition problem for clustering purposes. In the CF field, the use of MF provides some relevant clustering advantages [40]: 1) MF accurately models sparse data variations [41], 2) It can implement both hard and soft clustering: e.g.: by means of Nonnegative Matrix Factorization (NMF) and BNMF, and 3) MF simultaneously factorizes users and items. Co-clustering is a clustering approach used in CF; it can be used to discover space correlations in big data scenarios [42]. An NMF-based semi-supervised co-clustering is proposed in [43] involving CF link relations. When CF datasets incorporate constraint restrictions, performance can be improved by means of clustering approaches. The regularized NMF incorporates constraints support to the standard NMF method, such as neighborhood-based local learning regularization [44]. The constraint restrictions methods rely on datasets that incorporate this type of information: they cannot be considered generalized, such as the one we propose in this paper.

D. Recommendation to Groups Of Users and Proposed Approach

Once the above three blocks (CF RS, clustering and MF-based clustering) have been addressed we will focus on the CF recommendations to groups of users. The recommendations to groups of users arise from the convenience of being able to recommend a group of users about products or services that satisfy the entire group [45].

Group RS can be classified according to the group type [46]: a) ***Established group***: persons that belong to a stable group; b) **Occasional group**: persons that at times join to a group, c) ***Random group***: persons who share an environment in a particular moment, and d) ***Automatically identified group***: automatically detected groups, considering the preferences of users. From [47] we can distinguish between homogeneous and heterogeneous groups. ***Homogeneous groups*** are established by the System, whereas ***heterogeneous groups*** are dynamically created by the users. Most of the research has focused on established and heterogeneous groups [48], [49], [50], [51]. Our proposed approach is designed to make recommendations on homogeneous and automatically identified groups [46], [52], [53]. Homogeneous groups are particularly relevant for marketing processes, where companies want to recommend products or services to a broad target of similar users. Homogeneous groups are usually obtained by using non-supervised machine learning methods, such as diverse clustering approaches [5].

The ***main objective*** of this paper is to improve the quality of recommendations made to automatically detected homogeneous groups in a RS. With this purpose, it is essential to divide the RS set of users in the very best homogeneous groups. Our ***first hypothesis*** is that we can improve the detection of homogeneous groups by performing a clustering that combines the appropriate MF dimensionality reduction with the aggregation of factors. The ***second hypothesis*** is that the RS recommendation quality will be improved by combining the proposed aggregation model with the designed clustering approach, both based on factors. Our probabilistic approach aimed at groups of users allows predicting the probability that a virtual user (group) likes a specific item.

The ***proposed dimensionality reduction*** is based on BNMF method [18]. We have made experiments to compare BNMF, Principal Component Analysis (PCA), and TruncatedSVD dimensionality reduction techniques. Since PCA computes the covariance matrix, it operates on the entire sparse matrix, whereas BNMF and Truncated Singular Value Decomposition (TruncatedSVD) do not have this limitation. The obtained results show that to recommend homogeneous groups, BNMF outperforms PCA and TruncatedSVD: 1) It obtains better variance values than the baselines, so it needs fewer dimensions (factors) to provide the same information level, 2) It provides better within-cluster results and also a more equilibrated number of users in its clusters, and 3) BNMF returns factors with probabilistic meaning that can be used in the prediction stage. As we will see later, the BNMF hidden factors will be used both to feed the clustering process and to make virtual user models by aggregating factors of users.

In Fig. 1 we explain the ***current research context of the proposed approach***. This figure shows five states of the art methods, from a) to e), designed to make recommendations to groups of users. The method labeled as f) corresponds to the proposed one in this paper. Method a): ***Recommendation fusion*** [49] is known as RANK; it makes an aggregation of the set of individual recommendation made to the users of the group. Method b): ***Prediction aggregation*** [54] is known as PER an it makes an aggregation of the set of individual predictions made to the users of the group. Method c): ***User preferences aggregation*** makes a model of the group of users (virtual user); this is a synthetic user that represents to whole set of users in the group. Once the virtual user is obtained, the traditional recommendation to one user process is done; this approach is the more accurate one when applied to heterogeneous groups. We call it VUR (*Virtual User based Recommendation*) [45], [51], [55] and we use it as a baseline. Method d) [56], [52] makes predictions before clustering, it performs aggregation post-clustering and it does not use dimensionality reduction. We will use it as a baseline using the name PC (***Predict & Cluster***). Method e): this is a variant of the method "b". We call it as RAP (***Recommendation via Aggregation of Predictions***) [57]; using RAP, dimensionality reduction is made before clustering, and predictions are obtained from ratings. We will use it as baseline to test the dimensionality reduction impact on the results.

The proposed method, called RAF (Reduction and Aggregation of Factors) (Fig. 1f) adds a dimensionality reduction that is not included in methods a) to d). Both the proposed method and the RAP one (Fig. 1e) perform the clustering by using the obtained hidden factors in the dimensionality reduction process. Methods a) to b) run the clustering algorithm by using the dataset ratings information. Since both the baseline RAP and the proposed RAF methods make the same clustering, their homogenous groups are not different. The crucial differences between the proposed RAF and the baseline RAP methods are: 1) RAP characterizes users by their preferences (ratings), whereas RAF characterizes users by their hidden factors; consequently, RAF provides a higher semantic level than RAP, and 2) RAP first makes individual predictions and then aggregates them, whereas RAF creates a model (virtual user of the group) and then obtains its prediction. It is relevant to highlight that the RAF virtual user is an aggregation of hidden factors, whereas the RAP prediction, is an aggregation of user predictions based on ratings.

The rest of the paper is divided into the following sections (with the same numbering shown here):

II. Related work, in which a review is made of the most relevant contributions that exist in the CF aspects covered in the paper.

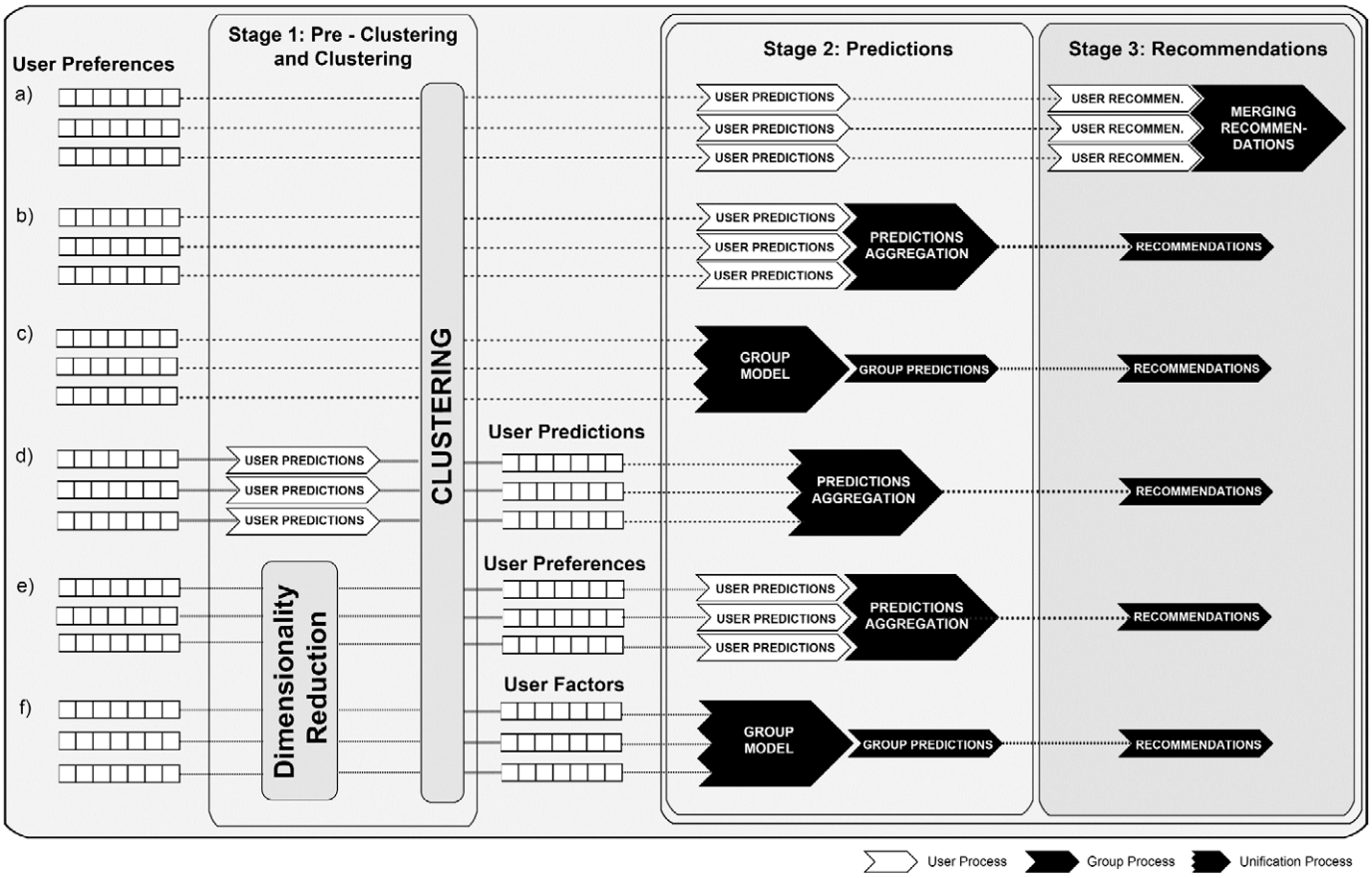


Fig. 1. State of the art approaches for recommendation to groups of users.

III. Formalization of the proposed CF method which specifies the way to predict and recommend. In addition, we provide a running example.

IV. Experiments and results: experiments set up, quality measures, model optimization of parameters, model performance, comparative results, and discussion.

V. Most relevant conclusions obtained and future works.

II. Related Work

In Fig. 1, from methods a) to d) we show the most relevant researches in the context of the proposed approach (designed to make recommendations to groups of users). This works are: a) Recommendation merging [49], b) Prediction aggregation [54], c) User preferences aggregation [45], [51], [55], and, d) Predict & Cluster [56], [52]. In addition to these methods, there are other models and methods based on matrix factorization, neural collaborative filtering, clustering methods, and graph approaches. From these methods we can highlight the following researches:

Virtual users can be extracted from MF factors [50] to recommend to each group of users by means of its corresponding group virtual user. Authors perform their experiments using groups of sizes: 2 to 4 users, 5 to 8 users, 9 to 12 users. For the state-of-the-art papers in the field, this is a typical range of user size in a group. Conversely, this paper makes use of much larger numbers of users in each group. Design alternatives for CF recommendation to a group of users are tested in [51]; conclusions point to a relevant increase of performance (execution time) when aggregation is made in early stages, whereas accuracy does not significantly change according to the stage where the aggregation is made. Our paper borrows this concept to design the proposed approach. A model based on the topic of argumentation [58] is used to recommend personalized items for groups. The argumentation subject is extracted and the users with similar views are clustered into groups. Uncertainty is used in [59] to model the way members might agree on a group ranking. Based on the observed member's individual rankings, they quantify the likelihood of group rankings.

To use embedding representations of CF data contributes to improving results. A Neural Collaborative Filtering design is presented in [60] where a neural network learns the interactions of groups and items; it uses factor embeddings. On the same line [61] uses group activity history information and recommender post-rating feedback to generate interactive preference parameters. In the same line that our proposal, [62] combines latent factor models to obtain improved group recommendations. In this case, a multi-layer perceptron is used to make this task. State of art in recommendation to groups of users includes a broad range of application fields: authors in [63] use a dragonfly algorithm to deal with sparsity; they provide a client-based collaborative filtering approach and apply it to restaurant recommendation. Considering repeat purchasing, [64] provides a group RS to optimize the offline physical store inventories. A large study with real groups of tourists [65] manages the problem of finding a sequence of points of interest that satisfies all group members. A travel RS for individual and group of users is also proposed in [66]; this RS provides a list of POI as recommendations. They also exploit relationships between users. Recommendation of clinics to patient groups is provided in [67], where it divides patients into multiple groups by mining their unknown preferences before recommending them suitable clinics. An academic venue RS is proposed [68], where academic venues are recommended for a group of researchers based on the venues attended by their co-authors, the group members and also on their co-citers. Group cohesion is a key factor in group

recommendation, and clustering is a powerful tool to reach cohesion. Heterogeneous information networks can contain rich information about entities and relationships: [69] provides an approach for group recommendation that appropriately captures group cohesion. Social information can also improve group recommendation results: To detect the inherent associations among group members, [70] incorporates user social networks into the random walk with restart model.

Recommendations explanation is a difficult task, particularly in group recommendations; [71] investigates which explanation best helps to increase the satisfaction of group members, to improve fairness perception and to obtain consensus perception. Visualization of group recommendations is also a challenging task; authors in [72] provide visual presentations and intuitive explanations. Finally, using social trust information, [73] identifies trustworthy users and it analyses the degrees of trust among users in a group.

III. Proposed Method

This section introduces the architecture of the proposed method and its relevant details. Explanations are reinforced by means of a data toy running example that helps to understand some internal functionalities. Fig. 2 shows the architecture of the method that we will use as a roadmap for the section explanations. From Fig. 2 we find the following procedural blocks:

1. Our starting point is the CF set of ratings: We are not proposing a hybrid RS [28]; we provide a pure CF RS method that can be used in all types of CF datasets.
2. The second block in Fig. 2 implements the currently most used model to obtain individual recommendations on RS: Matrix Factorization. We have selected the Bayesian Non-Negative Matrix Factorization (BNMF) [18], since we have made experiments (explained in the next section) and it has been found that BNMF outperforms PCA and TruncatedSVD in several aspects: mainly, it obtains better variance values and it provides both better within-cluster results and a more equilibrated number of users in their clusters. The BNMF mathematical formulation can be found in [18]; it is out of the scope of this paper.
3. Machine learning clustering methods return better results when the iterative algorithms are fed with appropriate initial values. Pre-clustering is a very dependent task on the data scope. RS clustering can be particularly improved by using power users [13], [39], and then we make this process. The algorithm *KMeansPlusLogPower* chooses K users from the dataset, as centroids. Its high-level algorithm is:

Algorithm: KMeansPlusLogPower

Input: U, hidden factors of the users in training set; k, total number of clusters

Output: k centroids, $\{c1, c2, \cdots ck\}$

1: Define desired number of clusters, k

2: Select the initial centroid $c1$ to be Pu

3: **repeat**

4: Select the next centroid ci where $ci = u' \in U$ with the probability: $Prob = distance(u) + log[\ (1/p(x)) + 1]$

5: **until** k centroids are found

6: **return** $\{c1, c2, \cdots ck\}$ "k centroids"

$p(x)$ is the relation between the number of items consumed by u'user and the number of items consumed by the Pu power user.

4. The proposed method runs the clustering algorithm by processing the user hidden factors (and not the user ratings). In this paper, we conclude that hidden factors provide better quality results than ratings. The reasons behind the result are: a) They offer a higher level of abstraction, b) Its dimensionality is much lower than the ratings one, c) They are not sparse.

   Clustering is made using the set of hidden factors from each RS user. It can be defined as:

   Let U be the set of users (1)

   Let I be the set of items (2)

   Let F be the set of Hidden Factors (3)

   Let f_j be the hidden factors of user j

   $$f_j = \{f_{j,1}, f_{j,2}, f_{j,3}, \dots, f_{j,F}\} \mid j \in [1..U] \quad (4)$$

5. The proposed method makes a model where each class is represented by a virtual user [50]. The virtual user is obtained by aggregating the hidden factors of all the users belonging to the class. The key concept here is that the virtual user accurately represents the users of its class because its limited number of factors contains the most relevant information. Aggregating ratings (instead of factors) cannot catch the relevant information and it does not provide a representative virtual user.

   Let K be the number of clusters (5)

   Let u_j be the user j $\mid j \in [1..U]$ (6)

   Let v_k be the virtual user k $\mid k \in [1..K]$ (7)

   Let C_k be the set of users in the cluster k (8)

   Let $f_{v_k,f}$ be the factor f of the virtual user k

   $$f_{v_k,f} = \frac{1}{\#C_k} \sum_{j \in C_k} f_{j,f} \quad (9)$$

   Let f_{v_k} be the hidden factors of virtual user k

   $$f_{v_k} = \{f_{v_k,1}, f_{v_k,2}, f_{v_k,3}, \dots, f_{v_k,F}\} \mid v_k \in [1..K] \quad (10)$$

6. To make predictions for each group (class), the proposed method is based on the BNMF individual predictions schema [18]. In this case, our method contains two parameters:
   - $\boldsymbol{\alpha}$, which controls the amount of overlapping of a user between user factors.
   - $\boldsymbol{\beta}$, that fixes the amount of evidence needed to determine that an item factor is associated with a factor of a virtual user.

   The association between item factors and virtual user factors allows obtaining the probability that a virtual user v_k likes a specific item i. This probability is obtained through the dot product of the virtual user factors and the set of item factors. Subsequently, this probability is transformed to the rating scale of the dataset, and the virtual user prediction to the item i is obtained, this is $p_{v_k,i}$.

   Let f'_i be the hidden factors of item i

   $$f'_i = \{f'_{i,1}, f'_{i,2}, f'_{i,3}, \dots, f'_{i,F}\} \mid i \in [1..I] \quad (11)$$

   Let $p_{v_k,i}$ be the prediction made to the virtual user k on item i

   $$p_{v_k,i} = \sum_{f \in F} f_{v_k,f}\ f'_{i,f} \mid k \in [1..K], i \in [1..I] \quad (12)$$

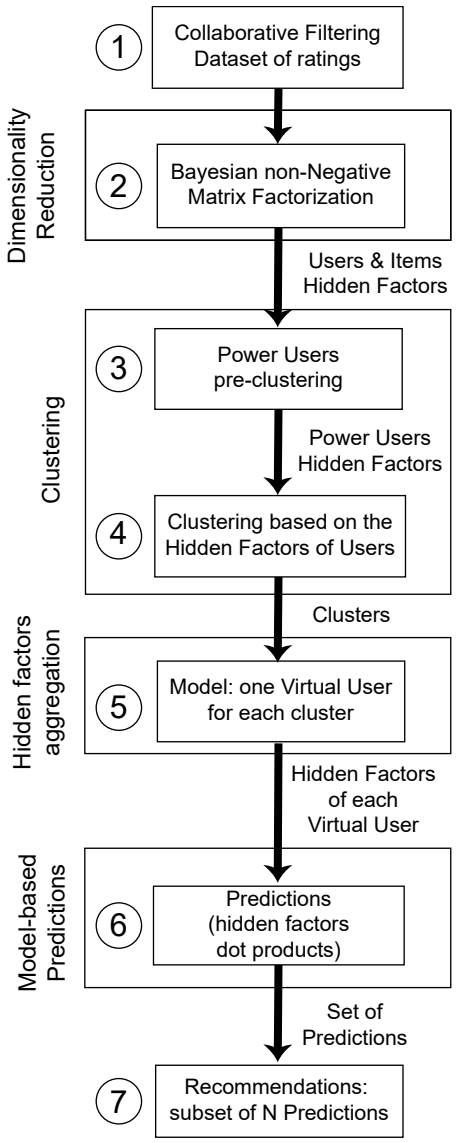


Fig. 2. Proposed method architecture.

In summary, this block of predictions is a probabilistic Bayesian approach of recommendation for groups of users. This approach allows us to compute:

- The probability that a virtual user likes a specific item.
- The prediction for the virtual user (group). The underlying idea is to predict to the set of users of a cluster the same predicted items for their common virtual user.

7. Recommendations, as usual, are selected from the N highest predictions [1].

Let R_k be the set of recommendations made to the class k and $\#R_k=N$

$$R_k = \{i \in I \mid p_{v_k,i} \geq p_{v_k,j} \ \forall i \in R_k, \forall j \in R_k^c\} \quad (13)$$

Previously to the explained method, the BNMF factorization approach has been selected to feed the clustering algorithm. The next section includes experimental results that show the BNMF is able to hold more information than PCA and TruncatedSVD, by using the same number of factors. Experiments have been run using RS data. The BNMF accumulated variance is obtained implementing (14) to (17).

Let σ_f be the set of factors variances

$$\sigma_f = \{\sigma(f_1), \sigma(f_2), \sigma(f_3), \dots, \sigma(f_F)\} \quad (14)$$

Let $\sigma(f_f)$ be the factor f variance

$$\sigma(f_f) = \frac{\sum_{u \in U}(f_{u,f} - \bar{f}_f)}{\#U} \quad (15)$$

Let θ be the threshold of required accumulated variance (16)

Let T be the set of factors that hold the required accumulated variance

$$T = \{i \in F \mid \sigma(f_i) \geq \sigma(f_j) \ \forall i \in T, \forall j \in T^c, \sum_{i \in T} \sigma(f_i) \geq \theta\} \quad (17)$$

To fix the proposed method concepts we provide a data toy example. Fig. 3.1 contains the CF dataset ratings casted for 10 users (u_1 to u_{10}) on 15 items (i_1 to i_{15}). To run the BNMF method we have chosen *F=3*. Fig. 3.2 shows the obtained factors for each user (on the right of the dataset) and the obtained factors for each item (on the bottom of the dataset). The clustering process has been run using a *K=5*; the obtained clusters (groups) are shown in Fig. 3.3. Fig. 3.4 contains the factors of each one of the five virtual users representing the five groups. The factor values are obtained by implementing (5) to (10). Making the dot product of each virtual user factors and each item factors predictions are obtained (Fig. 3.5): (11) and (12). Finally, from predictions we can obtain each group recommendations; e.g.: using *N=3*, group 1 recommendations are: i_1, i_2, i_3; group 2 recommendations are: i_6, i_7, i_9; ... group 5 recommendations are: i_{10}, i_{12}, i_{15}.

IV. Experiments and Results

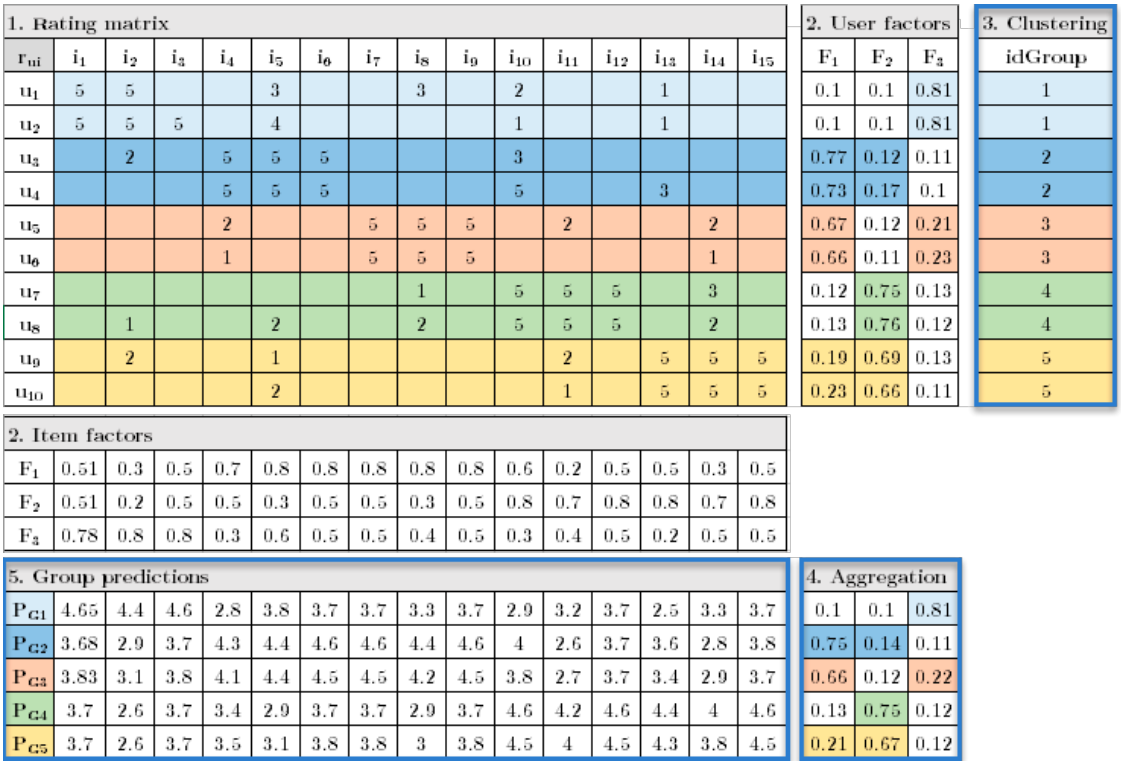
This section explains the design of the experiments: chosen datasets, quality measures, parameter values, etc., and their results. Experiments have been divided into two phases: (1) finding the optimal parameters of the proposed method to each tested dataset, (2) comparing the proposed method with state of art to make CF recommendations to homogeneous groups of users. All the experiments have been carried out using public datasets widely used by RS research papers. We have selected MovieLens [74] and FilmTrust [75] datasets. Table I contains the most relevant facts about these datasets. Finally, we test the *RMSE* prediction quality measure and the *F1* recommendation quality measure. Cross-validation values used in the experiments are abstracted in Table II.

A. Experiments Set Up

To perform the experiments, we have split users and items into test and training sets. To avoid fluctuations, we perform each experiment using 10-folds Monte Carlo cross-validation. Table I and Table II contain the main parameters of each dataset, chosen for the execution of experiments.

TABLE I. Main Properties of the Datasets Used in the Experiments

Dataset	Number of ratings	Number of items	Number of users	Rating values
MovieLens	1,000,209	3,706	6,040	5-star scale, with half-star increments
FilmTrust	33,470	2,059	1,227	0.5 to 4 with half increments



1. Rating matrix

r_{ui}	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8	i_9	i_{10}	i_{11}	i_{12}	i_{13}	i_{14}	i_{15}
u_1	5	5			3			3		2			1		
u_2	5	5	5		4					1			1		
u_3		2		5	5	5				3					
u_4				5	5	5				5			3		
u_5				2			5	5	5		2			2	
u_6				1			5	5	5					1	
u_7								1		5	5	5		3	
u_8		1			2			2		5	5	5		2	
u_9		2			1						2		5	5	5
u_{10}					2						1		5	5	5

2. User factors

F_1	F_2	F_3
0.1	0.1	0.81
0.1	0.1	0.81
0.77	0.12	0.11
0.73	0.17	0.1
0.67	0.12	0.21
0.66	0.11	0.23
0.12	0.75	0.13
0.13	0.76	0.12
0.19	0.69	0.13
0.23	0.66	0.11

3. Clustering

idGroup
1
1
2
2
3
3
4
4
5
5

2. Item factors

F_1	0.51	0.3	0.5	0.7	0.8	0.8	0.8	0.8	0.8	0.6	0.2	0.5	0.5	0.3	0.5
F_2	0.51	0.2	0.5	0.5	0.3	0.5	0.5	0.3	0.5	0.8	0.7	0.8	0.8	0.7	0.8
F_3	0.78	0.8	0.8	0.3	0.6	0.5	0.5	0.4	0.5	0.3	0.4	0.5	0.2	0.5	0.5

5. Group predictions

P_{G1}	4.65	4.4	4.6	2.8	3.8	3.7	3.7	3.3	3.7	2.9	3.2	3.7	2.5	3.3	3.7
P_{G2}	3.68	2.9	3.7	4.3	4.4	4.6	4.6	4.4	4.6	4	2.6	3.7	3.6	2.8	3.8
P_{G3}	3.83	3.1	3.8	4.1	4.4	4.5	4.5	4.2	4.5	3.8	2.7	3.7	3.4	2.9	3.7
P_{G4}	3.7	2.6	3.7	3.4	2.9	3.7	3.7	2.9	3.7	4.6	4.2	4.6	4.4	4	4.6
P_{G5}	3.7	2.6	3.7	3.5	3.1	3.8	3.8	3	3.8	4.5	4	4.5	4.3	3.8	4.5

4. Aggregation

0.1	0.1	0.81
0.75	0.14	0.11
0.66	0.12	0.22
0.13	0.75	0.12
0.21	0.67	0.12

Fig. 3. Data toy to explain the proposed method.

TABLE II. CROSS-VALIDATION VALUES USED IN THE EXPERIMENTS

Parameter	Values
Testing-Ratings	20%
Training-Ratings	80%
#clusters (K)	MovieLens: {20 to 200 step 5} FilmTrust: {20 to 200 step 5}
#recommendations (N)	15

B. Quality Measures

The quality measures we use in the experiments are:

- *Root Mean Square Error (RMSE)*: it is a collaborative filtering prediction quality measure. The *RMSE* can be formalized as:

$$RMSE = \frac{\sum_{i \in I}(r_{u,i} - P_{G,i})^2}{n} \quad (18)$$

Where $r_{u,i}$ is the user u rating to the item i. $P_{G,i}$ is the prediction for the item i for the group G in which user u is, hence $P_{G,i} = p_{v_k,i}$. I is the set of all items and n is the number of ratings available in the test set. Low *RMSE* values are better since it means that prediction errors are lower.

- *F1*: in order to evaluate the quality of recommendations to groups of users, we define *F1* as the harmonic mean that combines the values of precision and recall. We define precision and recall for group G, as:

$$precision_G = \frac{\sum_{u \in G} precision_u}{\#u} \quad (19)$$

$$recall_G = \frac{\sum_{u \in G} recall_u}{\#u} \quad (20)$$

Where G is the group in which user u is, $precision_u$ and $recall_u$ is the precision and recall for the user u, respectively. The $precision_u$ and $recall_u$ can be formalized as:

$$precision_u = \frac{\#TP}{\#(TP \cup FP)} \quad (21)$$

$$recall_u = \frac{\#TP}{\#T} \quad (22)$$

Where *TP*, *FP*, and *T* denote the true positive, false positive and expected recommendations sets, respectively:

$$FP = \{i \in L_G \mid r_{u,i} < \theta \wedge r_{u,i} \neq \bullet\} \quad (23)$$

$$TP = \{i \in L_G \mid r_{u,i} \geq \theta \wedge r_{u,i} \neq \bullet\} \quad (24)$$

$$T = \{i \in I \mid r_{u,i} \geq \theta\} \quad (25)$$

Where L_G is the set of items recommended to the group G in which user u is, $r_{u,i}$ is the test rating of the user u to the item i, • means that the test rating does not exist, and θ is a threshold to consider a rating as like or dislike.

Finally, we will denote *precision* and *recall* as the averaged precision and recall to each group of users, and *F1* combines the values of precision and recall.

$$precision = \frac{\sum_G precision_G}{\#G} \quad (26)$$

$$recall = \frac{\sum_G recall_G}{\#G} \quad (27)$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (28)$$

C. Model Optimization of Parameters

The proposed method contains two parameters: $\boldsymbol{\alpha}$, that controls the amount of overlapping of a user between user factors, and $\boldsymbol{\beta}$, that fixes the amount of evidence needed to determine that an item factor is associated with a user factor. This association between an item factor and user factor allows obtaining the probability that a virtual user likes a specific item. A proper adjustment of these parameters, for each dataset, is required in order to maximize the quality.

In this experiment, we will evaluate the proposed method

for different combinations of both α and β parameters. Table III contains the tested values for each parameter. To select the optimal configuration of these parameters we will: a) measure the *RMSE* and *F1*, b) recommend *15* items for each configuration of parameters, and c) select the best one. Table IV contains the remaining parameters required in this experiment.

TABLE III. TESTED MODEL PARAMETERS

Dataset	α	β
MovieLens	0.1 to 1.0 step 0.1	1 to 10 step 1
FilmTrust	0.1 to 1.0 step 0.1	1 to 5 step 1

TABLE IV. CONFIGURATION OF THE EXPERIMENT TO OPTIMIZE THE PROPOSED METHOD PARAMETERS

Parameter	MovieLens	FilmTrust
α	0.4	0.5
β	5	3

D. Model Performance

To measure the performance of the proposed model we will compare it with state-of-the-art group recommendation methods. The baselines selected for this comparison are *PC* [52], [56], *RAP* [57], and *VUR* [45], [51], [55]. Some of these recommendation methods require different parameters to work. We have configured these parameters in order to maximize the quality of the recommendation in each dataset. We will evaluate both the proposed and the baselines methods using the previously defined quality measures: *RMSE* and *F1*. Table V contains the parameters in each experiment.

E. Comparative Results and Discussion

The first set of experiments is designed to select the ***most appropriate reduction dimension technique for the CF RS*** field context. The tested methods are: 1) PCA, since it is the classical baseline for the machine learning reduction of dimensions field, 2) TruncatedSVD, because it is appropriate to be used in the CF sparse datasets, and 3) BNMF, since it has proved to accurately catch the CF non-linear features relations and to provide state of the art recommendation results.

To test and compare the three chosen methods we study their *cumulative explained variance* and their *user distribution*. Fig. 4a shows the cumulative explained variance results; as it can be seen, BNMF strongly outperforms TruncatedSVD and PCA. BNMF is able to catch more cumulative variance using the same number of dimensions; this means that using BNMF we expect better prediction results using the same number of dimensions (same number of hidden factors) than using PCA or TruncatedSVD. Fig. 4b shows another relevant measure: *user distribution*; it is valuable to avoid big differences in the number of users belonging to the groups (clusters) in the CF RS. Fig. 4b shows a much better BNMF distribution of users when the number of groups is small, and somewhat better results for larger number of clusters. Overall, in the CF context, we can conclude that BNMF outperforms the tested methods and it is able to compress more information in the same number of dimensions. This is expected behavior since BNMF has been designed to outperform MF approaches in CF environments. Because of the explained results, the proposed method in this paper uses the BNMF hidden factors both to feed the clustering algorithm and to get predictions by means of the aggregated virtual users.

To improve the ***group predictions accuracy*** is a relevant objective of the paper. Fig. 5 shows the *RMSE* error results both in the MovieLens dataset (Fig. 5a) and the FilmTrust dataset (Fig. 5b). As can be seen, the proposed method RAF returns lower errors than the chosen baselines (PC, RAP, and VUR). RAF outperforms the baselines for all the tested number of groups (*x*-axis: number of clusters). As expected, the lower the number of groups the higher the number of users in each cluster; and, consequently, the higher the error. This is the behavior for both the proposed method and the most competitive baselines: RAP and PC.

It is important to note that experiments were carried out with several methods designed from the VUR one. These methods are [45], [51], and [55], which have recently been published for recommendation to groups of users. The baseline from [55] offers better results than the [45], and [51] ones. The VUR results in Fig. 5 correspond to the results obtained with the baseline [55]. Results from [45], and [51] have not been incorporated because they are worse and do not help the visualization of the methods with relevant results. The VUR method offers the worst quality results; it demonstrates that to recommend to groups of users, a method cannot be based on a just aggregation of preferences, virtual users, and similarity measures for groups of users. This is not an optimal design to recommend to large groups of users. It is logical that a virtual user that represents a very large group will tend to generate very general and not accurate predictions. This points towards our proposed method do not achieve the best results only by an aggregation of factors. The key to the success of the proposed method has been the appropriate dimensionality reduction, the aggregation of factors, and especially our probabilistic approach for groups of users that allows predicting: 1) the probability that a virtual user likes a

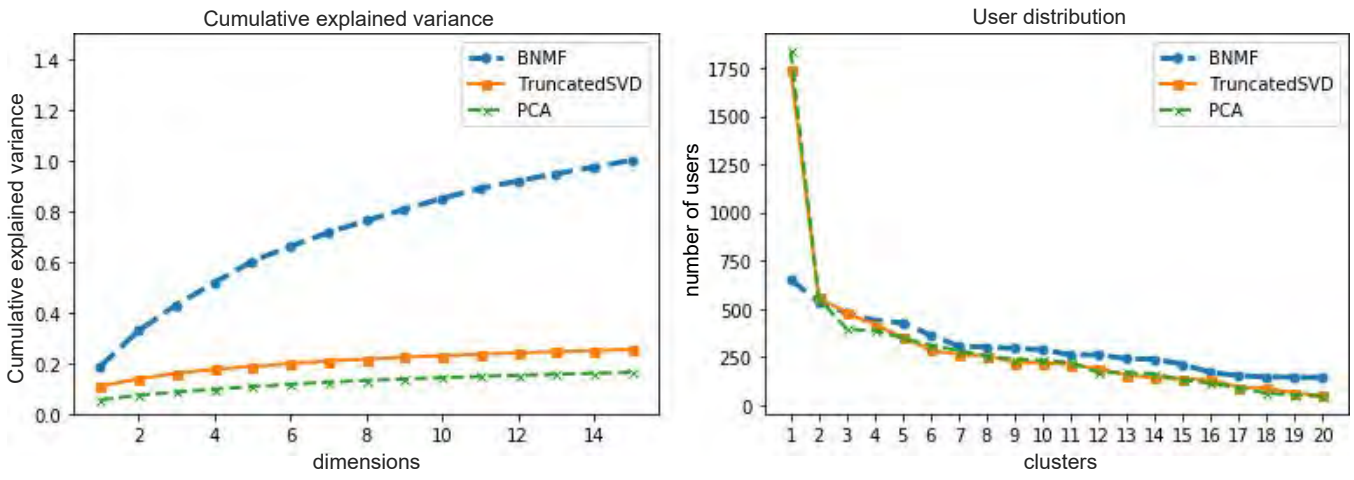


Fig. 4. Dimensionality dimensions results; a) Cumulative explained variance obtained on diverse number of dimensions; b) averaged number of users when diverse number of clusters is set.

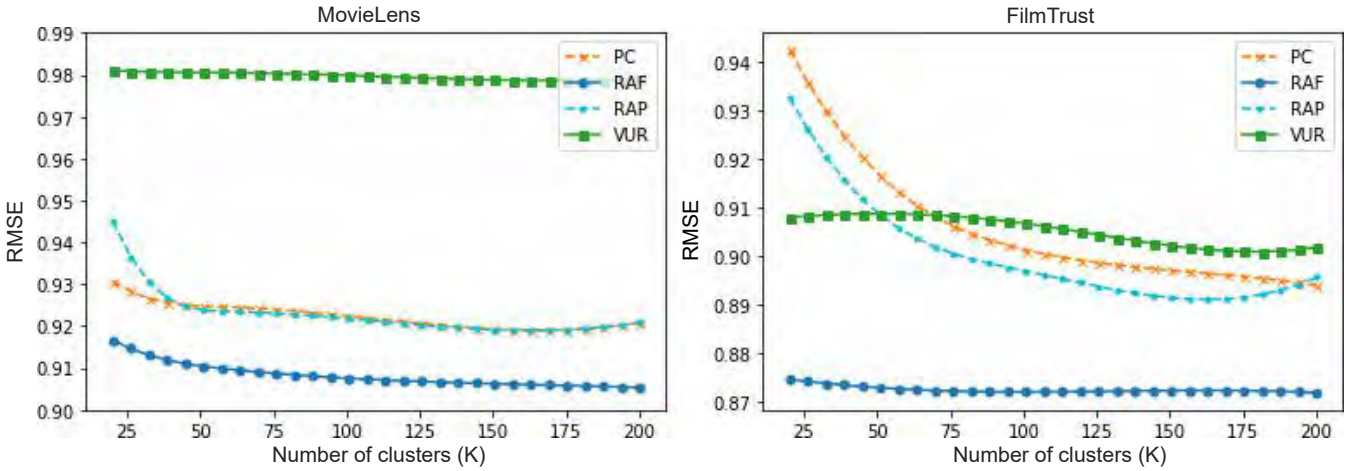


Fig. 5. Prediction accuracy results. RMSE obtained when the proposed method is run using diverse number of groups (clusters); a) MovieLens accuracy, b) FilmTrust accuracy. The lower the values, the higher the accuracy.

