An Improved Deep Learning Model for Electricity Price Forecasting

Rashed Iqbal[1], Hazlie Mokhlis[1], Anis Salwa Mohd Khairuddin[1,3]*, Munir Azam Muhammad[2]

[1] Department of Electrical Engineering, Faculty of Engineering, University Malaya, Kuala Lumpur (Malaysia)
[2] Department of Electrical Engineering, Main Campus, Iqra University, Karachi (Pakistan)
[3] Centre of Intelligent Systems for Emerging Technology, Faculty of Engineering, University Malaya, Kuala Lumpur (Malaysia)

* Corresponding author: anissalwa@um.edu.my

Received 6 January 2022 | Accepted 12 April 2023 | Published 6 June 2023

Abstract

Accurate electricity price forecasting (EPF) is important for the purpose of bidding strategies and minimizing the risk for market participants in the competitive electricity market. Besides that, EPF becomes critically important for effective planning and efficient operation of a power system due to deregulation of electricity industry. However, accurate EPF is very challenging due to complex nonlinearity in the time series-based electricity prices. Hence, this work proposed two-fold contributions which are (1) effective time series pre-processing module to ensure feasible time-series data is fitted in the deep learning model, and (2) an improved long short-term memory (LSTM) model by incorporating linear scaled hyperbolic tangent (LiSHT) layer in the EPF. In this work, the time series pre-processing module adopted linear trend of the correlated features of electricity price series and the time series are tested by using Augmented Dickey Fuller (ADF) test method. In addition, the time series are transformed using boxcox transformation method in order to satisfy the stationarity property. Then, an improved LSTM prediction module is proposed to forecast electricity prices where LiSHT layer is adopted to optimize the parameters of the heterogeneous LSTM. This study is performed using the Australian electricity market price, load and renewable energy supply data. The experimental results obtained show that the proposed EPF framework performed better compared to previous techniques.

Keywords

Intelligent System, LSTM, Smart Grid, Time Series, Forecasting.

DOI: 10.9781/ijimai.2023.06.001

I. Introduction

Smart grids (SG) are introduced to improve the performance of the traditional grid. With deregulation of electricity industry, electricity price forecasting (EPF) becomes critically important for effective planning and efficient operation of the power system. In several countries, deregulations of the electricity sector have been developed to enhance congestion control, facilitate renewable energy, and maximize the resource allocation of the power system [1]. Due to the significant volatility and intricate nonlinearity of electricity pricing, EPF has been a challenging issue. Accurate price forecasting has the ability to assist market participants to regulate their bidding strategies, production or consumption schedule with the intention to maximize their profits in the electricity market [2], [3]. Whenever demand is over- or under-predicted, inaccurate projections can have disastrous social and financial repercussions. Underestimating demand has a negative impact on supply, which leads to forced power interruptions and negative production effects. Meanwhile, overestimating demand may result in excessive investment in generation capacity, potential financial difficulty, and eventually increased electricity prices. Hence, this study plays an important role in the areas of power production and management with the aim to overcome the risk in electricity production investments and maintain affordable electricity price for the consumers [4].

Existing statistical techniques aim to reveal the specific pattern of historic power price by utilizing curve fitting. For instance, German electricity market has tested a k-factor Guégan Introduced Generalized Autoregressive Conditionally Heteroskedastic (GIGARCH) to forecast electricity price [5]-[6]. An iterative neural network methodology is also adopted along with this combinatorial neural network-based prediction technique to forecast upcoming electricity price. The advantages of this method include good precision, model functionality, and reliability. Meanwhile, Auto-regressive Integrated Moving Average (ARIMA) was proposed for short-term power load forecasting [7]. Application of statistical models had shown to be challenging when predicting multi-dimensional nonlinear price of electricity since they are mainly based on linear equations.

Please cite this article as:
R. Iqbal, H. Mokhlis, A. S. M. Khairuddin, M. A. Muhammad. An Improved Deep Learning Model for Electricity Price Forecasting, International Journal of Interactive Multimedia and Artificial Intelligence, vol. 9, no. 1, pp. 149-161, 2024, http://dx.doi.org/10.9781/ijimai.2023.06.001

On the other hand, shallow learning models have been proven to perform better compared to statistical models in terms of error minimization and some other factors. Due to nonlinearity and high volatility of the features in EPF, shallow learning models have shown to be feasible in electricity price forecasting [8]. In the field of load forecasting, Support Vector Machine (SVM) [9], [10] has been applied to predict ranges of nonlinear quantities and perform feature selection. Support vector regression (SVR) [11], artificial neural network (ANN) [18], [19], and regression tree are the main shallow machine learning models that have been commonly applied in forecasting system. Besides, the work in proposed a hybrid of SVR and gray wolf optimization to forecast life cost of power transformer. A hybrid model based on SVR and ANN is proposed in [16] by adopting new signal decomposition and correlation analysis technique to predict electricity price for next 24-hours. Furthermore, in [1] a hybrid approach of ANFIS and Backtracking Search algorithm (BSA) was proposed for electricity price forecasting and feature selection. Besides, a multi-objective binary-valued backtracking search algorithm (MOBBSA) and ANFIS approach have been employed. Nevertheless, over-fitting and gradient disappearance have been the common challenges in shallow machine learning models. It can be seen that previous techniques seemed to be less feasible for day-ahead EPF due to limited compatibility with big data and perplexing nonlinear problems [20]. The detail literature reviews related to this study is shown in the Table I.

Alternatively, deep learning algorithms have increasingly become popular in the disciplines of artificial intelligence and big data due to its ability to generate efficient classification approximations from a huge volume of input data and extract the data's underlying properties [21]-[23]. The model in [16] focused on distributed depiction, bidirectional gated recurrent unit (BiGRU) and learning algorithm with the BiGRU layer processing past and prospect information concurrently to fully extract chronological and nonstationary features from input data with the goal of improving forecasting performance. Meanwhile, to extract difficult nonlinear characteristics, [14] incorporated the deep belief network (DBN), LSTM RNN, and convolutional neural network (CNN). The DBN model was used in [24] to use signal processing and correlation analysis techniques. In addition, [25] created a multi-input and multi-output LSTM model for forecasting electricity demand. When evaluating the aerial correlation of dataset, it seems to be that a deep learning algorithm with a recurrent feedback framework called Recurrent neural network RNN has the capacity to accomplish more overarching and entire designing of time series than other traditional AI algorithms. The gradient inflation and gradient vanishing issues could be handled using LSTM through the RNN training procedure. As a result, LSTM has been used to anticipate day-ahead power prices for the Victoria region of Australia and the Singapore market [17]. Furthermore, the network topology of single gated recurrent units (GRU) has been explored for prediction purposes. When compared to an LSTM network, the GRU's simple neuron topology has been proven to lead to a faster processing time [26]. In a nutshell, LSTM has been demonstrated to perform better in terms of forecasting accuracy than SVM, ANN, and RNN [27], [28]. As a result, the analysis of time-series data for deep learning model in EPF has been an active subject of research for decades.

Based on the previous literatures, it can be seen that most of the works consider electricity supply, price and seasons to be the input features for the EPF system. Thus, in order to develop accurate prediction model, this work considers several inputs such as the price,

TABLE I. Comparisons of Recent Studies in Electricity Price Forecasting

Method	Application	RMSE	MAPE (%)	Limitations/Challenges
SVM[12] LSSVM[12]	Machine learning techniques are adopted to solve longer time horizon and highly nonlinear data for mid-term electricity market clearing price.	N/A	11.7491 10.9722	Accuracy in spike price forecasting considerably low by using the proposed machine learning methods. Optimization of forecasting accuracy in the spike price area is the main challenge of the study.
ANN PSO (Hybrid)[13]	Mid-Term Load Power Forecasting considering environment Emission using North American electricity market	N/A	1.9	ANN PSO method is not feasible to handle large data set of nonlinear data.
IFCM-SVM [9]	A dynamic parallel forecasting model using modified fuzzy time series and SVM. IFCM model is used to cluster the input data set, then the FTS model and SVM model are improved, finally the dynamic parallel model is used to forecast.	11.66	7.92	Computation of large data is a challenge. During forecasting process, the number of operation cycles need to be reduced in order to obtain good prediction accuracy.
k-factor GIGARCH [5]	This work combines several machine learning approaches to develop a novel hybrid forecasting model, namely EMD-SVR-PSO-AR-GARCH model. This work adopts the New South Wales (NSW, Australia) electricity market.	427 - 759	2.76 - 3.74	The nonlinearity and randomness of sequences of electricity consumption data.
GA-CNN [14]	The work is tested on Pennsylvania-New Jersey-Maryland (PJM) power market. The method integrates CNN with an evolutionary algorithm and utilizes spatiotemporal data.	0.007 – 0.02	3.5 – 4.9	Limited discussion on time series data analysis and statistical reliability.
EEMD-LSTM_SMBO [15]	An optimized heterogeneous structure LSTM model is proposed to solve the problems of the single network structure and hyperparameter selection. PJM electricity market is adopted in this work.	0.9 – 1.9	2.5 – 4.7	Uncertain accuracy due to limited variables considered in the prediction model.
Bi-GRU (EGA-STLF) [16]	Bi-GRU layer in EGA-STLF computes the past and the future data simultaneously to fully extract temporal and nonlinear features from input data. This work adopts Australia electricity market for short-term load forecasting (STLF).	255.12	3.06	Analyze influence factors from more complex environments.
SCAR-Dvine model [17]	A flexible class of drawable vine copula models is applied by incorporating the dependence parameters of the constituting bivariate copulae to be time-varying. This work adopts Australia electricity market for the one-day-ahead forecasting.	N/A	N/A	Modelling risk of the five markets as a complex interconnected system, as opposed to analyzing markets individually.

demand, seasons, fuel supply, renewable and non-renewable energy, peak, and off-peak hours. Predicting the unknown source of spike can be a challenge in EPF. In light of this, an optimized improved deep learning framework is proposed by incorporating linear scaled hyperbolic tangent (LiSHT) layer in the EPF. The LiSHT layer is adopted to optimize the hyper parameters of the heterogeneous LSTM, to further improve the performance of the forecasting model and predicting the spikes.

