Self-Supervised Attentive Feature Learning for Alzheimer's Disease Detection

Hela Elmannai[1]*, Nasir Saleem[2], Sami Bourouis[3], Reem Ibrahim Alkanhel[1]

[1] Department of Information Technology, College of Computer & Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671 (Saudi Arabia)
[2] Department of Electrical Engineering, FET, Gomal University (Pakistan)
[3] Department of Information Technology, College of Computers and Information Technology, Taif University (Saudi Arabia)

* Corresponding author: hselmannai@pnu.edu.sa

Received 25 February 2025 | Accepted 9 July 2025 | Early Access 25 September 2025

Abstract

Alzheimer's disease (AD) is a progressive neurodegenerative disorder that leads to memory loss and a decline in cognitive abilities. It primarily affects older adults and is the most common cause of dementia. Using deep learning, models can analyze brain imaging scans to detect specific patterns and biomarkers associated with the disease. Supervised learning models achieve high accuracy rates, but they require a large amount of data sets and labelled medical images. Self-supervised learning can achieve high accuracy rates with fewer training data. This study proposes a self-supervised attentive feature learning network (SSA-Net) for classifying Alzheimer's disease. The proposed approach leverages self-supervised learning and attention mechanisms to enhance the accuracy and reliability of the classifying model. We employ ResNet-50, incorporating attentive activation, which replaces the ReLU activation, improving the ability of the neural model to focus on the most relevant features in the input medical images. We use SimCLR (Simple Framework for Contrastive Learning of Visual Representations) with the ResNet-50 backbone as a self-supervised learning framework that effectively learns high-quality visual representations in brain MRI (Magnetic Resonance Imaging) scans without labelling. We used the Kaggle Alzheimer's classification dataset (KACD) containing brain MRI scans for training and testing. Experimental results on the KACD dataset show that the proposed attentive self-supervised ResNet50 reached 99.7% classification accuracy compared to the traditional ResNet50 with 98.1% accuracy. Evaluation metrics show the effectiveness of the proposed SSA-Net for the efficient classification of Alzheimer's disease.

Keywords

Alzheimer's Disease, Attentional Network, Brain MRI, Feature Learning, Intelligent Classification, SimCLR, Self-Supervising Learning.

DOI: 10.9781/ijimai.2025.09.002

I. Introduction

Alzheimer disease is a progressive neurodegenerative disorder that leads to the gradual deterioration of cognitive abilities and memory. It stands as the leading cause of dementia in older adults. Classifying AD into stages such as non-demented, very mildly demented, mildly demented, and moderately demented is essential. This classification allows for personalized medical care and effective monitoring of disease progression. The early and exact identification of the AD stage enables healthcare providers to tailor interventions more effectively, potentially slowing the progression of the disease. AD causes extensive neuronal damage, disrupting associations with critical brain regions like the cerebral cortex and hippocampus. However, the exact reason for Alzheimer's remains unidentified; it is often associated with the accumulation of neurofibrillary tangles and amyloid plaques [1]. The report discussing the increasing apprehensions regarding the growing trend of dementia across the world [2]. The number of dementia cases is predicted to increase from 57.4M to 152.8M by 2050. The World Health Organization (WHO) documents that there are 10 million new dementia cases annually worldwide, which means one new case every 3.2 seconds. These reports indicate a notable rise in dementia over time, emphasizing effective systems to manage this growing challenge.

Classification of AD using deep neural networks (DNNs) is an emerging domain that leverages the ability of artificial intelligence to improve detection accuracy. DNN models, especially convolutional neural networks (CNNs), have demonstrated significant potential in analyzing medical imaging data such as MRI (Magnetic Resonance Imaging) and PET (Positron Emission Tomography) scans. MRI provides high-resolution structural images of the brain, helping

Please cite this article as:
H. Elmannai, N. Saleem, S. Bourouis, R. I. Alkanhel. Self-Supervised Attentive Feature Learning for Alzheimer's Disease Detection, International Journal of Interactive Multimedia and Artificial Intelligence, (2025), http://dx.doi.org/10.9781/ijimai.2025.09.002

identify atrophy patterns typical of AD, while PET scans, using tracers like 18F-FDG, reveal metabolic activity indicative of the disease [3], [4]. These models are trained on large datasets to detect fine changes in brain structure and function, offering a non-invasive, automated approach to early detection and progression monitoring. Integrating DNNs in AD research improves early detection and personalized treatment procedures. MRI is considered the most advanced and precise radiological technique to examine the brain's structural transitions. MRI captures images from different angles, typically in axial, sagittal, and coronal planes, providing comprehensive insights into the brain's anatomy and potential abnormalities [5]. Detecting alterations in the corpus callosum is possible through axial plane MRI scans. This imaging technique is valuable for differentiating between various phases of AD [6], [7].