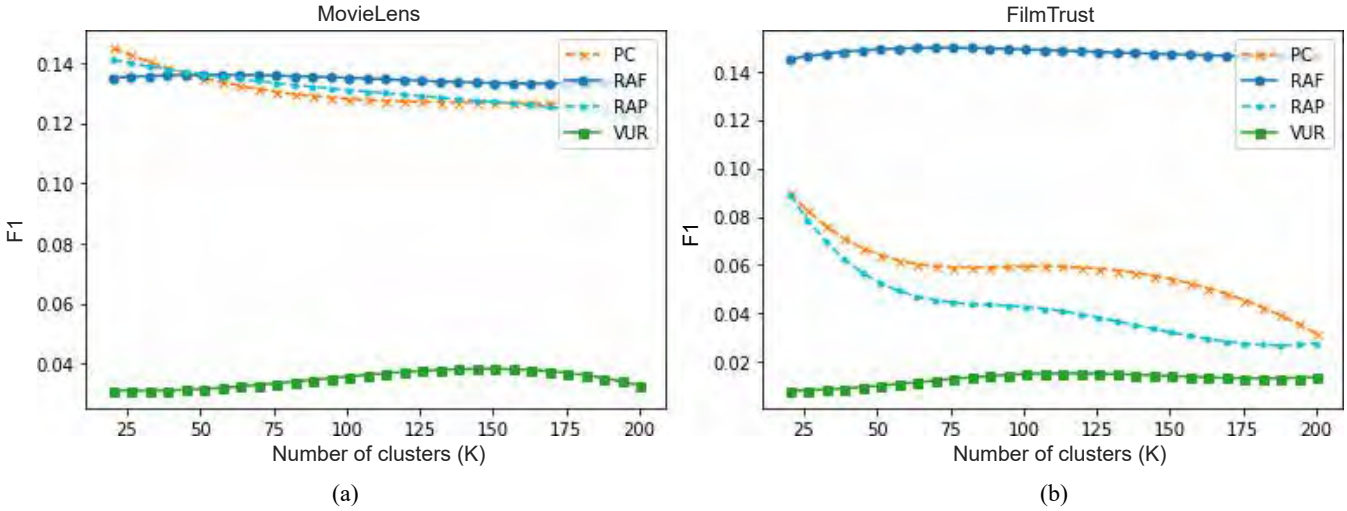


Fig. 6. Recommendation accuracy results. F1 obtained when the proposed method is run using diverse number of groups (clusters); a) MovieLens accuracy, b)

specific item and, 2) the prediction for the virtual user (group).

Finally, the ***recommendation accuracy*** is tested. Since prediction accuracy has been improved (Fig. 5), we also expect to improve recommendation accuracy, which is directly related to the quality of the highest predictions. To reduce the number of figures we use the F1 quality measure to join the precision and recall results. Fig. 6 is equivalent to Fig. 5, but it shows the recommendation F1 results instead of the prediction RMSE ones. In Fig. 6, higher values mean better accuracies. Overall, recommendations are improved by using the proposed method (RAF), particularly in the FilmTrust dataset, whose sparsity is much greater than the MovieLens one. The BNMF performance, when applied to the FilmTrust sparse dataset, is reflected both in the prediction and the recommendation results.

V. Conclusions and Future Works

To make recommendations on large homogeneous and automatically detected groups of users is a challenging task that is not adequately addressed using the existing approaches to recommend to little groups of stablished users. To accurately detect collaborative filtering groups is necessary to feed the clustering process with high-level information: hidden factors obtained from ratings. This approach makes use of the abstraction level provided by the chosen dimensionality reduction method. The Bayesian non-Negative Matrix Factorization (BNMF) has proved to be the most effective dimensionality reduction technique for this paper's objectives. Its superior cumulative deviation result shows that it provides more information, by using the same number of factors, than other representative dimensionality reduction methods. This is a key contribution of the paper since it opens the possibility to design alternative group recommendation methods based on the representative BNMF hidden factors.

The group recommendation model-based approach provided in this paper aggregates the hidden factors of each group of users to make a virtual user that represents the set of users of the group. This strategy provides two main advantages: 1) It makes the group model from the higher semantic representation of users: their factors and 2) It allows to simplify the next stages: prediction and recommendation since they can be done as if they were individual recommendations (made to the virtual users).

The proposed approach outperforms the state-of-the-art baselines used to recommend to groups of users. Prediction and recommendation results are particularly improved using the proposed method when it is applied to very sparse datasets. This is because BNMF has been designed to work on collaborative filtering sparse environments, and it provides suitable probabilistic hidden factors to feed the proposed

clustering stage.

The proposed method and their results open the door to different future works, such as: a) The use of weighted aggregation approaches to obtain the virtual users, or b) The prediction stage improvement by replacing the linear dot product of hidden factors by a neural network architecture that learns the complex non-linear relations that exist between hidden factors.

References

[1] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowledge-Based Syst.*, vol. 46, pp. 109–132, 2013.

[2] J. Bobadilla, F. Serradilla, and A. Hernando, "Collaborative filtering adapted to recommender systems of e-learning," *Knowledge-Based Syst.*, vol. 22, no. 4, pp. 261–265, 2009.

[3] C. Yang, L. Bai, C. Zhang, Q. Yuan, and J. Han, "Bridging Collaborative Filtering and Semi-Supervised Learning: A Neural Approach for POI Recommendation," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2017, pp. 1245–1254.

[4] E. Thomas, A. G. Ferrer, B. Lardeux, M. Boudia, C. Haas-Frangii, and R. A. Agost, "Cascaded Machine Learning Model for Efficient Hotel Recommendations from Air Travel Bookings," 2019.

[5] R. Hurtado, J. Bobadilla, R. Bojorque, F. Ortega, and X. Li, "A New Recommendation Approach Based on Probabilistic Soft Clustering Methods: A Scientific Documentation Case Study," *IEEE Access*, vol. 7, pp. 7522–7534, 2019.

[6] W.-T. Chu and Y.-L. Tsai, "A hybrid recommendation system considering visual information for predicting favorite restaurants," *World Wide Web*, vol. 20, no. 6, pp. 1313–1331, Nov. 2017.

[7] X. Wang, X. He, L. Nie, and T.-S. Chua, "Item Silk Road: Recommending Items from Information Domains to Social Users," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2017, pp. 185–194.

[8] X. Deng and F. Huangfu, "Collaborative Variational Deep Learning for Healthcare Recommendation," *IEEE Access*, vol. 7, pp. 55679–55688, 2019.

[9] F. Z. Benkaddour, N. Taghezout, F. Z. Kaddour-Ahmed, and I.-A. Hammadi, "An Adapted Approach for User Profiling in a Recommendation System: Application to Industrial Diagnosis," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 5, no. 3, pp. 118–130, 2018.

[10] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Adv. Artif. Intell.*, vol. 2009, 2009.

[11] J. Bobadilla, F. Ortega, A. Hernando, and Á. Arroyo, "A balanced memory-based collaborative filtering similarity measure," *Int. J. Intell. Syst.*, vol. 27, no. 10, pp. 939–946, 2012.

[12] B. Zhu, R. Hurtado, J. Bobadilla, and F. Ortega, "An efficient recommender system method based on the numerical relevances and the non-numerical structures of the ratings," *IEEE Access*, 2018.

[13] J. Bobadilla, R. Bojorque, A. H. Esteban, and R. Hurtado, "Recommender Systems Clustering Using Bayesian Non Negative Matrix Factorization," *IEEE Access*, vol. 6, pp. 3549–3564, 2018.

[14] X. Guan, C. T. Li, and Y. Guan, "Matrix Factorization with Rating Completion: An Enhanced SVD Model for Collaborative Filtering Recommender Systems," *IEEE Access*, vol. 5, pp. 27668–27678, 2017.

[15] R. Mehta and K. Rana, "A review on matrix factorization techniques in recommender systems," in *2017 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA)*, 2017, pp. 269–274.

[16] Z. Chen, L. Li, H. Peng, Y. Liu, H. Zhu, and Y. Yang, "Sparse General Non-Negative Matrix Factorization Based on Left Semi-Tensor Product," *IEEE Access*, vol. 7, pp. 81599–81611, 2019.

[17] K. Li, X. Zhou, F. Lin, W. Zeng, and G. Alterovitz, "Deep Probabilistic Matrix Factorization Framework for Online Collaborative Filtering," *IEEE Access*, vol. 7, pp. 56117–56128, 2019.

[18] A. Hernando, J. Bobadilla, and F. Ortega, "A non negative matrix factorization for collaborative filtering recommender systems based on a Bayesian probabilistic model," *Knowledge-Based Syst.*, vol. 97, pp. 188–202, 2016.

[19] X. He, H. Zhang, M.-Y. Kan, and T.-S. Chua, "Fast Matrix Factorization for Online Recommendation with Implicit Feedback," in *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2016, pp. 549–558.

[20] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, "Neural Collaborative Filtering," in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 173–182.

[21] D. Liang, M. Zhan, and D. P. W. Ellis, "Content-Aware Collaborative Music Recommendation Using Pre-trained Neural Networks.," in *ISMIR*, 2015, pp. 295–301.

[22] K. Yoshii, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno, "An Efficient Hybrid Music Recommender System Using an Incrementally Trainable Probabilistic Generative Model," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 16, no. 2, pp. 435–447, Feb. 2008.

[23] C. Musto, C. Greco, A. Suglia, and G. Semeraro, "Ask Me Any Rating: A Content-based Recommender System based on Recurrent Neural Networks.," in *IIR*, 2016.

[24] T. Ebesu and Y. Fang, "Neural Citation Network for Context-Aware Citation Recommendation," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2017, pp. 1093–1096.

[25] C. C. Aggarwal and C. K. Reddy, "Data clustering," *Algorithms Appl. Boca Rat. CRC Press*, 2014.

[26] Chan Young Kim, Jae Kyu Lee, Yoon Ho Cho, and Deok Hwan Kim, "VISCORS: a visual-content recommender for the mobile Web," *IEEE Intell. Syst.*, vol. 19, no. 6, pp. 32–39, Nov. 2004.

[27] L. H. Son, "A Novel Kernel Fuzzy Clustering Algorithm for Geo-Demographic Analysis," *Inf. Sci.*, vol. 317, no. C, pp. 202–223, 2015.

[28] Qing Li and Byeong Man Kim, "Clustering approach for hybrid recommender system," in *Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003)*, 2003, pp. 33–38.

[29] D. Rafailidis and P. Daras, "The TFC Model: Tensor Factorization and Tag Clustering for Item Recommendation in Social Tagging Systems," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 43, no. 3, pp. 673–688, May 2013.

[30] S. Gong, "A collaborative filtering recommendation algorithm based on user clustering and item clustering.," *JSW*, vol. 5, no. 7, pp. 745–752, 2010.

[31] R. Pan, P. Dolog, and G. Xu, "KNN-Based Clustering for Improving Social Recommender Systems," in *Agents and Data Mining Interaction*, 2013, pp. 115–125.

[32] C.-X. Zhang, Z.-K. Zhang, L. Yu, C. Liu, H. Liu, and X.-Y. Yan, "Information filtering via collaborative user clustering modeling," *Phys. A Stat. Mech. its Appl.*, vol. 396, pp. 195–203, 2014.

[33] X. Wang and J. Zhang, "Using incremental clustering technique in collaborative filtering data update," in *Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014)*, 2014, pp. 420–427.

[34] H. WU, Y. Wang, Z. WANG, X.-L. WANG, and S.-Z. DU, "Two-Phase Collaborative Filtering Algorithm Based on Co-Clustering: Two-Phase Collaborative Filtering Algorithm Based on Co-Clustering," *J. Softw.*, vol. 21, pp. 1042–1054, 2010.

[35] C. Rana and S. K. Jain, "An evolutionary clustering algorithm based on temporal features for dynamic recommender systems," *Swarm Evol. Comput.*, vol. 14, pp. 21–30, 2014.

[36] C.-L. Liao and S.-J. Lee, "A clustering based approach to improving the efficiency of collaborative filtering recommendation," *Electron. Commer. Res. Appl.*, vol. 18, pp. 1–9, 2016.

[37] Á. M. Navarro and P. Moreno-Ger, "Comparison of Clustering Algorithms for Learning Analytics with Educational Datasets," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 5, no. 2, pp. 9–16, 2018.

[38] M. K. Najafabadi, M. N. Mahrin, S. Chuprat, and H. M. Sarkan, "Improving the accuracy of collaborative filtering recommendations using clustering and association rules mining on implicit data," *Comput. Human Behav.*, vol. 67, pp. 113–128, 2017.

[39] S. Zahra, M. A. Ghazanfar, A. Khalid, M. A. Azam, U. Naeem, and A. Prugel-Bennett, "Novel centroid selection approaches for KMeans-clustering based recommender systems," *Inf. Sci. (Ny).*, vol. 320, pp. 156–189, 2015.

[40] X. Zhang, L. Zong, X. Liu, and J. Luo, "Constrained Clustering With Nonnegative Matrix Factorization," *IEEE Trans. Neural Networks Learn.*

Syst., vol. 27, no. 7, pp. 1514–1526, 2016.

[41] N. Del Buono and G. Pio, "Non-negative Matrix Tri-Factorization for co-clustering: An analysis of the block matrix," *Inf. Sci. (Ny).*, vol. 301, pp. 13–26, 2015.

[42] T. George and S. Merugu, "A scalable collaborative filtering framework based on co-clustering," in *Fifth IEEE International Conference on Data Mining (ICDM'05)*, 2005, pp. 4 pp.-.

[43] T. Li, Y. Zhang, and V. Sindhwani, "A Non-negative Matrix Tri-factorization Approach to Sentiment Classification with Lexical Prior Knowledge," in *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1 - Volume 1*, 2009, pp. 244–252.

[44] Q. Gu and J. Zhou, "Local learning regularized nonnegative matrix factorization," in *Twenty-First International Joint Conference on Artificial Intelligence*, 2009.

[45] F. Ortega, R. Hurtado, J. Bobadilla, and R. Bojorque, "Recommendation to Groups of Users Using the Singularities Concept," *IEEE Access*, vol. 6, 2018.

[46] L. Boratto and S. Carta, "State-of-the-Art in Group Recommendation and New Approaches for Automatic Identification of Groups," in *Information Retrieval and Mining in Distributed Environments*, A. Soro, E. Vargiu, G. Armano, and G. Paddeu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 1–20.

[47] Y.-D. Seo, Y.-G. Kim, E. Lee, K.-S. Seol, and D.-K. Baik, "An enhanced aggregation method considering deviations for a group recommendation," *Expert Syst. Appl.*, vol. 93, pp. 299–312, 2018.

[48] S. Feng and J. Cao, "Improving group recommendations via detecting comprehensive correlative information," *Multimed. Tools Appl.*, vol. 76, no. 1, pp. 1355–1377, 2017.

[49] L. Baltrunas, T. Makcinskas, and F. Ricci, "Group recommendations with rank aggregation and collaborative filtering," in *Proceedings of the fourth ACM conference on Recommender systems - RecSys '10*, 2010, p. 119.

[50] F. Ortega, A. Hernando, J. Bobadilla, and J. H. Kang, "Recommending items to group of users using Matrix Factorization based Collaborative Filtering," *Inf. Sci. (Ny).*, vol. 345, pp. 313–324, 2016.

[51] F. Ortega, J. Bobadilla, A. Hernando, and A. Gutiérrez, "Incorporating group recommendations to recommender systems: Alternatives and performance," *Inf. Process. Manag.*, vol. 49, no. 4, pp. 895–901, 2013.

[52] L. Boratto and S. Carta, "ART: group recommendation approaches for automatically detected groups," *Int. J. Mach. Learn. Cybern.*, vol. 6, no. 6, pp. 953–980, 2015.

[53] L. Boratto and S. Carta, "The rating prediction task in a group recommender system that automatically detects groups: architectures, algorithms, and performance evaluation," *J. Intell. Inf. Syst.*, vol. 45, no. 2, pp. 221–245, Oct. 2015.

[54] S. Berkovsky and J. Freyne, "Group-based Recipe Recommendations: Analysis of Data Aggregation Strategies," in *Proceedings of the Fourth ACM Conference on Recommender Systems*, 2010, pp. 111–118.

[55] J. Masthoff, "Group Recommender Systems: Combining Individual Models," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Boston, MA: Springer US, 2011, pp. 677–702.

[56] L. Boratto, S. Carta, and G. Fenu, "Investigating the role of the rating prediction task in granularity-based group recommender systems and big data scenarios," *Inf. Sci. (Ny).*, vol. 378, pp. 424–443, 2017.

[57] M. O'Connor, D. Cosley, J. A. Konstan, and J. Riedl, "PolyLens: A Recommender System for Groups of Users," in *ECSCW 2001: Proceedings of the Seventh European Conference on Computer Supported Cooperative Work 16--20 September 2001, Bonn, Germany*, W. Prinz, M. Jarke, Y. Rogers, K. Schmidt, and V. Wulf, Eds. Dordrecht: Springer Netherlands, 2001, pp. 199–218.

[58] C. Xiong, K. Lv, H. Wang, and C. Qi, "Personalized Group Recommendation Model Based on Argumentation Topic," in *Complex, Intelligent, and Software Intensive Systems*, 2019, pp. 206–217.

[59] D. Sacharidis, "Modeling Uncertainty in Group Recommendations," in *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, 2019, pp. 69–74.

[60] D. Cao, X. He, L. Miao, Y. An, C. Yang, and R. Hong, "Attentive Group Recommendation," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 2018, pp. 645–654.

[61] B.-H. Li *et al.*, "GRIP: A Group Recommender Based on Interactive Preference Model," *J. Comput. Sci. Technol.*, vol. 33, no. 5, pp. 1039–1055, Sep. 2018.

[62] J. Du, L. Li, P. Gu, and Q. Xie, "A Group Recommendation Approach Based on Neural Network Collaborative Filtering," in *2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW)*, 2019, pp. 148–154.

[63] A. Roy, S. Banerjee, M. Sarkar, A. Darwish, M. Elhoseny, and A. E. Hassanien, "Exploring New Vista of intelligent collaborative filtering: A restaurant recommendation paradigm," *J. Comput. Sci.*, vol. 27, pp. 168–182, 2018.

[64] J. Park and K. Nam, "Group recommender system for store product placement," *Data Min. Knowl. Discov.*, vol. 33, no. 1, pp. 204–229, 2019.

[65] D. Herzog and W. Wörndl, "A User Study on Groups Interacting with Tourist Trip Recommender Systems in Public Spaces," in *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*, 2019, pp. 130–138.

[66] R. Logesh, V. Subramaniyaswamy, V. Vijayakumar, and X. Li, "Efficient User Profiling Based Intelligent Travel Recommender System for Individual and Group of Users," *Mob. Networks Appl.*, vol. 24, no. 3, pp. 1018–1033, Jun. 2019.

[67] T.-C. T. Chen and M.-C. Chiu, "A classifying ubiquitous clinic recommendation approach for forming patient groups and recommending suitable clinics," *Comput. Ind. Eng.*, vol. 133, pp. 165–174, 2019.

[68] A. Zawali and I. Boukhris, "A Group Recommender System for Academic Venue Personalization," in *Intelligent Systems Design and Applications*, 2020, pp. 597–606.

[69] H. J. Jeong and M. H. Kim, "HGGC: A hybrid group recommendation model considering group cohesion," *Expert Syst. Appl.*, vol. 136, pp. 73–82, 2019.

[70] S. Feng, H. Zhang, J. Cao, and Y. Yao, "Merging user social network into the random walk model for better group recommendation," *Appl. Intell.*, vol. 49, no. 6, pp. 2046–2058, Jun. 2019.

[71] T. N. T. Tran, M. Atas, A. Felfernig, V. M. Le, R. Samer, and M. Stettinger, "Towards Social Choice-based Explanations in Group Recommender Systems," in *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*, 2019, pp. 13–21.

[72] W. Wang, G. Zhang, and J. Lu, "Hierarchy Visualization for Group Recommender Systems," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 49, no. 6, pp. 1152–1163, 2019.

[73] C.-H. Lai and Y.-C. Chang, "Document recommendation based on the analysis of group trust and user weightings," *J. Inf. Sci.*, vol. 0, no. 0, p. 0165551518819973.

[74] F. M. Harper and J. A. Konstan, "The MovieLens Datasets: History and Context," *ACM Trans. Interact. Intell. Syst.*, vol. 5, no. 4, pp. 19:1--19:19, 2015.

[75] J. Golbeck and J. Hendler, "FilmTrust: movie recommendations using trust in web-based social networks," in *CCNC 2006. 2006 3rd IEEE Consumer Communications and Networking Conference, 2006.*, 2006, vol. 1, pp. 282–286.

[76] J. Bobadilla, F. Ortega, A. Gutiérrez and S. Alonso, "Classification-based deep neural network architecture for collaborative filtering recommender systems", *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 1, pp. 68-77, 2020.

Remigio Hurtado

Remigio Hurtado was born in 1989. He received the B.S degree in Systems Engineering from Universidad Politécnica Salesiana, Ecuador in 2012, Master degree in Information and Software Technology, Instituto Tecnológico y de Estudios Superiores de Monterrey, México, 2014, and Master degree in Computer Science and Technology, Universidad Politécnica de Madrid, Spain, 2017. He is lecturer at the Universidad Politécnica Salesiana, Ecuador. His research interests include the recommender systems and natural language processing. He is currently member of research team in Artificial Intelligence and Assistive Technology, his team is collaborating with Universidad Politécnica de Madrid.

Jesús Bobadilla

Jesús Bobadilla received the B.S. and the Ph.D. degrees in computer science from the Universidad Politécnica de Madrid and the Universidad Carlos III. Currently, he is a lecturer with the Department of Applied Intelligent Systems, Universidad Politécnica de Madrid. He is a habitual author of programming languages books working with McGraw-Hill, Ra-Ma and Alfa Omega publishers. His research interests include information retrieval, recommender systems and speech processing. He is in charge of the FilmAffinity.com research team working on the collaborative filtering kernel of the web site. He has been a researcher into the International Computer Science Institute at Berkeley University and into the Sheffield University. Head of the research group.

Abraham Gutiérrez

Abraham Gutiérrez was born in Madrid, Spain, in 1969. He received the B.S. and the Ph.D. degrees in computer science from the Universidad Politécnica de Madrid. Currently, he is a lecturer with the Department of Applied Intelligent Systems, Universidad Politécnica de Madrid. He is a habitual author of programming languages books working with McGraw-Hill, Ra-Ma and Alfa Omega publishers. His research interests include P-Systems, social networks and recommender systems. He is in charge of this group innovation issues, including the commercial projects.

Santiago Alonso

Santiago Alonso received his B.S. degree in software engineering from Universidad Autónoma de Madrid and his Ph.D. degree in computer science and artificial intelligence from Universidad Politécnica de Madrid, in 2015, where he is currently an Associate Professor, participating in master and degree subjects and doing work related with advanced databases. His main research interests include natural computing (P-Systems), and did some work on genetic algorithms. His current interests include machine learning, data analysis and artificial intelligence.

Tree Growth Algorithm for Parameter Identification of Proton Exchange Membrane Fuel Cell Models

Hamdy M. Sultan[1,2], Ahmed S. Menesy[1,3], Salah Kamel[3,4], Francisco Jurado[5]*

[1] Electrical Engineering Department, Faculty of Engineering, Minia University, 61111 Minia (Egypt)
[2] Electrical power systems Department, Moscow Power Engineering Institute (MPEI), 111250 Moscow (Russia)
[3] State Key Laboratory of Power Transmission Equipment & System Security and New Technology, School of Electrical Engineering, Chongqing University, 400044 Chongqing (China)
[4] Electrical Engineering Department, Faculty of Engineering, Aswan University, 81542 Aswan (Egypt)
[5] Department of Electrical Engineering, University of Jaén, 23700 EPS Linares, Jaén (Spain)

Received 26 December 2019 | Accepted 5 March 2020 | Published 27 March 2020

Abstract

Demonstrating an accurate mathematical model is a mandatory issue for realistic simulation, optimization and performance evaluation of proton exchange membrane fuel cells (PEMFCs). The main goal of this study is to demonstrate a precise mathematical model of PEMFCs through estimating the optimal values of the unknown parameters of these cells. In this paper, an efficient optimization technique, namely, Tree Growth Algorithm (TGA) is applied for extracting the optimal parameters of different PEMFC stacks. The total of the squared deviations (TSD) between the experimentally measured data and the estimated ones is adopted as the objective function. The effectiveness of the developed parameter identification algorithm is validated through four case studies of commercial PEMFC stacks under various operating conditions. Moreover, comprehensive comparisons with other optimization algorithms under the same study cases are demonstrated. Statistical analysis is presented to evaluate the accuracy and reliability of the developed algorithm in solving the studied optimization problem.

Keywords

Proton Exchange Membrane Fuel Cell, Parameters Estimation, Tree Growth Algorithm, Total of the Squared Deviations (TSD).

DOI: 10.9781/ijimai.2020.03.003

I. Introduction

THE alarming increase in environmental pollution and the growing shortage in conventional energy sources, lead to searching for alternative environmentally friendly sources. Photovoltaic, wind and fuel cells (FC) are considered the most promising sources that are applied in small and large scales [1]-[5]. Thanks to its high efficiency, simple operating principle, low exhausts, durability and reliability, the fuel cell has attracted the attention of researches and decision-makers in the last decades [6]. Particularly, the proton exchange membrane fuel cell (PEMFC) has been demonstrated as a suitable solution for various applications, because of its high efficiency, short startup time and its good performance under low temperatures [7]-[9]. The operating temperature of PEMFC occurs in the range of 70-85°C.

The operation of the PEMFC can be well understood if an efficient reliable model is established. Accordingly, the design of the PEMFC stack can be improved and optimum operation can be reached as well as the integration of the fuel cell stack into other devices and systems [5]. In the last years, many research papers tried to model the operation of PEMFC and many models have been provided in literature [10]-[15]. The semi-empirical model proposed by Amphlett et al. is considered as the most acceptable model for many researchers [10]. In this model the problem of modelling the PEMFC has been turned into an optimization problem of some unknown parameters in the parametric equations that describe the model. Meanwhile, the model includes multi-variable nonlinear equation that makes it very difficult to estimate the values of these parameters using traditional optimization techniques.

Many researches have been demonstrated for extracting the optimal design parameters using optimization algorithms in both parameter and non-parameter modelling [16], [17]. The polarization curves of the PEMFC is highly nonlinear and the proposed model is based on a set of nonlinear equations that contain a set of semi-empirical adjustable parameters. The conventional optimization methods are not suitable for such complicated optimization problems, so metaheuristic optimization techniques are used for such issues [18], [19]. Based on parameter modelling, various algorithms have been proposed to solve the optimization problem of PEMFC parameters. In order to improve the parameters' accuracy of the PEM fuel cell a hybrid genetic algorithm was proposed in [20]. The particle swarm optimization (PSO) has been proposed for PEMFC parameters estimation problem in [21]. The differential evaluation (DE), as well as the hybrid adaptive differential evaluation (HADE) algorithms, have been introduced for solving the optimization problem presented in [22], [23]. More recent optimization methods have been applied to solve the problem of PEMFC's parameter such as: the harmony search algorithm (HAS) [24], the seeker optimization algorithm (SOA) [25], the multi-verse optimizer (MVO) [26], the adaptive RNA genetic algorithm [27], Eagle strategy based on JAYA algorithm and Nelder-Mead simplex method (JAYA-

* Corresponding author.
E-mail address: fjurado@ujaen.es

NM) [28], grey wolf optimizer (GWO) [29], hybrid Teaching Learning Based Optimization – Differential Evolution algorithm (TLBO-DE) [30], shark smell optimizer (SSO) [25], Cuckoo search algorithm with explosion operator (CS-EO) [31], selective hybrid stochastic strategy [32], bird mating optimizer [33], grasshopper optimizer (GHO) [34], Chaotic Harris Hawks optimization (CHHO) [35], and Modified Artificial Ecosystem Optimization (MAEO) [36].

To extract the optimal values of the unknown parameters on the proposed model, a metaheuristic technique inspired by the competition among trees in reaching the sources of food and light, Tree Growth Algorithm (TGA), is proposed to solve the optimization problems [37]. TGA is a very simple code that can be applied to various types of problems. For example, in [38], tree growth algorithm has been used to solve the localization problem in wireless sensor networks. The present work also applies TGA to the targeted problem. The main contributions of this study can be summarized as follows:

- Developing the TGA for extracting the optimal best parameters of different PEMFC stacks.
- Demonstrating an effective mathematical model of PEMFC stacks which imitates the principle operation of different commercial PEMFCs through estimating the optimal values of the unknown parameters of these cells;
- Studying the effect of cell temperature and reactants' pressure variations on the electrical characteristics of various PEMFC stacks;
- A comprehensive comparison between the results obtained by the TGA and those obtained by other metaheuristic optimization algorithms has been provided;
- Four different models of PEM fuel cells are introduced to verify the effectiveness of the TGA optimization method;
- Parametric and non-parametric statistical analysis have been demonstrated to validate the goodness of the proposed metaheuristic technique;
- The results obtained by the application of TGA prove its reliability and superiority in estimation of the effective parameters of PEMFC stacks.

The rest of the paper is arranged as follows: Section II introduces the model of the PEMFC as well as the objective function. The tree growth algorithm is briefly explained in section III. Section IV introduces the application of TGA for parameter estimation of different PEM fuel cell stacks under various operating conditions. Statistical analysis of the obtained results from the TGA is demonstrated in Section V. Section VI is dedicated to the conclusion.

II. Mathematical Description of the PEMFCs for Parameters Estimation

A. PEMFC Model

A PEMFC consists of two isolated electrodes (anode and cathode) separated by a thin solid membrane able to conduct protons as described in Fig. 1 [39].

The overall chemical reaction occurring in the fuel cell can be represented as follows [39]:

$$H_2 + \frac{1}{2}O_2 \rightarrow H_2O + Electricity + Heat \tag{1}$$

The voltage and current generated from a single fuel cell are too small, therefore a number of fuel cells are connected in parallel and/or series to form a fuel cell stack having a reasonable voltage and current rating. When N_{cells} of identical fuel cells are connected in series, the resultant voltage of the stack will be calculated as follows,

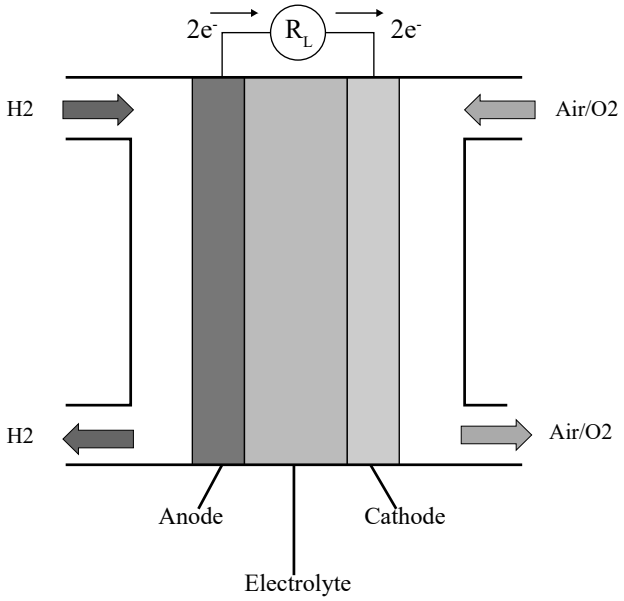


Fig. 1. Schematic configuration of the PEMFC.

$$V_{stack} = N_{cells}.V_{cell} \tag{2}$$

where, V_{cell} denotes the voltage of a single fuel cell.

Because of the three types of voltage loss occurred in the fuel cell, namely, the activation loss V_{act}, the concentration voltage loss V_{con}, and the ohmic voltage loss V_{ohm}, the terminal voltage of the fuel cell will be calculated as follows [40],

$$V_{cell} = E_{Nernst} - V_{act} - V_{ohm} - V_{con} \tag{3}$$

where, E_{Nernst} denotes the theoretic voltage of the fuel cell, which can be expressed as given in the following formula [33],

$$E_{Nernst} = 1.229 - 0.85\times10^{-3}(T-298.15) + 4.3085 \times10^{-5}T\times\left[\ln(P_{H_2}) - \frac{1}{2}\ln(P_{O_2})\right] \tag{4}$$

where, T is the cell temperature in Kelvin; and denotes the partial pressures of the reactants (i.e. hydrogen and oxygen) at the inlet channels of the fuel cell stack (atm).

During the operation of the fuel cell, if the inputs to the PEMFC stack are hydrogen and natural air, then can be expressed as follows [40],

$$P_{O_2} = P_c - RH_c P^{sat}_{H_2O} - \frac{0.79}{0.21}P_{O_2} \times \exp\left(\frac{0.291\left(I_{fc}/A\right)}{T^{0.832}}\right) \tag{5}$$

where, P_c represents the pressure of the input channel at the cathode (atm), RH_c denotes the relative humidity around the cathode, I_{fc} is the current generated by the cell (A), A is the area of the membrane surface (cm2), $P^{sat}_{H_2O}$ represents the water vapor pressure at saturation, which is defined as follows [40],

$$\log_{10}\left(P^{sat}_{H_2O}\right) = 2.95\times10^{-2}(T-273.15) - 9.18\times10^{-5}(T-273.15)^2 + 1.44\times10^{-7}(T-273.15)^3 - 2.18 \tag{6}$$

If the inputs of the fuel cell stack are hydrogen and pure oxygen, the partial pressure of oxygen will be calculated as follows [40],

$$P_{O_2} = RH_c P^{sat}_{H_2O}\left[\left(\exp\left(\frac{4.192\left(I_{fc}/A\right)}{T^{1.334}}\right)\times\frac{RH_c P^{sat}_{H_2O}}{P_a}\right)^{-1}-1\right] \quad (7)$$

In both previously mentioned conditions, the partial pressure of the hydrogen is calculated as follows,

$$P_{H_2} = 0.5RH_a P^{sat}_{H_2O}\left[\left(\exp\left(\frac{1.635\left(I_{fc}/A\right)}{T^{1.334}}\right)\times\frac{RH_a P^{sat}_{H_2O}}{P_a}\right)^{-1}-1\right] \quad (8)$$

where, P_a denotes the pressure at the anode side channel (atm), RH_a represents the relative humidity of water vapor in the side of the anode.

The voltage loss due to the activation process V_{act} can be calculated as,

$$V_{act} = -\left[\xi_1 + \xi_2 T + \xi_3 T\ln\left(C_{O_2}\right) + \xi_4 T\ln\left(I_{fc}\right)\right] \quad (9)$$

where, ξ_1, ξ_2, ξ_3, ξ_4 denote semi-empirical coefficients; C_{O2} represents the concentration of the oxygen at the cathode (mol.cm-3) and is calculated as [40]:

$$C_{O_2} = \frac{P_{O_2}}{5.08\times10^6\times\exp^{-\left(498/T\right)}} \quad (10)$$

The second type of voltage loss occurred in the fuel cell, the ohmic voltage loss V_{ohm} can be defined as,

$$V_{ohm} = I_{fc}\left(R_M + R_C\right) \quad (11)$$

where, R_M is the resistance of the membrane surface, RC denotes the resistance that the protons face when transferring through the membrane. The membrane resistance can be calculated as follows [40],

$$R_M = \frac{\rho_M . l}{A} \quad (12)$$

where, ρ_M denotes the specific resistance of membrane material (Ω.cm), l is the thickness of the membrane (cm). ρ_M can be expressed as follows,

$$\rho_M = \frac{181.6\left[1+0.03\left(\frac{I_{fc}}{A}\right)+0.062\left(\frac{T}{303}\right)^2\left(\frac{I_{fc}}{A}\right)^{2.5}\right]}{\left[\lambda-0.634-3\left(\frac{I_{fc}}{A}\right)\right]\times\exp\left[4.18\left(\frac{T-303}{T}\right)\right]} \quad (13)$$

where λ denotes an adjustable empirical parameter, which needs to be extracted.

The last type of losses that takes place in the fuel cell is the voltage loss due to concentration, which can be expressed as follows [40],

$$V_{con} = -b\ln\left(1-\frac{J}{J_{\max}}\right) \quad (14)$$

where b denotes the parametric coefficient that needs to be estimated; J and $Jmax$ are the current density and the maximum current density (A/cm2), respectively.

B. Formulation of the Objective Function

Through a closer look at the previous equations, it will be noticed that the operation of the PEMFC depends on a set of unknown adjustable parameters. For the optimization problem provided in this study, the target from the parameter estimation is to extract the optimal values of these parameters, which let the proposed model to match well with the experimentally measured data of the PEMFC. Therefore, the proposed objective function is a measure of the quality of the extracted parameters. The degree of matching between the calculated data and the data obtained from experiments can be formulated as the difference between the estimated output voltage based on the extracted parameters and the measured voltage. Accordingly, in this paper the Total Square Deviation (TSD) between the measured voltage of PEMFC and computed stack voltage is defined as the objective function (OF) [29]-[31], [41].

$$OF = \min TSD(X) = \sum_{i=1}^{N}\left(V_{meas} - V_{stack}\right)^2 \quad (15)$$

according to the following constraints,

$$\begin{aligned} &\xi_{k\min} \le \xi_k \le \xi_{k\max}, \quad k = 1:4 \\ &b_{\min} \le b \le b_{\max} \\ &R_{C\min} \le R_C \le R_{C\max} \\ &\lambda_{\min} \le \lambda \le \lambda_{\max} \end{aligned} \quad (16)$$

where, X denotes a vector of the parameters that have to be estimated, V_{meas} is the experimentally measured voltage, V_{stack} is the computed voltage from the proposed model of the PEMFC, and N denotes the length of the data points.