Meanwhile, the unique properties and characteristics of time-series data are important for forecasting and prediction purposes. Time series data can be challenging due to presence of noise, exhibit high volatility and extremal directional movements [29]. Generally, stationarity of time series data is vital because various analytical tools and statistical models rely on it. Therefore, a pre-processing module is required to contribute towards accurate forecasting performance in terms of reliability and accuracy. In order to overcome this challenge, this work proposed a pre-processing module to ensure the feasibility of the time-series data to fit the proposed deep learning model. In the pre-processing module, a proper transformation is performed in order to satisfy the stationarity property of the time series data by applying Augmented Dickey–Fuller test and transformations. This is to prevent autocorrelation in the prediction model's errors. This will then contribute towards more accurate EPF. The contributions of this work are as follow:

1. Propose pre-processing module to monitor the suitability of the time series data for the deep learning model.
2. Propose an optimized RNN-LSTM based algorithm by incorporating LiSHT layer.

This paper is organized as follows. Section II discusses on the time-series analysis; Section III explains on the proposed forecasting model; Section IV presents the experimental results and discussion on case studies of the Australian electricity market. Finally, Section V concludes this study.

II. Data Pre-processing

A. Autocorrelation of the Model's Forecasting Reliability

In this research, the time series dataset includes electricity price, demand, and renewable energy supply of Australia's most important five economic zones. The electricity market data covers the duration from 1 September 2020 to 31 May 31 2021 which is obtained from https://aemo.com.au website. Conventional time series data may contain missing values, outliers and high dimensional data. These factors contribute to unstable forecasting performance. Therefore, pre-processing is required to solve the abovementioned problems. This work emphasized on linear trend-based equation for features processing. The linear trend approach can perform effectively with the trend and depict it without any assumptions. Besides, the residual seasonality, peak-off peak hour and renewable energy trend can distinguish any time series dataset shown in Table II.

Let h_1, h_2, … …, h_n be the time-series data. Equation (1) is the definition of a nonlinear regression model of order m is denoted by:

$$h_t = f(g_t, \theta) + \epsilon_t \tag{1}$$

where $g_t = (h_{t-1}, h_{t-2} \ldots . h_{t-m}) \in \mathbb{R}^m$ made of m values of h_t, θ is the parametric vector and ϵ_t is the residual. After the model has been built, machine learning or deep learning approaches can be used to find the function $f(*)$. Root mean square error (RMSE) and mean absolute error (MAE) are the most often used indicators for regression performance evaluation of a forecasting model. Nevertheless, both regression performance evaluators only indicate the accuracy of the observed and estimated values. Since they are unable to analyse the fitness of time series data in the proposed forecasting model, the residuals are employed to assess this dedicatedly. In other words, the forecasting model's residuals of regression analysis for normal distribution and autocorrelation are estimated by function $\widehat{\epsilon_t}$, where $\hat{h}_t$ is the predicted value as equation (2).

$$\widehat{\epsilon_t} = h_t - \hat{h}_t \tag{2}$$

The presumption of no autocorrelation in the residuals might make the forecasting vulnerable as there may not be exploration on all available data in the training process. In other words, the reliance of the residuals indicates that the model did not well fit the time-series data and that there is important data remaining that must be investigated. Dataset used in this research showed in Table II. The autocorrelation function (ACF) plot and the Ljung–Box Q test for residual autocorrelation are two important techniques for - determining the presence of autocorrelation in the residuals [30]. More analytically, by calculating the linear correlation of every residual in various lags, $\hat{\epsilon}_{t-1}$, $\hat{\epsilon}_{t-2}$, … the ACF can be obtained in which the temporal autocorrelation is depicted by ACF, and the Ljung–Box Q test is a "portmanteau" test. The null hypothesis H_0 that ''a sequence of residuals does not exhibit autocorrelation for a specified number of lags L", is proved technically with respect to other hypothesis H1 that ''some autocorrelation coefficient is nonzero.'' Equation (3) defines the Ljung–Box Q test statistic in more detail,

$$Q = s(s+2)\sum_{k=1}^{M}\frac{\rho_k^2}{s-k'} \tag{3}$$

where equation (4) indicates at lag-k, autocorrelation coefficients ρ_k are,

$$\rho_k = \frac{\sum_{i=1}^{s-k}(h_i - \bar{h})(h_{i+k} - \bar{h})}{\sum_{i=1}^{s}(h_i - \bar{h})^2} \tag{4}$$

with $\bar{h} = \frac{1}{s}\sum_{i=1}^{s} h_i$ under H_0 the statistic Q asymptotically follows a $g_{(M)}^2$ distribution. The model shows autocorrelation and reject the zero hypothesis H_0 if as following equation (5),

$$Q > g_{(1-\alpha,M)}^2 \tag{5}$$

where the critical value of the Chi-square distribution is defined for significance level α, or critical level $p = 1 - \alpha$, known as p value.

TABLE II. Features Used in Time Series Data

Peak/ Off-Peak Hours	Time (Hour)	Main Electricity Supply (MWH)	Previous Hour Price (AUD)	Solar Power supply (MWH)	Hydro Power Supply (MWH)	Wind Power Supply (MWH)
1	5	7360.25	40.54	0	992	215.06
1	6	7066.01	43.59	0	992	192.84
1	7	6841.68	34.74	0	645.09	168.44
1	8	6732.33	17.15	0	325.58	146.70
1	9	6980.24	16.61	0	214.92	136.46
0	10	7661.34	31.56	0	125	186.70
0	11	8639.57	40.02	0	60	179.90
0	12	9890.74	49.99	0	2.5	810.19
0	13	9845.55	50.25	0	20	776.65
0	14	9446.45	54.32	0	94.58	819.08
0	15	8992.55	55.51	9.76	380.62	825.95
0	16	8547.70	47.99	189.73	705.43	760.85
0	17	8162.07	39.65	418.42	478.88	746.72
cont	cont	cont	cont	cont	cont	cont

B. Stationarity and No Stationarity

Autocorrelation, long memory, fractal and multi-fractal properties are the features of time-series that appear so frequently that they are referred to as stylized facts. The main disadvantage of working with values of price time series is that they follow a random walk process from the standpoint of stochastic processes. The coefficient of autocorrelation is ρ_k, with $k > 1$ are statistically remarkable for many lags L and the first-order autocorrelation coefficient ρ_1 is equal to one. This kind of time series are called unit root time-series or integrated of order one which are expressed by $I(1)$. Modelling the levels of these series under such conditions is unproductive since the residuals of the models show redundancy, which putting the entire framework of statistical validity in jeopardy. In order to examine these series effectively, they must be stationary which is essential for the advent of a new forecasting model.