Recently, countless studies have examined the potential use of DNNs in detecting AD. Classes of the convolutional neural networks (CNN), such as, ResNet [8], [9], VGG [10], AlexNet [11], GoogleNet [12], and recurrent neural networks (RNN) [13], are considered to improve the AD detection. A study [14] introduced the model to classify four dementia stages—very mild dementia, mild dementia, moderate dementia, and non-dementia. Dementia image features are effectively visualized by employing the occlusion sensitivity map alongside multifocal ground-glass opacities. Another study [15] devised a model that combines Inception and ResNet, achieving significant improvements against standalone ResNet and Inception. A study [16] classified distinct AD stages: cognitively normal (not-demented), early-stage (very mild-demented), later-stage (mild-demented), and AD (moderate-demented). Another study employed different CNN models, including ResNet-50, DenseNet-201, VGG-16, and AlexNet to develop a hybrid model based on ResNet50 with additional layers. The proposed hybrid model showed improved performance in detecting Alzheimer's disease. Another study [17] formulated an automated neural detection approach operating on the sagittal plane of the brain MRI. The study operated transfer learning, utilizing ResNet for feature extraction and a SVM classifier for AD detection. Another study [18] proposed an Alzheimer's detection using a pre-trained AlexNet for feature extraction. These features are then classified using different classifiers, including KNN, random forest, and SVM. A study [19] presents a multistage CNN model to detect AD. The framework has developed a 26-layer convolutional neural network specifically designed to distinguish between individuals without dementia and those with dementia. The model is reutilized via transfer learning to sub-classify dementia into severe, moderate, and mild stages. Exploiting the fixed weights, transfer learning has facilitated precise sub-classification of AD. In another study [20] a CNN architecture is formulated that utilized MRI data for AD classification. The network consists of two CNN models with different filter sizes and pooling layers, concatenated in a classification layer. Another study [21] uses MobileNet to classify five AD stages. Another study [22] employs a combination of deep CNN and ensemble learning, specifically MobileNetV2 and LSTM, using magnetic resonance images (MRI). In another study [23], a particle swarm optimization-based model is proposed to optimize hyperparameters for classifying Alzheimer's disease severity. The CNN model achieved a highly accurate classification of Alzheimer's disease, achieving 99.53% accuracy and a 99.63% F1-score. The study [24] introduces four different models for classifying manifold dementia stages including custom-built CNNs, VGG-16 using extra convolutional layers, graph convolutional network, and a hybrid of CNN and graph convolutional network. CNNs are first formed and their flattened layer outputs are subsequently fed to the graph convolutional network classifier. Another recent study [25] presents an architecture combining Multi-View Separable Residual CNNs, capable of processing total volumes with spatial complexity related to 2D CNNs. The model integrates voxel spatial relationships and achieves a significant 50% reduction in memory usage compared to 3D CNNs. Evaluated on a dataset of 540 patients (191 CN, 145 EMCI, 122 LMCI, 82 AD, 397 SMCI, and 61 PMCI), the model achieved accuracies of 86.97% for CN vs. EMCI vs. LMCI vs. AD and 95.73% for SMCI vs. PMCI classifications. A study [26] examines reinforcement learning to classify AD.