III. Tree Growth Algorithm (TGA)

The TGA optimization algorithm was firstly proposed by Mostafa Hajiaghaei-Keshteli and Armin Cheraghalipour, which is inspired by the competition among trees in a certain area for absorbing the resources of light and food [37]. TGA in its process is divided into four phases. The first phase N1, which includes the best trees that have a good opportunity for acquiring food and light. Thus these trees will focus their effort on obtaining food, as the light source is already guaranteed by their height. In the second group, called the competition for the light group N_2, some trees will move a distance by changing their angle to catch the light source through the tall elder ones. In the third group N_3, which is called remove and replace group, the trees which do not have a good chance for growth are cut by the foresters and replaced by small new ones. In the final group called the reproduction group N_4, the tall good trees begin to generate new small trees around the elder mother.

TGA algorithm is mathematically executed in the following manner. Firstly, an initial generation of trees N is randomly generated with respect to the predefined upper and lower boundaries and the fitness function is calculated for each individual search agent. After that, the initial generation is arranged with respect to the value of fitness function and the present best solution T^j_{GB} in the j-th iteration of the search space is determined. Then equation (17) is used to execute the local search to the individual agents in the first population $N1$ [37],

$$T_i^{j+1} = \frac{T_i^j}{\theta} + rT_i^j \quad (17)$$

where, θ denotes the reduction rate of trees due to their age and reduction of food sources in the surrounded area. r is a randomly distributed number between [0, 1]. T_i^j and T_i^{j+1} are the current solution and the next solution in the population N_1.

When the new solution is produced, a comparison with the previous solutions is conducted, in with the worst solutions will be ignored. The solutions of the second group N_2 will be moved towards the best solutions of the first group N_1 under different angles α. Equation (18) is used to define the distance between the selected solutions in the second group and the other solutions [38], [42],

$$d_i = \sqrt{\sum_{i=1}^{N_1+N_1}\left(T_{N_2}^{j} - T_i^{j}\right)^2}, \quad where \quad d_i = \begin{cases} d_i, \ if \ T_{N_2}^{j} \neq T_i^{j} \\ \infty, \ if \ T_{N_2}^{j} = T_i^{j} \end{cases} \tag{18}$$

Once the distance d_i has been calculated, two solutions x_1 and x_2, having the minimum distance according to any solution, form a linear combination between the trees as described in the following expression (Fig. 2),

$$y = \lambda x_1 + (1-\lambda)x_2 \tag{19}$$

where, λ denotes a control parameter in the range [0,1].

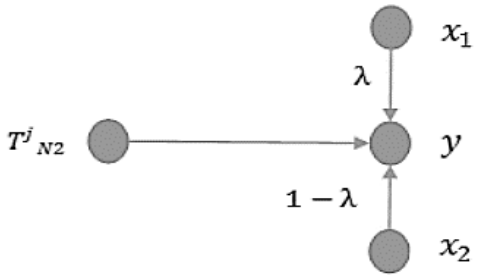


Fig. 2. Linear combination [37].

All possible solutions from the second group N_2 are moved between the two adjacent solutions with an angle α_i as shown in Fig. 3 and formulated as follows,

$$T_{N_2}^{j} = T_{N_2}^{j} + \alpha_i y \tag{20}$$

When the third subpopulation N_3 is reached, the worst solutions are removed and new solutions are generated randomly. Then a new population is created depending on the previous three populations, $N = N_1 + N_2 + N_3$. The new generated subpopulation N_4 is created and modified according to the best solution in the first group N_1 using a masked operator. Then the new solutions from the randomly created subpopulation are combined with the N population.

The fitness of the created population $N + N_4$ is determined and the best solutions are stored and considered as the initial population in the next iteration of the tree growth algorithm search space. The roulette wheel or tournament is employed for this process. The flowchart that describes the procedure of the TGA algorithm is shown in Fig. 4.

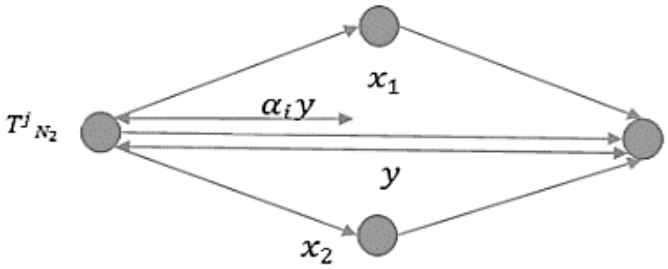


Fig. 3. Moving between two adjacent trees [37].

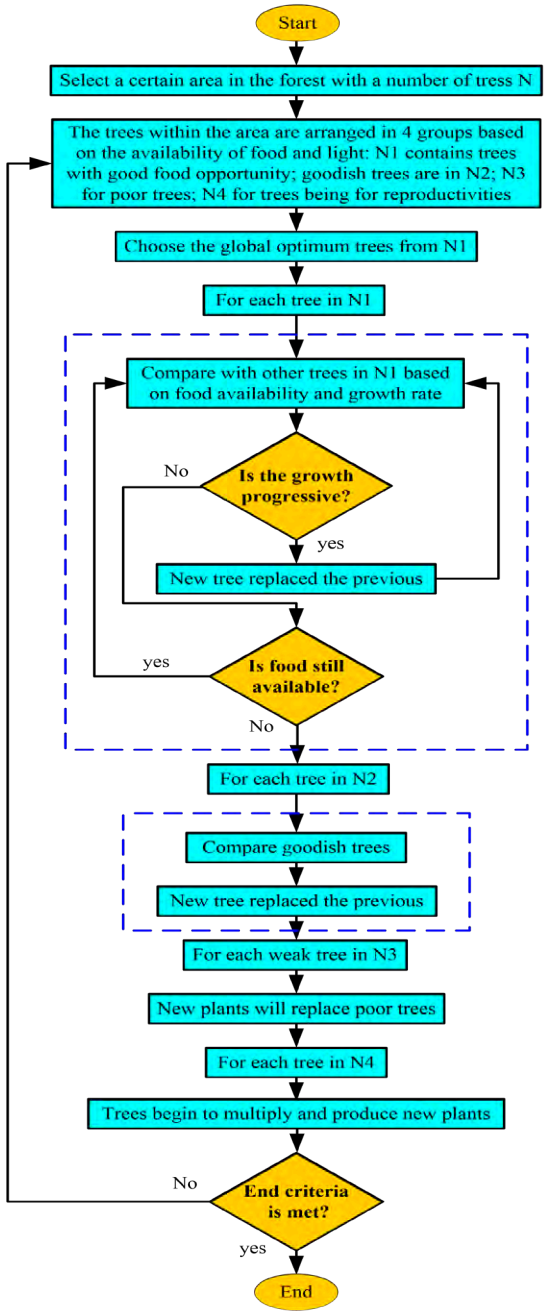


Fig. 4. Flowchart of the TGA algorithm.

IV. Simulation Results and Discussions

The simulation process has been performed using MATLAB software simulation package. The simulation was implemented using Intel Core i3-M370 CPU@2.40GHz and 4.00MB RAM Laptop. In order to validate the ability of TGA algorithm in identifying the unknown parameters of the PEM fuel cell, four different PEM stacks have been adopted from [30] - [33], [41]; namely BCS-500W FC, 250W PEMFC stack, SR-12-500W FC and Temasek 1kW FC stack. The estimation process has been used to identify the seven unknown parameters (ξ_1, ξ_2, ξ_3, ξ_4, β, Rc and λ). The upper and lower limits of the parameters, which have been optimized in this work are shown in Table I [30], [31]. The characteristics of the four PEMFC stacks included in this case study are summarized in Table II.

TABLE I. Upper and Lower Ranges of the Unknown Parameters

Parameter	ξ_1	$\xi_2\times10^{-3}$	$\xi_3\times10^{-5}$	$\xi_4\times10^{-4}$	λ	$R_c\times10^{-4}$	b
Min.	-1.1997	1	3.6	-2.6	10	1	0.0136
Max.	-0.8532	5	9.8	-0.954	23	8	0.5

TABLE II. Characteristics of Different FC Stack Types Under Study

PEMFC type	250 W stack	BCS 500W	SR-12 PEM 500W	Temasek 1kW
N (cells)	24	32	48	20
A (cm^2)	27	64	62.5	150
l (μm)	127	178	25	51
J_{max} (mA/cm^2)	860	469	672	1500
P_{H2} (atm)	1	1	1.47628	0.5
P_{O2} (atm)	1	0.2095	0.2095	0.5
T(K)	343.15	333	323	323

The TGA has been used in this study, while the following control variables have been adopted: the maximum number of iterations is adjusted at 500 iterations and the population size is 20. Due to the randomness nature of the technique used in this study, the best solution is taken as the minimum value obtained within 30 independent executions for the optimization algorithm. In addition, to validate the effectiveness of the developed algorithm, the dynamic response of the studied PEMFC stacks have been introduced. The results obtained from the proposed model is compared with the experimental data given in the datasheet of each fuel cell type.

A. Parameters' Estimation of the Proposed PEMFC Stacks

The TGA was applied for extracting the optimal values of the unknown seven parameters under the upper and lower boundaries given in Table I. The results of the minimum values of the objective function with the lowest SSE over the 30 runs as well as the optimized parameters of the four fuel cell stacks under study are given in Table III. Once again, by a deep closer look to Table III, it can be seen that small absolute deviations of the estimated values proved the agreement between the experimental datasheet values and the computed values of voltage. Moreover, the results obtained by the application of TGA are compared with the results introduced in literature [23], [28]-[31], [41]. The minor values of the objective function point out the significance of the suggested TGA in solving the optimization problem. Fig. 5 shows the convergences curves of the SSE of the proposed optimization technique for all PEMFC stacks under consideration. It can be seen that the value of TSD is minimized after 98 iterations in the case of 250W FC stack, 220 iterations for BCS 500W stack, 310 iterations for SR-12 500W PEMFC and finally after 435 iterations in the case of Temasek 1kW fuel cell stack. It may be noticed that the convergence curves prove smooth, rapid and steadily progress to the optimal final values for all fuel cell stacks under the demonstration.

The *I-V* and *I-P* polarization curves obtained from applying the best solution obtained from TGA-based method compared with the measured data for the four types of PEMFC stacks are shown in Fig. 6(a) – Fig. 6(h). Through a closer look to Fig. 6, it can be admitted that the computed curves based on the estimated values of the unknown parameters give a good matching with experimentally measured ones, which validate the insignificant value of the TSD given in Table III, IV, V and VI. Table III presents the obtained results from the developed TGA-based PEMFC model for 250W FC stack compared with the results introduced in literature; HGA and SGA [20], HADE [23], ARNA-GA [27], JAYA-NM [28], TLBO_DE [30]. It is clearly seen from Table III that the developed algorithm successes to find the minimum value of the TSD of 0.7496, which is less than the corresponding values obtained from other algorithms. The values of the optimized parameters as well as the value of TSD for BCS 500W PEMFC stack compared with that identified using GWO [29], SSO [41] and CS-EO [31] are provided in Table IV. Also, in this case, the developed algorithm provides the optimal values of the unknown parameters with an insignificant value of TSD of 0.083525, which ensures the coincidence between the polarization curves based on the proposed model and the measured ones. The estimated parameters compared with that obtained based on GWO [29], SSO [41] and CS-EO [31] for SR-12 500W FC stack are shown in Table V. Table VI shows the values of the seven unknown parameters using the TGA, GWO [29] and SSO [41] for Temasek-1kW PEMFC stack. Similarly, for SR-12 500W FC and Temasek 1kW PEMFC stack, TGA provides a superior performance over all optimization techniques as it is obviously seen from the small values of TSD. Although there are some deviations between the computed polarization characteristics and the measured ones provided in the datasheet of the manufacturers, they are acceptable in engineering standards.

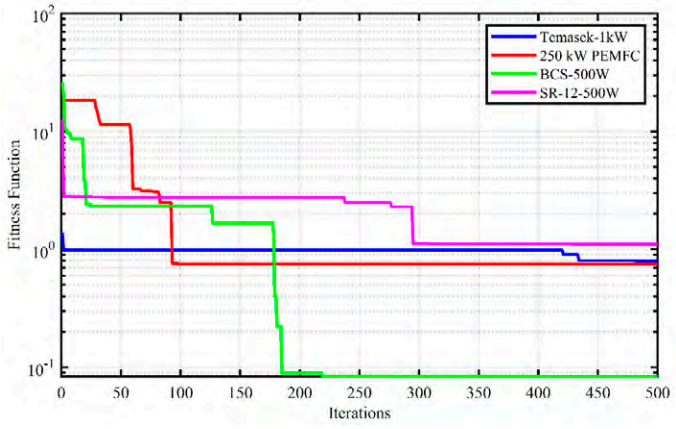


Fig. 5. Convergence trends of TSD of the TGA method for the PEMFC stacks under study.

TABLE III. Results of the TGA Compared with Other Literature for 250W PEMFC Stack

	TGA	HGA	SGA	HADE	ARNA-GA	JAYA-NM	TLBO-DE
ξ_1	-1.1914	-0.944	-0.947	-0.853	-0.947	-1.199	-0.853
$\xi_2\times10^{-3}$	4.1129	3.018	3.0641	2.8100	3.0586	3.55	2.6505
$\xi_3\times10^{-5}$	6.0573	7.401	7.7134	8.0920	7.6059	6	8.0015
$\xi_4\times10^{-4}$	-1.7090	-1.88	-1.939	-1.287	-1.88	-1.2	-1.360
λ	18.689	23	19.765	14.044	23	13.228	15.651
b	0.0544	0.029	0.024	0.0335	0.0329	0.0333	0.0364
$R_c\times10^{-4}$	4.8527	1.00	2.7197	1.00	1.1026	1.00	1.00
TSD	0.7496	4.846	5.653	7.990	2.951	5.2513	7.2776

TABLE IV. Results of the TGA Compared with Other Literature for BCS-500W PEMFC Stack

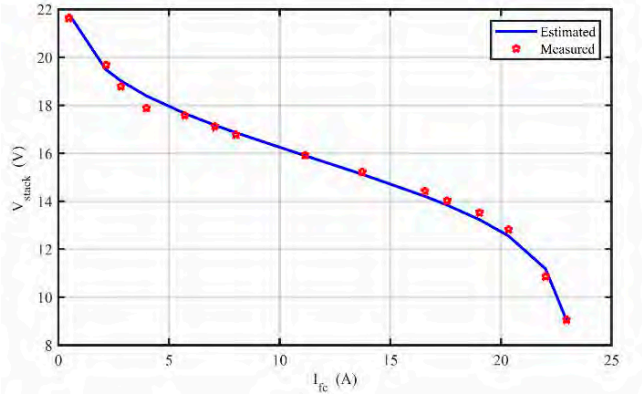
	TGA	GWO	SSO	CS-EO
ξ_1	-0.970482	-1.018	-1.018	-1.1365
$\xi_2\times10^{-3}$	2.952169	2.3151	2.3151	2.9254
$\xi_3\times10^{-5}$	5.9528506	5.24	5.24	3.7688
$\xi_4\times10^{-4}$	-1.838608	-1.2815	-1.2815	-1.3949
λ	22.50299	18.8547	18.8547	18.5446
b	0.018229	0.0136	0.0136	0.0136
$R_c\times10^{-4}$	3.8311999	7.503	7.5036	8.00
TSD	0.083525	7.1889	7.1889	5.5604

TABLE V. RESULTS OF THE TGA COMPARED WITH OTHER LITERATURE FOR SR-12 500W PEMFC STACK

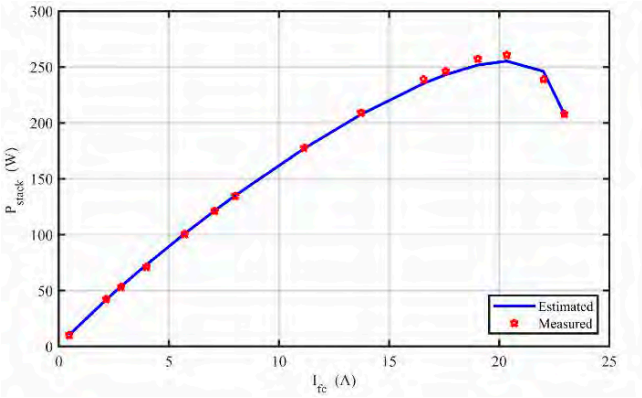
	TGA	GWO	SSO	CS-EO
ξ_1	-1.112395	-0.9664	-0.9664	-1.0353
$\xi_2\times10^{-3}$	3.8546635	2.2833	2.2833	3.354
$\xi_3\times10^{-5}$	4.3698573	3.4	3.4	7.2428
$\xi_4\times10^{-4}$	-0.964482	-0.954	-0.954	-0.954
λ	23	15.7969	15.7969	10
b	0.18307	0.1804	0.1804	0.1471
$R_c\times10^{-4}$	2.188689	6.6853	6.6853	7.1233
TSD	1.1040851	1.517	1.517	7.5753

TABLE VI. RESULTS OF THE TGA COMPARED WITH OTHER LITERATURE FOR TEMASEK-1KW PEMFC STACK

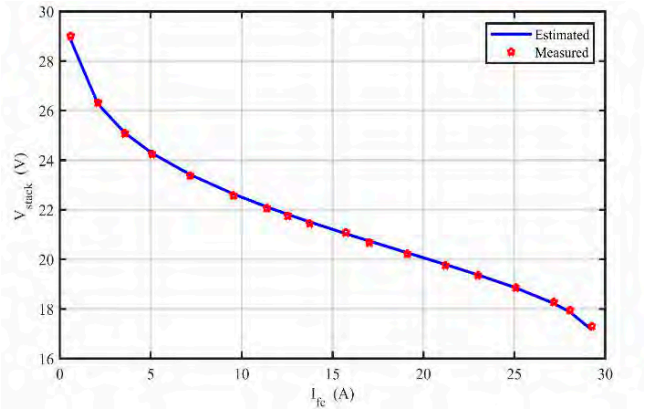
	TGA	GWO	SSO
ξ_1	-0.872182	-1.0299	-1.0299
$\xi_2\times10^{-3}$	2.5265567	2.4105	2.4105
$\xi_3\times10^{-5}$	3.818959	4.00	1.00
$\xi_4\times10^{-4}$	-2.42319	-0.954	-0.954
λ	14.79207	10.0005	10.0005
b	0.066572	0.1274	0.1274
$R_c\times10^{-4}$	0.894339	1.0873	1.0873
TSD	0.796926	1.6481	1.6481



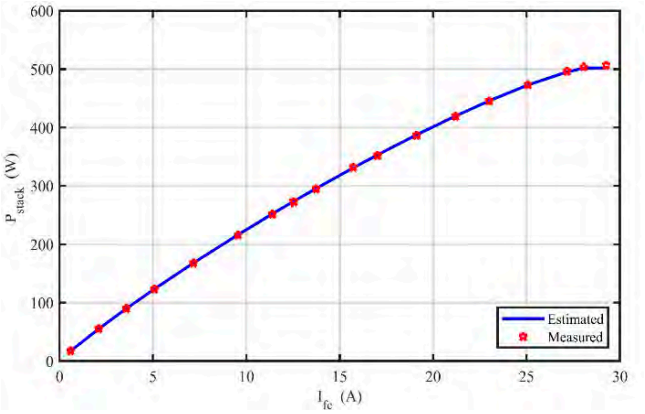
(a)



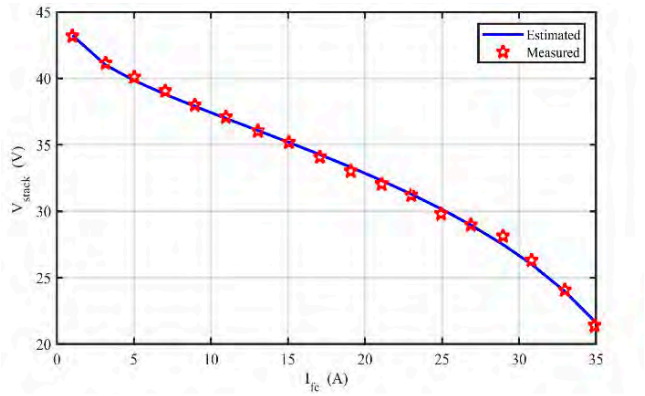
(b)



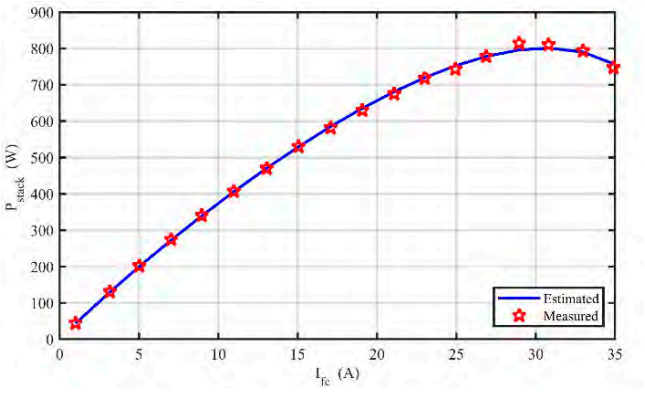
(c)



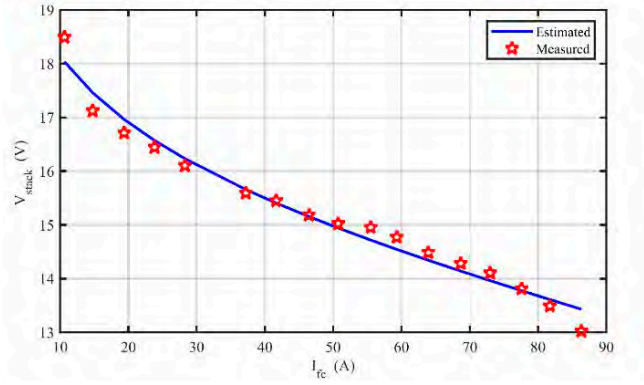
(d)



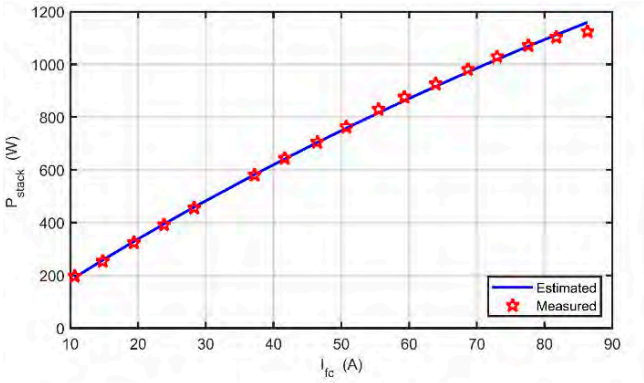
(e)



(f)



(g)



(h)

Fig. 6. *I-V* and *I-P* polarization curves: (a) 250W FC stack *I-V* plot, (b) 250W FC stack *I-P* plot, (c) BCS 500W stack *I-V* plot, (d) BCS 500W stack *I-P* plot, (e) SR-12 500W FC *I-V* plot (f) SR-12 500W FC *I-P* plot, (g) Temasek 1kW stack *I-V* plot, (h) Temasek 1kW stack *I-P* plot.

B. Simulation Under Different Operating Conditions

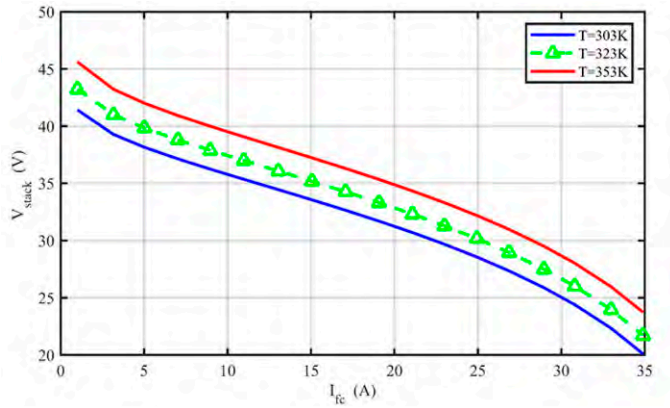
In this section, various combinations of cell temperature and the inlet pressure of oxygen and hydrogen are proposed to demonstrate the performance of the fuel cell stacks under study. Accordingly, the polarization characteristics of the PEMFC stacks are elucidated and the behavior of the stack efficiency is demonstrated as well. Starting from the importance of the PEMFC efficiency and according to [43], the efficiency of fuel cell stack is calculated as follows,

$$\eta_{stack} = \mu_F \times \frac{V_{stack}}{N_{cells} \times v_{\max}} \tag{21}$$

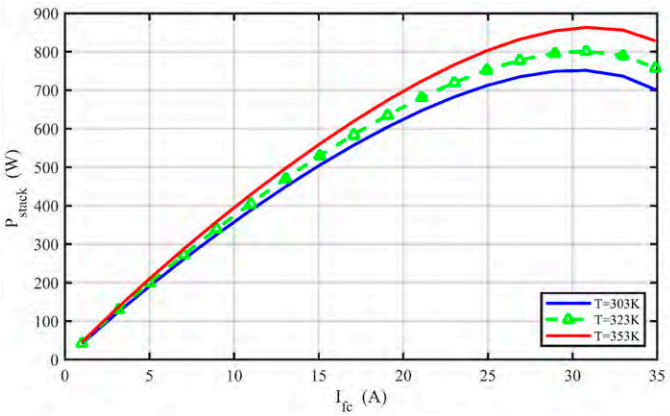
This equation is an approximation form for the exact efficiency called the voltage efficiency. v_{max} denotes the maximum value of the output voltage generated from the PEMFC under hydrogen higher heating values, which equals to 1.48 V/cell. μ_F is the utilization factor. It is assumed that the flow rate of hydrogen is controlled depending on the load condition, which leads to a constant utilization factor that equals to 95%.

Under the optimal values of the unknown parameters of the PEMFC stacks, the polarization curves (*I-V* and *I-P* characteristics) under different cell temperatures, while keeping the partial pressures of the reactants (P_{H2}/P_{O2}) constant at the values given in datasheets, are produced to validate the effectiveness of the developed TGA-based model. To avoid repeated figures, only two of the four proposed stacks are demonstrated in this section. The *I-V*, *I-P*, and efficiency of the SR-12 PEM 500W stack at 303, 323, and 353K are described in Fig. 7(a) – Fig. 7(c), respectively. Fig. 8(a) – Fig. 8(c) show the *I-V*, *I-P* polarization characteristics as well as the efficiency of 250W FC stack at 323, 353, and 383K, respectively.

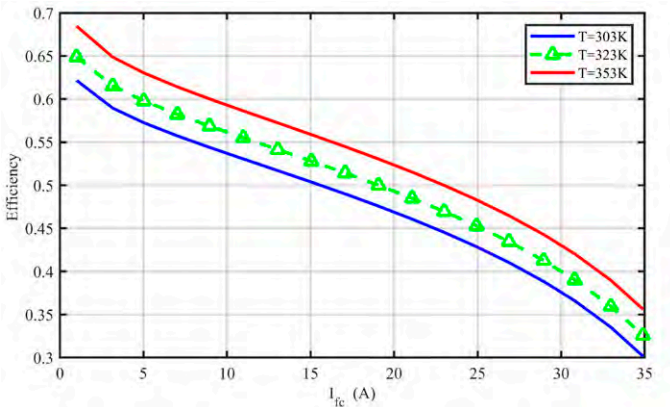
The impact of changing the pressures of the reactants in the inlet channels (P_{H2}/P_{O2}) under constant cell temperature, described in the datasheet of the manufactures, is introduced. Fig. 9(a) – Fig. 9(c) and Fig. 10(a) – Fig. 10(c) depicted the *I-V, I-P*, and efficiency of the BCS 500W PEM stack and Temasek 1kW FC stack at pressures of (1/0.2075bar), (1.5/1bar) and (2.5/1.5bar), respectively.



(a)

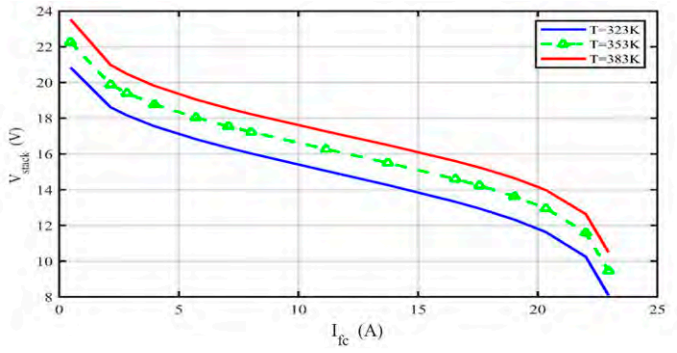


(b)

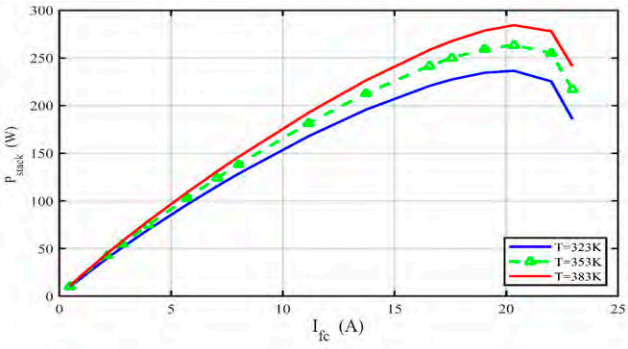


(c)

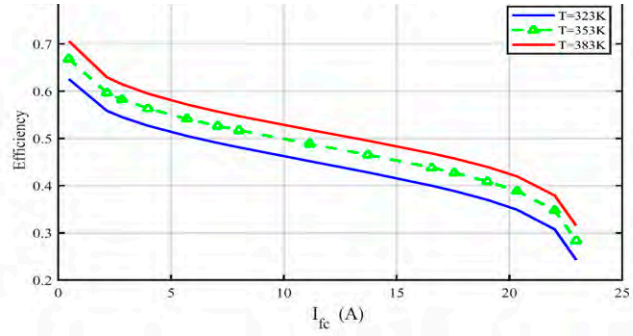
Fig. 7. Performance of SR-12 500W PEMFC stack under temperature variation: (a) *I-V* polarization curves, (b) *I-P* polarization curves, (c) efficiency.



(a)

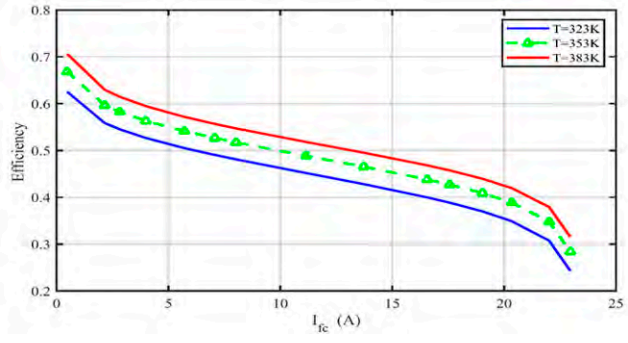


(b)

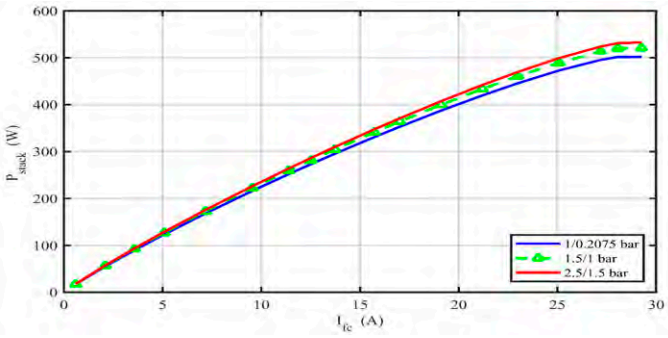


(c)

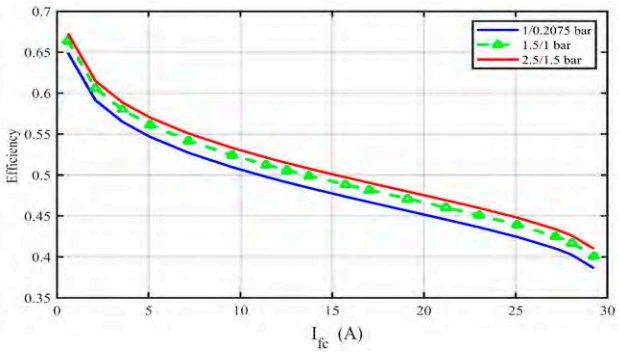
Fig. 8. Performance of 250W PEMFC stack under temperature variation: (a) *I-V* polarization curves, (b) *I-P* polarization curves, (c) efficiency.



(a)

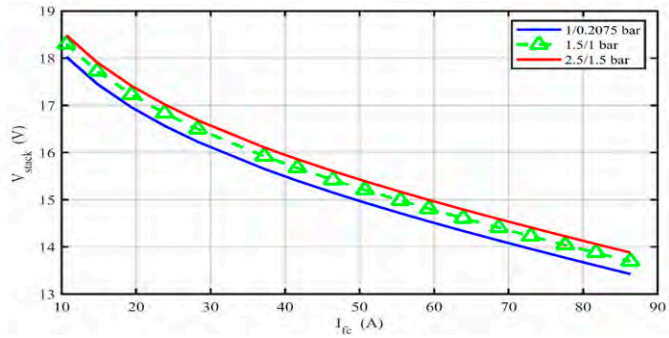


(b)

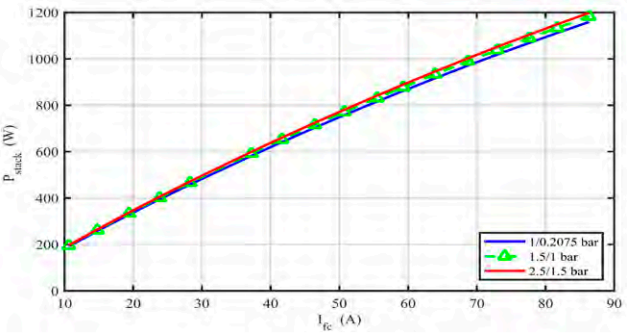


(c)

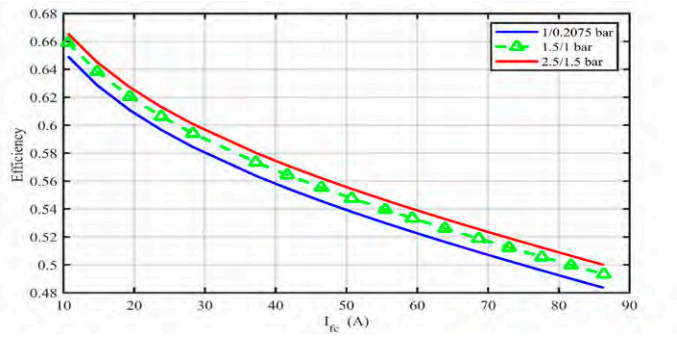
Fig. 9. Performance of BCS 500W PEMFC stack under varying supply pressures: (a) *I-V* polarization curves, (b) *I-P* polarization curves, (c) efficiency.



(a)



(b)



(c)

Fig. 10. Performance of Temasek 1kW PEMFC stack under varying supply pressures: (a) *I-V* polarization curves, (b) *I-P* polarization curves, (c) efficiency.

V. Performance and Statistical Measures

To validate the accuracy of the developed TGA-based model for parameter identification of the PEMFC stacks, statistical analysis of the minimum values of the TSD over 30 individual runs is demonstrated. Fig. 11 summarizes the change of the optimal value of the fitness function over the 30 executions for the four types of PEMFC stacks under consideration. It is evidently shown from the figure that the results proved more convergence in the case of 250 W PEMFC stack and Temasek 1kW stack, while a poor agreement between the final values of the individual runs is obtained in the case of the other two types.

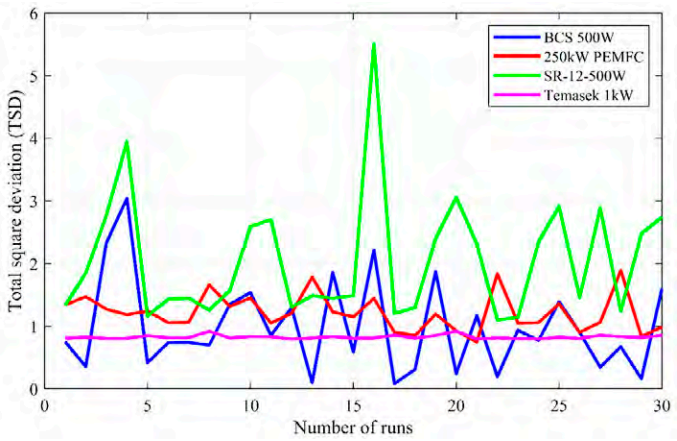


Fig. 11. Variation of the final value of the objective function (TSD) over 30 individual runs.

In this section, a deep statistical analysis has been demonstrated to give a clear assessment of the developed algorithm. In addition, a sensitivity analysis is provided as a measure of the stability of the optimization algorithm proposed in this study. The comparisons between the different PEMFC stacks are based on many metrics, mainly minimum and maximum values of TSD, mean value of TSD, Median.

In addition to the previously mentioned metrics, standard deviation (SD), Relative error of the objective function (RE), Root mean square error (RMSE), Mean absolute error (MAE) and efficiency have been provided to examine the accurateness of the developed TGA-based simulation model, which are arithmetically calculated based on (22) to (26), respectively.

$$SD = \sqrt{\frac{\sum_{i=1}^{30}\left(TSD_i - \overline{TSD}\right)}{30-1}} \tag{22}$$

$$RE = \frac{\sum_{i=1}^{30}\left(TSD_i - TSD_{\min}\right)}{TSD_{\min}} \tag{23}$$

$$MAE = \frac{\sum_{i=1}^{30}\left(TSD_i - TSD_{\min}\right)}{30} \tag{24}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{30}\left(TSD_i - TSD_{\min}\right)^2}{30}} \tag{25}$$

$$efficiency = \frac{TSD_{\min}}{TSD_i} \times 100\% \tag{26}$$

where, *TSDi* refers to the fitness function at each run. TSD_{min} denotes the minimum fitness function obtained over the 30 executions. $\overline{TSD}$ denotes the average value of the observed *TSD* over the simulation period. The summary of the studied metrics for the different PEMFC stacks is depicted in Table VII. It can be observed that the insignificant values of MAE and RMSE proved a well matching between the calculated values based on the estimated parameters and the measured ones. The values given in Table VII introduces a clear explanation of Fig. 11, in which the values of TSD in the case of Temasek 1kW stack over the 30 runs are changing in a narrow range.