Assume that $F_h(h_{t_{1+\tau}}, \ldots, h_{t_{n+\tau}})$ is the total distribution algorithm of the intrinsic joint distribution of $\{h_t\}$ at times $t_1 + \tau, \ldots, t_1 + n$ then the stochastic process $\{h_t\}$ is strictly stationary if (6),

$$F_h(h_{t_1+\tau}, \ldots, h_{t_n+\tau}) = F_h(h_{t_1}, \ldots, h_{t_n}) \tag{6}$$

for all τ, $t_1 \ldots t_n \in \mathbb{R}$ and $n \in N$. Nevertheless, the stationarity of time series is reduced resulting to weak covariance stationarity [30]. A stochastic process becomes covariance stationary when the mean is constant, the second moment is finite, and the covariance function relies on the difference between t_1 and t_2. Hence, in equation (7) the auto-covariance needs to be denoted with one variable, i.e.,

$$cov_{hh}(t_1, t_2) = cov_{hh}(t_1 - t_2, 0) \tag{7}$$

where cov_{hh} is the auto-covariance of the y_r series to summarize stationarity based on statistical features of the stochastic process. It has been a general hypothesis that many procedures such as statistical assessment, modelling and prediction become simpler when adopted the stationary processes. The partial autocorrelation function provides a resolution once the problem has been detected, where the lag-k coefficient $\phi_{k,k}$ is displayed by the indicated formula in equation (8),

$$\begin{cases} \phi_{k,k} = \dfrac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}} \\ \phi_{kj} = \phi_{k-1,j} - \phi_{k,k}\phi_{k-1,k-j}, \end{cases} \tag{8}$$

for $k > 1$ and $\phi_{1,1} = \rho_1$. Clearly, if there is unit root throughout the series, that is $\rho_1 = 1$, the first-order partial autocorrelation coefficient $\phi_{1,1}$ will become one. Commonly, the initial coefficient is statistically significant while the rest are insignificant. Then, the first series should be characterized by the first differences as sown in equation (9) of the series, defined by (9)

$$\triangle_t = h_t - h_{t-1} \tag{9}$$

Therefore, the first difference of the time series in stationarity obtained can be represented with integrated of order zero which is $I(0)$. However, crossover of different non-stationarities could present while computing the time-series data there such as unit-roots, structural pause, level up-downs, seasonal trend or a shifted variance. When the series is non-stationary ($I(1)$), the typical transformation is to take the first differences and transform it to stationary series ($I(0)$), whereas if the series contains structural breaks or a changing variance due to crises, a nonlinear BoxCox transformation will be the best solution [31]. As normality is an essential criterion for various statistical procedures, a BoxCox transformation provides a mechanism to turn non-normal data into a normal pattern. The following equation (10) defines one-parameter Box-Cox transformation as,

$$h_t = \begin{cases} \dfrac{h_t^{\lambda} - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \ln h_t, & \text{if } \lambda = 0. \end{cases}; \tag{10}$$

where nonzero Box-Cox transformations are used for $\lambda = -3, -2, -0.5, 0, 0.5, 1$ and 2. The rule $\lambda = 0$ is followed by majority of the time series; therefore, the returns which are the first logarithmic differences are used to attain stationarity in these series as equation (11),

$$r_t = \ln h_t - \ln h_{t-1} \approx \frac{h_t - h_{t-1}}{h_{t-1}} \tag{11}$$

the last expression being the percentage change or returns [30].

C. Augmented Dickey Fuller (ADF) Test

The proposed pre-processing module for greatly improving the accuracy and durability of a deep learning algorithm for time series prediction is discussed in this part, based on well-known statistical concept and estimation for stationarity and non-stationarity qualities. Generally, the components of the dataset are not-stationary when a machine learning or deep learning model is applied to estimate the time-series. This implies that they may have unit roots and some order of integration. It is worth noting that the Augmented Dickey–Fuller (ADF) test has the ability to identify a unit root in a time series data [30], [32]. The model is subjected to the testing method as following equation (12),

$$\triangle_{h_t} = \alpha + \beta t + \gamma h_{t-1} + \sum_{i=1}^{k-1} \delta_i \triangle_{h_{t-i}} + \epsilon_t \tag{12}$$

where ρ_1 denotes the first-order autocorrelation coefficient and α is a constant, β is the coefficient of trend, and $\gamma = (\rho_1 - 1)$. It is notable that k is the lag order of the autoregressive determined so that the residuals ϵ_t have no serial correlation. There has a stochastic random walk process, if $\alpha = 0$ and $\beta = 0$, while if $\alpha \neq 0$ and $\beta = 0$, here the stochastic process is with drift. The unit root test is employed to evaluate statistical importance under the null hypothesis. H_0:$\{\gamma = 0$ that is $\rho = 1\}$ versus the nonzero hypothesis H_1: $\{\gamma < 0$ that is $\rho < 1\}$.

Recursively taking the first differences in (9) or returns in (11) until the sequence is made stationary depending on the nature of the series. The autocorrelation in the model's residuals will be reduced when using a series of transformations based on the first difference and returns. This means that the forecasting method will be considerably better at explaining the data because it captures all conceivable nonlinearities, assuring the model's accuracy and efficacy.

The pseudo-code for the framework is shown below pseudo-code. Firstly, the time-series data is imported. The ADF test is then used to determine if the sequence levels are non-stationary, or if they have a unit root in Step 2. If the sequence is stochastic, the dataset will be continuously converted using first differences or returns till the resultant series becomes stationary in Steps 4–7. The newly modified time-series data is then utilised to train the forecasting model in Step 8.

Pseudo-code of proposed algorithm

```
1. Input time series data
2. Assess unit root test (ADF)
 3. If (Unit root exist in time-series) then
 4. Repeat
  5. Convert the time series based on differences (9) or returns (11).
  6. Assess unit root test (ADF).
  7. Until (Stationarity exists in time-series.)
  8. By using converted time-series, train the prediction model.
 Else
    9. By using real time-series, train the prediction models.
    10. On the training sample, estimate the residuals.
    11. If (Autocorrelation exists in residuals.) then
    12. Convert the time series based on differences (9) or returns (11).
    13. Using the converted time-series, retrain the forecasting model.
      End
 End
```

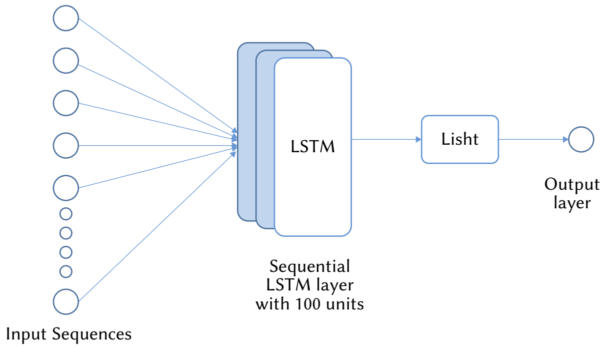

Fig. 1. LSTM+LiSHT prediction model architecture.

If the data is stationary, then levels of time-series are employed to train the forecasting model in Step 10. The errors of the estimation method on the learning algorithm are employed for further analysis and testing. It is noticeable that a training is performed with a series that has a unit root then the predicted values become near to the real values for any realistic model then, the presence of strong autocorrelation factors marks the model as unproductive [33]. Therefore, ACF plots and/or the Ljung–Box Q test are used to investigate autocorrelation within residuals of the dataset in step 11. Eventually, if the residuals have autocorrelation, the recommended transformation is performed to the training phase and the algorithm is retrained utilizing the newly transformed dataset according to steps 12–13. It is noticeable that if the series levels are stationary and the residuals on the training dataset indicate no autocorrelation, there is no need to reform the series because it will result in catastrophic phenomena of over-differencing. To put it in another way, over-differencing makes the entire mechanism "non-invertible," and thus lacked an endless autoregressive expression. In the form of a flowchart, Fig. 2 depicts an insight into the intended structure.

Finally, if the classifier is trained with a transformed series with no autocorrelation in residuals, the inverse transformation will be used in the model's forecasts to obtain the forecast for the levels of the exact time-series and reliable for parametric and no parametric tests.

III. The Proposed Improved LSTM Forecasting Model

In order to process the long sequence of time series data, LSTM Recurrent Neural Network (RNN) is proposed with the aim to overcome the problem of vanishing gradient and gradient explosion that can occur in conventional RNN. The input gates and output gates are replaced by memory/forget gates in the hidden layer of LSTM RNN which include memory space and information flow process for long historical time series [14]. A sequential layer followed by a fully connected layer, lstm layer, tanh layer and regression layer are applied in this algorithm Fig.1. In this study, sgdm optimization is applied with max epoch 1500. Gradient threshold is 1, learning rate schedule piecewise. In lstm layer, number of hidden units is 100, state activation function tanh and sigmoid gate activation function are adopted in this algorithm. The parameters of LSTM are shown in Table III.