For AD classification, integrating ResNet architectures with self-attention can be extremely beneficial. ResNet can capture complex hierarchical features from MRI scans for classifying fine structural changes signifying AD stages. By using pretrained or fine-tuned ResNet models, the network can efficiently extract features that signify early signs of AD. A study [27] introduces an AlzhiNet model formulated for detecting AD using 3D volumetric MRI data for multi-class classification. AlzhiNet incorporates self-attention to differentiate between stages such as mild cognitive impairment (MCI) and AD, with cognitively normal patients helping as the control group. A study suggested using a self-attention method to divide the brain into 90 regions by employing automatic anatomical labelling [28]. The next step involves calculating intermediate fractional anisotropy scores for each pair of brain areas. These scores are then used to construct a detailed and complex brain network. This network was then fed to the graph convolutional neural network, which integrated a self-attention pooling process. Additionally, the study leveraged the number of voxels with fibers passing through each brain region as node features. The researchers assessed various preprocessed brain networks and node features for their classification performance and reported a remarkable accuracy rate of 87.5% in distinguishing between individuals with Alzheimer's disease (AD) and normal controls (NC). A study [29] employed the ResNet50 architecture, modified to include self-attention for extracting meaningful features. The process initiated with the optimization of both the hyperparameters and the model using Bayesian optimizers. Once optimized, the model was utilized to extract features, which were then further optimized through a self-attentional approach. The most effective features were selected and fed into the classifiers for the ultimate classification task. As a result of these experiments, a significant improvement was achieved with an accuracy of 99.9%.

This study uses the Simple Framework for Contrastive Learning of Visual Representations (SimCLR) framework to train an attentional ResNet-50 model in a self-supervised learning manner. Since unsupervised learning operates without explicit class labels during training, this study formulates a neural model to accurately classify AD stages from MRI images as cognitively normal (non-dementia), early-stage (very mild-dementia), later-stage (mild-dementia), and AD (moderate-dementia). During the training process using the SimCLR learning framework, all input data points exclusively contain MRI images without label information. For ResNet-50, the architecture largely remains unchanged in terms of layer structure (convolutional layers and skip connections), but replaces ReLU activation with an attentive activation (AttAct) [30]. Attentive activation focuses on important features in the MRI images, allowing the model to selectively emphasize relevant information while de-emphasizing less relevant features. This potentially leads to more learned discriminative features, improving the ability of the model to determine AD classes (non-dementia, very mild dementia, mild dementia, and moderate dementia). The model learns to extract features that capture high-level abstract information regarding the input MRI images, guided by the contrastive loss function. After training on the unlabeled MRI images, the SimCLR framework produces learned feature representations (embeddings) of these images. Once the model learns feature representations from unlabeled images, the representations are fine-tuned for AD classification. The contributions of the study are twofold.

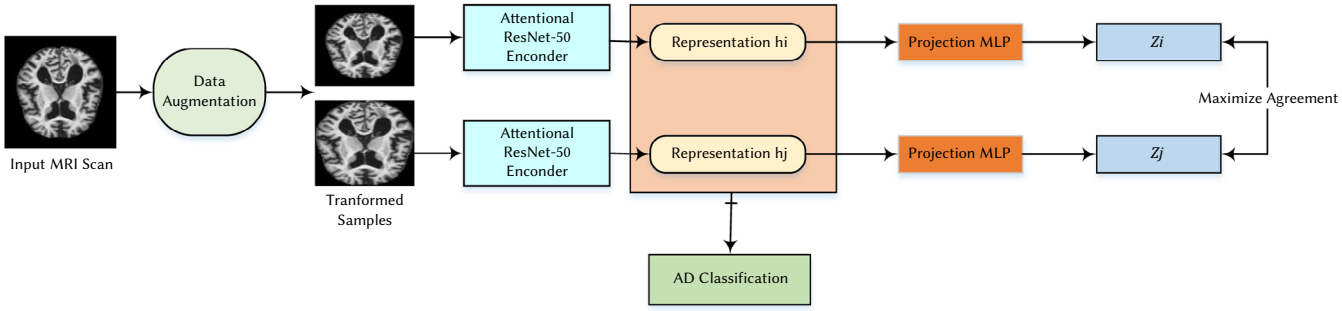

Fig. 1. Proposed Methodology with unsupervised feature learning using attentional ResNet-50.

First, we employ the SimCLR framework to train an attentional ResNet-50 model using a self-supervised learning approach that focuses on learning effective representations from unlabeled MRI images. The SimCLR applies data augmentation to create multiple views of MRI images, which helps the model learn robust features that are consistent with interpretations in the input data. Secondly, for ResNet-50, we make a few architectural changes and replace standard ReLU activation with an attentive activation. Unlike ReLU, which applies a fixed non-linearity, attentive activation adaptively adjusts the responses based on the input and context. This adaptability leads to more robust feature representations that are adjusted to the specific characteristics of MRI images, potentially improving generalization. The remaining paper is arranged as follows: Section II provides a detailed description of the proposed model (SSA-Net). The experimental setup is explained in Section III. Section IV presents the experimental results and discussions. Finally, Section V concludes the study.