TABLE VII. Statistical Results of TGA with Different PEMFC Stacks

	250 W stack	SR-12 500w stack	BCS 500W stack	Temasek 1kW stack
Min	0.74960694	1.1040851912	0.0835249987	0.796926618
Max	1.89347288	5.5041090089	3.0380153867	0.924600759
Mean	1.22002412	2.0635224515	0.9858239071	0.829441968
Median	1.19146585	1.5314876479	0.7636287517	0.816662317
SD	29.7779324	99.239232136	73.889767032	3.165014838
RE	18.8265535	26.069652993	324.08222277	1.224028021
MAE	0.47041717	0.959437260	0.9022989084	0.032515350
RMSE	0.55408401	1.368405748	0.9022989084	0.045006542
Eff.	64.8641850	63.44063220	19.0859501	96.2071929

VI. Conclusions

Extracting the values of seven unknown parameters of the PEMFC model is one of the most challenging points that attract the attention of many researchers. An effective PEMFC model based on TGA has been proposed in this paper, which is considered a suitable tool for simulation and performance evaluation of the PEMFC stacks under a wide range of operating scenarios. Many case studies have been performed, from which the estimated data, based on the optimal values of the unknown parameters, provide a good matching with the experimental data of different commercial fuel cell stacks. The results obtained from TGA-based model have been compared with different optimization methods. Different steady-state tests scenarios have been demonstrated to validate the effectiveness of the developed TGA-based technique. Moreover, statistical analysis has been conducted to measure the significance and precision of the optimal values obtained based on TGA method. Simulation results as well as statistical measurements emphasizes the superiority of the TGA over many optimization

algorithms in extracting the parameters of proton exchange membrane fuel cells. In the future studies, the developed algorithm can be applied for simulating the dynamic behavior of PEMFC and solid oxide fuel cell (SOFC).

Acknowledgment

The authors gratefully acknowledge the contribution of the NSFC (China)-ASRT (Egypt) Joint Research Fund, Project No. 51861145406 for providing partial research funding to the work reported in this research.

References

[1] A. H. Elkasem, S. Kamel, A. Rashad, and F. J. Melguizo, "Optimal Performance of Doubly Fed Induction Generator Wind Farm Using Multi-Objective Genetic Algorithm," *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 5, no. 5, pp. 48-53, 2019.

[2] Y. Ibrahim, S. Kamel, A. Rashad, L. Nasrat, and F. Jurado, "Performance Enhancement of Wind Farms Using Tuned SSSC Based on Artificial Neural Network," *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 5, no. 7, pp. 118-124, 2019.

[3] W. Ahme, A. Selim, S. Kamel, J. Yu, and F. J. Melguizo, "Probabilistic Load Flow Solution Considering Optimal Allocation of SVC in Radial Distribution System," *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 5, no. 3, pp. 152-161, 2018.

[4] S. Kamel and H. Youssef, "Voltage Stability Enhancement Based on Optimal Allocation of Shunt Compensation Devices Using Lightning Attachment Procedure Optimization," *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 5, no. 7, pp. 125-134, 2019.

[5] A. Selim, S. Kamel, L. Nasrat, and F. Jurado, "Voltage Stability Assessment of Radial Distribution Systems Including Optimal Allocation of Distributed Generators," *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 6, no. 1, pp. 32-40, 2020.

[6] J. Larminie, A. Dicks, and M. S. McDonald, *Fuel cell systems explained*. J. Wiley Chichester, UK, 2003.

[7] T. Manikandan and S. Ramalingam, "A review of optimization algorithms for the modeling of proton exchange membrane fuel cell," *Journal of Renewable and Sustainable Energy,* vol. 8, no. 3, p. 034301, 2016.

[8] A. Shamel and N. Ghadimi, "Hybrid PSOTVAC/BFA technique for tuning of robust PID controller of fuel cell voltage," *Indian Journal of Chemical Technology*, vol. 23, no.3, pp. 171-178, 2016.

[9] H.-W. Wu, "A review of recent development: Transport and performance modeling of PEM fuel cells," *Applied Energy,* vol. 165, pp. 81-106, 2016.

[10] J. C. Amphlett, R. F. Mann, B. A. Peppley, P. R. Roberge, and A. Rodrigues, "A model predicting transient responses of proton exchange membrane fuel cells," *Journal of Power sources,* vol. 61, no. 1-2, pp. 183-188, 1996.

[11] S. V. Puranik, A. Keyhani, and F. Khorrami, "Neural network modeling of proton exchange membrane fuel cell," *IEEE Transactions on Energy Conversion,* vol. 25, no. 2, pp. 474-483, 2010.

[12] C. E. Damian-Ascencio, A. Saldaña-Robles, A. Hernandez-Guerrero, and S. Cano-Andrade, "Numerical modeling of a proton exchange membrane fuel cell with tree-like flow field channels based on an entropy generation analysis," *Energy,* vol. 133, pp. 306-316, 2017.

[13] A. Kheirandish, F. Motlagh, N. Shafiabady, M. Dahari, and A. K. A. Wahab, "Dynamic fuzzy cognitive network approach for modelling and control of PEM fuel cell for power electric bicycle system," *Applied energy,* vol. 202, pp. 20-31, 2017.

[14] Y. Li, J. Yang, and J. Song, "Structure models and nano energy system design for proton exchange membrane fuel cells in electric energy vehicles," *Renewable and Sustainable Energy Reviews,* vol. 67, pp. 160-172, 2017.

[15] G. Zhang, L. Fan, J. Sun, and K. Jiao, "A 3D model of PEMFC considering detailed multiphase flow and anisotropic transport properties," *International Journal of Heat and Mass Transfer,* vol. 115, pp. 714-724, 2017.

[16] K. Priya, K. Sathishkumar, and N. Rajasekar, "A comprehensive review on parameter estimation techniques for Proton Exchange Membrane fuel cell modelling," *Renewable and Sustainable Energy Reviews,* vol. 93, pp. 121-144, 2018.

[17] L. Zhang and N. Wang, "Application of coRNA-GA based RBF-NN to model proton exchange membrane fuel cells," *International Journal of Hydrogen Energy,* vol. 43, no. 1, pp. 329-340, 2018.

[18] F. Bader, E.-M. Schön, and J. Thomaschewski, "Heuristics Considering UX and Quality Criteria for Heuristics," *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 4, no. 6, pp. 48-53, 2017.

[19] S. Arora and S. Singh, "An Effective Hybrid Butterfly Optimization Algorithm with Artificial Bee Colony for Numerical Optimization," *International Journal of Interactive Multimedia and Artificial Intelligence,* vol. 4, no. 4, pp. 14-21, 2017.

[20] Z. J. Mo, X. J. Zhu, L. Y. Wei, and G. Y. Cao, "Parameter optimization for a PEMFC model with a hybrid genetic algorithm," *International Journal of Energy Research,* vol. 30, no. 8, pp. 585-597, 2006.

[21] M. Ye, X. Wang, and Y. Xu, "Parameter identification for proton exchange membrane fuel cell model using particle swarm optimization," *International journal of hydrogen energy,* vol. 34, no. 2, pp. 981-989, 2009.

[22] W. Gong and Z. Cai, "Accelerating parameter identification of proton exchange membrane fuel cell model with ranking-based differential evolution," *Energy,* vol. 59, pp. 356-364, 2013.

[23] Z. Sun, N. Wang, Y. Bi, and D. Srinivasan, "Parameter identification of PEMFC model based on hybrid adaptive differential evolution algorithm," *Energy,* vol. 90, pp. 1334-1341, 2015.

[24] A. Askarzadeh and A. Rezazadeh, "A grouping-based global harmony search algorithm for modeling of proton exchange membrane fuel cell," *International Journal of Hydrogen Energy,* vol. 36, no. 8, pp. 5047-5053, 2011.

[25] C. Dai, W. Chen, Z. Cheng, Q. Li, Z. Jiang, and J. Jia, "Seeker optimization algorithm for global optimization: a case study on optimal modelling of proton exchange membrane fuel cell (PEMFC)," *International Journal of Electrical Power and Energy Systems,* vol. 33, no. 3, pp. 369-376, 2011.

[26] A. Fathy and H. Rezk, "Multi-verse optimizer for identifying the optimal parameters of PEMFC model," *Energy,* vol. 143, pp. 634-644, 2018.

[27] L. Zhang and N. Wang, "An adaptive RNA genetic algorithm for modeling of proton exchange membrane fuel cells," *International Journal of Hydrogen Energy,* vol. 38, no. 1, pp. 219-228, 2013.

[28] S. Xu, Y. Wang, and Z. Wang, "Parameter estimation of proton exchange membrane fuel cells using eagle strategy based on JAYA algorithm and Nelder-Mead simplex method," *Energy,* vol. 173, pp. 457-467, 2019.

[29] M. Ali, M. El-Hameed, and M. Farahat, "Effective parameters' identification for polymer electrolyte membrane fuel cell models using grey wolf optimizer," *Renewable energy,* vol. 111, pp. 455-462, 2017.

[30] O. E. Turgut and M. T. Coban, "Optimal proton exchange membrane fuel cell modelling based on hybrid Teaching Learning Based Optimization–Differential Evolution algorithm," *Ain Shams Engineering Journal,* vol. 7, no. 1, pp. 347-360, 2016.

[31] Y. Chen and N. Wang, "Cuckoo search algorithm with explosion operator for modeling proton exchange membrane fuel cells," *International Journal of Hydrogen Energy,* vol. 44, no. 5, pp. 3075-3087, 2019.

[32] M. Guarnieri, E. Negro, V. Di Noto, and P. Alotto, "A selective hybrid stochastic strategy for fuel-cell multi-parameter identification," *Journal of Power Sources,* vol. 332, pp. 249-264, 2016.

[33] A. Askarzadeh and A. Rezazadeh, "A new heuristic optimization algorithm for modeling of proton exchange membrane fuel cell: bird mating optimizer," *International Journal of Energy Research,* vol. 37, no. 10, pp. 1196-1204, 2013.

[34] A. A. El-Fergany, "Electrical characterisation of proton exchange membrane fuel cells stack using grasshopper optimiser," *IET Renewable Power Generation,* vol. 12, no. 1, pp. 9-17, 2017.

[35] A. S. Menesy, H. M. Sultan, A. Selim, M. G. Ashmawy, and S. Kamel, "Developing and Applying Chaotic Harris Hawks Optimization Technique for Extracting Parameters of Several Proton Exchange Membrane Fuel Cell Stacks," *IEEE Access,* vol. 8, no. 1, pp. 1146-1159, 2020.

[36] A. S. Menesy, H. M. Sultan, A. Korashy, M. G. Ashmawy, and S. Kamel, "Effective Parameter Extraction of Different Polymer Electrolyte Membrane Fuel Cell Stack Models Using a Modified Artificial Ecosystem Optimization Algorithm," *IEEE Access,* vol. 8, no. 1, pp. 31892 - 31909, 2020.

[37] A. Cheraghalipour and M. Hajiaghaei-Keshteli, "Tree growth algorithm (TGA): An effective metaheuristic algorithm inspired by trees behavior,"

in *13th International Conference on Industrial Engineering*, 2017, vol. 13: Scientific Information Databases Babolsar.

[38] I. Strumberger, M. Minovic, M. Tuba, and N. Bacanin, "Performance of Elephant Herding Optimization and Tree Growth Algorithm Adapted for Node Localization in Wireless Sensor Networks," *Sensors,* vol. 19, no. 11, p. 2515, 2019.

[39] H. H. EL-Tamaly, H. M. Sultan, and M. Azzam, "Control and Operation Of a Solid Oxide Fuel-Cell Power Plant In an Isolated System," in 9 th International Conference on Electrical Engineering (9 thICEENG), Military Technical College Kobry El-Kobbah, Cairo, Egypt, 2014, pp. 1-13.

[40] C. Panos, K. Kouramas, M. Georgiadis, and E. Pistikopoulos, "Modelling and explicit model predictive control for PEM fuel cell systems," *Chemical Engineering Science,* vol. 67, no. 1, pp. 15-25, 2012.

[41] Y. Rao, Z. Shao, A. H. Ahangarnejad, E. Gholamalizadeh, and B. Sobhani, "Shark Smell Optimizer applied to identify the optimal parameters of the proton exchange membrane fuel cell model," *Energy conversion and management,* vol. 182, pp. 1-8, 2019.

[42] J. Too, A. Abdullah, N. Mohd Saad, and N. Mohd Ali, "Feature Selection Based on Binary Tree Growth Algorithm for the Classification of Myoelectric Signals," *Machines,* vol. 6, no. 4, p. 65, 2018.

[43] J. M. Corrêa, F. A. Farret, L. N. Canha, and M. G. Simoes, "An electrochemical-based fuel-cell model suitable for electrical engineering automation approach," *IEEE Transactions on industrial electronics,* vol. 51, no. 5, pp. 1103-1112, 2004.

Hamdy M. Sultan

Hamdy M. Sultan received a B.Sc. of Electrical Power Engineering from Minia University, and M.Sc. in Electrical Engineering from Minia University, Minia, Egypt in 2014. He is an Assistant Lecturer in the Electrical Engineering Department, Minia University, Minia, Egypt. He is currently working toward the Ph.D. degree at National Research University (Moscow Power Engineering Institute), Moscow, Russia. The key research area includes optimization techniques, power system plan and operation, power system transient stability and renewable energy.

Ahmed S. Menesy

Ahmed S. Menesy received the B.Sc. degree in Electrical Engineering from Minia University, Minia, Egypt, in 2014. He is a Teaching Assistant in the Electrical Engineering Department, Minia University, Minia, Egypt. He is currently working toward the M.Sc. degree at Chongqing University, Chongqing, China. His research interests include optimization techniques, power system analysis, high voltage external insulation, renewable energy, and smart grids.

Salah Kamel

Salah Kamel received the international PhD degree from University of Jaen, Spain (Main) and Aalborg University, Denmark (Host) in Jan. 2014. He is an Associate Professor in Electrical Engineering Department, Aswan University. Also, he is a leader for power systems research group in the Advanced Power Systems Research Laboratory (APSR Lab), Aswan, Egypt. He is currently a Postdoctoral Research Fellow in State Key Laboratory of Power Transmission Equipment and System Security and New Technology, School of Electrical Engineering, Chongqing University, Chongqing, China. His research activities include power system modeling, analysis and simulation, and applications of power electronics to power systems and power quality.

Francisco Jurado

Francisco Jurado obtained the MSc and PhD degrees from the UNED, Madrid, Spain, in 1995 and 1999 respectively. He is Full Professor at the Department of Electrical Engineering of the University of Jaén, Spain. His research activities have focused on two topics: power systems and renewable energy.

Time-Dependent Performance Prediction System for Early Insight in Learning Trends

Carlos J. Villagrá-Arnedo*, Francisco J. Gallego-Durán*, Faraón Llorens-Largo, Rosana Satorre-Cuerda, Patricia Compañ-Rosique, Rafael Molina-Carmona

University of Alicante, Alicante (Spain)

Received 3 April 2020 | Accepted 20 May 2020 | Published 27 May 2020

Abstract

Performance prediction systems allow knowing the learning status of students during a term and produce estimations on future status, what is invaluable information for teachers. The majority of current systems statically classify students once in time and show results in simple visual modes. This paper presents an innovative system with progressive, time-dependent and probabilistic performance predictions. The system produces by-weekly probabilistic classifications of students in three groups: high, medium or low performance. The system is empirically tested and data is gathered, analysed and presented. Predictions are shown as point graphs over time, along with calculated learning trends. Summary blocks are with latest predictions and trends are also provided for teacher efficiency. Moreover, some methods for selecting best moments for teacher intervention are derived from predictions. Evidence gathered shows potential to give teachers insights on students' learning trends, early diagnose learning status and selecting best moment for intervention.

Keywords

E-learning, Education, Learning Analytics, Learning Management Systems, Prediction, Support Vector Machine.

DOI: 10.9781/ijimai.2020.05.006

I. Introduction

Knowing students' learning trends is relevant to diagnose learning performance and early detect situations where teachers' intervention would be most effective. Prediction systems represent one of the bests tools for this purpose. Predicting performance is the basis for student diagnostics, learning trends projection and early detection.

Most performance prediction systems output numerical grades or performance class memberships. Research tends to focus on prediction accuracy. Accuracy is relevant, because it helps improving diagnostics, but it should not be confused with the main goal: improving learning. To help teachers improve student performance many other aspects can be considered: more accessible prediction data, better graphical representations, methods for detecting learning trends and most suitable moments for intervention, etc. Most of these improvements rely on the ability to consider learning data evolution over time. This is particularly relevant due to cumulative nature of learning and so it is one of the main characteristics considered in this work.

This work is an empirical research in the search for practical systems to help teachers in their guidance duties. It relays on teachers receiving in-depth information on student learning trends during semester. This information is elaborated from an automatic system which yields predictions on expected student performance. Main contribution of this work is a custom-designed practical prediction system. Main innovations of the proposed system are its time-dependent nature and the use of probabilistic predictions. The proposed system delivers by-weekly probabilistic performance predictions and analytical time-dependent graphs that help gaining insight in students' learning trends. The proposed system is tested during a complete semester in the subject *Mathematics I* at the University of Alicante. Data gathered is used as initial evidence to empirically test the system and results are shown and discussed. Usefulness, convenience and advantages of the time-dependent nature of learning data are also tested and discussed. As an additional consequence derived from these tests, some initial methods for selecting the best moments for teacher intervention are proposed and discussed.

Performance predictions are shown as point graphs over time, along with calculated trends. This information is summarized and organized to help teachers explore and analyse student learning performance efficiently. Some case examples are presented and analysed using these graphs, showing their potential to help teachers understand beyond raw data. Teachers can use this information to diagnose students, understand learning trends, early detect intervention situations and act accordingly to help students improve their learning results. This research considers only learning trend diagnosis and detection of most suitable moments for teacher intervention. Intervention strategies and their results are out of scope.

This paper is structured in seven sections. Section II analyses some relevant background works. First, several reviews which describe the most appropriate techniques in prediction are presented. Then, some related works on early detection and on providing insightful, graphical representations are explained. Lastly, a discussion drawing conclusions of this review is performed. As a result, research questions are proposed in section III. A custom automated learning system, in which the proposed prediction system is included, is presented in section IV. Section V explains how data from the system is used to perform student diagnosis and to select the best intervention moment. Section VI analyses some

* Corresponding author.

E-mail addresses: fjgallego@ua.es (F. J. Gallego-Durán), villagra@ua.es (C. J. Villagrá-Arnedo).

paradigmatic student case examples, showing how prediction graphs and calculated trends help understanding student learning trends. Finally, section VII covers conclusions and further work.

II. Related Work

A. Prediction Techniques

Several prediction systems focused on student academic performance have been developed in recent years. Hellas et al. [1] perform a great survey on prediction techniques, predicted factors and prediction methods. Authors find that most predicted values are course grades and individual exam grades. Most studies used statistical correlations and regression, followed by machine learning techniques such us Decision Trees and Naive Bayes classifiers and clustering. Hämäläinen and Vinni [2] carry out a comprehensive study about classification methods in the discipline of Educational Data Mining. They organize predictive classifiers in education into four groups depending on the aim of the prediction: academic success, course outcomes, success in the next task and meta-cognitive skills, habits and motivation. They conclude that the main concerns are the choice of a discriminative or probabilistic classifier, the estimation of the real accuracy, the tradeoff between overfitting and underfitting and the impact of data preprocessing. According to this work [2], the most used classification techniques are Decision Trees, Bayesian Networks, Neural Networks and Support Vector Machines, in this order.

Kotsiantis [3] also makes an interesting review of different techniques in Learning Analytics for educational purposes (classification and regression algorithms, association rules, sequential patterns analysis, clustering and web mining). Kotsiantis indicates that the use of Machine Learning techniques is an emerging field that aims to develop educational methods of data exploration and meaningful patterns finding. He also notes that professionals tend to build a model once in time, not considering data evolution over time, and that the general trend focuses on predicting students' final grades (i.e. learning performance).

Prediction accuracy is the main concern of most works. [4] predict academic success of students as low, medium and high risk. They use two data mining techniques: Decision Trees and Neural Networks. After analytically comparing various techniques, [5] achieved high precision results in student performance, using Decision Trees and ranking students as fail, pass, good or very good. [6] also use Decision Trees to predict students' dropout, achieving comparable precision to more sophisticated techniques. However, it is important to consider that they perform flat classifications, not accounting for probabilistic belongings to classes. Hamound et al. [7] compare tree classifiers that try to predict student's success from questionnaires regarding their social activity, health, relationships and academic performance. They find the J48 algorithm to give better performance results than Random Tree and RepTree. In another work [8], authors compare four Machine Learning techniques to predict student performance. Their research compares quality of predictions based on two features: average precision and percentage of accurate predictions. They conclude that the simplest linear regression model is enough to predict average academic performance on groups of students, whereas individual performance is best predicted using Support Vector Machines (SVM). Importantly, these authors [8] train their predictive models once with static past data: they do not take into account data evolution over time.

B. Early Detection

Other works stress the importance of how system outputs are shown. They consider it highly relevant for teacher understanding and their improved ability to help in the learning process. [9] attempts to predict students' dropout or failure as earliest as possible. They use two pairs of descriptive-predictive techniques to achieve 80% accuracy: 1) Correlations / Linear Regression and 2) Association Rules / Bayes Model. They conclude that these techniques can help teachers understand and interpret course progress on two levels: 1) the whole group of students or 2) individually. [10] cluster students in three ways: 1) in nine classes by ranges of marks, 2) classified in high, medium or low performance, and 3) classified in pass or fail. They find that accuracy of their predictions improves when they use Genetic Algorithms. [11] also designs two partitions: 1) per mark as fail, pass, good or excellent, and 2) classified in underperforming, medium or high. They use student interaction data from Moodle and final marks. They combine different Data Mining techniques (Statistical Classification, Decision Trees, Neural Networks and Induction of Rules). They conclude that a classifier must be not only accurate, but also understandable by trainers to be useful as a guide in learning.

Early detecting learning performance issues is one of the most relevant goals in this field. Main intention is helping teachers to guide students towards academic success. Freund et al. [12] present a prototype of a performance prediction system, combining classification techniques based on Decision Trees, which achieves an accuracy close to 98%. It consists of a set of decision rules that automatically detect at risk students and trigger alerts based on most significant variables. Alerts materialize into emails sent to both student and teacher. In [13] authors propose using students' online activity data in a web-based Learning Management System. The system provides an early indicator of predicted academic performance and results of a test assessing student motivation for the online course. They also try to help at risk students by providing information on students who successfully finished the course and links to assess their willingness to virtual classes. Similarly, in [14], authors propose Feed-Forward Neural Networks to predict final marks of students in an e-Learning course. They use predictions to classify students into two performance groups. Their results show that accurate predictions are viable at early stages (in their case, in the third week). However, the proposed system failed predicting certain specific students. This is expectable with early predictions, but also opens up discussion on the convenience of intervention based on predictions at early stages. They conclude that their proposal can help teachers assist students in a more personalized manner.

[15] present a final marks prediction system. They argue that most previous works perform predictions after their corresponding courses, which neglects the possibility of early predictions and detecting at risk students amid lessons. They gather activity data in a Learning Management System during three different periods: weeks four, eight and thirteen. They use three classification techniques based on Decision Trees, obtaining an overall accuracy of 95% at week four. This is one of the few works that consider data evolution over time, but it does it with quite coarse granularity (only 3 big course periods). However, their results are quite relevant and an improvement of their work with more focus on enabling teacher diagnosis through appropriate data presentation (maybe using carefully designed graphs) would have breakthrough potential. Following a similar path, Akçapınar et al. [16] develop an early prediction system using student's eBook reading data to detect at-risk students. Their system uses 13 prediction algorithms using data from different weeks of the course. They obtained best performance using Random Forests and Naive Bayes and analysed different details on raw data versus elaborated features.

C. Graphical Representation

Graphically representing system outcomes has the potential to help teachers better understand student learning trends. It also may help them early diagnose students, detect at risk scenarios and relevant time-frames for intervention. However, not many works focus on the

importance of graphical representations with respect to predictions. Most of them simply present predictions as raw values.

Some works use the graphing tools that come embedded in their learning platforms. [17] present Learning Analytics Enriched Rubric (LAe-R), a new cloud based assessment tool integrated into Moodle. They use GISMO, a visualization tool for Moodle that gathers and processes log data to produce graphical representations that can be used by teachers for assessing students' performance. They conclude highlighting the importance of data visualization for teachers and propose future work on this matter. [18] also graphically show prediction accuracy through Receiver Operating Characteristic (ROC) curves. [19] analyse accuracy on early identification of students who are at risk of not completing Massive Online Open Courses (MOOC) courses. They compare four weekly prediction models in terms of Area Under Curve (AUC), and graphically visualize student learning trends.

[20] have the goal to enhance Reactive Blended Learning with a control system including prediction features. Authors remark that their work is not focused on obtaining a complete student model, but on improving student diagnosis to help teachers act on low performance risks. This drives them to show results in learning evolution graphs during their courses. They also compare traditional methods with their approach for two consecutive years with interesting results.

D. Discussion About Background

Analysis undertaken in this work yields the following conclusions about performance prediction systems:

- Most of the works focus on prediction, specially on accuracy. Many different algorithms, methods, data types, and data sources are used. Many works also perform algorithm comparison, almost always using accuracy as measure. There seems to be no consensus on which algorithms, methods or data sources are better. However, some algorithms seem to give good results in general, including Decision Trees, Random Forests and Support Vector Machines. Although more research in this area is clearly justified, there seems to be too much emphasis on accuracy, sometimes forgetting that students and learning should be the major goal.
- System predictions are mostly plain classifications, with very few works modeling uncertainty and/or probability, and even fewer considering evolution over time of data and predictions. None of the works analysed considered everything at once. Much more research on producing progressive and probabilistic predictions, and analysing predicted learning trends is expected to follow.
- There is definitely no common way of representing predictions. A great majority of works output predictions as raw numbers or similar. Some give several values, probabilities or classes. Only a few works give importance to visual representation and its key role on teacher understanding and student diagnosis. More work on this matter is encouraged, as powerful representations may constitute the basis for actual improvements on the teaching-learning process.

III. Research Questions

After reviewing works in section II, questions arise about the static and punctual nature of performance predictions in most of them:

1. Are there benefits on exploiting time-dependent nature of learning data, by yielding frequent students' performance predictions over time?
2. Could frequent time-dependent predictions help teachers deduce students' performance trends?
3. Could this give early insights in student learning trends?
4. Could these deduced trends help teachers identify most relevant time-spots in the learning process?
5. Could this information be used to detect best moments for teacher intervention?

It seems plausible that consecutive, frequent predictions over time could yield additional information. As an example, let us compare a punctual performance prediction to a picture. Depending on the circumstances, it could be deduced that a person is running. A set of consecutive and frequent pictures would probably make it evident, also yielding information on distance, velocity, running technique... The idea behind this work is equivalent: from a graph of consecutive, frequent predictions, additional information could be deduced about student learning trends. Concretely, performance trends, better estimations on future performance and most relevant time-spots in the learning process.

This work will address these five research questions from an empirical point of view. The performance prediction system presented in next section will be tested within a semester and results will be analysed. Data gathered along with student case analyses will be presented as initial evidence related to these research questions. Authors aim is to present this initial evidence results to show that the proposed system is promising in this field and to encourage following studies to gather more evidence.

IV. Automated Learning and Performance Prediction System

The main contribution of this work is an improved insight in learning trends provided by frequent, consecutive performance predictions. This additional information has the potential to help teachers to early diagnose student issues and schedule interventions. To achieve this result, the first step is gathering student data to make predictions. In this work, we use a custom learning system both for student assessing and for data gathering. This section describes this custom web-based system which was initially developed to automate processing of student learning activities. The system was also designed with data gathering in mind, to help understanding students learning progress. The system supports *Mathematics I*, a first-year subject in Computer Science Engineering and Multimedia Engineering degrees at the University of Alicante. *Mathematics I* introduces students into Computational Logic and Logic Programming through Prolog programming language.

The automated learning system consists of four main components:

- *PLMan*[1], a Pacman-like, custom-developed videogame which is the students' central activity.
- A custom web-based automated learning system which manages student homework, assessment and progress, and lets teachers supervise the process.
- A performance prediction system based on Support Vector Machines (SVM) which classifies students every week according to their expected performance.
- A representation module which graphically shows predictions, current status and future trends about students.

Next subsections describe each one of these system components in greater detail.

A. Learning Activity: PLMan, the Game

PLMan is a cross-platform, text-mode, Pacman-like videogame implemented in Prolog programming language. It was created to support the learning of Prolog programming, Computational Logic and Reasoning. As it is part of the context of this work, this section briefly introduces the game. PLMan is described in depth in [21].

[1] PLMan can be downloaded from https://rua.ua.es/dspace/handle/10045/103447. Accessed: Mar, 30th 2020.

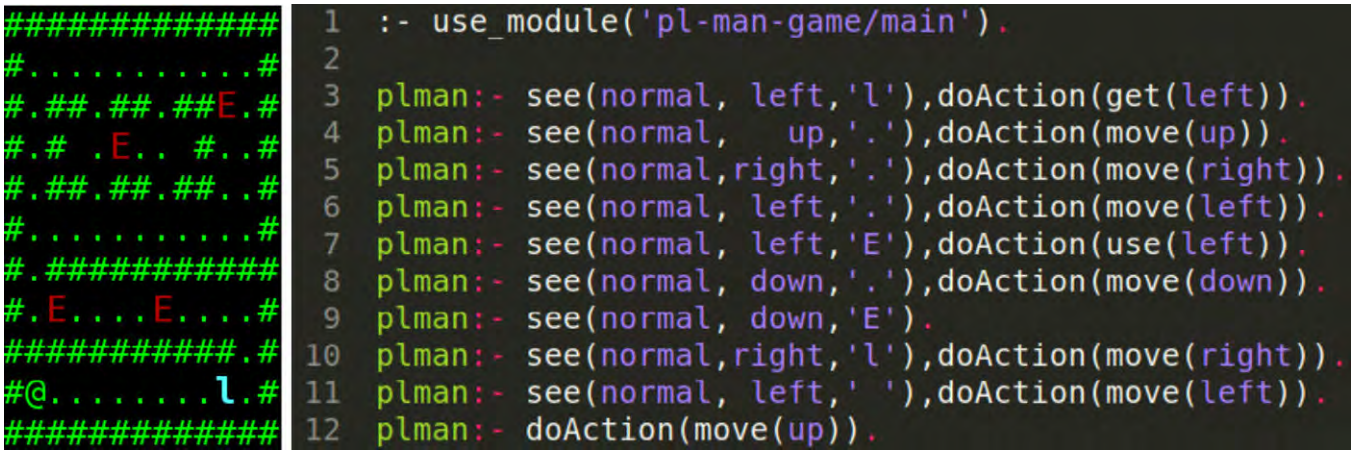


Fig. 1. Left. An example *PLMan* maze with walls (**#**), dots (**.**), enemies (**E**), a gun (**l**) and Mr. PLMan (**@**). *Right*. A Prolog AI code that solves the maze.

Students program the Artificial Intelligence (AI) of Mr. PLMan, a Pacman-like character, in Prolog. The goal is to make Mr. PLMan eat all the dots in a give maze (see Fig. 1). The game works like a simulator: unlimited different mazes can be created for *PLMan*. Each new maze becomes a different exercise for which students develop their AIs. Teacher-designed mazes are classified in increasing complexity to encourage students learn more about Prolog and be creative programming their AIs. These mazes are organized into four main stages and up to five levels of difficulty per stage.

Each AI program created by the students and aimed to solve a given PLMan maze is called a *solution*. The students send their solutions to a web-based system that evaluates them, based on the percentage of dots their solutions manage to eat when simulated. Each maze is worth different marks depending on its stage and difficulty. Students get cumulative marks for each solution sent, modulated by the percentage achieved. For instance, if maze A1 is worth 1 point, and student S1 sends a solution achieving 70%, student S1 will add 0,7 points to his/her cumulative marks. A solution achieving 75% or more unlocks the next maze for the student. Students have no limit on the number of times they can resend solutions, nor they are penalized. The system always considers the best solution sent, not the last. The only limits are stage deadlines and ten minutes delay between sent solutions.

Formative assessment has been considered as the basis to design this learning process: mistakes and partial progress are encouraged rather than penalized, to let students learn from their mistakes and evolve. Also, freeing students from fear to fail makes them more willing to participate, increasing motivation. Students also follow their own path by selecting difficulty levels that make them feel more comfortable. The greater the difficulty, the more the marks. They may also stop whenever they want. For instance, a student into the 3rd stage with 65 out of 100 marks accumulated may stop solving mazes and those will be his/her final marks (65%), or continue solving mazes to achieve better marks.

B. Web Site

Similar to *PLMan*, the web site is very briefly introduced in this section as context for this work. Complete details on the web system can be found in [22]. General behaviour of the web site is similar to many learning systems (like Moodle, for instance) but specifically adapted to the needs of the subject. For the purposes of this paper, there is no need to deepen into the details of this general part. The main contribution comes from the Progressive prediction system and representation modules, which are described in next subsections.

The web system is private and can only be accessed by students and teachers of *Mathematics I*. The public area lets anyone download the *PLMan* game and some utilities. Once students sign in to the private area they see their current profile, along with their progress and status (Fig. 2, left). Their status includes their accumulated marks, their assigned mazes and all details for each maze: completion percentage, acquired marks, total marks, download button, send solution button and results section. The results section (not shown in the figure) contains all the information about solution assessment: global results of execution, details on marks calculation, comparison rankings and execution logs that let students repeat exact executions that have been performed on the server.

For teachers there is an administration panel (Fig. 2, right). This panel lets them supervise the evolution of their students and groups of students. Teachers can explore all details of any given student: mazes assigned, solutions sent, results of the solutions, actions performed in the system, marks acquired, code from sent solutions, etc. They can also manage the basic parts of the course like group creation, student sign up, assignment and deadlines, system marks reviewing, etc.

C. Progressive Prediction System

The progressive prediction system is briefly described in this section, with emphasis on its progressive nature. Present description is aimed to give a general understanding of what the system does without including mathematical and computational details. Complete details can be found in [22].

The system general purpose is to predict final students' performance. For this purpose, the system collects all data from students' participation and solutions sent to PLMan mazes. Every week of the semester, the system uses up-to-date information and generates performance predictions for every student. These output predictions are comprised of three real numbers for each student. These numbers predict the probabilities for the student to end up the semester pertaining to one of the three student classes defined in Table I. For example, an output from the system like this

```
prediction = (studentA, 0.40, 0.35, 0.25)
```

would mean that *studentA* has a predicted probability of 40% to end up the semester in the *High performance* class, which means his/her marks would be in the range]80,5% - 100%]. Similarly, *studentA* has 35% predicted probability of ending up in *Medium performance* class (marks in]57.5% - 80.5%]), and a 25% predicted probability of ending up in *Low performance* class (marks in [0% - 57.5%]).

TABLE I. Designed Student Classes as Output for the SVM Classifier

Class	Expected final marks	Label
1	]80.5% - 100%]	High performance
2	]57.5% - 80.5%]	Medium performance
3	[0.0% - 57.5%]	Low performance

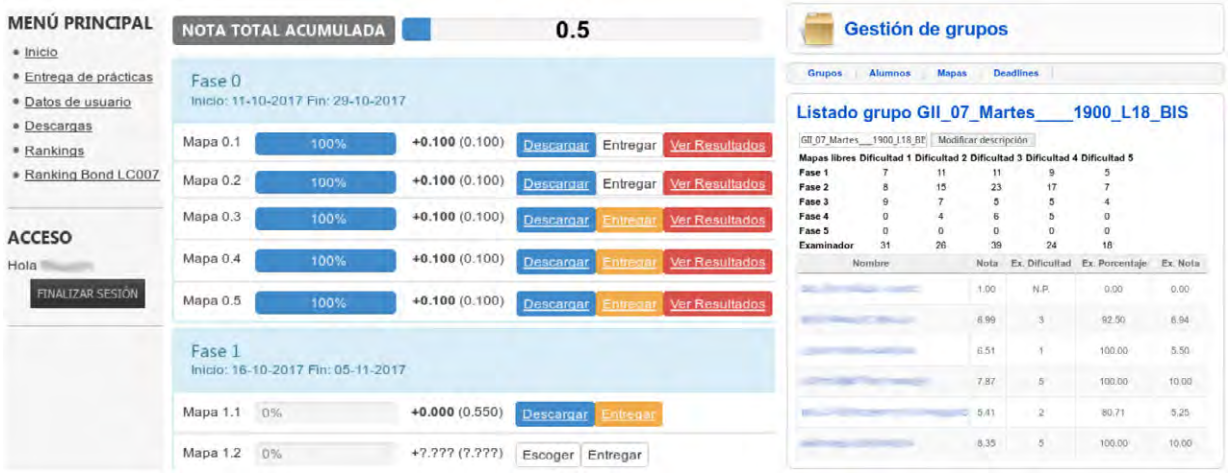


Fig. 2. *Left*. Student web interface showing stages 0 (complete) and 1 (incomplete). *Right*. Teachers web interface showing a group of students along with their marks.

The system is based on Support Vector Machines (SVM) [23] as Machine Learning model and low-level prediction technique. It is designed to predict student performance every week, using all past cumulative information. For instance, when predicting expected performance on week five, the first five weeks of input data are used, and not only data from week five. In this sense, predictions are cumulative and progressive. As a complete semester has eleven working weeks[2], eleven consecutive predictions are performed.

Although it would be much preferable to obtain predictions in the form of real-valued final marks, that practice results inviable for this study from the computational point of view. To obtain that kind of prediction within an acceptable error range the system would require input data in the order of the hundreds to miles of thousands of students. As results section shows, this work started with data from three hundred and thirty six students. Although it is a normal sized sample for two iterations of a first year semester, it is three to four orders of magnitude less than required for real-valued final marks as output prediction. Therefore, to obtain predictions within an acceptable error range, the system was designed as a classifier with the three classes presented in table I. This design decision ensures that the system will achieve a high probability of generalization in computational learning phases and minimal over/underfitting problems. This reasoning is similar to previous works found in literature, as many of them have samples of similar sizes.