TABLE III. Parameters Used in Lstm

LSTM Parameters	
Hidden layers	3
Neurons per layer	100
Type of layer	LSTM
Activation layers	Sigmoid
Epochs	1500
Optimizer	SGDM

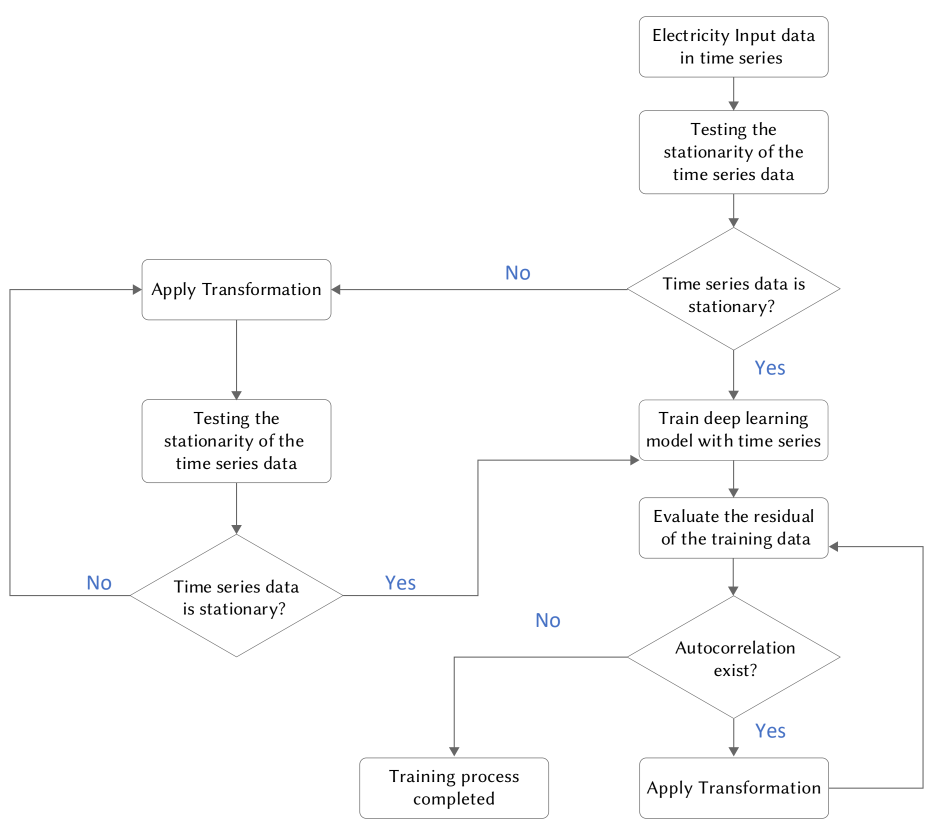

Fig.2. Flowchart of the proposed algorithm.

Conventional activation functions such as ReLU and Swish are less feasible for large negative input values and also may suffer from the dying gradient problem due to zero-hard rectification. Therefore, it is essential to adopt a better activation function to overcome those limitations. In this work, a non-parametric function, called Linearly Scaled Hyperbolic Tangent (LiSHT) for Neural Networks (NNs) is employed in this model as referred (13). The LiSHT activation function is utilized to scale the non-linear Hyperbolic Tangent (Tanh) function through a linear function and tackle the dying gradient problem.

Let an input vector be a $\in R^d$, and each hidden layer is capable to transform its input vector by applying a nonlinear mapping from the q^{th} layer to the $(q + 1)$ th layer as equation (13):

$$a = \tau^0$$
$$\sum_{l=1}^{N^q} w_{kl}^q \tau_l^q + o_k^q = c_k^{q+1}$$
$$\phi\left(c_k^{q+1}\right) = \tau_k^{q+1} \quad (13)$$

LiSHT is a non-parametric linearly scaled hyperbolic tangent activation layer that has unrestricted upper limits property on the right-hand side of the activation curve. LiSHT has the advantage of positive activation that does not identically propagate for all inputs, which solves the gradient problem at back propagation and contributes to faster training of the deep neural network. The LiSHT activation function is calculated by multiplying the *Tanh* function by its input x and defined as the equation (14) and (15). where $g(x)$ is a hyperbolic tangent function [32].

$$\phi(x) = x \cdot g(x) \quad (14)$$

$$g(x) = Tanh(x) = \frac{exp^x - exp^{-x}}{exp^x - exp^{-x}} \quad (15)$$

IV. Results and Discussions

In this work, the data were divided into training and test set consisting of hourly electricity price data as tabulated in Table IV:

TABLE IV. Seasonal Training Dataset

Season	Training set	
	1 day forecasting	1 month forecasting
Sep-Oct-Nov (Spring)	Oct (696 hours)	Sep-Oct (1440 hours)
Dec-Jan-Feb (Summer)	Jan (696 hours)	Dec-Jan (1440 hours)
Mar-Apr-May (Autumn)	Apr (696 hours)	Mar-Apr (1440 hours)

There have been no missing data in any of the time-series and outlier prices were not eliminated in order to preserve the characteristics of every series, even though these prices are the consequence of rare events presents the descriptive analysis for every training dataset and testing dataset, such as the measurements of minimal, max, average, sample variance (std. dev.), median, skewness, and kurtosis for illustrating the distribution's nature. Using the ADF unit root test, the proposed framework employed the National Electricity Market (NEM) price time-series in Australia to determine whether the training data are stationary or not. The outputs of the ADF unit root test for the training data of Australia's five states series under investigation are shown in Table V. Considering the t-statistics (t-stat) and the corresponding p values the null hypothesis H_0: ‘‘the levels possess a unit root and are non-stationary’’ is accepted for time series.

TABLE V. ADF Unit Root Test for the Training Data

Series	NSW	QLD	SA	TAS	VIC
***t* static**	-37.16	-71.36	-28.82	-39.42	-24.58
***p* value**	0.000*	0.000*	0.000*	0.000*	0.000*

In the sequel, the ADF test was run on a time-series to see if the unit root existed, as per the provided framework. The outcomes of the ADF unit root test for the training data of all time-series datasets are shown in Table IV. The (*) indicates statistical impact at the 5% critical threshold. Clearly, it's worth noting that all p values are almost zero, the null hypothesis H_0 is rejected.

As a result, the time series are “appropriate” for training a deep learning model with minimal autocorrelation in the errors, and a significant boost in forecasting accuracy is anticipated when comparing to same model trained with the non-transformed series. In order to evaluate the performance of the proposed model, the regression ability is assessed using mean absolute error (MAE) and root mean square error (RMSE). Besides that, another four key performance indicators are also employed: Accuracy (Acc), F1-score (F1), Sensitivity (Sen), Specificity (Spe), Positive Predicted Values (PPR) and Negative Predictive Values (NPV) which are indicated by the following equations (16)-(21).

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{FP}'} \quad (16)$$

$$\text{F}_1 = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}'} \quad (17)$$

$$\text{Sen} = \frac{\text{TP}}{\text{TP} + \text{FN}'} \quad (18)$$

$$\text{Spe} = \frac{\text{TN}}{\text{TN} + \text{FP}'} \quad (19)$$

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}'} \quad (20)$$

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (21)$$

In this case, TP represents the frequency of prices that were successfully identified as raised, the number of prices that were successfully detected as having a decreasing value is denoted by TN, FP is the amount of prices that were incorrectly detected as being increased, whereas FN denotes the quantity of prices that were incorrectly detected as being dropped. Furthermore, the area under curve (AUC) statistic, considered one of the most important classification metrics which has been incorporated in the assessment and is shown using the receiver operating characteristic (ROC) curve. The ROC curve is made by comparing the true positive rate (Sensitivity) against the false positive rate (Specificity) at different cut-off values.

A. Pre-processing of Time Series Data for the Prediction Model

In the following section, the predictability of all prediction techniques is investigated by using the Auto-Correlation Function (ACF) plot and the Ljung–Box Q test to detect autocorrelation in the residuals. This is to ensure that each trained model adequately fits the time-series and if they are uniformly distributed evenly and monotonically independent. The Ljung– Box Q test is a ‘‘portmanteau’’ test which analyse the null hypothesis H_0 that ‘‘a series of residuals exhibits no autocorrelation for a fixed number of lags L,’’ which is the opposite of another hypothesis H_1 that ‘‘some autocorrelation coefficient is nonzero coefficient is nonzero’’. Fig. 3 – 7 displayed the ACF graphs of LSTM+LiSHT model for the electricity price data. The ACF graphs of the prediction model are trained with typical time-series defying the presumption of the free autocorrelation residuals. The high spikes which observed in several lags (fig. 3(a) – 7(a)), show that such model's estimation may be inaccurate. The ACF plot spikes at lag 1 then slowly decays to lag 10. From lag 1 to 4 spikes are too high and cut off at the significant band 0.2, Which shows that the significant autocorrelation presents in the residual of trained data. On the contrary, from the fig 3(b)-7(b) it is shown that the spikes from lag 2 immediately go down under or between the significant band. Therefore, the autocorrelation in the residual does not exist in the trained data and is statistically sound for the evaluation of time series. In summary, all ACF plots of the LSTM+LiSHT generated using the converted time series (figs 3(b) – 7(b)), show that the residuals do not have autocorrelation. This can be further verified by the results obtained from the Ljung– Box Q test

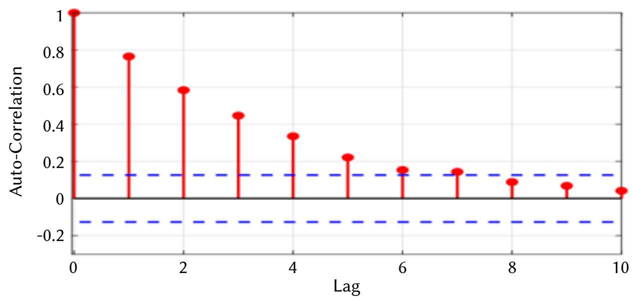

Fig. 3(a). Autocorrelation of residuals for NSW time series.

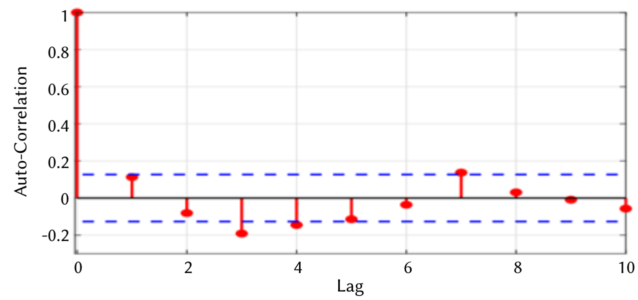

Fig. 3(b). Autocorrelation of residuals for transformed (box-cox) NSW time series.

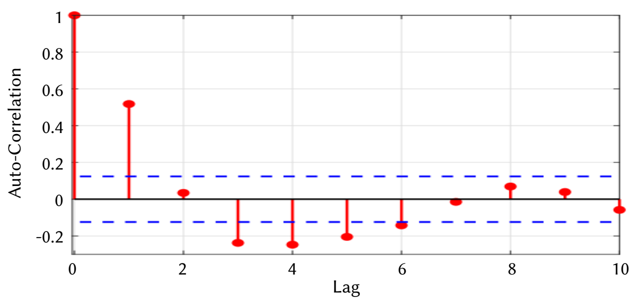

Fig. 4(a). Autocorrelation of residuals for QLD time series.

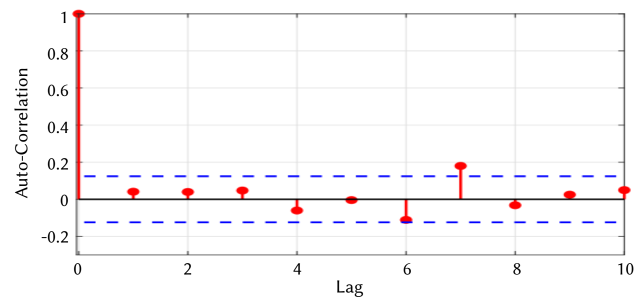

Fig. 4(b). Autocorrelation of residuals for transformed (box-cox) QLD time series.

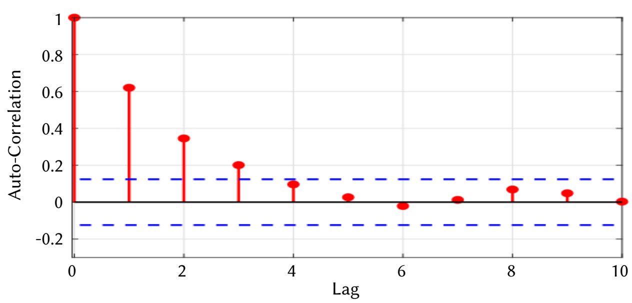

Fig. 5(a). Autocorrelation of residuals for SA time series.

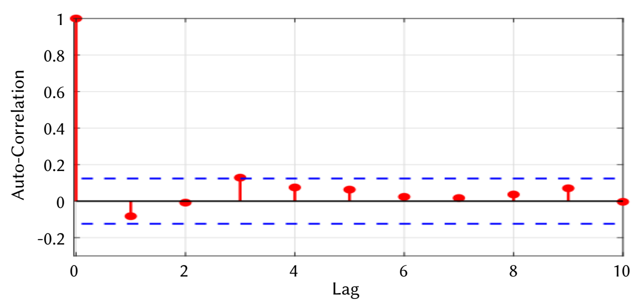

Fig. 5(b). Autocorrelation of residuals for transformed (box-cox) SA time series.

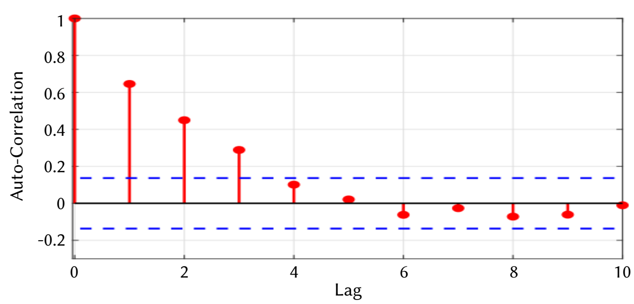

Fig. 6(a). Autocorrelation of residuals for TAS time series.

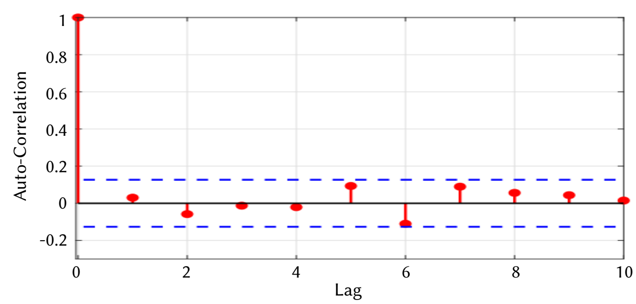

Fig. 6(b). Autocorrelation of residuals for transformed (box-cox) TAS time series.

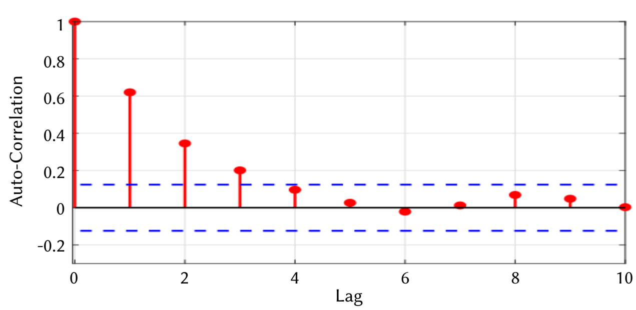

Fig. 7(a). Autocorrelation of residuals for VIC time series.

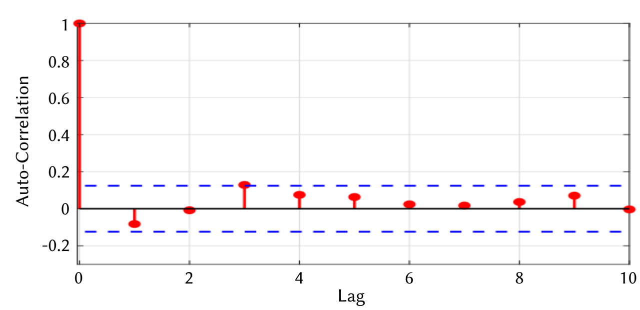

Fig. 7(b). Autocorrelation of residuals for transformed (box-cox) VIC time series.

(Table VI) where the transformed time series data using the BoxCox transformation is free from autocorrelation.

TABLE VI. Represent the Result of the Ljung–Box Q test Using $L = 10$

Condition	Forecasting	Autocorrelation existence
Before transformed	1 day	Yes
	1 month	Yes
After transformed	1 day	No
	1 month	No

In this work, it has been established theoretically and experimentally that the time-series data are ‘‘appropriate’’ for developing a deep learning model, which is one of the contributions of this study. In another way it can be said that this work has developed a new framework that can discover effective time series data for training a deep learning model. This will lead to a stable and reliable forecasting model. The term "appropriate" denotes that the time-series data has satisfied the stated scientific requirements and is adequate for training a forecasting model. If, on the other hand, the series fails to meet the desired criteria, it is deemed "unsuitable," and any attempts to develop a solid prediction model would most likely be useless. Therefore, this work is a beginning point for the development of any prediction methodology for various time series forecasting. If the starting dataset is unstable or non-stationary, the work done for developing the forecasting model could be meaningless. Furthermore, it can be justified that this work has developed an innovative and comprehensive framework that allows any unstable time-series to be transformed to a stable condition by conducting a boxcox transformation method. It can be seen that the proposed transformation has successfully eliminated the "unsuitable" data, avoiding the costly and time-consuming "trial and error" method. Besides that, it is noticeable that one of the most interesting properties of our suggested framework is that this method can be simply modified to encompass a broader scientific domain of time-series forecasting operations without requiring any further adjustments or limits. More specifically, the recommended method uses statistic and economic tests to conduct an optimal pre-processing phase for utilising the internal structure of the timeseries. Finally, it is seen that while deep learning models are well accepted for time series, the proposed framework significantly enhances performance. However, more study is being done to see which of these approaches may be implemented more effectively a priori based on the properties for every time-series in order to get better forecasting performance. For accomplishing the prerequisite diagnosis and appropriate time transformation methodology, a complex pre-processing framework that refers to the inherent time-series particular traits such as stationarity, heteroskedasticity, seasonal cycles, and shifting variance can be used.

B. Forecasting Performance of the Proposed LSTM+LiSHT Model

The efficacy of the proposed LSTM+LiSHT prediction model for the energy price dataset during spring season is presented in Table VI, while the results for other seasons are presented in Appendix A. In spring, the accuracy of the proposed LSTM+LiSHT model is above 0.95 and 0.87 for the case of one day forecasting and one month forecasting respectively. Commonly, Medium-Term Forecast (MTF) studies horizons from a few days to months ahead. MTF is normally used for risk management, balance sheet calculations, and derivatives pricing. Meanwhile, Short-Term Forecast (STF) covers horizons from a few minutes up to a few days ahead has become an essential tool for the daily market operations.