The input information used for prediction comes directly from the interaction between the students and the system. This includes difficulties selected, mazes assigned, number of tries to solve a maze, time taken to develop solutions, etc. It also includes number of accesses to the website, time between accesses, downloads of mazes, time spent on different views of maze and solution information, etc. Data is collected, organized, normalized and finally input into one of eleven SVMs in the prediction system. Each SVM is specialized in predicting a final performance class for a specific week. For this purpose, each SVM uses the data corresponding to the previous **n** weeks. Detailed description on exact information used and input features constructed is specified in [22].

As discussed previously, each SVM outputs three prediction probabilities, one for each final performance class. The student is finally considered to pertain to the class with greatest probability. However, all probabilities are taken into account and given to teachers. This gives much more information than the single class the student is considered to pertain to. As a simple example, there are cases in which one class has 0.38 and next one has 0.375 probability. That means both may be almost equally probable, and this information is important to take into account when diagnosing a student. This information is given mainly in the form of graphs, but also numerically if requested.

Fig. 3 exemplifies what can be shown having predicted performance probabilities over time. This graph could not have been drawn if students where merely classified in high, medium and low predicted performance classes, and it shows valuable information on student trends over time.

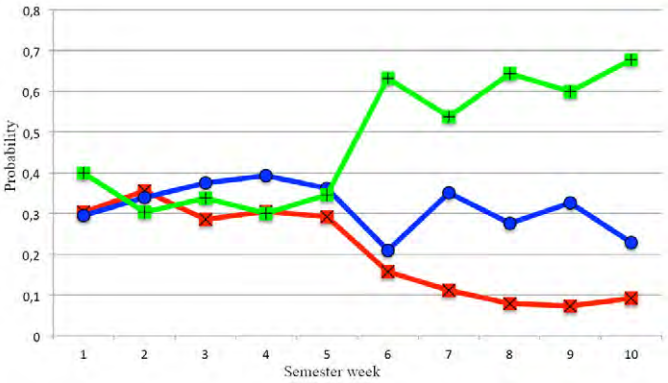


Fig. 3. Example graph with predicted performance probabilities over 10 weeks from a random student. Probabilities are: green) high performance, blue) medium performance, red) low performance.

In Fig. 3, predicted performance probabilities come from the same student. Therefore, a simple visual glance shows that the student had similar probabilities of ending up as high/medium/low performance up to week five. In week six, there is a great change and student is predicted to end up in the high performance range with ~0.6 probability. Predicted probabilities maintain this trend up to week ten and, finally, the student got 91% marks, which effectively enters in the high performance range, confirming that predictions were accurate in this case.

As shown, the design of the system takes into account time-dependent nature of data and predictions. Predictions are frequently made (every week), using cumulative data from previous weeks, and with probabilistic output. With all this information, progression graphs are created (see section VI). All these steps are done as a consequence from the first research question. Fig. 3 suggests that this additional

[2] The complete semester contained fifteen weeks, but two of them were required for introducing students, one was public holidays and the last one was required for exams.

information coming from probabilities, time-dependent predictions and cumulative data can be valuable to get more insights in student learning trends.

D. Representation Module

As many authors previously identified, it is highly relevant to find an appropriate graphical way to show system outputs to teachers. The proposed system, built on previous work [24], has a representation module able to show raw data to teachers as well as several graphs designed with evolution over time in mind.

Main visual outputs of the system are a set of point graphs and a control panel (see Fig. 5). Every student has three point-graphs with by-weekly predictions probabilities for high, medium and low performance groups, including all available information up to current week. For instance, leftmost point graph (green points) shows all system's probability predictions for the student ending up in high performance class at the end of the semester. Similarly, central and rightmost point-graphs show predictions for medium and low performance classes. Each point represents a single prediction made by the system and corresponds to a probability (y-axis) estimated at the week the prediction is made (x-axis). From all these predictions, a trend line is calculated by common linear regression and depicted dashed. This trend line visually shows if predicted probabilities are increasing or decreasing and how fast. In example Fig. 5, the student being analyzed is increasing (green) his/her probability of ending up the semester with high performance marks (]80.5%-100%]) and decreasing probabilities (grey/red) of ending up with medium or low performance marks.

The control panel on the right side of the graphs (Fig. 5) summarizes predictions for current week (seventh week on Fig. 5). There are three predictions in the form of value-arrow-color. Let us understand the value-color pair first, as it is most important. Value represents probability (0-1) and color identifies a performance group (green-high, grey-medium, red-low). So, 0.9-green means 0.9 probability of ending up in high performance group, whereas 0.9-red means 0.9 probability of ending up in low performance group. The arrow indicates the probability of increase/decrease. So an increasing green probability is a good sign (greater probability of high performance) whereas an increasing red probability is a bad sign (greater probability of low performance). The greater the increase/decrease velocity, the greater the angle for the arrow. Angles are discretized to nine possible values to simplify visual interpretation and comparison.

The representation module lets teachers study the evolution of individual students and groups. For individual students, several rows with point-graphs and control panels like those in Fig. 5 can be shown at once. This lets teacher select, analyse and study the evolution of any student with respect to system's performance predictions. Section VI shows three selected case studies from three model students to show how these analyses are performed. Section V shows most important group information that teachers use to select which students to analyse individually.

V. Diagnosing Students and Finding Time-spots for Intervention

Main research questions focus on diagnosing students, inducing and understanding learning trends and detecting relevant time-frames and spots for teacher intervention. Section IV has introduced the proposed performance prediction system and representation modules, which directly address student diagnosis. This section details the system processes and tools to help teachers perform diagnosis, and proposed ways to find the most important time-spots for intervention.

A. Student Diagnosis

Predictions, trends and student information are forms of generated and aggregated information that help teachers efficiently understand general student statuses. In the absence of this information, they would have to manually analyse all ground-work produced by students. By-weekly analysing every bit of student ground-work quickly becomes impractical. It is important to scale this information up, as teachers typically supervise tens to hundreds of students at once. This is the main purpose of student diagnosis tools and generated information.

The concrete process to produce student diagnosis information follows these steps:

1. The system estimates by-weekly probabilities for each student to end up as high, medium or low performance.
2. Predictions are accumulated into point-graphs that show student progression over time (Fig. 5).
3. Performance trend lines are estimated applying linear regression to probabilities (Fig. 5).
4. Latest predictions and trend estimations are summarized as three arrow-value pairs (Fig. 5).

The system has been designed considering teacher time as a scarce resource. Therefore, it should be assigned with higher priority to students that can benefit most from it. In the absence of a proper estimation, this work assumes that less performing students can benefit most, as they have greater improvement margin. A more accurate estimation would work as triage, removing uninterested students and leaving only those at-risk but willing to improve. However, data produced by the system cannot directly identify these cases. Therefore, the present system leaves this task to teachers.

To help teachers focus on students that can benefit most, the system performs a visual classification, beginning with student summaries that are shown in Fig. 4. Summaries reduce student information to their highest probability value-arrow pair, and sorts them from worst to best success probability.

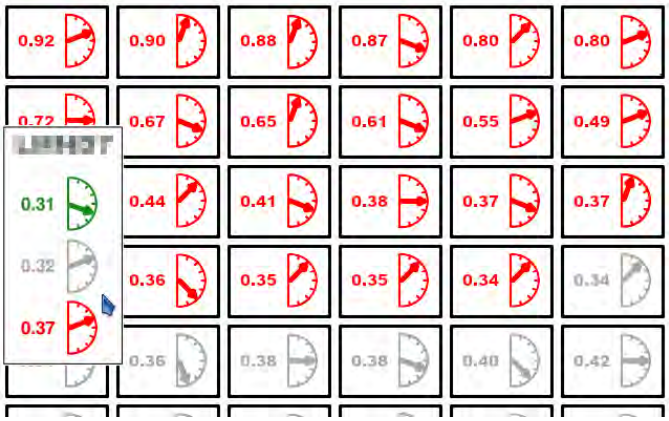


Fig. 4. A teacher is navigating student statuses. Each number+arrow represents one student by his/her highest probability. When pointed with the mouse, a pop-up shows present control panel for the pointed student with all probabilities. Student ID has been anonymized.

Fig. 4 shows a screenshot of the classification for a group of students, while a teacher is navigating their status. First row of students in the figure are those with worst prediction: they show great probabilities (0.80 to 0.92) of ending up as low performance (red). Arrows help knowing if these students are increasing or decreasing this low performance probability. A decrease in this probability would mean an improvement, as it would be less probable for them to end up as low performance students.

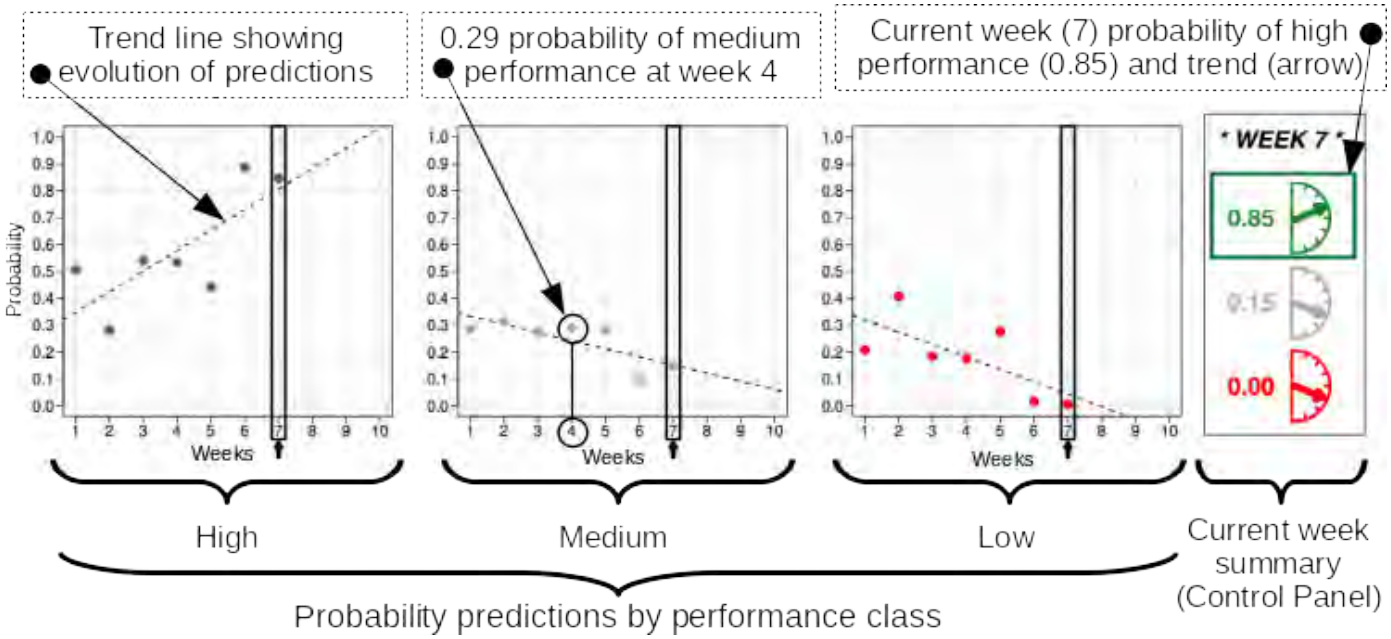


Fig. 5. Visual representation module for a random student at semester week 7 (example). Three point-graphs show by-weekly predicted probabilities for high/medium/low performance (current week, 7, highlighted with a vertical rectangle). Each point represents a probability prediction. Regression trend lines (dashed) are calculated from individual probabilities to show evolution over time. On the right side, control panel summarizes student status on current week (7): probabilities {0.85, 0.15, 0.00} for {high/medium/low} performance {green/grey/red} and arrows indicating whether each probability tends to increase or decrease. Inclination of the arrow represents increase/decrease velocity.

Once teachers detect candidate students, they can click and get detailed information to diagnose them in detail. For this task, the system provides weekly point graphs and trend information described in section IV (representation). Section VI deeply analyses the information provided by point graphs and trends for some typical students. The system also provides access to complete student activity logs including all student accesses, realized tasks, assigned mazes, solutions sent, code from solutions, etc. This is the lowest level information and can represent tons of information just for a single student. Teachers have always relied in this information to diagnose students and is always required for proper diagnosis. The proposed system does automatic processing of this information, along with described predictions and graphs. This process helps teachers navigate information faster, diagnose easier and be more efficient on helping students, but does never substitute ground-level information.

B. Best Moment for Intervention

Student diagnosis is highly dependent on subject and tasks time-frames. Patterns are different for a single-project-based subject with only one final submission, than for another subject requiring by-weekly task submissions. Considered subject asks students to submit solutions to many mazes one by one, but with no specified time-frame for individual mazes. Instead, mazes are grouped into stages with two intermediate deadlines for stages one and two, and a final subject deadline for the rest at the end of the semester. Intermediate deadlines where placed in weeks five-to-six and eight-to-nine.

Time-frames are highly important because they condition student workload. Students tend to accumulate work near deadlines. Although the system was designed with incentives to prevent this behaviour [21], it was only slightly mitigated. This greatly influences predictions and their importance. For instance, some students may not work at all during initial weeks, and perform great later. Early discriminating these students from those not willing to work could be very difficult. Moreover, students with difficulties may work from the start and have confusing results and predictions, which could difficult teacher diagnosis at first.

Similarly to a virus infection, symptoms may not be clear until an initial time-frame has passed. Understanding these time-frames and detecting spots where diagnosis could be most accurate is relevant for teacher intervention. Intervention could be most effective when performed on time: too early or too late interventions may target students not requiring it or may be ineffective due to lack of remaining time.

To find best moments for intervention, Fig. 6 shows all performance predictions for fifty test students. These test students have been selected randomly from the three hundred and thirty six that form our complete sample. Fifty is approximately 15% of the sample, and is a standard proportion to use for Machine Learning algorithms. For this study, this means that our Machine Learning SVM models have been trained with two hundred and eighty six students and these fifty have been left out for out-of-sample tests. This is a common practice to have an estimation on how well trained Machine Learning algorithms perform with new, not previously seen data. As these fifty students come from the main sample at random, it is appropriate to assume that both represent the same distribution. We use only test students because they represent the actual accuracy of the prediction system. Predictions are shown using three different symbols for performance groups: **x** low, · medium, / high. These symbols have been selected to help visually identify predictions in Fig. 6. Weeks one to ten are semester weeks, whereas week eleven shows the final result of students. Students are identified by an anonymous number and visually grouped by their final marks to simplify analysis. Although performance predictions vary over time, there exists a week for every student from which predictions stabilize. This week is highlighted with a background colour: **red** low, **grey** medium, **green** high performance.

Student \ Week	W 1	W 2	W 3	W 4	W 5	W 6	W 7	W 8	W 9	W 10	W 11
21	×	×	·	×	·	×	×	×	×	·	×
34	×	×	×	×	·	×	·	×	·	·	×
50	↗	·	·	·	×	↗	↗	↗	·	×	×
38	↗	↗	↗	·	×	×	·	·	×	·	×
22	↗	·	↗	·	×	↗	↗	·	↗	·	×
11	×	×	×	×	×	×	×	·	·	·	×
49	↗	×	↗	·	↗	×	↗	·	·	↗	×
43	×	·	×	×	×	×	×	·	·	×	×
3	↗	×	×	×	×	×	·	·	·	×	×
35	·	×	·	·	·	↗	↗	↗	·	·	×
41	×	×	×	×	×	×	×	×	·	·	·
2	↗	·	·	·	·	↗	·	·	·	·	·
18	↗	↗	·	×	×	×	×	·	·	·	·
10	↗	×	·	×	·	×	·	·	·	·	·
7	↗	×	↗	·	·	↗	↗	↗	·	·	·
37	↗	·	↗	·	↗	↗	↗	↗	↗	·	·
16	×	×	×	×	×	×	×	×	·	×	·
14	↗	↗	·	·	·	×	·	·	↗	·	·
13	×	×	×	×	×	×	×	↗	·	·	·
51	↗	·	↗	·	×	×	×	↗	↗	·	·
46	↗	·	·	×	×	×	·	·	↗	↗	·
31	↗	×	×	×	×	×	×	×	·	×	↗
9	×	×	×	×	×	×	↗	↗	↗	·	↗
28	↗	×	↗	·	↗	↗	↗	↗	↗	↗	↗
29	↗	·	·	×	·	×	·	·	·	↗	↗
45	↗	·	·	·	·	↗	·	·	↗	↗	↗
36	×	×	·	×	×	·	·	·	↗	·	↗
47	×	×	×	×	↗	×	·	·	·	↗	↗
24	×	×	×	×	↗	↗	↗	↗	↗	↗	↗
15	↗	·	↗	↗	·	↗	↗	↗	↗	↗	↗
40	↗	×	·	·	·	↗	×	·	↗	↗	↗
48	↗	×	×	·	↗	×	×	×	·	↗	↗
20	×	×	↗	·	·	↗	↗	·	↗	↗	↗
27	↗	×	↗	·	·	↗	↗	↗	↗	↗	↗
8	×	×	×	×	×	↗	↗	↗	↗	↗	↗
44	×	×	↗	·	↗	·	↗	↗	↗	↗	↗
6	↗	×	·	·	·	↗	↗	↗	↗	↗	↗
5	×	×	·	·	·	↗	↗	↗	↗	↗	↗
19	↗	×	↗	·	·	↗	↗	↗	↗	↗	↗
17	↗	·	·	×	×	↗	·	·	·	·	↗
23	×	·	×	·	↗	↗	↗	↗	↗	↗	↗
30	↗	×	·	·	·	↗	↗	↗	↗	↗	↗
25	×	×	↗	·	·	↗	↗	↗	↗	↗	↗
26	×	·	·	·	·	↗	↗	↗	↗	↗	↗
33	×	↗	·	·	↗	↗	↗	↗	↗	↗	↗
42	↗	·	↗	·	·	↗	↗	↗	↗	↗	↗
39	·	↗	↗	·	↗	↗	↗	↗	↗	↗	↗
12	↗	×	↗	↗	↗	↗	↗	↗	↗	↗	↗
4	↗	·	↗	·	↗	↗	↗	↗	↗	↗	↗

Fig. 6. Weekly performance predictions for all test students (× low, · medium, / high). Highlighted cells indicate predictions becoming stable, revealing earliest moments for accurate student classification.

Analysing Fig. 6 some visual rules can be inferred:

- Best performance students tend to stabilise their prediction during weeks five-to-six, coinciding with first deadline. Most students classified as medium or high in both weeks five and six end up as high performing. Moreover, only student 3 ends up as low performance with this classification. This simple visual rule is a great candidate for identifying candidate students.
- Most students with two consecutive low performance predictions at weeks five and six end up in medium or low performance groups. Only students 31 and 9 end up as high performance, with borderline result, as they are firsts in Fig. 6 with this classification (students are ordered by final marks).
- It seems quite difficult to identify students that will end up as low performance. On weeks seven to ten they seem to increase their efforts trying to save their final result. That is clearly shown in Fig. 6 with an increase in medium and high classifications. This also seems to happen with students that end up as medium performance. It might be due to a lack of information to get better predictions or, most probably, to an actual impossibility to predict which borderline students will be able to save their course with a final effort.

From this analysis, it seems that weeks five-to-six represent a great moment for discrimination between high performance students and medium-to-low performers. This could also represent a great moment for teacher intervention, as symptoms seem to be highly descriptive in many cases. Teachers could use those weeks to deeply analyse described cases and seek for student problems they can address to give them an effective impulse upwards. These conclusions are also supported by prediction accuracy, as shown in Fig. 7. Concretely, week six has greatest accuracy results for weeks five and six. Both weeks show 70% accuracy for low performance predictions, whereas week 6 shows 84% accuracy for high performance. Reasonably, medium performance is most difficult to predict. However, this problem is minimized by high accuracy of predictions for high performance group and visual aid of Fig. 6.

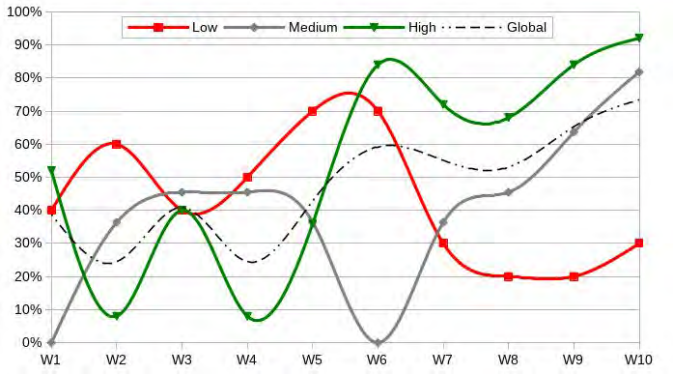


Fig. 7. Prediction accuracy by performance class. Predictions are based only on test students and their highest probability class. Accuracy is calculated as proportion of correctly classified students for each class, with respect to their actual final class.

Accuracy results from Fig. 7 are obtained only from test students. They are the proportion of correctly classified students, comparing their highest probability classification with their actual final class. Goodness of this accuracy results is bound to discussion. They are probably affected by a great variance, as N=50 is a small sample. Moreover, they could have been improved considering SVM classifiers second options by probability. On misclassified students this second option is usually correct and tends to be in narrow probability margin respect to the first option (typically < 0.05). However, further work on this topic has been left out, as this was not the focus of this research.

VI. Case Examples Discussion

Results presented on this paper have been obtained by the system on past courses of the subject *Mathematics I*. These results include a total of 400 first-year students, 336 of which actively participated in the practical lessons and used the system. 286 students were used to train SVM classifiers and 50 were reserved for 12 validation tests. All results presented on this paper refer to validation tests, as they represent out-of-training-sample probabilities that can better estimate actual application results. Original 336 students sample was composed of students with ages **A** ranging from 18 to 21, **A**~N(18.8, 1.33), of which 56 were female (16.6%) and 280 were male (83.3%).

To exemplify the inner working of the system and to discuss its utility three paradigmatic student cases have been selected for more detailed analysis. For each selected student, probability point graphs and summaries for weeks three, five and seven are shown and discussed. Results in these weeks provide an idea on how progression can help in

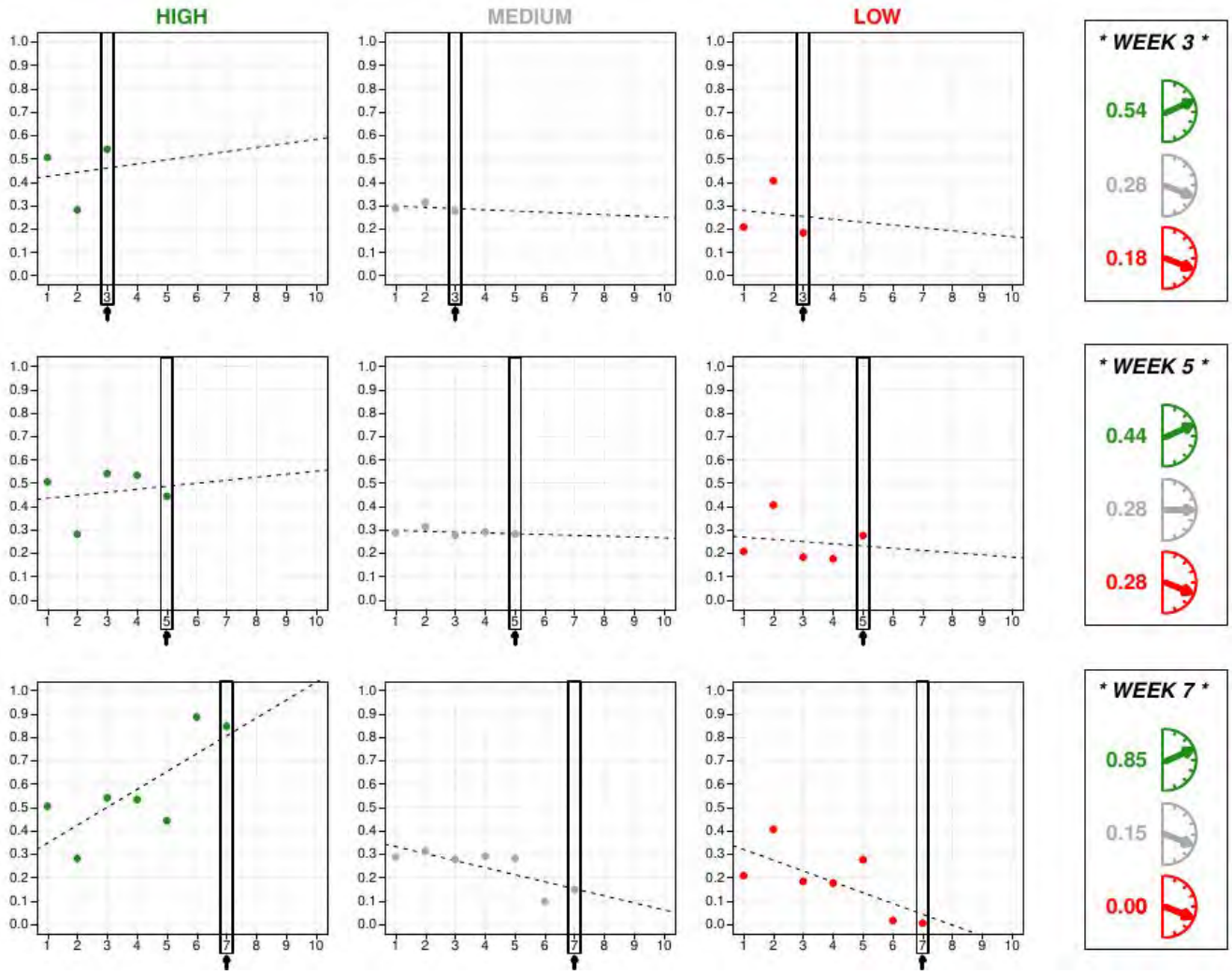


Fig. 8. Probability prediction graphs and summaries for student 12 in weeks 3, 5 and 7.

student diagnosis. As discussed previously, the system provides access to teachers to all this information besides students' ground work (tasks, solutions to mazes, etc.).

The three examples have been grouped into two subsections: stable students and an unsteady student. Stable students are two of them that end up in high and low performance groups respectively. Both students exemplify the observed behaviour norm regarding these groups. They either work hard to get best marks or are not interested in the subject and do some minimal attempts. On the other side, the unsteady student represents most of the students. Although this example ends up as medium performance, many other students behave similarly and end up as low performance, and some of them as high performance. It clearly exemplifies why it results so difficult to accurately predict their behaviour and, consequently, their expected performance.

A. Stable Students

Fig. 8 shows graphs for a high performance student who finds the right path very early and follows it up to the finish. This is student ID 12 from Fig. 6. The student achieves the final classification label at the 3rd week of the course, which is quite remarkable and similar to other high performing students.

The student shows a clear trend to high performance right from the 3rd week, after just one week of lower performance (the 2nd one). Trend predictions from the 3rd week clearly show that probabilities are not casual, but aligned with what probably is a great student: high performance increasing, both medium and low performance probabilities decreasing. The 5th week confirms the prediction, but with one worrying detail. Although proportions are comparable to the 3rd week, the 5th week has introduced a slight increase in low performance probability, to the cost of high performance. There is still nothing to worry about, but this detail might signal an excess in confidence from the student who could be just partially exploiting capabilities. It could be a hint for the teacher to just ask the student about his progress and then induce some extra motivation for hard work.

However, the 7th week clears all doubts about the progress. The student has achieved 0% probability of failure. Performance has great change to end up high, and could be medium with quite small probability. These results help teacher not to worry about the student, as is clearly well focused.

In contrast with these results, Fig. 9 shows predictions for a low performing student (ID 43 from Fig. 6) whose working attitude is almost null since the beginning.

This student shows clear trend to failure right from the 1st week, with 21% probability of high performance versus 40% of low performance. These probabilities are maintained and worsened by the 3rd week. A 50% low performance probability along with a serious tendency to increase. Although very early in the course, it would be interesting for the teacher to consider if the student has problems and can be helped. However, values seem to point to a lack of interest.

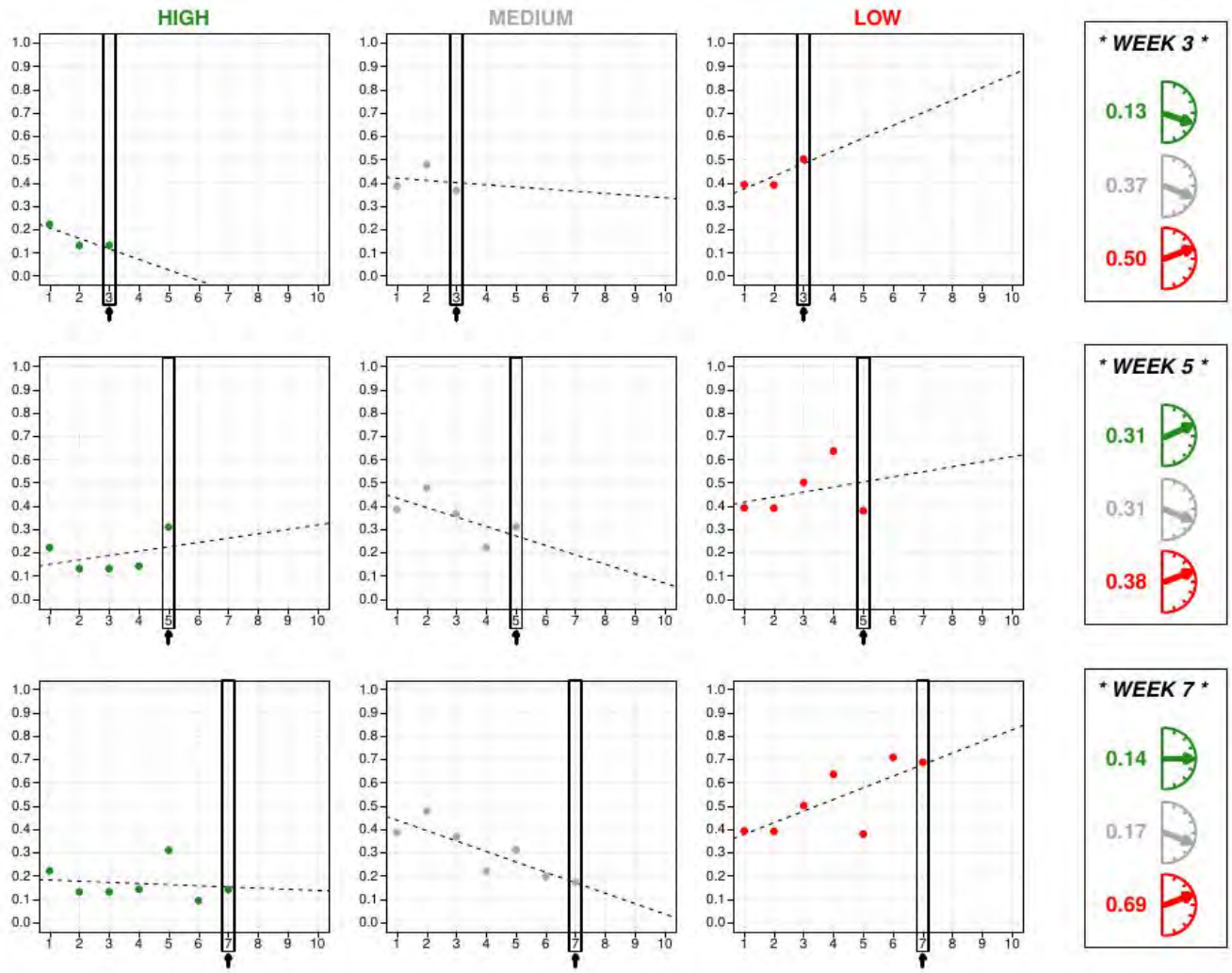


Fig. 9. Probability prediction graphs and summaries for student 43 in weeks 3, 5 and 7.

The 5th week shows an attempt at changing direction. Clearly, first four weeks are horrible for the trend of the student, whereas exactly the 5th shows a change. This change identifies the student performing a before-deadline crash work. Judging by the probabilities, this work has not been enough to get great marks at the first deadline. However, this situation is appropriate for the teacher to verify student logs to see what has been achieved with respect to the deadline. This information could be valuable to help guide the student, in case there is some interest.

As in the previous example, the 7th week gives clear evidence of the kind of student and where are the trends going to. The small increase in the 5th week seems an illusion. Most probably, the student felt incapable of recovering lost time, partially failed at the first deadline and abandoned work. It remains unclear if some teacher intervention could have helped the student, either at the 3rd or the 5th week. However, prediction graphs and trends clearly help predict student progress beyond individual predictions, and give some interesting hints that teachers could use to diagnose and help some of these students. In any case, intervention strategies and their results lie outside the scope of this work.

B. An Unsteady Student

As stated before, next case shown in Fig. 10 represents one of the most general cases of students. Most students do not follow a clear pattern, but instead their numbers change and evolve in complicated ways. These behaviours justify the difficulties the SVM has to predict them, making the medium performance class the most difficult to accurately predict.

In the 1st week the student seems to start well getting a 52% probability of high performance. But this start rapidly decays and medium-to-low classes gain much momentum. Summary for the 3rd week clearly shows a 21% for high performance with an arrow that indicates a fast down tendency. Due to the interesting start of the 1st week, tendencies are sharper and the student seems to go direct to a low-to-medium performance.

However, the 5th week shows more balanced probabilities with flatter tendencies. The student seems to be climbing again and recovering. Numbers for this week are not conclusive, identifying a difficult to classify student. These kind of students probably represent the group that could benefit most from teacher intervention. Evolution shows interest in the subject, as the student is clearly working to pass, but it is unclear what exactly does happen. The student may have a lack of proper scheduling, may need help with some concepts or problems, may have temporal problems... It is interesting for the teacher to deepen in the knowledge about the student to try and help.

After the first deadline and getting into the 7th week, the student has overall improved all probabilities, gaining much more momentum towards high performance. However, the difference between 6th and 7th week indicates that these numbers are much based on first deadline crash effort. After first deadline, student is again losing momentum, probably due to some relaxation after achieving an adequate result.

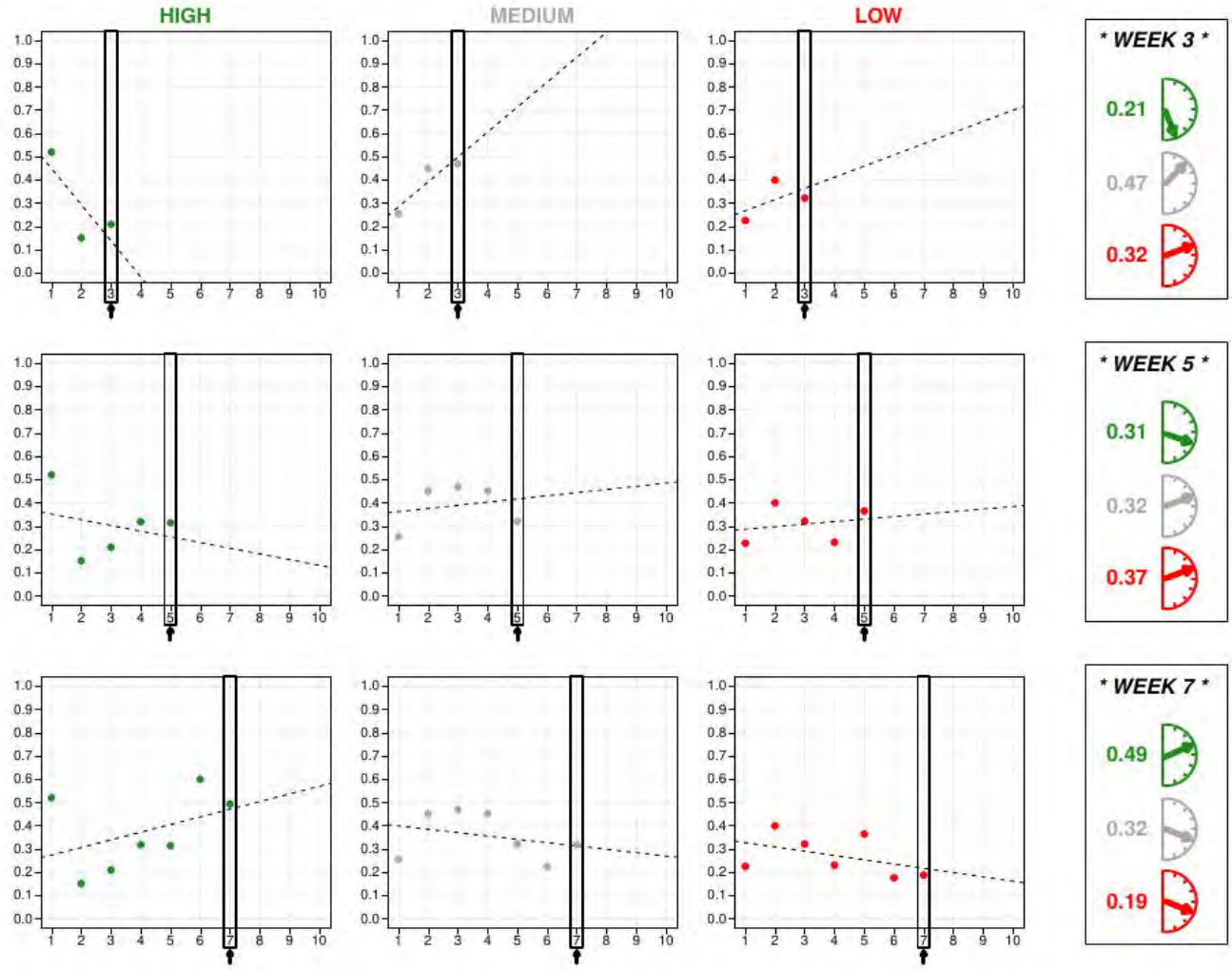


Fig. 10. Probability prediction graphs and summaries for student 50 in weeks 3, 5 and 7.

As previously told, the student ends up in medium performance, which finally identifies with a general student. The analysis of the graphs results very interesting, because even with the difficulties to classify the student, there are many valuable clues. This suggests again that exploiting the time-related nature of predictions has great potential for providing insights in learning trends beyond the mere values of predictions.

VII. Conclusions and Further Work

This research started by creating a custom automated learning system to support *Mathematics I*, a first-year subject that introduces students into Computational Logic and Prolog Programming. This system included a performance prediction system based on Support Vector Machines. The complete system consisted on 4 main components:

- A computer game called *PLMan*, whose many different mazes are learning activities students have to solve programming in Prolog language.
- A custom web-based automated learning system for teachers and students to interact based on *PLMan* mazes.
- A performance prediction system based on Support Vector Machines.
- A representation module for graphically showing performance predictions and student learning trends.

The performance prediction system and representation module have been designed to exploit the time-dependent nature of student data. The system produces probabilistic, consecutive, by-weekly predictions which are added into progression graphs. With these predictions and graphs, learning trends are calculated using linear regression. All this information is shown and summarized in visual ways designed to help teachers diagnose students.

Moreover, filling up a table with by-weekly individual class predictions, some ways for understanding learning time-frames and selecting better moments for teacher intervention have been presented. Identifying most accurate by-weekly predictions for each class, prediction patterns in the table and when different student classifications stabilize, rules for selecting appropriate intervention moments are deduced.

Presented evidence produces some tentative initial answers to the research questions. First, it suggests that exploiting time-dependent nature of student data is viable and desirable. It also suggests that frequent, probabilistic and cumulative predictions have potential for giving early insight into student learning trends. Example cases analysed have shown how student status, progression and trends can be induced from presented graphs, providing more information than the mere performance probabilities. Moreover, evidence presented also indicates that there are methods to identify most relevant time-spots for teacher intervention. The presented method is simple yet effective. However, much research is required in this topic to develop proper,

more elaborate and adaptive methods to select appropriate intervention moments.

Although this is just a first step in this direction, results are promising as an aid for teachers to be more efficient and effective in diagnosing and helping students. However, intervention strategies and their results have not been covered and are left for future research.

There are many debatable points in the presented research that represent valuable features for future development and improvement. There is a question about accuracy of predictions. Other works seem to obtain much greater accuracy results. Although this has not been a problem for this research, it remains to be analysed if accuracy could be actually improved. In the same line, getting more data could help in creating classifications with greater granularity. Performance could also be broken into separate features and individual progression could be predicted and tracked among these. Better and distinct graphs can be designed to give teachers different views that could help them understand faster and deeper about student learning trends. Finally, analysing teacher intervention and their results in learning trends would be a step beyond.

VIII. Limitations

This is an initial empirical research that has gathered evidence to support the capabilities of the presented performance prediction system. Although gathered evidence shows that student performance trends can be inferred from by-weekly performance predictions, it is important to acknowledge that it has been done with a small sample of students (N=336) all from the same university and all first years. Bias and size of the sample are inevitable in this study, and so more studies and more data are required to empirically assess the validity of the proposed system as a way to predict student performance, trends and best moments for teacher intervention.

References

[1] A. Hellas, P. Ihantola, A. Petersen, V. V. Ajanovski, M. Gutica, T. Hynninen, A. Knutas, J. Leinonen, C. Messom, and S. N. Liao, "Predicting academic performance: A systematic literature review," in Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education, ITiCSE 2018 Companion, (New York, NY, USA), p. 175–199, Association for Computing Machinery, 2018.

[2] W. Hämäläinen and M. Vinni, "Classifiers for educational data mining," Handbook of Educational Data Mining, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, pp. 57–71, 2010.

[3] S. B. Kotsiantis, "Use of machine learning techniques for educational proposes: A decision support system for forecasting students' grades," Artificial Intelligence Review, vol. 37, pp. 331–344, Apr. 2012.

[4] J.-F. Superby, J. Vandamme, and N. Meskens, "Determination of factors influencing the achievement of the first-year university students using data mining methods," in Workshop on Educational Data Mining, pp. 37–44, 2006.

[5] N. T. Nghe, P. Janecek, and P. Haddawy, "A comparative analysis of techniques for predicting academic performance," in Frontiers In Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE '07. 37th Annual, pp. T2G–7–T2G–12, Oct 2007.

[6] G. Dekker, M. Pechenizkiy, and J. Vleeshouwers, "Predicting students drop out: a case study," in Educational Data Mining 2009, 2009.

[7] A. K. Hamoud, A. S. Hashim, and W. A. Awadh, "Predicting student performance in higher education institutions using decision tree analysis," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 2, p. 26, 2018.

[8] S. Huang and N. Fang, "Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models," Computers & Education, vol. 61, no. 0, pp. 133 – 145, 2013.

[9] W. Hämäläinen, J. Suhonen, E. Sutinen, and H. Toivonen, "Data mining in personalizing distance education courses," in Proceedings of the 21st ICDE World Conference on Open Learning and Distance Education, pp. 18–21, 2004.

[10] B. Minaei-Bidgoli, D. Kashy, G. Kortemeyer, and W. Punch, "Predicting student performance: an application of data mining methods with an educational web-based system," in Frontiers in Education, 2003. FIE 2003 33rd Annual, vol. 1, pp. T2A–13, Nov 2003.

[11] C. Romero, S. Ventura, P. G. Espejo, and C. Hervás, "Data mining algorithms to classify students," in Educational Data Mining 2008, 2008.

[12] Y. Freund, R. E. Schapire, et al., "Experiments with a new boosting algorithm," in ICML, vol. 96, pp. 148–156, 1996.

[13] A. Y. Wang and M. H. Newlin, "Predictors of performance in the virtual classroom: Identifying and helping at-risk cyber-students," The Journal of Higher Education, vol. 29, no. 10, pp. 21–25, 2002.

[14] I. Lykourentzou, I. Giannoukos, G. Mpardis, V. Nikolopoulos, and V. Loumos, "Early and dynamic student achievement prediction in e-learning courses using neural networks," J. Am. Soc. Inf. Sci. Technol., vol. 60, pp. 372–380, Feb. 2009.

[15] Y.-H. Hu, C.-L. Lo, and S.-P. Shih, "Developing early warning systems to predict students' online learning performance," Computers in Human Behavior, vol. 36, pp. 469 – 478, 2014.

[16] G. Akçapınar, M. N. Hasnine, R. Majumdar, B. Flanagan, and H. Ogata, "Developing an early-warning system for spotting at-risk students by using eBook interaction logs," Smart Learning Environments, vol. 6, p. 4, may 2019.

[17] O. Petropoulou, K. Kasimatis, I. Dimopoulos, and S. Retalis, "Lae-r: A new learning analytics tool in moodle for assessing students' performance," Bulletin of the IEEE Technical Committee on Learning Technology, vol. 16, no. 1, p. 1, 2014.

[18] T. Fawcett, "An introduction to {ROC} analysis," Pattern Recognition Letters, vol. 27, no. 8, pp. 861 – 874, 2006. ROC Analysis in Pattern Recognition.

[19] J. He, J. Bailey, B. I. Rubinstein, and R. Zhang, "Identifying at-risk students in massive open online courses.," in AAAI, pp. 1749–1755, 2015.

[20] J. A. Méndez and E. J. González, "A control system proposal for engineering education," Computers & Education, vol. 68, pp. 266 – 274, 2013.

[21] C. Villagrá-Arnedo, F. J. Gallego-Durán, R. Molina-Carmona, and F. Llorens-Largo, PLMan: Towards a Gamified Learning System, pp. 82–93. Cham: Springer International Publishing, 2016.

[22] C. J. Villagrá-Arnedo, "Sistema predictivo progresivo de clasificación probabilística como guía para el aprendizaje," 2016.

[23] C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, pp. 273–297, Sep 1995.

[24] C. J. Villagrá-Arnedo, F. J. Gallego-Durán, F. Llorens-Largo, P. Compañ-Rosique, R. Satorre-Cuerda, and R. Molina-Carmona, "Improving the expressiveness of black-box models for predicting student performance," Computers in Human Behavior, vol. 72, pp. 621–631, jul 2017.

Smart Learning Research Group

All authors in this paper are members of the Smart Learning Research Group on Intelligent Technologies for Learning, including Adaptive Learning, Learning Analytics and Predictive Systems, Gamification, Videogames and Digital Transformation of Educational Institutions.

Carlos J. Villagrá-Arnedo

Professor of Computer Science and Artificial Intelligence at the University of Alicante. PhD Special Award (programme in Computer Engineering and Computing, 2018/19 call). He has held the position of Head of studies of Multimedia Engineering and has coordinated the Digital Creation and Entertainment pathway, which is taught in the 4th year. His research work focuses on the areas of Artificial Intelligence and the application of video games in the field of education.

Francisco J. Gallego-Durán

Lecturer and researcher at Dept. of Computer Science and Artificial Intelligence. Technical Director at ByteRealms computer games development trademark. Has designed and developed game engines like WiseToad Framework or CPCtelera, and games like PLMan or MindRider. Has developed educational innovation projects combining Automated Learning Systems and Project Based Learning, even on non-technical subjects. His present research includes games for education, innovative programming teaching, Gamification, Machine Learning and Neuroevolution.

Faraón Llorens-Largo

Professor of Computer Science and Artificial Intelligence of the University of Alicante. PhD in Computer Science. Director of the Polytechnic School (2000-2005) and Vice-rector of Technology and Educational Innovation (2005-2012), both at the UA and Executive Secretary of the ICT Sector Commission of the CRUE (2010-2012). His work is in the fields of artificial intelligence, adaptive learning, gamification and video games, IT governance and digital transformation of educational institutions.

Rosana Satorre-Cuerda

Professor of Computer Science and Artificial Intelligence at the University of Alicante. PhD in Computer Science. Very involved in teaching issues, she participates in educational innovation projects related to the EHEA, belonging to the REDES program of the University of Alicante. She is a member of AEPIA (Spanish Association for Artificial Intelligence) and AENUI (Association of University Teachers of Computer Science).

Patricia Compañ-Rosique

PhD in Computer Science. Her research lines are within the application of AI techniques and the application of digital technologies to education. She is a professor in the Department of Computer Science and Artificial Intelligence at the University of Alicante. She has held the position of Deputy Head of Computer Engineering of the Polytechnic School and deputy director of the Polytechnic School. She participates in many educational innovation projects related to the EHEA.

Rafael Molina-Carmona

Professor of Computer Science and Artificial Intelligence at the University of Alicante, PhD in Computer Science. Director of "Smart Learning", research group on Intelligent Technologies for Learning. His research focuses mainly on the use of information technologies to transform society and technological innovation. In particular, he works on Artificial Intelligence applications in different fields: computer-aided design and manufacturing, computer graphics, adaptive learning, gamification, learning analytics, IT governance and information representation.

Two-Stage Human Activity Recognition Using 2D-ConvNet

Kamal Kant Verma[1]*, Brij Mohan Singh[2], H. L. Mandoria[3], Prachi Chauhan[4]

[1] Research Scholar, Uttarakhand Technical University, Dehradun (India)
[2] Department of Computer Science and Engineering, College of Engineering Roorkee, Roorkee (India)
[3] Department of Information Technology, G.B Pant University of Agriculture and Technology, Pantnagar (India)
[4] Research Scholar, Department of Information Technology, G. B Pant University of Agriculture and Technology, Pantnagar (India)

Received 25 October 2019 | Accepted 11 March 2020 | Published 24 April 2020

unir
LA UNIVERSIDAD EN INTERNET

Abstract

There is huge requirement of continuous intelligent monitoring system for human activity recognition in various domains like public places, automated teller machines or healthcare sector. Increasing demand of automatic recognition of human activity in these sectors and need to reduce the cost involved in manual surveillance have motivated the research community towards deep learning techniques so that a smart monitoring system for recognition of human activities can be designed and developed. Because of low cost, high resolution and ease of availability of surveillance cameras, the authors developed a new two-stage intelligent framework for detection and recognition of human activity types inside the premises. This paper, introduces a novel framework to recognize single-limb and multi-limb human activities using a Convolution Neural Network. In the first phase single-limb and multi-limb activities are separated. Next, these separated single and multi-limb activities have been recognized using sequence-classification. For training and validation of our framework we have used the UTKinect-Action Dataset having 199 actions sequences performed by 10 users. We have achieved an overall accuracy of 97.88% in real-time recognition of the activity sequences.

Keywords

Activities Recognition, Random Forest, 2D Convolution Neural Network, Intelligent Monitoring System.

DOI: 10.9781/ijimai.2020.04.002

I. Introduction

RECOGNITION of human activities have been a hot research field in computer vision for more than two decades and researchers are still working in this domain due to unavailability of perfect human activity recognition system. Still images give less knowledge for action recognition as compared to the videos. Videos give temporal information as an additional ingredient, which is an important indicator for action recognition. A large number of different activities may be correctly identified depending on the motion component found in videos.

Action recognition is an active ingredient of many applications such as automatic video surveillance [2]-[6], object detection and tracking [7], video retrieval [8], etc. Other applications are strongly connected to the activities and actions recognitions, like human motion analysis [9]-[15], analysis of dynamic scene activities [16], classification of human actions [17], or understanding human behavior [18]. Human activity recognition comprises various steps, which define the features that represent low level activities. The activities of interest and their details may vary depending on the applications. For example, from the last few years Automatic Teller Machine (ATM) has become one of the prime facilities for cash dispense, cash withdrawal, balance enquires, etc. For this reason ATM has become an unsafe site and if the security issues of ATM are concern then, it requires an intelligent video surveillance system that not only captures the scene information at the time of abnormality, but also recognizes single and multi-limb human abnormal activities so that the intelligent system could warn to the security-in-charge in real time and the corrective action could be taken at the time when the abnormal activity happens either by single or multi human limb.

A basic model of human activity recognition in video frame sequences consists of mainly two levels. In the first level, handcrafted features have been extracted from raw input data and in the second level a classifier model is built depending on these features. Here some of the most frequently used feature detectors for human activity recognition have been discussed, which include Histogram of optical flow (HOF), Spatial-Temporal Interest Points (STIP), Histogram of Oriented Gradients (HOG), and dense trajectories [43] etc. However, the extraction of these features is a really difficult and time consuming process as well as it is challenging to know which kind of feature is relevant to the problem because feature selection varies from problem to problem in real time. Therefore a deep learning based model has been proposed and discussed in below section to attend the demand for handcrafted features and reduce the complexity of this process.

At a recent time, deep learning has arisen as a group of deep architecture based learning models that render high-level abstraction of data. A deep learning model is a systematic presentation of multiple

* Corresponding author.
E-mail address: kkv.verma@gmail.com

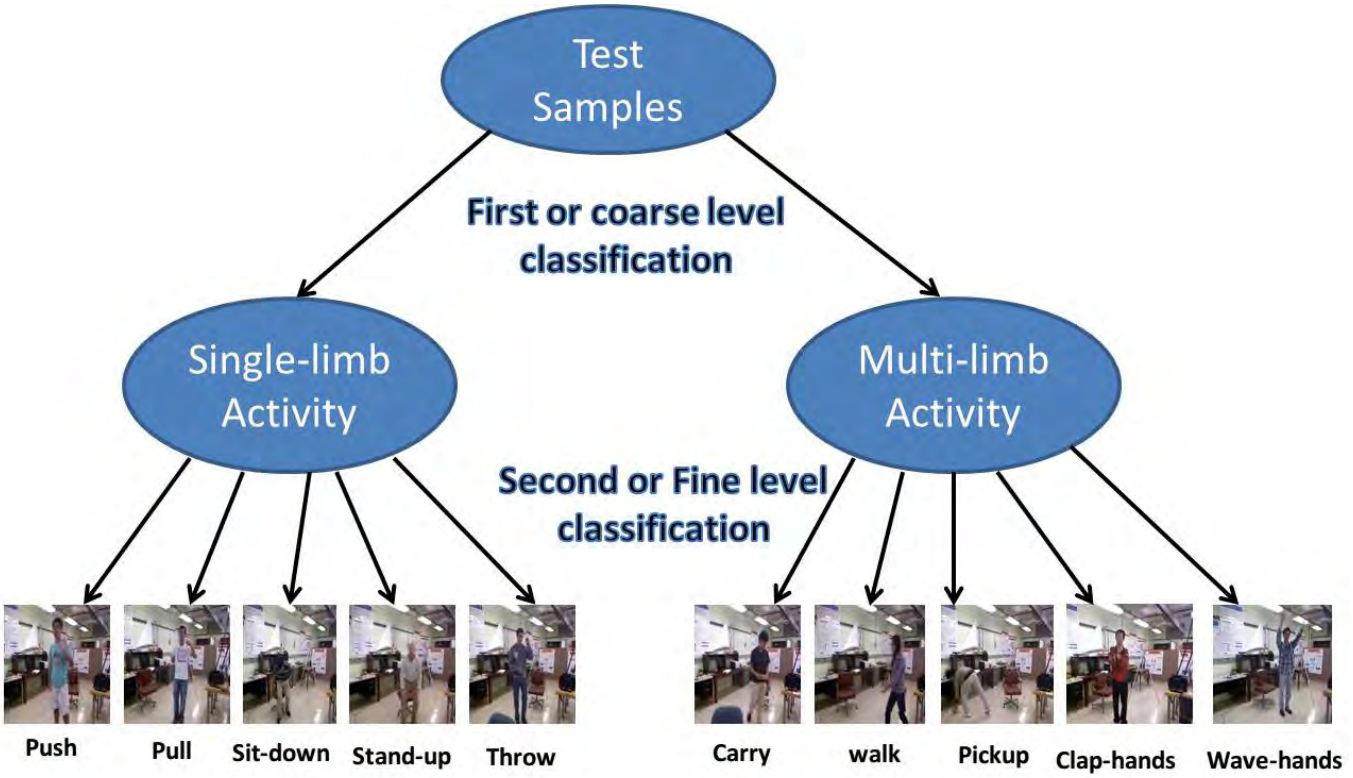


Fig. 1. Two-Stage Classification of human single-limb and multi-limb activities.

layers that are organized for automatically learning of features. Moreover, every layer in the deep architecture model receives output from the previous layer and implements nonlinear transformation such that the input data are transformed into systematic order of low level features to more advanced level features. The most common types of deep learning models are convolutional neural networks, recurrent neural networks, auto encoders, deep belief networks etc. For the labeled input data the deep learning model is trained based on supervised learning and, in case of unlabeled data, the deep architecture model is trained via unsupervised learning. Due to its outstanding performance in various areas like bio signal recognition, gesture recognition, computer vision, bio-informatics, etc. it could be fully deployed in human activity recognition.

II. Related Work

Action recognition in video frames, images sequences or in still images have become a hot research area over the past several years. Since it is not possible to discuss complete literature of action recognition, hence major focus has been given on action recognition in (a) RGB video frames (b) depth frames using deep neural network. Due to the increasing demand of computer vision, latest research work has shifted towards applying convolution neural networks (CNNs) for activity recognition because it is able to learn spatio-temporal information [20], [21], [22] from the videos. Li et al. in [23] used CNN for recognition of human activity captured using a smart phones data set. However activity recognition is not the only field where convolution neural network has achieved wonderful results but it outperforms in various areas such as human facial recognition [24], image recognition [25]-[27] and human pose estimation [28].

The first work on MSR3DAction dataset was given by L. Xia et al. in [29], a histogram of 3D joints (HOJ3D) has been computed and redirected by linear discriminant analysis and divided in k different visual words and temporal evaluations of these visual words are formed using HMM. L. Xia et al. validated their approach on MSR UT kinect-Action 3D joints dataset and achieved 90.92 % accuracy. The recently proposed work in [30] used CNNs, Long Short-Term Memory (LSTM) units and a temporal-wise attention model for action recognition. To learn visual features using CNNs [25] have given the benefit over hand-crafted features for recognition in still images [31]-[33] moreover it overcomes the limitation of the manual feature extraction process. Modified CNNs for recognition of activities in video frames was suggested in various approaches [20], [22], [26], [34]-[39]. A couple of methods used single video frames with spatial features [20], [26], [34]. Multi-channel inputs to 2-dimensional convolution neural networks have also been used in [20], [26], [37], [39]. In [26] author divided the temporal and spatial video parts by using the optical flow method of RGB frames. Each of the separated parts was placed into different deep convolution neural network for learning spatial and temporal features of appearance and movement of object in a frame. Moreover, activity recognition can be performed not only by finding spatial temporal information but a state-of-the-art video representation in [40]-[42] also uses dense point trajectories. The first work was given in [43], which uses dense points of each frame and follows these points based on displacement information after finding a dense optical flow field. The method proposed in [43] was validated on four bench mark datasets such as KTH, YouTube, Hollywood2, and UCF-sports. The better implementation of the approach based on trajectory was given in the Motion based Histogram [44] which was calculated on vertical and horizontal parts of optical flow [50].

However the recent CNN approaches use 2D-convolution architecture of the video which allows learning the shift-invariant representations of the image scene. Meanwhile 2D-convolution is unable to incorporate the depth volume of video which vary with time and is also an important ingredient for activity recognition from the beginning to the end of the activity. 3D-Convolution addresses this issue and incorporates spatio-temporal information of videos and provides a real extension of 2D convolution. A different way to deal with spatio-temporal features is

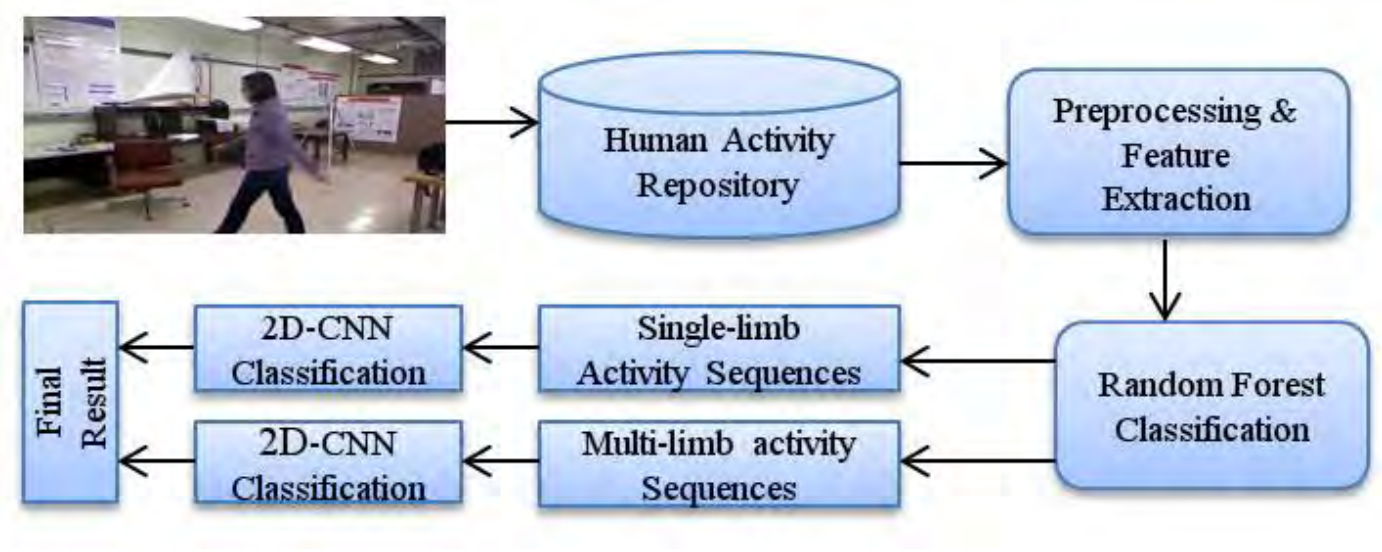


Fig. 2. Flow diagram of the proposed framework.

suggested in [45], where the author factorized original 3D convolution into 2D spatial in the lower layer followed by one-dimensional temporal convolution in the upper layer. The proposed framework was tested on two bench marked datasets (UCF-101 and HMDB-51) and outperformed over existing CNN based methods.

Liu et al. [46] proposed an approach for human activity recognition using coupled hidden conditional random fields (CRF) by combining RGB and depth data together rather than CRF because it had the limitation that it cannot capture the intermediate hidden state using variables. Liu et al. implemented their method on three datasets named DHA, UT kinect-Action 3D and TJU, which produced 95.9%, 92% and 92.5 % accuracies respectively. Zhao et al.[47] proposed a multi-model architecture using 3D Spatio-temporal CNN and SVM using raw depth sequences, depth motion map and human 3D skeleton body joints for recognition of human actions, the proposed approach was validated on UTKinect-Action 3D dataset and MSR-Action 3D dataset and gave 94.15% and 97.29% accuracies respectively. Arrate et al. in [48] suggested 3D geometry of different human parts by taking translation and rotation in the 3D space and tested the approach on UTKinect-Action 3D dataset, achieving 97.02% of accuracy. Siirtola et al. in [51] discussed the personal human activity recognition model using incremental learning and smartphone sensors. Authors in [51] have also discussed that how human activity recognition system has changed since 2012, and how this human activity recognition method can be used in healthcare application. Jalal et al. in [52] suggested a novel framework for human behavior modeling using 3D human body posture captured from Kinect sensor based on clue parameters. In [52] authors extracted human silhouettes from noisy data and tracked body joints values by considering spatio-temporal, motion information and frame differentiation, then angular direction, invariant feature and spatio-temporal velocity features have been extracted in order to find the clue parameters, at last clue parameters are mapped into code-words and recognize the human behavior using advanced Hidden Markov Model (HMM), the proposed approach is tested on three benchmark depth datasets: IM-DailyDepthActivity, SRDailyActivity3D and MSRAction3D, and got 68.4%, 91.2% and 92.9% accuracies respectively.

Simonyan et al. [26] used two convolution neural networks for two streams, one for spatial features and another for temporal features and validated the proposed system on different bench-mark datasets, UCF-101 and HMDB-51. The limitation of this method [26] is that they applied a complete dataset for spatial stream and temporal stream, both for 101 classes' data in UCF-101 dataset and 51 class's data for HMDB-51 and achieved 88% and 59.4 % accuracies, respectively. There may be a scope of research to improve the results in terms of recognition accuracies by applying our two stage methodologies. Similarly we can apply our proposed approach to any other challenging large scale action recognition dataset.

III. Motivation

In our proposed work, we have used a novel multi-stage classification framework based on the color information retrieved from a RGB camera in an intelligent video surveillance system. In our work we have used two-stage classification. The advantage of two-stage classification is that we can handle a large complex problem in a better way in real time since it is difficult to train the network when the number of classes are high.

Therefore in this approach we have divided the classes into two categories at subsequent level. In two-stage classification the first stage is also known as coarse level classification and the second stage is called fine level classification. The two stage classification is given in Fig. 1, as Wang et al. [30] used CNN-LSTM based attention model for human action recognition on UCF-101 dataset and achieved 84.10% accuracy. Similarly Karpathy et al. [20] used CNN on sports dataset for sport action recognition and got 63.3% accuracy. Both, Wang et al [30] and Karpathy et al. [20] used complete dataset only at once and performed classification, which was the limitation of the work given in [20] and [30]. If our proposed two-stage approach had been used on the datasets mentioned in [30] and [20], we could get better results.

However, our approach divides the number of classes into multi-levels and reduces the loss first at initial level, and then at subsequent levels, overcoming the limitation of work given in [20] and [30] in which whole dataset was taken at once, hence by splitting the dataset into a multi-level scenario we applied the proposed technique and achieved good results in terms of accuracy compared to the state of art literature.

Our proposed method may be applied to large scale human action datasets such as UCF-101 [22], [34], [36]-[37], HMDB-51[26], [38], [39], [41]-[42] which divides the whole dataset into a subsequent level of hierarchy in downward direction, which leads to the reduction in losses at each level. Our proposed method is applied and validated on UTKinect-Action 3D dataset, giving promising result and advances as compared to the previous literature.

In this work, a 2D-Convolution Neural Network has been used due to its excellent performance on object detection, activity classification and recognition. It has the power of automatically learning the features from the input video data. The work recently published in [19] used a stack of 3D-CNN on spatio- temporal activation re-projection (STAR-Net) using RGB information. Most of the previous approaches have utilized hand crafted features, which is a really time consuming process. Therefore this work proposed a two-stage-DNN based model to recognize the human activities. In the first stage we divided all activity types into two categories based on human limbs which are human single limb and human multi limb activities. In the second stage two 2D-Convolutions Neural Network have been used for activity recognition.

From the experimental point of view, UTKinect-Action3D Dataset has been used which contains the following ten human activity classes types: wave hands, pull, walk, sit down, push, stand up, throw, pick up, carry, and clap hands. Based on our proposed approach we have divided sit down, standup, pull, push and throw into human single limb activities category, and walk, pickup, carry, wave hands, clap hands, into human multi-limb activities category. Table I shows some of the state-of-art techniques, datasets used, accuracies, number of classes in datasets used alongside with their applications.

In summary, the main work of this paper is given as:

- This paper proposes a two-stage activity recognition framework for RGB information. In the first stage human activities are distinguished based on human single and multi-limb categories.
- In the second stage two Deep-CNN models are used to recognize the separated single and multi-limb human activities.

TABLE I. Some Of Art-Of-The-State Approaches With Their Datasets And Accuracy

Name	Approach	Datasets	Accuracy	Classes	Applications
Zhang. et al.[9]	Affine Transformation, SVM	-	-	-	Pedestrian Detection
Wang et al.[30]	CNN, LSTM and Attention Model	UCF-11,UCF-Sports and UCF-101	98.45%,91.9% and 84.10%	11,10,101	Video Action Recognition
Karpathy et al. [20]	CNN	Sports-1M	63.3%	487	Recognition of Sports actions
Taylor et al.[21]	Convolution Gated RBM	KTH action dataset, Hollywood2 dataset	90% and 47.4%	6,12	Human action recognition
Tran et al.[22]	3DCNN+SVM	UCF101	52.8%	101	Action Recognition
Xue et al.[23]	CNN	HARUSP dataset	96.1%	6	Sensor based activity
Schroff. et al.[24]	CNN+L2 Normalization +Triplet Loss	YouTube Face DB Dataset	99.63%	-	Face Recognition
Simonyan et al.[26]	Two Stream ConvNet	UCF-101,HMDB-51	88% and 59.4%	101,51	Action Recognition
Xia. et al.[29]	HOJ3D+Lin. Discriminant +HMM	MSR3DAction Dataset	90.92%	10	Human Action Recognition
Girshick. et al.[32]	Region-CNN	ILSVRC2013 detection dataset	47.9%	200	Object detection and region segmentation
Taigman. et al. [33]	Deep CNN	(LFW) and You-tube datasets	97.35% and 91.4%	-	Face Recognition
Donahue. et al. [34]	CNN+LSTM	Face(YTF) UCF-101	68.2%	101	Action Recognition
Xu. et al. [35]	3DCNN	TRECVID 2008 And KTH Dataset	72.9% and 67.52%	3,6	Action Recognition
Wang. et al. [36]	trajectory-pooled deep-convolutional descriptor (TDD)	UCF-101	84.7%	101	Action Recognition
Wang. et al. [37]	GoogleNet+VGG16 based Two Stream ConvNet	UCF-101	91.4%	101	Action Recognition
Bilen. et al. [38]	Dynamic Image + CNN	UCF-101 and HMDB-51	89.1% and 65.2%	101,51	Action Recognition
Feichtenhofer. et al. [39]	Two Stream Network Fusion CNN	UCF-101 and HMDB-51	93.5%and 69.2%	101,51	Action Recognition
Peng. et al. [41]	Stacked Fisher Vector(SFV)	YouTube, J-HMDB and HMDB51dataset datasets	93.77%,69.03% and 66.79%	11,51,21	Action Recognition
Wang. et al. [42]	Dense trajectories using SURF and optical flow Features + RANSAC	Hollywood2, HMDB51,Olympic Sports and UCF051 Dataset	64.3%,57.2%, 91.1% and 91.2%	12,51,16 and 50	Action Recognition
Sun et al.[45]	Factorized Spatio-temporal CNN	UCF-101 and HMDB-51	88.1% and 59.1%	101 and 51	Activity Recognition

The paper is arranged as follows: section II contains the related work on activity recognition in RGB videos, depth data and skeleton information: section III contains motivation of the proposed work: section IV describes our suggested method for activity recognition: experimental result and discussion is given in the section V and section VI, respectively, and at last conclusion and future work has been described in section VII.

IV. Proposed Approach

To represent human single-limb and multi-limb activities sequence recognition, this paper proposes a novel framework based on the color information retrieved from a RGB camera in an intelligent video surveillance system. We have done our proposed work in two stages. In the first stage input data have been preprocessed by resizing into the new scale, then we have extracted the histogram of oriented gradients (HOG) features from all input frames and we have classified the activities into two categories named single-limb and multi-limb activities, using a random forest (RF) classifier. Then in the second stage two 2D convolution neural networks have been applied to each category for recognition of actual activity type under single and multi-limb category. The flow diagram of the proposed two stage human activity recognition is depicted in Fig. 2.

A. Preprocessing and Feature Extraction

Resizing: Collected RGB frames from MSR UTKinect Action 3D dataset have size of 480x640. The proposed approach read all single and multi-limb activity sequences for resizing the initial raw color information to a new size (80x120) using the OpenCV library.

Histogram of Oriented Gradients: HOG features are widely used in various motion based applications and activity recognition systems. These features gradients have been calculated over preprocessed image. We have divided the image into 8x8 cells. All the gradient values of a cell are divided into 9 equal bins histogram. To estimate the feature value, a block size of [16, 16] has been taken. A [16, 16] block has 4 histogram which will form a one dimensional vector of size 36. In addition to this, a detection window size of [16, 16] is used having a stride of [8, 8] leading to the14 horizontal and 9 vertical positions where the detection window moves for constructing a total of 126 positions. To calculate the final feature vector of an entire image, all 126 features having a size of 36, are combined together to form a final large vector, which is a 4536(126 * 36) dimensional vector. The final feature vector (F) can be defined using (1) where d_1, d_2, d_3…… d_k are the dimensions.

$$F = (d_1, d_2, d_3, \ldots\ldots\ldots\ldots d_k) \qquad \text{where } k = 4536 \tag{1}$$

B. Initial Classification using Random Forest

Random Forest is a type of ensemble based classifier proposed by Breiman [1] in 2001. The RF method works on different models increasing the accuracy (bagging) and improving the performance of previous trees by the subsequent trees (boosting). The basic principle of Random Forest is that it takes the decision based on de-correlated decision trees. It can be used with multi-class classification purposes. An RF model is non-parametric in nature. For an ensemble of classifiers C1(θ), C2(θ),.....Cn(θ), and having a training set chosen arbitrarily from the distribution of the random vector P, Q, then the margin function given in equation (2) can be defined as,

$$mg(\mathbf{P}, Q) = av_n I(C_n(\mathbf{P}) = Q) - \max_{i \neq Q} av_n I(C_n(\mathbf{P}) = i) \tag{2}$$

Where $i \neq Q$

Where I (*) is the indication function. The margin determines the limit to which the average number of votes at **P**, Q for the right class exceeds the average vote for any other class. In our work, we have used the RF classifier to distinguish two classes, i.e. either single-limb or multi-limb activity. Since Random Forest is an ensemble classifier, a HOG feature has been computed for each video frame in an activity. This final feature vector (F) of dimension 4536 has been used to train the classifier.

C. 2D-CNN Classifier Based Activity Recognition

A 2D-CNN or ConvNet is a type of deep neural network, often used for video analysis and recognition and image classification. ConvNet is a biological inspired deep-network whose connectivity designs between the neurons are similar to the animal visual cortex.

CNN is a sequence modeling classifier and has a sequence of layers where every layer transforms input volume of activation into another form. In ConvNet, three main layers are used to make a ConvNet model, a convolution layer, pooling or sub sampling and fully connected layer. We keep the layers one after another to build the ConvNet architecture. CNN uses minimal preprocessing compared to other classification algorithms. Automatic feature detection is the considerable advantage. In this work, we have used 2D-ConvNet classifiers for recognition of activities that belong to the single limb and multi-limb classes.

Therefore, two 2DconvNet models have been trained using the categorical cross entropy (CCE) based objective function. The architecture of 2D-ConvNets for single and multi-limb activity is given in Fig. 3.

CCE based 2DconvNet: The Network has C output values, corresponding to one value of each class for the activity sequences. The Categorical Cross entropy (CEE) for C classes is defined in (4)

$$CCE = -\frac{1}{N}\sum_{j=1}^{N}\sum_{k=1}^{C} l_{yi \in Ck} \log Prob_{model}[yi \in Ck] \tag{4}$$

Where N represents the number of samples. Each sample belongs to a category k (the total of categories is C). $l_{yi \in Ck}$ *is the indicator function of j samples for belonging to* k^{th} category and logProb model $[y_j \in C_k]$ is the predicted-probability of the jth sample belonging to k^{th} category.

V. Experimental Work

First, the description of the dataset used in this study is given. Then results have been evaluated by splitting the dataset into a 80% training set and a 20% test set using the train test split method. For training and validation of our framework, we used the public dataset: UTKinect-Action Dataset. The model was trained on Intel Core i5 7th Gen, 2.4GHz processor, 8GB of RAM and a 2GB 940MX NVIDIA GPU support on Ubuntu 16.04 LTS(Linux) operating system.

A. UTKinect-Action3D Dataset

To validate our proposed framework we composed a dataset having 10 types of indoor human activity sequences. The dataset was taken using a stationary Kinect sensor with Kinect for Windows SDK Beta version having a frame rate of 30 fps. The Kinect sensor has a capacity to capture about 4 to 11 feet. The Dataset contains 10 activities performed by 10 different persons two times: 9 male and 1 female, out of them one person is left-handed and the rest are right-handed, with a total of 199 activity sequences. The label of the carry activity performed by the persons the 2nd time is not given, hence frames for this activity cannot be identified. The 10 activity sequences are wave hands, pull, walk, sit down, push, stand up, throw, pick up, carry, and clap hands. All these activity sequences are given in the three different formats RGB, Depth frames and skeleton information. The dataset contains all the actions in indoor scenario.