Table VII also computed the sensitivity and specificity of the proposed forecasting model for both forecasting horizons. The sensitivity analysis of the proposed model is computed to permit the analysis of changes in expectations used in forecasting the electricity price. By studying all the variables and the possible outcomes, important decisions can be made about businesses, the economy, and making investments. On the other hand, high specificity indicates the good capability of the proposed model to avoid false alarms or false spikes in forecasting the electricity prices. The sensitivity and specificity of the proposed forecasting model is considered good performance which is above 0.7 and 0.8 respectively for both 1 day forecasting and 1 month forecasting. As such, sensitivity and specificity analysis are very useful methods to be applied in investment appraisal, sales and profit forecasting and other quantitative aspects of business management.

Table VIII presented the performance of the forecasting model without applying the transformation method in the pre-processing module. More specifically, the proposed model had shown to be biased when it was trained using the conventional time-series data which resulted to low forecasting performance. The forecasting accuracies are in the range between 0.4 to 0.8 and 0.3 to 0.8 for 1 day forecasting and 1 month forecasting respectively. The sensitivity and specificity are significantly low as well. The forecasting sensitivity is below 0.6 for both 1 day forecasting and 1 month forecasting. The sensitivity and specificity are significantly low as well. Hence, it is important to adopt the proposed pre-processing module to transform the conventional time series data in order to improve the forecasting performance.

As compared to table VII, the proposed forecasting model exhibits better performance in terms of ACC, sensitivity, and specificity when the data is trained with the transformed time series in the proposed pre-processing module. Furthermore, the interrelation between sensitivity and specificity has been significantly improved. In summary, the performance of the proposed LSTM+LiSHT forecasting model has been considerably improved after adopting the first transformed box-cox time series data, instead of the conventional time-series data. This justifies the contribution of the proposed pre-processing module in this work.

TABLE VII. Performance of the Proposed LSTM+LiSHT Model for Spring Season (After Transformation)

Australia's region	Forecasting	ACC	AUC	F1	Sen	Spe	PPV	NPV
NSW	1 day	1	0.945	1	1	1	1	1
	1 month	0.984	0.802	0.852	1	0.983	0.742	1
QLD	1 day	1	0.957	1	1	1	1	1
	1 month	0.870	0.768	0.678	0.76	0.894	0.612	0.944
SA	1 day	0.958	0.896	0.909	1	0.947	0.833	1
	1 month	0.923	0.853	0.927	1	0.851	0.864	1
TAS	1 day	0.958	0.965	0.933	0.875	1	1	0.941
	1 month	0.970	0.910	0.977	0.997	0.921	0.958	0.995
VIC	1 day	1	0.962	1	1	1	1	1
	1 month	0.997	0.955	0.9583	1	0.997	0.92	1

TABLE VIII. Performance of the Proposed LSTM+LiSHT Model for Spring Season (Before Transformation)

Australia's region	Forecasting	ACC	Sen	Spe
NSW	1 day	0.583	0	0.636
	1 month	0.885	0.090	0.936
QLD	1 day	0.800	0.500	0.826
	1 month	0.761	0.183	0.818
SA	1 day	0.400	0.500	0.380
	1 month	0.332	0.574	0.226
TAS	1 day	0.680	0.500	0.695
	1 month	0.785	0.333	0.825
VIC	1 day	0.800	0.500	0.826
	1 month	0.850	0.065	0.976

TABLE IX. One Day Forecasting for New South Wales

Seasonality	Error	BILSTM	LSTM+GRU	GRU	LSTM	The proposed work
Spring	RMSE	2.7418	2.1190	1.5455	1.3222	4.197x10-6
	MAE	2.0013	1.6080	1.0110	0.8655	0.023
	MBE	0.6502	0.2198	0.1456	0.0443	0.0067
	MSE	7.5173	4.4903	2.3884	1.7481	0.0004
	R	0.9998	0.9999	0.9984	0.9993	0.9998
Summer	RMSE	4.3231	0.8878	0.8254	0.6725	2.598x10-6
	MAE	4.1249	0.7626	0.6944	0.4768	0.0094
	MBE	4.1249	0.5250	0.3768	0.2045	0.1822
	MSE	18.6895	0.7882	0.6814	0.4523	0.0003
	R	0.9967	0.9988	0.9992	0.9992	0.9997
Autumn	RMSE	3.1852	0.4122	0.2081	0.2430	1.730x10-6
	MAE	2.8768	0.3114	0.1596	0.1762	0.0018
	MBE	2.8550	-0.1127	0.0659	0.0427	0.0599
	MSE	10.1452	0.1699	0.0433	0.0590	0.0009
	R	0.9972	0.9992	0.9998	0.9997	0.9999

TABLE X. One Month Forecasting for New South Wales

Seasonality	Error	BILSTM	LSTM+GRU	GRU	LSTM	The proposed (LSTM+LiSHT)
Spring	RMSE	3.5388	0.5678	0.5630	0.4507	0.4065
	MAE	3.1164	0.3811	0.3236	0.2492	0.2530
	MBE	2.6034	0.0205	0.0192	0.0352	0.0134
	MSE	12.5229	0.3224	0.3170	0.2032	0.1652
	R	0.9991	0.9997	0.9997	0.9997	0.9998
Summer	RMSE	3.5682	0.7065	0.7134	0.5407	0.4207
	MAE	3.3058	0.3294	0.3274	0.2634	0.2329
	MBE	3.2019	0.0254	0.0498	8.6863e-04	0.0255
	MSE	12.7318	0.4991	0.5090	0.2924	0.2285
	R	0.9956	0.9964	0.9978	0.9978	0.9984
Autumn	RMSE	16.2570	7.6945	7.5842	5.5981	5.4047
	MAE	14.4163	2.6332	2.2764	2.3837	2.4569
	MBE	13.0280	0.1859	0.9894	0.5336	0.3663
	MSE	264.290	59.2051	57.520	31.3386	29.2107
	R	0.9983	0.9980	0.9980	0.9989	0.9990

In addition, the proposed forecasting model is compared with several deep learning models as tabulated in Table IX and Table X for one day forecasting and one month forecasting respectively. More comparison results which cover other regions and seasons are presented in the Appendix. Besides that, the graphical representation of the actual electricity price and predicted electricity price for one day (24 hours) forecasting and 1 month forecasting for five different states over the spring season are presented in Figs. 8 (a-e) and 9 (a-e), respectively. The regression analysis has been performed to quantify the relationship between variables used in the forecasting models. As shown in Table VIII and Table IX, the regression (R) values for all the forecasting methods are approximately 1.0 which indicates a good relationship between the forecast variable of interest and the predictor variables. Generally, RMSE which represents the standard deviation of residuals (forecasting errors), is computed to determine the concentration of data around the line of best fit. Meanwhile, MAE computed the average of all differences between actual and forecast absolute value. Another common metric applied is the mean

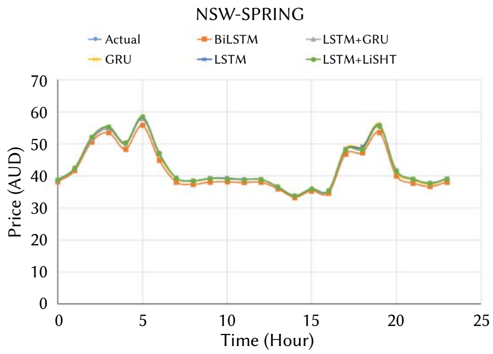

Fig. 8(a). One day prediction results for NSW.

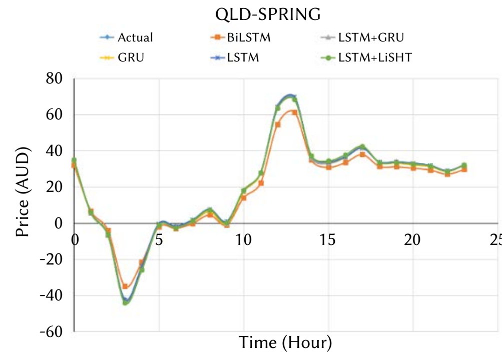

Fig. 8(b). One day prediction results for QLD.

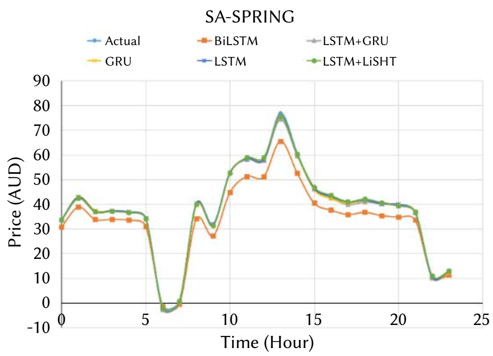

Fig. 8(c). One day prediction results for SA.

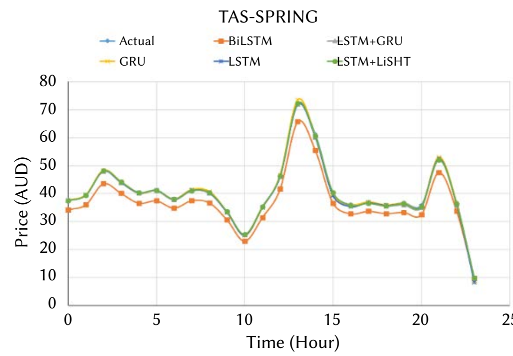

Fig. 8(d). One day prediction results for TAS.

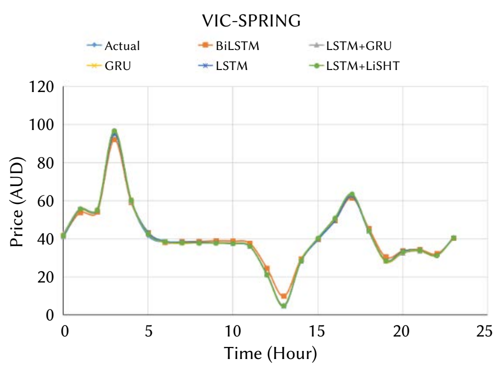

Fig. 8(e). One day prediction results for VIC.

bias error (MBE) which could provide indication whether the model overestimates or underestimates the output. The smaller the values of RMSE, MBE and MAE indicate better performance of the forecasting model. It can be seen that the lowest RMSE, MBE and MAE are achieved by the proposed LSTM+LiSHT forecasting model when compared to other forecasting models. In contrast, BiLSTM forecasting model produces the highest RMSE, MBE and MAE for spring, summer and autumn respectively. This shows that BiLSTM is the least preferable forecasting model to be used in this work followed by LSTM+GRU, GRU and LSTM. Hence, this can be justified from figs 8 (a-e) and 9 (a-e), where the curves of the proposed model coincide with the curve of the actual data which shows that the proposed model is able to forecast the electricity price effectively unlike the BiLSTM curves.

The proposed model is benchmarked with previous works as tabulated in Table XI. The work in [34] shows that the proposed Bi-GRU and Gated-FCN obtains RMSE of 8.23 and 3.12 respectively. Besides, the work in [35] that applied CNN-LSTM obtains RMSE of 6. The work in [36] applied BP, CNN, LSTM-NN, WT-TDLSTM model for electricity price forecasting and achieved the considerably low RMSE of 0.012798, 4.697257×10^{-5}, 0.008360, and 3.940309×10^{-6} respectively. Moreover, the work in [36] shows the WT_TDLSTM model is superior compared to the neural network system that exclude discrete wavelet transform and pre-processing of multiscale data. As for this work, the proposed model is tested on a smaller dataset that spans from year 2020 to 2021 and has produced considerably low

TABLE XI. Performance Comparison of the Proposed EPF Model With Recent Works

Work	MSE	RMSE	MAE
Bi-GRU [34]	N/A	8.23	N/A
CNN-LSTM [35]	N/A	5.92	N/A
Gated-FCN [34]	N/A	3.12	N/A
BP [36]	0.113129	0.012798	0.281345
CNN [36]	0.006854	4.697257×10^{-5}	0.071783
LSTM-NN [36]	0.091435	0.008360	0.22594
WT_TDLSTM [36]	0.001985	3.940309 × 10-6	0.035024
The proposed work	0.00053	2.8438× 10-6	0.0114

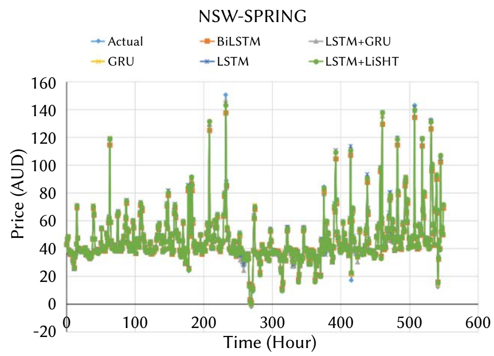

Fig. 9(a). Monthly prediction results for NSW.

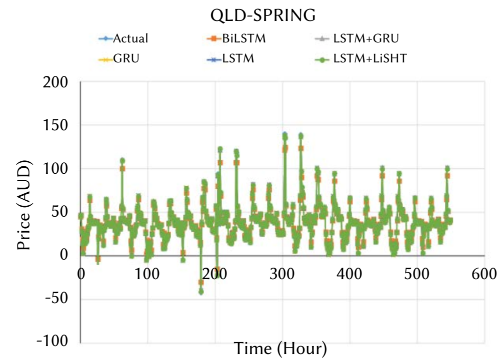

Fig. 9(b). Monthly prediction results for QLD.

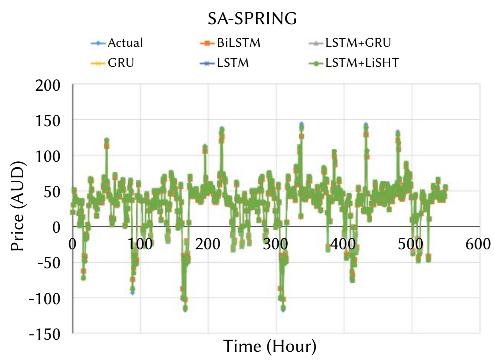

Fig. 9(c). Monthly prediction results for SA.

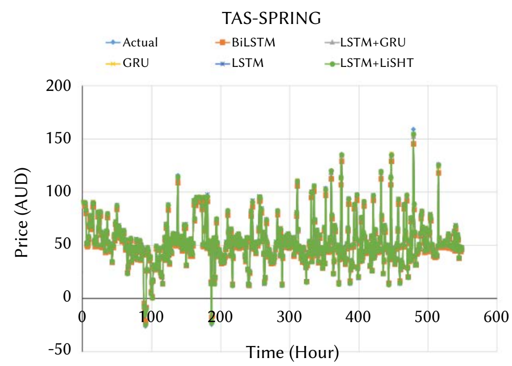

Fig. 9(d). Monthly prediction results for TAS.

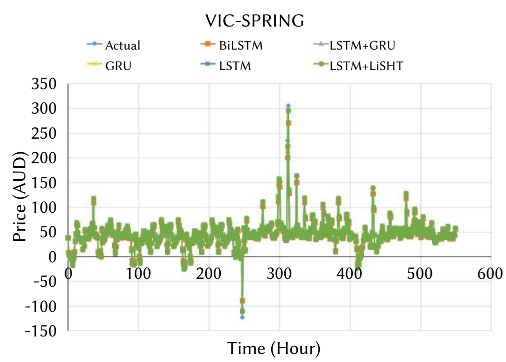

Fig. 9(e). Monthly prediction results for VIC.

RMSE, MAE and MSE as compared to previous works. The RMSE of the proposed LSTM+LiSHT framework varies for different season and region with the lowest value of 2.8438 × 10^{-6}. This justifies that the proposed forecasting model is suitable to be applied in the EPF application under various seasons in Australian electricity market.

V. Conclusion

In this work, time series data analysis has been performed and improved deep learning method has been proposed for short term electricity price forecasting. The developed forecasting model consists of pre-processed and post trained data analysis which incorporates time series statistical reliability method. An augmented dickey fuller test is performed to examine the stationarity and nonstationary data before the training process. Then, autocorrelation of the residuals is computed after the training process. In this work, the autocorrelation of the residuals has been evaluated to ensure the feasibility of the data for the deep learning approach. The autocorrelation in residuals has been fixed by transforming the data through box-cox transformation technique. Finally, the forecasting of electricity price is performed by applying the proposed deep learning module which has been modified to optimize the parameters of the heterogeneous LSTM. The performance of the proposed forecasting model has been benchmarked with previous works to justify the feasibility of the proposed method. Based on the results obtained, it can be seen that the proposed model has shown superior results compared to other methods in terms of RMSE, MSE and MAE.

In future works, further analysis can be performed such as comparing the proposed method for new profit and return-based performance measurements. Moreover, long-term electricity price forecasting can be explored. The proposed methodology can also be applied in other time series forecasting data.

Acknowledgment

This work is funded by Universiti Malaya research grant from Malaysia under the project name ‘Intelligent Price Forecasting System for Optimal Energy Market’ with grant number ST005-2021.

References

[1] A. Pourdaryaei, H. Mokhlis, H. A. Illias, S. H. A. Kaboli, and S. Ahmad, "Short-term electricity price forecasting via hybrid backtracking search algorithm and ANFIS approach," *IEEE Access,* vol. 7, pp. 77674-77691, 2019.

[2] I. Ozer, S. B. Efe, and H. Ozbay, "A combined deep learning application for short term load forecasting," *Alexandria Engineering Journal,* vol. 60, pp. 3807-3818, 2021.