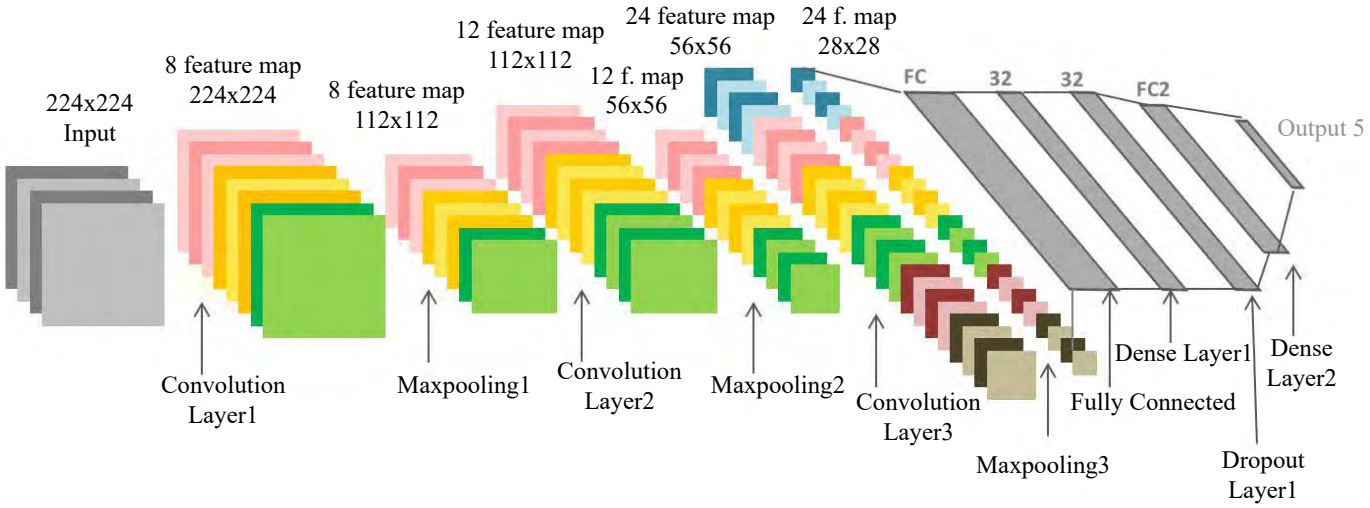
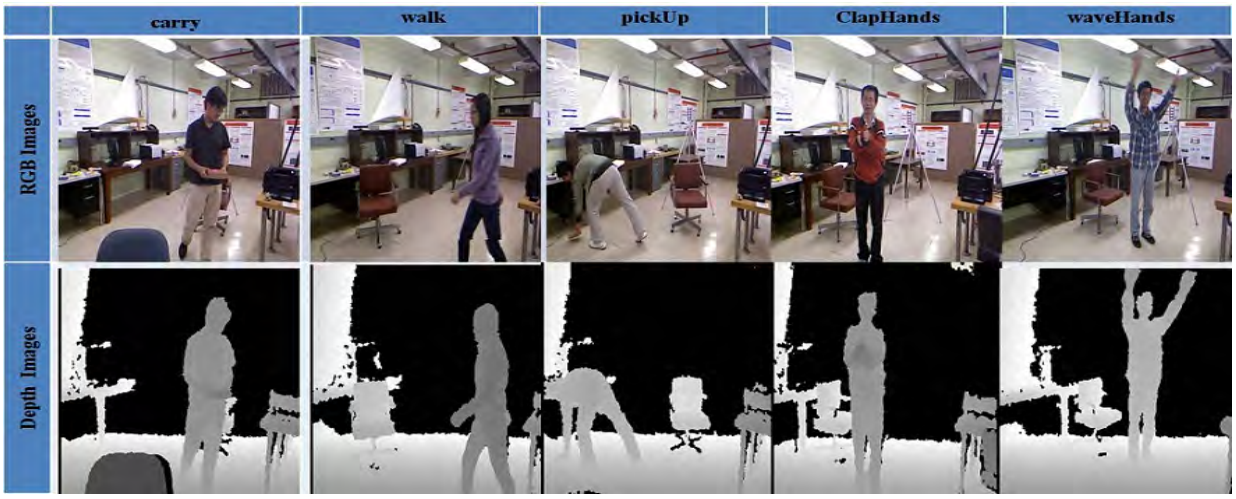
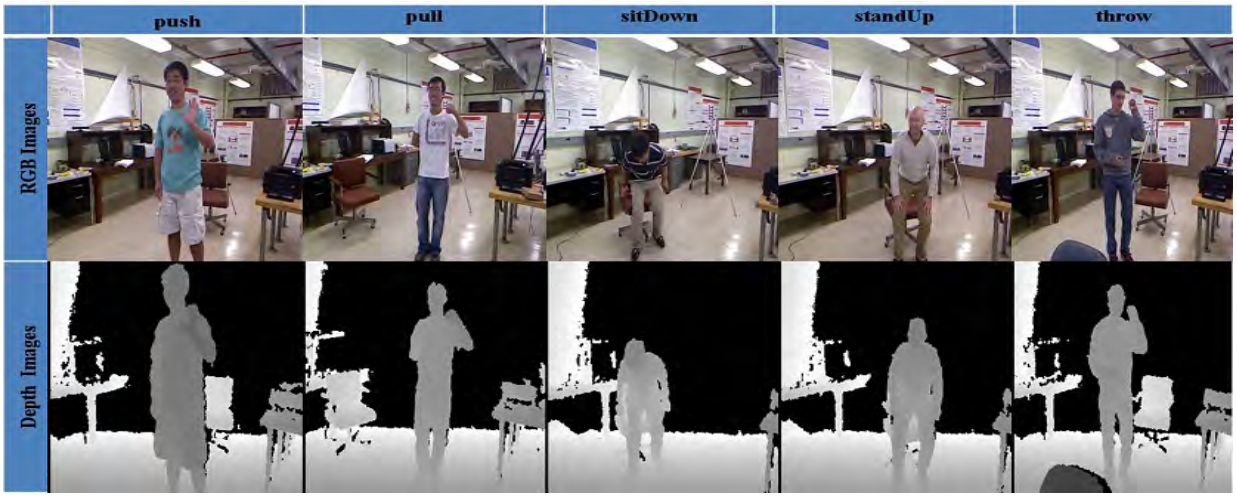


Fig. 3. 2D-ConvNet Architecture for human single and multi-limb activity recognition.



(a)



(b)

Fig. 4. Few sample frames of 10 different activity sequences are shown from the dataset. (a) Shows multi limb activities in RGB frames and their corresponding depth frames. (b) Shows single limb activities RGB frames and their corresponding depth frames. In this work only RGB frames are used for action recognition and depth frames are just for depiction.

The number of frames for each activity ranges from 5 to 120. The resolution of RGB and Depth images is 480x640. The total numbers of frames are 5869 for 199 activity sequences. The proposed frame work uses activity recognition only in RGB frames sequences and the rest of the sequences are just for demonstration of the dataset. Few RGB frames and their corresponding depth images are given in the Fig. 4.

B. Activity-Type Recognition using Random Forest

To Recognize single limb and multi limb types of activities we trained the random forest classifier from Scikit-Learn Python Library. The classification has been performed by varying the number of decision trees (n_estimators) from 1 to 100.An accuracy of 99.92% has been recorded in classification of single limb and multi-limb activity types for n=46.This is shown in the confusion matrix given in Fig. 5, that all samples are classified except one sample from single limb class.

C. Activity Recognition using 2D-ConvNet

Two 2D-ConvNet classifiers have been trained for single limb and multi-limb activities using the output of initial Random Forest Classification with the help of a feature vector F. For Single limb activity, the network has been trained with categorical cross-entropy objective function with a learning rate of 10^{-3} and decay of $5x10^{-6}$. The architecture of 2D-ConvNet is given in Fig. 3 and the learning curve of the network for single-limb activities is shown in the Fig. 6(a) and, for multi-limb activities the learning curve is given in Fig. 6(b).

It can be seen from the learning curve in Fig. 6(a) of single limb activity that, after 146 epochs, there is no change in the validation network. Thus, it has been pointed as the Best-Network. An Accuracy of 97.9% has been recorded in the recognition of single-limb activities as shown in Fig. 7(a). The Confusion matrix corresponding to classification of single limb activities is given in Fig. 8(a) Recognition performance has also been marked against each class of activities as shown in Fig. 9(a) where accuracies vary from 89% to 100% for different activities and 100 percent accuracies have been recorded for pull, stand up and throw activities. It also has been noticed that one more activity sit down is having approximately 100 percent recognition.

The learning curve corresponding to the multi-limb activities is shown in the Fig. 6(b). After 42 epochs, the best network is found because further there is no change in the validation results. An accuracy of 98% has been recorded in recognition of multi-limb activities given in Fig 7(b). The confusion matrix of multi-limb activities is given in Fig. 8(b). Recognition performance has also been recorded for every individual class of activities as depicted in Fig. 9(b). It is also noticed that 100 percent recognition is achieved for clap hands activity.

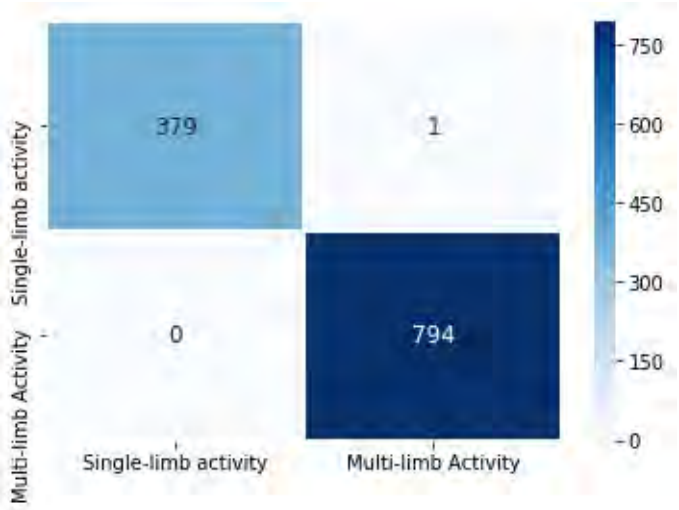
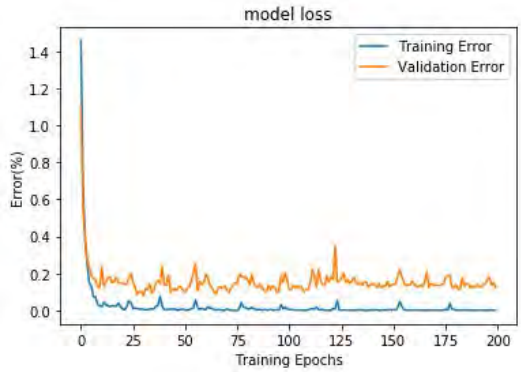
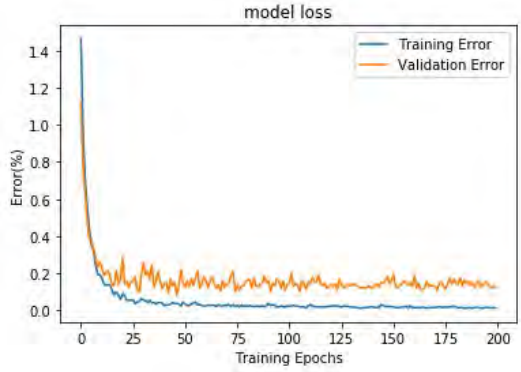


Fig. 5. Confusion Matrix of activity-type recognition.

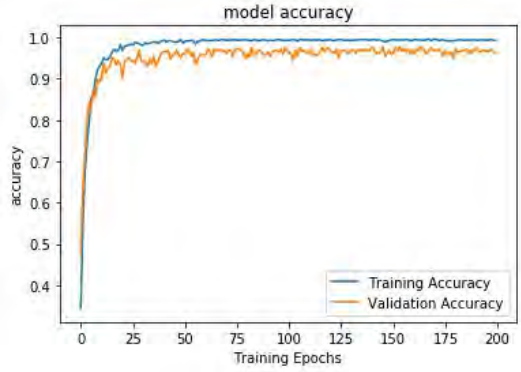


(a)

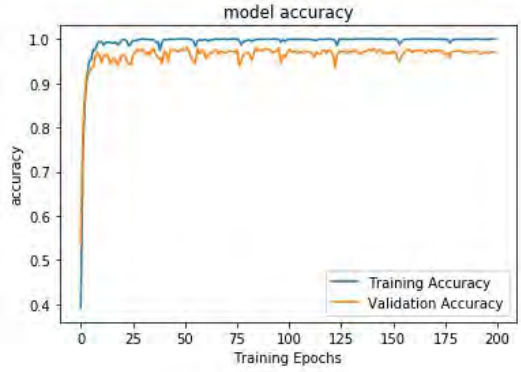


(b)

Fig. 6. Learning curves of 2DConvNet showing variation in training and validation (a) For single-limb activity (b) For multi-limb activity.

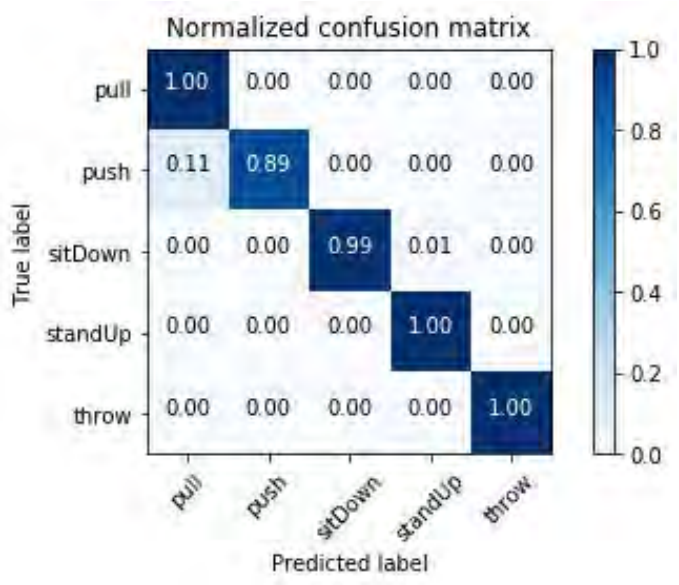


(a)

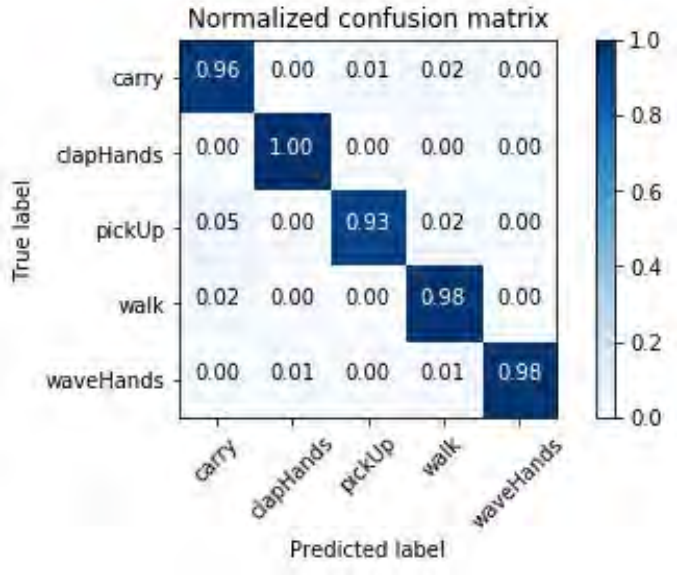


(b)

Fig. 7. Accuracy curves of 2DConvNet (a) shows accuracy curve of single-limb activity (b) shows accuracy curve of multi-limb activity.

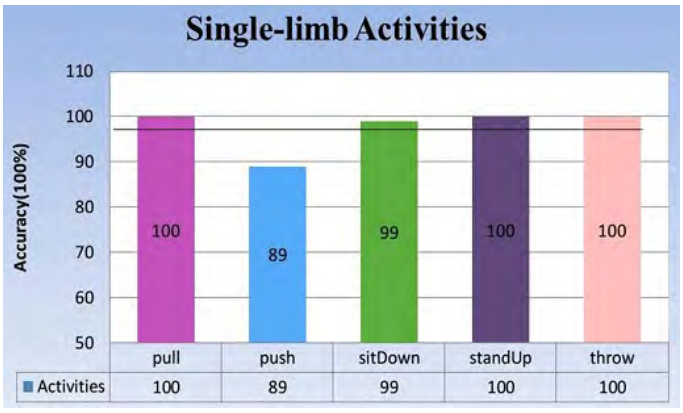


(a)

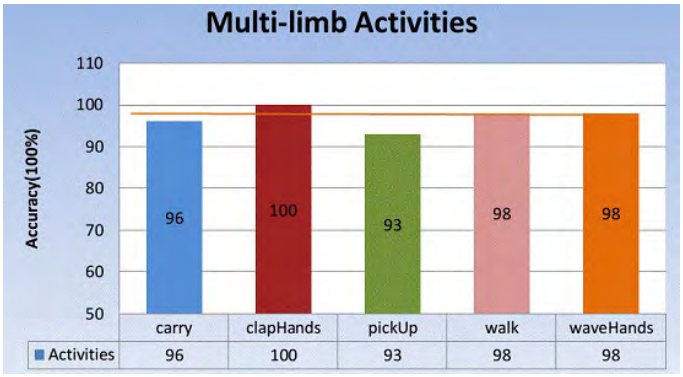


(b)

Fig. 8. Confusion matrix of 2DConvNet model (a) For single-limb activities (b) Multi-limb activities.



(a)



(b)

Fig. 9. Activity recognition performance for each activity class: (a) single limb activities (b) multi limb activities.

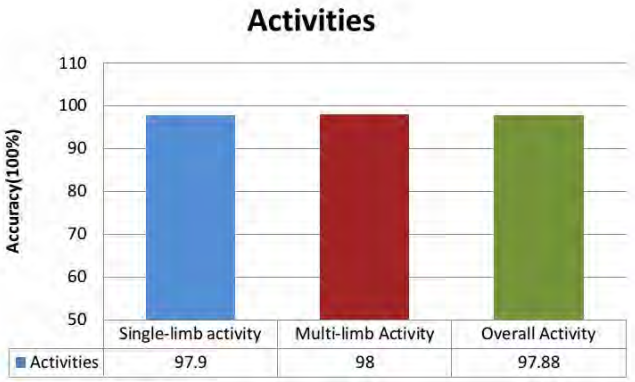


Fig. 10. Comparative performance analysis between recognition rates of single-limb and multi-limb human activities along with complete system performance..

VI. Results and Discussions

This section presents the obtained experimental results and discussions about them. In this paper we adopted two-stage classification in which we obtained the results in two phase called coarse and fine level classification. First stage classification divides the activity types in two categories, single-limb and multi-limb respectively. Second stage classification has been performed to recognize actual type of activities in both. Finally a randomly train test split method has been used to validate our proposed approach.

A. First Stage Classification Results

In the first level classification, human activity types are categorized into two type's, single-limb and multi-limb activities. This classification makes easy the process of recognition at the second level. In this phase classification has been performed by Random Forest classifier changing the number-of-trees in the forest. It is illustrated from the confusion matrix given in Fig. 5 that all test samples have been classified correctly into single-limb and multi-limb classes except for one sample and 99.92% is the classification accuracy.

B. Second Stage Classification Results

In the second stage, actual recognition of individual activities takes place. Recognition has been performed using human activities for both single-limb and multi-limb categories. The overall recognition accuracy of single-limb activities is 97.9%. Individual activity classes named throw, standup and pull have been recognized 100 percent correctly, and for the remaining two activities push and sitdown the obtained accuracy is 89% and 99 %, respectively. On the other hand the

overall recognition accuracy of multi-limb activities is 98% in which claphands activity is recognized 100% and the remaining individual classes named carry, pickup, walk and wavehands have 96%, 93%, 98% and 98% accuracies, respectively. It has also been illustrated from Fig. 10 that the overall accuracy of the system is 97.88 %, which is better than some of the previous state-of-art results.

The approach proposed in the paper is better than the existing methods in the sense that in a two-stage strategy the larger class problem may be divided into the sub class problems where individual sub classes are recognized at next sub level in downward stream, since real time recognition is difficult if larger number of classes are taken into account. Hence, the problem of large classes may be improved by dividing the classes into two or more levels. The recognition performance of multi-level classification depends on the training losses incurred at each level. Therefore we try to minimize the training losses at each level of classification. For example the author in [49] performed classification task on EEG signals dataset by considering the whole dataset at a time and got 39.34% accuracy, but when they used two-stage strategy (coarse-fine level classification) they got 85.20 % accuracy at coarse level and 65.03 % at fine level and combined accuracy was 57.11% (85.20 % * 65.03 %), a major increment in accuracy by a factor of 17.77% using two-stage classification.

C. Error Analysis

The proposed human activity recognition system is working in two stages. In the first stage, the proposed system is classifying single limb and multi limb activities with an accuracy of 99.92% which is approximately 100% except only one sample whose actual class was single-limb, which was being predicted as multi-limb.

In the next stage, after classifying single and multi-limb activities when each individual category is being recognized, 11% of the push activity in single limb was erroneously recognized as pull, owing to the similarity in both activities because the captured video frames show similarity during these two activates. Simultaneously, 1% error was detected (during sit down activity) as standup activity due to the high degree of similarity between the two activities.

While in case of multi-limb activity, an error of 4% was found in carry activity, which is misclassified 1% as pick up and 2% as walk, because carry activity is somewhat a combination of walk and pick up activities. Similarly, 7% of misclassification error was found in the pickup activity, being 5% of instances predicated as carry and 2% as walk, because in pickup activity video frames the subject is carrying some goods and some pickup activity video frames are similar to walk activity. At the same time, an error of 2% was found in walk activity and model predicted it as a carry activity because both activities contain the motion information. Similarly 2% misclassification was in wave hands activity which is predicted 1% as wave hands activity and 1% as walk activity due to the similarity between them.

D. Comparison with State-of-art

We compared classification-accuracy of our proposed system with other approaches given in previous methodologies. The result comparison has been shown in Table II. Our method achieved the highest accuracy among the methods given in Table II. Starting from [29] where the authors have taken human posture as histogram of 3D joints (HOJ3D) as a novel descriptor and got 90.92% classification accuracy while our two-stage strategy produces 97.98 % on the same dataset. Liu et al. [46] used both RGB and depth information of human activities and fussed this information together with coupled hidden conditional random field model and generated 92% accuracy. Zhao. et al. [47] used raw depth sequences, depth motion map and RGB information and fussed together all this information and applied 3DSTCNN with SVM for human action recognition. The proposed approach in [48] produces 97.29% accuracy on UTKinect-Action 3D dataset, 94.15% accuracy on MSR-Action 3D dataset. Vemulapalli. et al. [48] used 3D geometry of different body parts using translation and rotation in 3D space and generated 97.08% accuracy on the UTKinect-Action 3D dataset. It is clear that our proposed approach generates good results on the UTKinect-Action 3D dataset as compared to the methods given in Table II. Thus our methodology advances some of the methodology as discussed above.

TABLE II. Performance Comparison with Other Methods on UTKinect-Action3D Dataset

Methods	Accuracy
Xia. et al., (2012) [29]	90.92 %
Liu et al., (2015) [46]	92.00 %
Zhao. et al., (2019)[47]	97.29 %
Vemulapalli. et al., (2014) [48]	97.1 %
Proposed Approach	**97.88%**

Based on the comparison with previous state-of-art results discussed in Table II our proposed approach has some advantages which are mentioned below:

- The multi-stage method facilitates the classification task by reducing large training losses given in complex problems into the low training losses given at the different levels.
- Complexity of the system may be reduced by making multi stages.
- Better recognition using initial and subsequent levels.
- Suitable for human computer interaction application.

Although our proposed multi-stage strategy has good results as compared to the state-of-art results given in Table II, it also has some limitation as the system will produce good results if and only if the classification accuracy at the initial stages is high.

VII. Conclusion

This paper presents a novel framework to recognize human single-limb and multi-limb activities using video frames. This framework facilitates to analyze human limb activities in real-time. The recognition process has been done in two stages. Firstly a Random Forest classifier has been used to distinguish input activities into two classes of activities, such as human single-limb and multi-limb. In the second phase, two 2D Convolution neural network classifiers have been trained for recognition of separated activities using a sequence classification based approach. The UTKinect-Action Dataset of 199 activities sequences has been used by the proposed framework. An accuracy of 99.92% was achieved using the Random Forest classifier. An overall accuracy of 97.88% has been recorded by our system for both types of activity classes. The major components of this proposed approach are real time, computation of HOG feature and classification. Obtained experimental results show the major advantage of deep convolution neural network implementation in activities recognition. This work also proposes the advantages of applying RGB information to recognize human activity types. In future work, Depth frames and Skeleton joints data may be combined with RGB information to form a large amount of data and generate a robust approach for better human activity recognition.

Acknowledgment

We are thankful to Uttarakhand Technical University Dehradun for maintaining the research facilities for this work. We also give special thanks to Microsoft Research team to maintain UTKinect-Action Dataset repository to complete this research work. We also thanks to

Dr. Pradeep Kumar Postdoctoral Fellow at UNB Canada for revising and improving this manuscript

References

[1] L. Breiman, "Random forests". Machine Learning, vol. 45, Jan. 2001 pp. 5-32.

[2] N. Haering, P.L. Venetianer and A. Lipton, "The evolution of video surveillance: an overview", Machine Vision and Applications, vol. 19 May 2008, pp. 279-290.

[3] W. Hu, T. Tan, L. Wang and S. Maybank, "A survey on visual surveillance of object motion and behaviors", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol 34, 2004, pp. 334-352.

[4] I.S. Kim, H.S. Choi, K.M. Yi, J.Y. Choi and S.G. Kong, "Intelligent visual surveillance a survey", International Journal of Control, Automation, and Systems, vol. 8, 2010, pp. 1598- 6446.

[5] T.Ko, "A survey on behavior analysis in video surveillance for homeland security applications", in: 37th IEEE Applied Imagery Pattern Recognition Workshop, 2008 (AIPR 08), October 2008, pp. 1-8.

[6] O.P. Popoola and K. Wang, "Video-based abnormal human behavior recognition review", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 42, 2012, pp. 865-878.

[7] K. K. Verma, P. Kumar and A. Tomar, "Analysis of moving object detection and tracking in video surveillance system", In 2015 IEEE 2nd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 1758-1762.

[8] P. Geetha, V. Narayanan, "A survey of content-based video retrieval", Journal of Computer Science, vol. 4 2008, pp. 474-486.

[9] J. Zhang, Z. Liu, (2008, June). "Detecting abnormal motion of pedestrian in video", In IEEE International Conference in Information and Automation ICIA, 2008, pp.81-85.

[10] C. C. Hsieh, and S. S. Hsu, "A simple and fast surveillance system for human tracking and behavior analysis." In Third IEEE International Conference on Signal-Image Technologies and Internet Based System SITIS'07, December 2007, pp. 812-818.

[11] J.K. Aggarwal and Q. Cai, "Human motion analysis: a review", Computer Vision and Image Understanding, vol. 73, 1999, pp. 428-440.

[12] J.K. Aggarwal, Q. Cai, W. Liao and B. Sabata, "Non rigid motion analysis: articulated and elastic motion", Computer Vision and Image Understanding, vol. 70, 1998, pp. 142-156.

[13] X. Ji and H. Liu, "Advances in view-invariant human motion analysis: a review", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 40, 2010, pp. 13-24.

[14] R. Poppe, "Vision-based human motion analysis: an overview", Computer Vision and Image Understanding, vol. 108, 2007, pp. 4-18.

[15] W. Wei and A. Yunxiao, "Vision-based human motion recognition: a survey", in: Second International Conference on Intelligent Networks and Intelligent Systems, 2009 (ICINIS09), November 2009, pp. 386-389.

[16] H. Buxton, "Learning and understanding dynamic scene activity: a review", Image and Vision Computing, vol. 21, 2003, pp. 125-136.

[17] M. Del Rose and C. Wagner, "Survey on classifying human actions through visual sensors", Artificial Intelligence Review, vol. 37, 2012, pp. 301-311.

[18] M. Pantic, A. Pentland, A. Nijholt and T. Huang, "Human computing and machine understanding of human behavior: a survey", in: Proceedings of the 8th International Conference on Multimodal interfaces, ICMI 06, ACM, New York, NY, USA, 2006, ISBN 1-59593-541-X, pp. 239248.

[19] W. McNally, A. Wong, and J. McPhee, "STAR-Net: Action Recognition using Spatio-Temporal Activation Reprojection", 16th IEEE Conference on Computer and Robot Vision (CRV), 2019, pp. 49-56.

[20] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. "Large-scale video classification with convolutional neural networks". In Proc. CVPR, 2014, pp. 1725-1732.

[21] G. W. Taylor, R. Fergus, Y. LeCun, and C. Bregler. "Convolutional learning of spatio-temporal features". In Proc.ECCV, 2010, pp. 140-153.

[22] Tran, L. Bourdev, R. Fergus, L. Torresani, and M.Paluri. "Learning spatiotemporal features with 3D convolutional networks". In Proc. IEEE International Conference on Computer Vision (ICCV), 2015, pp. 4489-4497.

[23] L. Xue, S. Xiandong, N. Lanshun, L. Jiazhen, D. Renjie, Z. Dechen, and C. Dianhui, "Understanding and Improving Deep Neural Network for Activity Recognition" .arXiv preprint arXiv: 1805.07020, 2018.

[24] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: "A unified embedding for face recognition and clustering". In Proc. IEEE International Conference on Computer Vision and Pattern Recognition CVPR, 2015, pp. 815-823.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet classification with deep convolutional neural networks". In Advances in Neural Information Processing System (NIPS), 2012, pp. 1097-1105.

[26] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos". In Advances in Neural Information Processing System (NIPS), 2014, pp. 568-576.

[27] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions". In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1-9.

[28] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using convolutional networks", In IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 648-656.

[29] L. Xia, C. C. Chen, and J. K. Aggarwal, "View invariant human action recognition using histograms of 3d joints". In 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, June 2012, pp.20-27.

[30] L. Wang, Y. Xu, J. Cheng, H. Xia, J. Yin, and J. Wu, "Human action recognition by learning spatio-temporal features with deep neural networks". IEEE access, vol. 6, pp. 17913-17922.

[31] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A.Oliva, "Learning deep features for scene recognition using places database", in Proc. Advances in Neural Information Processing System, 2014, pp. 487-495.

[32] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2014, pp. 580-587.

[33] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deep Face: Closing the gap to human-level performance in face verification", in Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701-1708.

[34] J. Donahue, A. Hendricks, L. Guadarrama, S. Rohrbach, M. Venugopalan, S. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description", in Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2625-2634.

[35] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition", IEEE Transaction on Pattern Analysis Machine Intelligence, vol. 35, no. 1, pp. 221-231, Jan.2010.

[36] L. Wang, Y. Qiao, and X. Tang, "Action recognition with trajectory-pooled deep-convolutional descriptors", in Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4305-4314.

[37] L. Wang, Y. Xiong, Z. Wang, and Y. Qiao, "Towards good practices for very deep two-stream convnets, CoRR", In IEEE International Conference on Computer Vision and Pattern Recognition, July 2015, pp. 1-5.

[38] H. Bilen, B. Fernando, E. Gavves, A. Vedaldi, and S. Gould, "Dynamic image networks for action recognition", in Proc. IEEE Conference on Computer Vision and Pattern Recognition., 2016, pp.3034-3042.

[39] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition", in Proc. IEEE Conference on Computer Vision and Pattern Recognition., 2016, pp. 1933-1941.

[40] X. Peng, L. Wang, X. Wang, and Y. Qiao, "Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice", Computer Vision and Image Understanding, vol. 150, 2016, pp. 109-125.

[41] X. Peng, C. Zou, Y. Qiao, and Q. Peng, "Action recognition with stacked fisher vectors". In Proc. European Conference on Computer Vision (ECCV), 2014, pages 581-595.

[42] H. Wang and C. Schmid, "Action recognition with improved trajectories, In Proc. International Conference on Computer Vision (ICCV), 2013, pp. 3551-3558.

[43] H. Wang, A. Klser, C. Schmid, and C.-L. Liu, "Action recognition by dense trajectories", In Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 3169-3176.

[44] N. Dalal, B. Triggs, and C. Schmid, "Human detection using oriented histograms of flow and appearance", In Proc. European Conference on

Computer Vision (ECCV), 2006, pp. 428-441.
[45] L. Sun, K. Jia, D.-Y. Yeung, and B. Shi, "Human action recognition using factorized spatio-temporal convolutional networks", In Proc. International Conference on Computer Vision (ICCV), 2015, pp. 4597-4605.
[46] A. A. Liu, W. Z. Nie, T. Su, L. Ma, T. Hao, and Z. Yang, "Coupled hidden conditional random fields for RGB-D human action recognition." Signal Processing, vol. 112, 2015, pp. 74-82.
[47] C. Zhao, M. Chen, J. Zhao, Q. Wang, and Y. Shen, "3D Behavior Recognition Based on Multi-Modal Deep Space-Time Learning." Applied Sciences, vol. 9.no. 4, 2019, pp. 7-16.
[48] Vemulapalli, Raviteja, Felipe Arrate, and Rama Chellappa. "Human action recognition by representing 3d skeletons as points in a lie group." Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 588-595.
[49] P. Kumar, R. Saini, P. P. Roy, P. K. Sahu, and D. P. Dogra, "Envisioned speech recognition using EEG sensors." Personal and Ubiquitous Computing, 2018, vol. 22, no. 1, pp. 185-199.
[50] P. Sewaiwar and Kamal Kant Verma. "Comparative study of various decision tree classification algorithm using WEKA." International Journal of Emerging Research in Management & Technology, vol. 4 2015, pp. 2278-9359.
[51] Siirtola, Pekka, and Juha Röning. "Revisiting" Recognizing Human Activities User-Independently on Smartphones Based on Accelerometer Data"-What Has Happened Since 2012?" International Journal of Interactive Multimedia & Artificial Intelligence, 2018, vol. 5, no. 3, pp. 17-21.
[52] A. Jalal and S. Kamal. "Improved behavior monitoring and classification using cues parameters extraction from camera array images." International Journal of Interactive multimedia and Artificial Intelligence, 2018, vol. 5, no. 3, pp. 2-18.

Kamal Kant Verma

Kamal Kant Verma is a research scholar in Uttarakhand Technical University Dehradun. He is currently working as an Assistant Professor in the Department of CSE, College of Engineering Roorkee Roorkee Uttarakhand India. He has 13 years of Teaching and research experience. His research area is Computer Vision, Video Processing, Machine Learning, and Deep Learning.

Brij Mohan Singh

Brij Mohan Singh is Dean Academics & Professor in Department of CSE, COER Roorkee. He has published more than 35 research papers in International Journals such as Document Analysis and Recognition-Springer, CSI Transactions on ICT-Springer, IJIG-World Scientific, IJMECS, EURASIP Journal on Image and Video Processing etc. His research areas are Digital Image Processing and Pattern Recognition. He has guided 3 PhD Thesis of Uttarakhand Technical University (UTU) Dehradun India and currently 6 are in process.

H. L. Mandoria

H. L. Mandoria is Professor & Head in Department of IT, Govind Ballabh Pant University of Agriculture and Technology, Pantnagar. He has published more than 70 research papers in International Journals and conferences. His research area is Computer Network, Information Security and Cyber Security. He has guided more than 20 M.Tech and presently three phd are in progress.

Prachi Chauhan

Prachi Chauhan is a research scholar in Govind Ballabh Pant University of Agriculture and Technology, Pantnagar. Her research area is Cyber Security, Machine Learning, and Deep Learning.

Guidelines for performing Systematic Research Projects Reviews

Alicia García-Holgado*, Samuel Marcos-Pablos, Francisco José García-Peñalvo

GRIAL Research Group, Computer Science Department, University of Salamanca (Spain)

Received 14 April 2020 | Accepted 19 May 2020 | Published 27 May 2020

Abstract

There are different methods and techniques to carry out systematic reviews in order to address a set of research questions or getting the state of the art of a particular topic, but there is no a method to carry out a systematic analysis of research projects not only based on scientific publications. The main challenge is the difference between research projects and scientific literature. Research projects are a collection of information in different formats and available in different places. Even projects from the same funding call follow a different structure in most of the cases, despite there were some requirements that they should meet at the end of the funding period. Furthermore, the sources in which the scientific literature is available provide metadata and powerful search tools, meanwhile most of the research projects are not stored in public and accessible databases, or the databases usually do not provide enough information and tools to conduct a systematic search. For this reason, this work provides the guidelines to support systematics reviews of research projects following the method called Systematic Research Projects Review (SRPR). This methodology is based on the Kitchenham's adaptation of the systematic literature review.

Keywords

Guidelines, Method, Research projects, Reviews, Systematic Review.

DOI: 10.9781/ijimai.2020.05.005

I. Introduction

THE number of scientific articles published, regardless of the academic discipline, has dramatically increased in the last decades. The publication in impact journals is considered one of the KPI (key performance indicators) in research centres and one of the measures to get funds. Moreover, in the current information society, most of the published works are available in online journals, repositories, databases, so researchers have access to them.

One of the first tasks before conducting a research, regardless of the field of study, is to identify related works and previous studies as a way to support the need to conduct new research on a particular topic. Likewise, the review of available research provides answers to particular research questions and a knowledge base to learn from previous experiences and identify new research opportunities. Nevertheless, although the need to synthesise research evidence has been recognised for well over two centuries, it was not until the end of the last century that researchers began to develop explicit methods for this form of research.

In particular, a literature review allows for achieving this objective. According to Grant and Booth [1], it involves some process for identifying materials for potential inclusion, for selecting included materials, for synthesizing them in textual, tabular or graphical form and for making some analysis of their contributions or value. There are different review types and associated methodologies. Specifically, before 1990, narrative reviews were typically used, but they have some limitations such as the subjectivity, coupled with the lack of transparency, and the early expiration because the synthetization process becomes complicated and eventually untenable as the number of studies increases [2].

The systematic review or systematic literature review method seeks to mitigate the limitations of narrative reviews. Systematic reviews have their origin in the field of Medicine and Health. Nevertheless, the logic of systematic methods for reviewing the literature can be applied to other areas of research such as Humanities, Social Sciences or Software Engineering; therefore there can be as much variation in systematic reviews as is found in primary research [3], [4].

A systematic review is a protocol-driven comprehensive review and synthesis of data focusing on a topic or related key questions. It is typically performed by experienced methodologists with the input of domain experts [5]. The systematic review methods are a way of bringing together what is known from the research literature using explicit and accountable methods [4]. According to Kitchenham [6]-[8], a systematic review is a means of evaluating and interpreting all available research relevant to a particular research question, topic area, or phenomenon of interest by using a trustworthy, rigorous, and auditable methodology.

The analysis of related works and previous studies is not only associated with scientific literature. Another KPI in research centres is the number of projects funded in competitive calls. Project proposals, like other formal studies, have to justify the need to conduct them. Furthermore, most of the calls for funding projects require to justify the innovation of the proposal against other developed projects.

Although it might be expected that the results of all funded projects are available in scientific publications, this is not always the norm. Determining the progress made through a research project requires the

* Corresponding author.
E-mail address: aliciagh@usal.es

analysis of the project itself, not only through its scientific publications but also through the information available (implementation details, results, etc.). Besides, as in any scientific study, it is important to ensure that the new project proposal does not repeat work previously done, or a research design found previously to be ineffective.

There is no established methodology that allows carrying out a systematic analysis of the studies and progress made through research projects in a specific area or topic. According to a systematic review conducted, even though there exists literature related to projects' review, the process is not fully systematised.

The research projects generate many types of results, and scientific publications are a small part of them. There are initiatives to facilitate open access to scientific data at national levels [9], and increased efforts to co-ordinate and support global data networks are necessary [10], but not all the projects are data-driven and data sets are not the only results in data-driven projects. At European level, solutions such as OpenAIRE supports the access to open science connecting open repositories across Europe, including Zenodo which enables the deposit of research documents, research data and software. However, not all European funding calls include sharing the results in OpenAIRE as a requirement.