[3] E. Almeshaiei and H. Soltan, "A methodology for electric power load forecasting," *Alexandria Engineering Journal,* vol. 50, pp. 137-144, 2011.

[4] A. Pourdaryaei, H. Mokhlis, H. A. Illias, S. H. A. Kaboli, S. Ahmad, and S. P. Ang, "Hybrid ANN and artificial cooperative search algorithm to forecast short-term electricity price in de-regulated electricity market," *IEEE Access,* vol. 7, pp. 125369-125386, 2019.

[5] G.-F. Fan, X. Wei, Y.-T. Li, and W.-C. Hong, "Forecasting electricity consumption using a novel hybrid model," *Sustainable Cities and Society,* vol. 61, p. 102320, 2020.

[6] Y.-Y. Hong, J. V. Taylar, and A. C. Fajardo, "Locational Marginal Price Forecasting Using Deep Learning Network Optimized by Mapping-Based Genetic Algorithm," *IEEE Access,* vol. 8, pp. 91975-91988, 2020.

[7] F. Wu, C. Cattani, W. Song, and E. Zio, "Fractional ARIMA with an improved cuckoo search optimization for the efficient Short-term power load forecasting," *Alexandria Engineering Journal,* vol. 59, pp. 3111-3118, 2020.

[8] H. Wang, Y. Liu, B. Zhou, C. Li, G. Cao, N. Voropai, *et al.*, "Taxonomy research of artificial intelligence for deterministic solar power forecasting," *Energy Conversion and Management,* vol. 214, p. 112909, 2020.

[9] Y. Shuai, T. Song, and J. Wang, "Integrated parallel forecasting model based on modified fuzzy time series and SVM," *Journal of Systems Engineering and Electronics,* vol. 28, pp. 766-775, 2017.

[10] J. H. Zhao, Z. Y. Dong, X. Li, and K. P. Wong, "A framework for electricity price spike analysis with advanced data mining methods," *IEEE Transactions on Power Systems,* vol. 22, pp. 376-385, 2007.

[11] J. Dhillon, S. A. Rahman, S. U. Ahmad, and M. J. Hossain, "Peak electricity load forecasting using online support vector regression," in *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2016, pp. 1-4.

[12] X. Yan, Y. Song, and N. A. Chowdhury, "Performance evaluation of single SVM and LSSVM based forecasting models using price zones analysis," in *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2016, pp. 79-83.

[13] A. Heydari, F. Keynia, D. A. Garcia, and L. De Santoli, "Mid-term load power forecasting considering environment emission using a hybrid intelligent approach," in *2018 5th International Symposium on Environment-Friendly Energies and Applications (EFEA)*, 2018, pp. 1-5.

[14] R. Zhang, G. Li, and Z. Ma, "A Deep Learning Based Hybrid Framework for Day-Ahead Electricity Price Forecasting," *IEEE Access,* vol. 8, pp. 143423-143436, 2020.

[15] S. Zhou, L. Zhou, M. Mao, H.-M. Tai, and Y. Wan, "An optimized heterogeneous structure LSTM network for electricity price forecasting," *IEEE Access,* vol. 7, pp. 108161-108173, 2019.

[16] P. Lv, S. Liu, W. Yu, S. Zheng, and J. Lv, "EGA-STLF: A hybrid short-term load forecasting model," *IEEE Access,* vol. 8, pp. 31742-31752, 2020.

[17] H. Manner, F. A. Fard, A. Pourkhanali, and L. Tafakori, "Forecasting the joint distribution of Australian electricity prices using dynamic vine copulae," *Energy Economics,* vol. 78, pp. 143-164, 2019.

[18] Y. Elfahham, "Estimation and prediction of construction cost index using neural networks, time series, and regression," *Alexandria Engineering Journal,* vol. 58, pp. 499-506, 2019.

[19] G. Hamilton, A. Abeygunawardana, D. P. Jovanović, and G. F. Ledwich, "Hybrid model for very short-term electricity price forecasting," in *2018 IEEE Power & Energy Society General Meeting (PESGM)*, 2018, pp. 1-5.

[20] M. Alazab, S. Khan, S. S. R. Krishnan, Q.-V. Pham, M. P. K. Reddy, and T. R. Gadekallu, "A multidirectional LSTM model for predicting the stability of a smart grid," *IEEE Access,* vol. 8, pp. 85454-85463, 2020.

[21] S. K. Gupta, M. Tripathi, and J. Grover, "Hybrid optimization and deep learning based intrusion detection system," *Computers and Electrical Engineering,* vol. 100, p. 107876, 2022.

[22] S. K. Gupta, M. Tripathi, and J. Grover, "Towards an Effective Intrusion Detection System using Machine Learning techniques: Comprehensive Analysis and Review," in *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 2021, pp. 1-6.

[23] S. Patidar, M. Tripathi, and S. K. Gupta, "Leveraging LSTM-RNN combined with SVM for Network Intrusion Detection," in *Proceedings of the International Conference on Data Science, Machine Learning and Artificial Intelligence*, 2021, pp. 26-31.

[24] M. R. Haq and Z. Ni, "A new hybrid model for short-term electricity load forecasting," *IEEE Access,* vol. 7, pp. 125413-125423, 2019.

[25] J. Bedi and D. Toshniwal, "Deep learning framework to forecast electricity demand," *Applied energy,* vol. 238, pp. 1312-1326, 2019.

[26] U. Ugurlu, I. Oksuz, and O. Tas, "Electricity price forecasting using recurrent neural networks," *Energies,* vol. 11, p. 1255, 2018.

[27] U. Ugurlu, O. Tas, A. Kaya, and I. Oksuz, "The financial effect of the electricity price forecasts' inaccuracy on a hydro-based generation company," *Energies,* vol. 11, p. 2093, 2018.

[28] C. Fan, Y. Sun, Y. Zhao, M. Song, and J. Wang, "Deep learning-based feature engineering methods for improved building energy prediction," *Applied energy,* vol. 240, pp. 35-45, 2019.

[29] A. S. Weigend, *Time series prediction: forecasting the future and understanding the past*: Routledge, 2018.

[30] P. J. Brockwell, P. J. Brockwell, R. A. Davis, and R. A. Davis, *Introduction to time series and forecasting*: Springer, 2016.

[31] J. Osborne, "Improving your data transformations: Applying the Box-Cox transformation," *Practical Assessment, Research, and Evaluation,* vol. 15, p. 12, 2010.

[32] A. Pal and P. Prakash, *Practical time series analysis: master time series data processing, visualization, and modeling using python*: Packt Publishing Ltd, 2017.

[33] I. E. Livieris, S. Stavroyiannis, E. Pintelas, and P. Pintelas, "A novel validation framework to enhance deep learning models in time-series forecasting," *Neural Computing and Applications,* vol. 32, pp. 17149-17167, 2020.

[34] A. Naz, N. Javaid, M. Asif, M. U. Javed, A. Ahmed, S. M. Gulfam, *et al.*, "Electricity Consumption Forecasting Using Gated-FCN With Ensemble Strategy," *IEEE Access,* vol. 9, pp. 131365-131381, 2021.

[35] T.-Y. Kim and S.-B. Cho, "Predicting residential energy consumption using CNN-LSTM neural networks," *Energy,* vol. 182, pp. 72-81, 2019.

[36] X. Xie, M. Li, and D. Zhang, "A Multiscale Electricity Price Forecasting Model Based on Tensor Fusion and Deep Learning," *Energies,* vol. 14, p. 7333, 2021.

Rashed Iqbal

Rashed Iqbal received his B.Sc. degree in electrical and electronic engineering from the BRAC University, Bangladesh, in 2016, currently he is pursuing his degree in Master of Engineering Science in power system from the Universiti Malaya (UM), Malaysia. His current research interests include artificial intelligence in power system, renewable energy, smart and microgrids.

Hazlie Mokhlis

Prof. Ir. Dr Hazlie Mokhlis is currently a Professor at Department of Electrical Engineering, Universiti Malaya. His research interests include power system analysis (distribution automation), power quality, and fault location. His detail CV can be obtained from https://umexpert.um.edu.my/hazli

Anis Salwa Mohd Khairuddin

Anis Salwa Mohd Khairuddin is currently working as a senior lecturer at Department of Electrical Engineering, Faculty of Engineering, Universiti Malaya, Malaysia. Her research interests include Expert system (machine learning, optimization, prediction), signal and image processing. Her detail CV can be obtained from https://umexpert.um.edu.my/anissalwa.

Munir Azam Muhammad

Munir Azam Muhammad received a B.E. degree in electronic engineering from Hamdard University in 2010 and an M.E. degree in Industrial Electronics from N.E.D University, Pakistan, in 2013. He is currently pursuing a Ph.D. in Electrical Engineering in the University of Malaya (UM), Malaysia. His research interests include distribution automation, load shedding, distributed generation and development of optimization strategies for a smart grid.