The main problem to implement a systematised review process with research projects is their differences with the scientific literature. The research projects are a compilation of several types of documents, data sets, software, with different purposes. Even in projects inside the same funding call, that have to follow the same guidelines and meet the same requirements at the end of the funding period, the information available is entirely different. In fact, most of the information generated inside a research project is not publicly accessible, although it depends on the regulations of the funding call. Moreover, unlike the scientific literature, not all the research projects are available in accessible databases. Likewise, the project databases usually do not enhance projects findability with metadata, and the search tools provided are simple.

This work aims to present a set of guidelines to support systematic reviews of research projects as a way to summarise, synthesize, critique, and use that information to identify trends and lacks, justify the innovation of a new research project or collect valuable results that could be applied in other context beyond the project in which they were developed.

The paper has been divided into five sections. The second section describes a review of other works focused on methods and guidelines to carry out systematic reviews, as a way to justify the added value of this work. The third section describes the set of guidelines to apply systematic reviews of research projects. The fourth section describes two examples in which the guidelines were applied. Finally, last section summarises the main conclusions of this work.

II. Related Works

A. Identifying the Need of a Systematic Methodology

Literature reviews, and more specifically Systematic Literature Reviews (SLR) have reached a considerable level of adoption in many research fields. A literature review is a search and evaluation of the available literature in a given topic, providing state of the art about previous research and giving an overview of what are the strengths of the area of interest and which weaknesses need further improvement. Different methodologies have been proposed for performing literature reviews but most of them share the same base structure defined by the SALSA (Search, AppraisaL, Synthesis and Analysis) framework [11].

Within this framework, the aim of the Search phase is to gather a preliminary list of publications to analyse. During the Appraisal stage the papers collected during the previous stage are analysed, eliminating those that are irrelevant. Finally, the Synthesis and Analysis stages extract relevant data from the selected papers and draw conclusions from them in order to produce a final report.

In order to investigate the previous works carried out in the field of the systematic review of research projects, we have performed a systematic review of the related literature. The following subsections describe the process undertaken, which follows the recommendations of Kitchenham [7] and Petersen [12] regarding the methodologies for conducting systematic literature reviews and mapping studies.

1. Database Selection

In terms of the information sources of the papers to be included in the search process, we selected the Web of Science and Scopus electronic databases, as they fulfil the following requirements:

- The database is available for us through our institution.
- The database can use logical expressions or a similar mechanism.
- The database allows full-length searches or searches only in specific fields of the works.
- The database allows additional filtering options such as publication year or publication language.

Also, as there are many entities that fund research projects in Europe through different programs and calls, we considered including them in the search for related work. The motivation for this is that during the preparation and execution of such projects it is mandatory to carry out searches for related projects. The results in the form of reports and publications are published in different databases and webpages. However, the primary public repository and portal to disseminate information on all EU-funded research projects and their results is the Community Research and Development Information Service (CORDIS), (https://cordis.europa.eu). This database fulfils the following requirements:

- Many of the results are publicly available.
- It is a reference database in the research scope.
- It allows using a search string equal or similar to the ones used in the selected scientific databases.
- Provides means to filter the obtained results.

2. Inclusion Criteria

A set of inclusion criteria (IC) was defined to select those works that are relevant in the scope of the considered related work:

- IC1: The document focuses on a systematic review of research projects AND
- IC1: The document describes the followed review methodology AND
- IC2: The document is written in English AND
- IC3: For research paper, it was published in peer-reviewed Journals, Books, Conferences or Workshops OR
- IC4: For project reports, they were publicly available through CORDIS as part of the peer-reviewed project results

3. Query String

To create the search string, we identified the main terms related to the review scope and the possible alternative spellings and synonyms. Based on them the employed query for all the databases was:

("systematic review method" OR "systematic review methods" OR "systematic review guideline" OR "systematic review guidelines") AND (projects OR "R&D")

4. Review Process

After defining the sources, search string and appraisal criteria, the following steps were followed for the literature review:

1. Collect raw results in two different spreadsheets for the scientific and CORDIS databases (https://bit.ly/3dRkLbn, https://bit.ly/2z2IA1a). After removing all the duplicates across the databases, we obtained 509 results.
2. Analyse the resultant documents based on the title and abstract and the inclusion/exclusion criteria. In those cases where the title and abstract were not sufficient to decide, the authors quickly assessed the entire content of the paper. The resultant candidate papers (118) were added to another two spreadsheets (https://bit.ly/2WFc9Pi, https://bit.ly/2z2IC9i).
3. The documents then passed a quality assessment, looking for those that clearly describe a methodology for performing Systematic Research Projects Reviews. Results were collected in https://bit.ly/3bIAGHE and https://bit.ly/2LzA3FJ.

5. Results

After following the abovementioned systematic search, appraisal and analysis of documents, it was observed that the majority of papers focused on systematic methods for scientific literature reviews of research documents. However, we did not find any document in the form of research paper nor project deliverable or report in the consulted databases that fully describes a systematic methodology in order to review research projects.

B. Identifying Reviews of Projects that Partially Follow the SALSA Framework

As we did not find any systematic methodology proposal for reviewing research projects, we performed a second literature review. The objective was to gain a better perspective on other authors' approaches to systematic reviews of projects, in terms of how they had followed any of the steps described within the SALSA framework.

We opted for broadening the search and softening the inclusion and quality criteria in order to find proposals that had made a systematic search of projects, even if the purpose is not to propose a unified method of systematization of the review. The search also included the grey literature, provided that it was available. The process undertaken was as follows:

1. We decided to conduct this second search in the Google Scholar database and employed the search string: "systematic review of projects". The search returned 3,390,000 results.
2. We further filtered the results using the advanced search options and limited the output to documents: "with the exact phrase | anywhere in the article: systematic review of projects ". After applying the advanced filtering, we obtained 99 results.
3. The inclusion criteria were limited to documents that focused on a systematic review of research projects, even though they do not fully describe the methodology. Also, we removed duplicates,

TABLE I. Summary of the Application of the SALSA Framework to the Selected Documents

Ref.	SALSA			
	Search	**Appraisal**	**Synthesis**	**Analysis**
[11]	Scrutinized AMIF and ESF databases (related to integration, social inclusion of migrants and refugees into the labour market)	Period was limited to January 2014 through May 2019. Used words in advanced search: e.g. "Migrant", "refugee", "integration"	Not described	Organize projects (calculate % of the total) by keywords, regions, topics, integration level
[13]	Different databases: sector-specific databases, NGO websites, and peer-reviewed academic literature and gray literature. Used different search terms for each database	Not described	Created two indices (ASI, MSI) and 4 criterions: if criterion is present => scored as 1.0. If referred to without explicit identification => scored as 0.5. For the ASI, the four scores were summed together, and for the MSI they were multiplied	Kruskal—Wallis nonparametric one-way analysis of variance tests to test for differences in ASI and MSI values
[14]	All 152 projects from the CSHGP database	Projects that began after October 1, 2000; had some level of effort for malaria, pneumonia, and/or diarrhoea; provided curative interventions; and had project documentation with CCM and contextual detail	Tabulated 11 self-made-items (indicators) fulfilled with quantitative and qualitative information	Calculated the "indicator yield" across projects (proportion of projects measuring a given indicator) and "indicator density" within projects (proportion of recommended indicators measured by a given project)
[15]	Self-created database of projects through consultation with the study oversight panel	Describe a list of criteria to select the appropriate cases, including project characteristics, reviewer, project stage, number of review comments and geographic distribution	Define a set of quantitative metrics: benefit from project level evaluation, estimated savings, etc. and qualitative metrics: "extracted from review comments of the project"	Probabilistic and regression analysis
[16]	Review programs from agencies, reports and a Lesson Learned database. Not systematic	Not described	Authors define a set of categories to be identified within the projects	Identify the most frequently used categories within the existing practices (in %)
[17]	A systematic review of projects in all county councils in Sweden was performed	The final sample consisted of documents from the planning and design process and the target organizations' strategic operational plans. Overall, 45 various documents were reviewed from five building projects that had a budget of over 50 million € (US$5,493,000.00) and were executed between 2010 and 2014 in Sweden	Qualitative content analysis. Data from documents were organized into a matrix with five different levels of headings: meaningful unit, condensing meaningful unit, code, subcategory, and category	Not described

quotes, not available documents and those not written in English, 68 final references were obtained.

4. The resultant documents were read in detail and analysed in order to assess which of the steps of the SALSA framework were systematized and described. The results from this final step can be further consulted in https://bit.ly/3g2Zu0D. Likewise, the Table I summarises the documents which fulfil the SALSA framework.

It can be observed that none of the studied documents completely describe a systematic process throughout all the SALSA framework. In terms of the search stage, only one document fully describes the database selection process and criteria, along with the employed search string and its motivation [13]. Only two documents, [14] and [15], describe a systematic procedure for the appraisal, synthesis and Analysis steps. However, in those cases the search phase is either performed over all documents of a particular database [14], or performed over a self-created database which construction process is not detailed in the document [15]. On the other hand, in [11] the appraisal and analysis of documents is systematised and properly described. However, the search phase is vaguely mentioned, and the synthesis phase is omitted. In [16] authors define a set of categories during the synthesis phase and calculate their percentage of occurrence in the considered projects during the analysis stage. However, in this case, the search phase is only partially described and there is no appraisal phase to systematise and validate the selection of considered projects. Finally, the work of [17] follows SALSA methodology during the search appraisal and synthesis stages, but omits analysis of the data obtained.

As a conclusion of the systematic process of reviewing the related work, we can conclude that no works were found that describe a methodology or procedure for the systematic review of research projects. Also, even though there exists extensive literature related to projects' review, it is not fully systematised, and although some works were found which take into account some of the SALSA framework steps, none of them completely describes the SALSA framework.

III. The Method

The systematic review of research projects, also called Systematic Research Projects Review (SRPR), is based on the Kitchenham's adaptation of the systematic literature review (SLR) [6]-[8] and the Petersen's proposal to carry out systematic mapping studies [12], [18]. The high number of projects that are developed annually makes it impossible to analyse all of them. SRPR enables the selection of a set of projects that fulfil a particular criterion; likewise, SLR facilitates the review of scientific papers that would be impossible to handle otherwise. The main objectives that can be achieved through SRPR are:

- To identify trends in research projects.
- To identify lacks to define new research projects.
- To justify the innovation of new research projects proposals.
- To collect valuable results that could be applied in another context beyond the project in which they were developed.

This work does not seek to reinvent the protocols of the systematic review, but to adapt them in order to review the compendium of resources, documents, information, which form a research project. Table II shows the main differences at the macro level between a SLR and a SRPR.

The SRPR is divided into four phases with a set of steps inside each one. These phases are quite similar to those defined in the SLR, but the main differences are related to the implementation of the steps. The first and second phases are focused on the definition of the review, and the third and fourth phases are related to the retrieval, appraisal, synthesis and analysis of the research projects. Fig. 1 shows the main phases and steps that compose the SRPR method.

TABLE II. Main Characteristics of SRPR versus SLR

	SLR	SRPR
Context	Publications	Research projects
Sources	Databases prepared to support searches and metadata	Databases are not always available and are heterogeneous with no support to search or metadata sometimes
Review	Review process focused on reading	Review process focused on searching resources, documents, publications

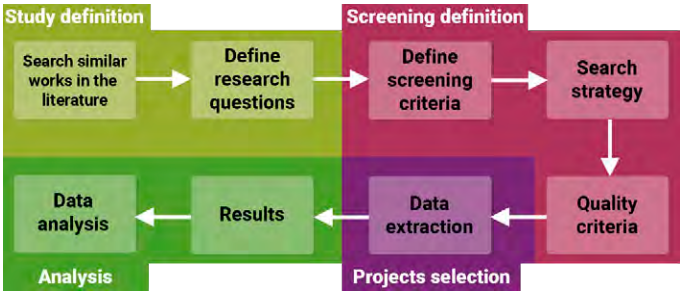


Fig. 1. Definition and implementation phases of the SRPR method.

A. Study Definition

The first phase is focused on the preparation of the research projects' review. In particular, it is focused on the objective of the SRPR. This phase does not differ from a typical systematic literature review. It is composed of two steps: identifying the need to conduct the review and defining the research questions.

Before carrying out a systematic review, regardless of whether it is scientific literature or projects, we must ensure that it is necessary to carry out the review. Two questions arise in the first step:

- Is there already a systematic review that pursues the same objectives?
- Are the necessary resources (people, time, money) available to carry out the review?

According to Petticrew and Roberts [19], it has no sense to conduct a systematic review that has already been done unless the previous systematic reviews are biased or outdated. These guidelines also apply for research projects.

Regarding the second question, an SRPR usually implies more time and effort per reviewer than a SLR, due to the intrinsic characteristics of the research projects. It is recommended to involve a minimum of two reviewers working together, at least, in the second and third phases.

The second step analyses the objective of the review through the PICOC framework [19] to formulate a set of research and mapping questions. In particular, PICOC stands population, intervention, comparison, outcomes and context:

- Population (P): the scope of the projects (i.e. local, regional, national, international) and the main topic of the projects (i.e. mental health, educational technology, gender gap, etc.).
- Intervention (I): the intervention applied in the research projects.
- Comparison (C): to which the intervention is compared. For example, a comparison between various calls for projects or between nationally and internationally funded projects.
- Outcomes (O): what the review seeks to achieve, such as identifying trends or lacks or selecting a set of project results focused on a particular objective.
- Context (C): the context of the research projects must be defined, which is an extended view of the population, including in which

sectors are developed, i.e. academia or industry.

Regarding the questions, their definition depends on the objective of the review. Despite this, the SRPR provides a set of meta-questions for defining the mapping questions:

- What are the trends?
- What types of institutions are involved in the projects?
- In which countries were the projects implemented?
- Which calls fund this kind of research projects?
- Which years cover the projects?
- How much money was invested in the projects?
- In which contexts are the projects carried out?
- Which kind of outcomes are provided by the projects?

B. Screening Definition

The second phase continues the preparation of the research projects' review. In particular, it covers the protocol definition through the identification of the inclusion, exclusion and quality criteria (screening criteria) and devises search strategy. The main differences between SRPR and SLR during the screening definition falls on the search strategy and the quality criteria.

First, it is necessary to define the inclusion and exclusion criteria in order to apply then during the appraisal of the research projects. In particular, these criteria are applied to the title, keywords, and the summary of the project. For this reason, these criteria have to take into account information that could be available in most of the sources, such as the objective of the projects, the publication dates or the official language of the project.

The search strategy is divided into three tasks: define the search terms, build the search string and select the sources. The definition of the search terms follows the same process as an SLR. The PICOC allows identifying the main terms, and this first cohort is extended with synonyms and words used in the scientific context to describe the same concept. Regarding the search string, it is important to keep it simple to ensure that the search string will be the same in all the selected sources. The databases of research projects do not usually have powerful search tools, so a search string that combines different logical operators and nests may be impossible to use.

The most difficult part of the search strategy is the selection of the sources. The databases of research projects are not always accessible and most of them do not have powerful search tools and do not support metadata, so the search strategy in SRPR depends on the available databases much more than the SLR. Furthermore, there is not an international database of research projects such as Scopus and Web of Science. Usually, the databases are related to particular funding calls. There is a set of recommendation to select them:

1. Identify the scope of the research projects. It is possible to combine different scopes in the same review, but it is important to ensure completeness. For example, it is possible combining national and European projects if we take into account the national projects of all the European countries, not only a part.
2. Identify the main funding calls in the selected scope. Although there are projects that obtain funds by other means, it is not possible to access them unless we get access to the research projects databases inside each institution. It is recommended to include different calls and funding periods in those reviews focused on identifying trends and lacks.
3. Identify the databases with the granted projects for each funding call.
4. Search on the Internet other databases of research projects of the same scope.
5. Navigate through the database, run the search string and apply the following requirements to the identified databases. They are an adaptation of [20]:
   - It is a reference database in the selected scope.
   - It is a relevant database in the research area of this review.
   - It allows carrying out searches and downloading the results in some accessible format.
   - It is a database available through the authors' institution or an authors' membership to an association.
6. The selected sources will be those that fulfil the requirements and return different research projects using the search string.
7. Identify and document the search filters that will be used in each database. Each database has a totally different search tool, so it is necessary to provide all the filtering details and the search string used in each database in order to ensure the replication and transfer of the review.

Finally, to complete the screening definition, it is necessary to define a set of quality criteria or quality questions. The quality criteria for projects are totally different from the criteria used for scientific literature. The information available about the projects are not always the same. Usually, the databases provide title, summary and funding information. For this reason, we must complete the information provided by the databases with information of the project available on the Internet, such as the project website or scientific papers published about the project. The quality criteria should ensure that the screened projects have enough accessible information to answer the research questions. The following quality questions can be combined with other questions more focused on the objectives and the results of the projects:

- Is the website of the project available?
- Are the outputs of the project available?
- Is there more information available about the project than the project summary in the languages you speak?
- Are there scientific publications associated with the project?

C. Projects' Selection

The third phase has only one step, the data extraction, which covers the retrieval and appraisal of the research projects. It is a screening process based on several iterations following the guidelines defined in the previous phase:

1. Search: the search strategy is applied in each selected source to collect the projects according to the search terms and the filters described per each source. In those cases, in which the search tool does not support logic operators, the OR operator will be replaced by a set of individual searches that will be combined manually. The reviewers download the search results and integrate them into a spreadsheet shared between all the reviewers involved in the process.
2. Duplicates: remove the projects that are duplicated. It is crucial to ensure that the projects are the same, not only using the title, but also the funding call and/or the summary. We have to take notes of the duplicated projects to document the process.
3. Screen summaries: the inclusion and exclusion criteria are applied at least to title and summary. Not all the databases provide the keywords. The projects that fulfil the criteria are copied in another spreadsheet.
4. Screen full project: each project that passed the inclusion and exclusion criteria is in-depth analysed using the quality criteria. It is a discovery process, in which reviewers could spend unlimited time due to the project data is not standardised, and it is not always available. This process involves a search protocol to identify the

available information of the project outside the database and to bound the time invested [21]:

a) Identify the project website.
b) Identify the final project report in both the project page included in the databases or the own project webpage (if available).
c) Identify the published content on the project website: description, deliverables, multimedia material, etc.
d) Search on the Internet more information. Perform two searches, one with the project title and one with the project reference, both in Google and Google Scholar.
e) Optionally, launch a search in the main scientific databases using the project title and the reference. Although it is possible to conduct a formal search selecting terms and following a strict protocol, the process would be too long and complex. Moreover, although Scholar search engine is not so rigorous as scientific databases, the result provided should cover a large part of the scientific articles available in the databases.

5. Snowball: it is possible including research projects identified in the information available of the selected projects, usually in their websites. These new projects are included together with the search results and we apply all the previous tasks.
6. Documenting: it is possible to include a PRISMA flow [22] to summarise the screening process.

Regarding the quality criteria, each question is answered with a score between 0 (means the quality criterion is not fulfilled) and 1 (means the quality criterion is fulfilled by the project information available). Moreover, a score of 0.5 is used in those cases in which the criterion is partially fulfilled. Likewise, we will use the symbol "-" in those quality criteria which cannot be inferred from any of the identified resources. Although it is possible to use another scale such as a Likert scale with five levels, we recommend only three levels due to the difficulty to conduct an in-depth analysis of each available document or information of a project. Only those projects which achieve a score over a previously defined minimum will be the selected projects. In a systematic literature review, this minimum should be defined before starting the full review of the papers, but in the review of research projects, it is not possible to estimate a number due to the inherent uncertainty about the information available of each project.

D. Analysis

The last phase is focused on the synthesis and analysis of the selected projects. This phase is no different from the analysis carried out in an SLR. The synthesis is achieved in the step related to the presentation of the results. We extract the main characteristics of each research project using quantitative indicators. In those cases, in which there are qualitative characteristics, a set of dimensions are identified.

Finally, we have to write up the review, including the description and overall assessment of the results found. We have to answer each research and mapping question.

IV. Examples

The guidelines for performing Systematic Research Projects Review were applied in previous works. In particular, two examples are described in this section. First, to identify trends in research projects focused on technological ecosystems in the health sector as a way to identify lacks to develop new products [23], [24]. On the other hand, as a previous step to prepare a proposal about the gender gap in STEM (Science, Technology, Engineering and Mathematics) for a European call [21], [25]. The main characteristics of the method are highlighted in the next sections.

A. Trends in Research Projects

The European Union has a strong investment in R&D and demand side-measures in the health sector. In particular, there is a research area focused on defining and developing technological ecosystems for improving different aspects of the health sector, with a particular focus on the elderly population.

In this scenario, we used the SRPR method to get an overview of the current trends and identify the lacks and opportunities in the development of the technological ecosystem in the health sector. In particular, the mapping questions were the following [24]:

- MQ1: What are the trends in the development of technological ecosystems focused on health in Europe?
- MQ2: What is the application domain of the research conducted?
- MQ3: What types of institutions are involved in the project?
- MQ4: How are the stakeholders involved in the technological ecosystems developed?
- MQ5: Which calls fund this kind of research projects?
- MQ6: Which period do the projects cover?
- MQ7: How much money was invested in these projects?

We conducted two reviews. A first review focused on finished research projects which involved different European countries and were funded in health or technology calls. This review includes projects from 2004 to 2018. Moreover, we conducted a second review which updates the previous one with the ongoing projects that started from 2015 to 2018. This allowed testing the replicability of the SRPR method. Furthermore, the comparison between both reviews allows getting an overview of the evolution of technological ecosystems in the health sector.

We conducted the reviews in three databases after a selection process which involved eight sources: Community Research and Development Information Service (CORDIS (https://cordis.europa.eu), Active and Assisted Living (AAL) programme (http://www.aal-europe.eu), and KEEP Database (https://www.keep.eu/keep/search). Noteworthy that some of the sources were identified through a mapping study about cross-border cooperation in healthcare [26], which we found searching on the Internet other databases of research projects of the same scope.

The selected sources comply with the minimum requirements, but there are big differences between the search and download tools. In particular, the AAL projects do not allow downloading or filtering, but its inclusion was justified because it is one of the most important European programs that combine health and technology. Likewise, the search string does not match in all the databases. In particular, KEEP Database does not allow logic operators, so two searches were conducted separately with the different parts of the search string and after the results were combined.

Also, the quality criteria are highlighted. Only those projects which achieved a 6 over 8 score were selected. In addition to the pre-defined quality criteria to ensure that the project has enough accessible information, a set of criteria focused on the objectives and the results of the projects were included [23]:

- Does the project provide a full definition of the ecosystem? (It implements part or the whole ecosystem)
- Was (or will be) the ecosystem developed? (Proposal, proof of concept or real system)
- Does the website show the activity of the project?
- Does the ecosystem support evolution through the integration of new components?

Regarding the projects' selection, the search provided 718 projects in the first review. After removing duplicates, 707 projects

were screened. A total of 102 projects passed to screen full project and, finally, only 19 projects were selected. Concerning the second review, the search provided 368 research projects (344 after removing duplicates). 79 projects passed the inclusion and exclusion criteria and 23 were finally selected after applying the quality criteria. The full data sets are available on http://bit.ly/2uyLeWn and http://bit.ly/2TBm8RH.

Finally, we would like to highlight two main threats to validity directly related to the SRPR method. First, the number of projects that reached the final stage after applying the quality criteria were quite low. Although we applied the search protocol to identify the available information of the project, most of the projects provide scarce or none information for answering the quality criteria. On the other hand, the selection of databases based on funding programs introduce a bias in the results because each funding program guides the scope and goals of the financed research. To reduce this bias, the SRPR includes a recommendation during the selection of sources: "It is recommended to include different calls and funding periods in those reviews focused on identifying trends and lacks".

B. Justify the Innovation

The low female participation in all areas of society is one of the Sustainable Development Goals (SDG) of UNESCO. There are some areas in which female participation is lower than others. This is the case of the STEM areas. Among the different objectives of the European Union, increase the number of women in those areas has been a priority in the last years.

The number of European research projects related to gender and STEM has increased due to the priorities of the European Union. The definition of new proposals to foster this research line requires an adequate justification of the added value with regard to other proposals. For this reason, we used the SRPR method to get an overview of the research projects funded by the European Union about the gender gap in STEM areas, both academia and industry, in the last five years. This review covers three research questions to answer following a qualitative analysis of the selected projects [21]:

- RQ1: What are the trends in Europe on the study about the gender gap in STEM?
- RQ2: Which kind of outcomes are provided by the projects?
- RQ3: Which kind of solutions or initiatives are developed?

And a set of mapping questions based on the meta-questions provided by SRPR [25]:

- MQ1: What are the trends in Europe on the study about the gender gap in STEM?
- MQ2: What types of institutions are involved in the project?
- MQ3: In which countries the project was implemented?
- MQ4: Which calls fund this kind of research projects?
- MQ5: Which years cover the projects?
- MQ6: How much money was invested in these projects?
- MQ7: In which context is carried out the study?
- MQ8: Which kind of studies – diagnosis or intervention - are developed?
- MQ9: Which kind of outcomes are provided by the projects?

In this case, the selection of the databases was also based on the previous experiences applying the SRPR method. In particular, six databases were identified based on the identified scope and funding calls. After applying the database selection requirements, we conduct the review in three databases: CORDIS, which provides projects under the H2020 framework programme; Erasmus+ Project Results Platform (https://ec.europa.eu/programmes/erasmus-plus/projects/); and KEEP, to cover cross-border projects under the different interregional and cross-border programmes. The main problem of Erasmus+ and KEEP databases is that they do not allow logic operators. For this reason, we follow the SRPR guideline related to decompose the search string to carry out several searches with different terms and apply the boolean OR operator manually.

Regarding the quality criteria, we included five criteria related to the implementation and the results of the research projects:

- Is the gender gap the main focus of the project?
- Was the study carried out in different countries?
- Does the project carry out any evaluation process focused on the gender gap?
- Does the project provide a toolkit, framework, materials focused on STEM?
- Does the activity of the project continue (or is planned to continue) after the funding period?

In this case, the minimum score was lower than the previous example (5.5 over 9), due to the low number of projects which achieve highest score. Since the quality phase does not depend so much on the quality of the projects themselves but on the information that can be found about them, the minimum limit for selecting a project is defined once the full review of the projects is completed.

Regarding the projects' selection, we collected 580 results from the selected databases and 16 projects were identified as part of the snowball task. After removing duplicates, 435 projects were screened. 84 projects passed the inclusion and exclusion criteria and 31 were finally selected. The full data set is available on http://bit.ly/2IjUJlE.

Finally, it is important to highlight the limitation of the quality criteria. The lack of information available about the research projects introduces a big limitation in the results. On the other hand, this limitation serves as a learning outcome to improve access to our research projects.

V. Conclusion

This paper has explained the SRPR method as a way to conduct systematic reviews of research projects, facing the challenges related to analysing a compendium of different heterogeneous information available in different formats. Although there are previous works related to projects' review, the review conducted as part of this paper concludes that those reviews are not fully systematised.

According to Codina [27], a systematic review has four dimensions: systematic, complete, explicitly and reproducible. The proposed SRPR method fulfils these dimensions:

- Systematic: it is not arbitrary, biased, nor subjective. The method provides guidelines for examining the best available research projects using the best sources of information.
- Complete: the guidelines to conduct an SRPR provide criteria to select sources that facilitate access to the bulk of research projects in a discipline. Moreover, the project selection phase includes the application of the screening criteria.
- Explicitly: the inclusion and exclusion criteria, the search strategy and the quality criteria are known.
- Reproducible: all the process is documented to ensure that other researchers can follow the steps and compare the results obtained to determine their accuracy or degree of success. SRPR allows including ongoing research projects, so it is crucial to ensure it is replicable to update the review when those projects finish, or new projects are developed.

This proposal will serve as a base for the definition of future

research projects, providing a rigorous method to identify the lacks in previous research projects, justify the innovation of the new projects and also reuse the results of related projects previously developed by other research teams.

Regarding the examples described, we can state that the research project reviews that follow the SRPR method are rigorous. According to Onwuegbuzie and Frels [28], by rigorous we mean conducting a literature review that is defensible (i.e. integrates a rationale for decisions of inquiry, strategies, and designs), systematic (i.e. follows a set of guidelines), evaluative (i.e. whereby every step of the process is evaluated for relevance and credibility) and transparent (i.e. documenting beliefs, values, and philosophical assumptions and stances pertaining to decisions).

Finally, a number of caveats need to be noted regarding the present study. On the one hand, the successful application of the method relies on the availability of research project databases. For example, it is not possible to conduct the SRPR at national levels, if the information about research projects funded in public and/or private calls are not available in an accessible repository or database. On the other hand, the method was used in a European context, so it would be interesting to apply it in other regions and funding calls.

Acknowledgment

This research work has been carried out within the GRIAL Research Group of the University of Salamanca [29].

This research was partially funded by the Spanish Government Ministry of Economy and Competitiveness throughout the DEFINES project grant number (TIN2016-80172-R) and the Ministry of Education of the Junta de Castilla y León (Spain) throughout the T-CUIDA project (SA061P17).

With the support of the Erasmus+ Programme of the European Union in its Key Action 2 "Capacity-building in Higher Education". Project W-STEM "Building the future of Latin America: engaging women into STEM" (Ref. 598923-EPP-1-2018-1-ES-EPPKA2-CBHE-JP). The content of this publication does not reflect the official opinion of the European Union. Responsibility for the information and views expressed in the publication lies entirely with the authors.

References

[1] M. J. Grant and A. Booth, "A typology of reviews: an analysis of 14 review types and associated methodologies," *Health Information & Libraries Journal,* vol. 26, no. 2, pp. 91-108, 2009, doi: 10.1111/j.1471-1842.2009.00848.x.

[2] M. Borenstein, L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein, *Introduction to Meta-Analysis*. UK: John Wiley & Sons, 2009.

[3] D. Gough and J. Thomas, "Commonality and diversity in reviews," in *Introduction to Systematic Reviews*, D. Gough, S. Oliver, and J. Thomas Eds. London: Sage, 2012, pp. 35-65.

[4] D. Gough, J. Thomas, and S. Oliver, "Clarifying differences between review designs and methods," *Systematic Reviews,* vol. 1, no. 1, p. 28, 2012/06/09 2012, doi: 10.1186/2046-4053-1-28.

[5] R. Russell *et al.*, "Systematic Review Methods," in *Issues and Challenges in Conducting Systematic Reviews to Support Development of Nutrient Reference Values: Workshop Summary: Nutrition Research Series, Vol. 2.*, (Technical Reviews, No. 17.2. Rockville (MD): Agency for Healthcare Research and Quality (US), 2009.

[6] B. Kitchenham, "Procedures for Performing Systematic Reviews," Keele, UK, TR/SE-0401, 2004. [Online]. Available: http://bit.ly/2yW3zCg

[7] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering. Version 2.3," EBSE-2007-01, 2007. [Online]. Available: http://bit.ly/2Kr7M6l

[8] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering – A systematic literature review," *Information and Software Technology,* vol. 51, no. 1, pp. 7-15, 2009/01/01 2009, doi: 10.1016/j.infsof.2008.09.009.

[9] OECD, "Main policygaps hindering access to data," in *Enhanced Access to Publicly Funded Data for Science, Technology and Innovation*. Paris: OECD Publishing, 2020.

[10] OECD, "Business models for sustainable research data repositories," *OECD Science, Technologyand Industry Policy Papers,* p. No. 47, 2017, doi: 10.1787/302b12bb-en.

[11] E. Taskin, "Analysis of Projects related to the Integration of Migrants," *Horizon Insights,* vol. 2, no. 2, 2019, doi: 10.31175/hi.2019.02.

[12] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Information and Software Technology,* vol. 64, pp. 1-18, 2015, doi: 10.1016/j.infsof.2015.03.007.

[13] C. M. Prager *et al.*, "An assessment of adherence to basic ecological principles by payments for ecosystem service projects," *Conservation Biology,* vol. 30, no. 4, pp. 836-845, 2016, doi: 10.1111/cobi.12648.

[14] D. R. Marsh, L. Tsuma, K. Farnsworth, K. Unfried, and E. Jenkins, "What Did USAID's Child Survival and Health Grants Program Learn about Community Case Management and How Can It Learn More?," Maternal and Child Health Integrated Program (MCHIP), 2012. [Online]. Available: https://www.mchip.net/sites/default/files/CSHGP-CCM-Report.pdf

[15] N. Stamatiadis, R. Sturgill, and K. Amiridis, "Benefits from Constructability Reviews," *Transportation Research Procedia,* vol. 25, pp. 2889-2897, 2017, doi: 10.1016/j.trpro.2017.05.275.

[16] N. Stamatiadis, R. Sturgill, P. Goodrum, E. Shocklee, and C. Wang, "Tools for Applying ConstructabilityConcepts to Project Development (Design)," University of Kentucky, Kentucky Transportation Center, Kentucky, USA, Research Report KTC -13-15/FRT190-11-1F, 2013. [Online]. Available: http://uknowledge.uky.edu/cgi/viewcontent.cgi?article=1326&context=ktc_researchreports

[17] M. Elf, G. Lindahl, and A. Anåker, "A Study of Relationships Between Content in Documents From Health Service Operational Plans and Documents From the Planning of New Healthcare Environments," *HERD: Health Environments Research & Design Journal,* vol. 12, no. 3, pp. 107-118, 2019, doi: 10.1177/1937586718796643.

[18] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," presented at the Proceedings of the 12th international conference on Evaluation and Assessment in Software Engineering, Italy, 2008.

[19] M. Petticrew and H. Roberts, *Systematic Reviews in the Social Sciences: A Practical Guide*. Malden, USA: Blackwell Publishing, 2005.

[20] A. García-Holgado and F. J. García-Peñalvo, "Mapping the systematic literature studies about software ecosystems," in *Proceedings of the 6th International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM 2018) (Salamanca, Spain, October 24-26, 2018)*, F. J. García-Peñalvo Ed., (ACM International Conference Proceeding Series (ICPS). New York, NY, USA: ACM, 2018.

[21] A. García-Holgado, S. Verdugo-Castro, C. S. González, M. C. Sánchez-Gómez, and F. J. García-Peñalvo, "European Proposals to Work in the Gender Gap in STEM: A Systematic Analysis," *IEEE-RITA,* in press.

[22] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and PRISMA Group, "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," *PLoS medicine,* vol. 6, no. 7, p. e1000097, 2009, doi: 10.1371/journal.pmed1000097.

[23] A. García-Holgado, S. Marcos-Pablos, R. Therón, and F. J. García-Peñalvo, "Technological ecosystems in the health sector: A mapping study of European research projects," *J. Med. Syst.,* vol. 43, no. 100, 2019, doi: 10.1007/s10916-019-1241-5.

[24] S. Marcos-Pablos, A. García-Holgado, and F. J. García-Peñalvo, "Trends in European research projects focused on technological ecosystems in the health sector," in *Proceedings of the 6th International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM 2018) (Salamanca, Spain, October 24-26, 2018)*, F. J. García-Peñalvo Ed., (ACM International Conference Proceeding Series (ICPS). New York, NY, USA: ACM, 2018.

[25] A. García-Holgado, S. Verdugo-Castro, M. C. Sánchez-Gómez, and F. J. García-Peñalvo, "Trends in studies developed in Europe focused on the gender gap in STEM," in *Proceedings of the XX International Conference on Human Computer Interaction*. New York, NY, USA: ACM, 2019, p. Article 47.

[26] Gesundheit Österreich Forschungs und Planungs GmbH, "Study on Cross-Border Cooperation," European Commission, Directorate-General for Health and Food Safety, Luxembourg, 2018.

[27] L. Codina. "Revisiones bibliográficas y cómo llevarlas a cabo con garantías: systematic reviews y SALSA Framework." http://bit.ly/2lKiDct (accessed 16 May, 2020).

[28] A. Onwuegbuzie and R. Frels, *7 Steps to a Comprehensive Literature Review: A Multimodel & Cultural Approach*. London: Sage, 2016.

[29] F. J. García-Peñalvo, M. J. Rodríguez-Conde, R. Therón, A. García-Holgado, F. Martínez-Abad, and A. Benito-Santos, "Grupo GRIAL," *IE Comunicaciones. Revista Iberoamericana de Informática Educativa,* vol. 30, no. 33-48, 2019.

Alicia García-Holgado

She received the degree in Computer Sciences (2011), a M.Sc. in Intelligent Systems (2013) and a Ph.D. (2018) from the University of Salamanca, Spain. She is member of the GRIAL Research Group of the University of Salamanca since 2009. Her main lines of research are related to the development of technological ecosystems for knowledge and learning processes management in heterogeneous contexts, and the gender gap in the technological field. She has participated in many national and international R&D projects. She is a member of IEEE (Women in Engineering, Education Society and Computer Society), ACM (and ACM-W) and AMIT (Spanish Association for Women in Science and Technology).

Samuel Marcos-Pablos

He received a Telecommunication Engineer's Degree in 2006, a M.Eng. in robotics in 2009, and a Ph.D. in robotics in 2011 from the University of Valladolid (Spain). He has worked as a researcher at CARTIF's Robotics and Computer Vision Division from 2007 - 2018, where he combined theoretical and field work in the research and development of projects in the area of Social and Service robotics and computer vision. He is currently with the GRIAL research group, and focuses his efforts in the development of ecosystems for the health sector and teaching. Among others, he has authored papers for the journals of Interacting With Computers or Sensors MDPI, as well as conferences such as the IEEE International Conference on Intelligent Robots and Systems and the IEEE International Conference on Robotics and Automation.

Francisco José García-Peñalvo

He received the degrees in computing from the University of Salamanca and the University of Valladolid, and a Ph.D. from the University of Salamanca (USAL). He is Full Professor of the Computer Science Department at the University of Salamanca. In addition, he is a Distinguished Professor of the School of Humanities and Education of the Tecnológico de Monterrey, Mexico. Since 2006 he is the head of the GRIAL Research Group GRIAL. He is head of the Consolidated Research Unit of the Junta de Castilla y León (UIC 81). He was Vice-dean of Innovation and New Technologies of the Faculty of Sciences of the USAL between 2004 and 2007 and Vice-Chancellor of Technological Innovation of this University between 2007 and 2009. He is currently the Coordinator of the PhD Programme in Education in the Knowledge Society at USAL. He is a member of IEEE (Education Society and Computer Society) and ACM.

LA UNIVERSIDAD
EN INTERNET

Rectorado
Avenida de la Paz, 137
26006 Logroño (La Rioja)
t (+34) 941 21 02 11

www.unir.net