II. Methodology

We utilize the SimCLR framework to train an attentional ResNet-50 model through a self-supervised learning method aimed at deriving effective representations from unlabeled MRI images. SimCLR employs data augmentation to generate multiple views of MRI images, enabling the model to learn robust features consistent with the input data interpretations. For the ResNet-50 model, we introduce several architectural modifications and replace the standard ReLU activation with an attentive activation [31]. Unlike ReLU, which applies a fixed non-linearity, attentive activation adaptively adjusts responses based on input and context. This adaptability results in more robust feature representations. The network is further fine-tuned using SimCLR. This indicates that pre-trained weights are used as initial weights for training, where input data includes label information. Fig. 1 depicts the proposed methodology.

A. Training Protocol

SimCLR learns representations by maximizing the similarity between different augmented views of the same MRI samples using a contrastive loss function in the latent space. As illustrated in Fig. 1, the framework contains four components. First, a stochastic data augmentation module randomly transforms the given MRI image, resulting in two correlated views of the same MRI sample, $\hat{s}_i$ and $\hat{s}_j$, which are considered a positive pair. In this study, we apply simple random cropping augmentation. Secondly, the ResNet-50 neural network base encoder $f(\cdot)$ extracts feature representations from the augmented MRI samples; i.e.,

$$h_i = f(\hat{s}_i) = \text{ResNet50}(\hat{s}_i) \tag{1}$$

where $h_i \in \mathbb{R}_d$ represents the outputs after the average pooling layer.

After this, the representations are then processed by a projection head $g(\cdot)$ to map them into the space where the contrastive loss is applied. A single hidden-layered MLP is used for this projection, which is given as:

$$z_i = g(h_i) \tag{2}$$

$$z_i = W^{(2)}\,\sigma(W^{(1)}h_i) \tag{3}$$

where z_i denotes the projected feature used for computing the contrastive loss and σ is the ReLU non-linearity. Following the projection head, a contrastive loss function is applied for contrastive prediction. Given a set $\{\hat{s}_m\}$ that includes the positive pair samples $\hat{s}_i$ and $\hat{s}_j$, the contrastive prediction aims to identify $\hat{s}_j$ in $\{\hat{s}_m\}_{m\neq i}$ for a given $\hat{s}_i$. We follow the standard methodology for computing the contrastive loss. The loss for the positive pair samples (i, j) can be defined as:

$$\mathcal{L}_{(i,j)} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{m=1}^{2N} \mathbf{1}_{[m\neq i]}\ \exp(\text{sim}(z_i, z_m)/\tau)} \tag{4}$$

where $\mathbf{1}_{[m\neq i]} \in \{0,1\}$ is an indicator function and τ denotes the temperature parameter. The final loss function is computed over all positive pairs in a mini-batch. This loss is known as the normalized temperature-scaled cross-entropy loss.

B. Model: Attentional ResNet-50

ResNet-50 is a deep convolutional neural network that has been effectively utilized in Alzheimer's disease analysis to improve the accuracy of early detection and progression prediction. By leveraging a 50-layer architecture, ResNet-50 captures complex patterns and fine changes in brain MRI images. The model's ability to perform residual learning allows it to address the vanishing gradient problem, ensuring robust feature extraction even in deep layers. This capability is important for identifying the fine structural and functional alterations associated with Alzheimer's disease, thus aiding in early intervention and personalized treatment planning.

We use ResNet-50 as the base model with a few architectural changes and replace the standard ReLU activation with an attentive activation (AttAct). Attention and activation are distinct functions, but both are nonlinear gating functions. Drawing on this similarity, the ResNet-50 encoder combines attention and activation as a single approach, called attentional activation (AttAct).

The local channel attention block in AttAct enables simultaneous nonlinear activation and feature refinement. This block collects local point-by-point cross-channel feature contexts, improving the interpretation and performance of ResNet-50. Fig. 2 demonstrates the architecture of the AttAct block integrated into ResNet-50 for a fully attentional ResNet-50 backbone.

When considering a particular intermediate feature map denoted as f, the transformation brought about by attention can be expressed as follows:

$$f' = \text{AttG}(f) \otimes f \tag{5}$$

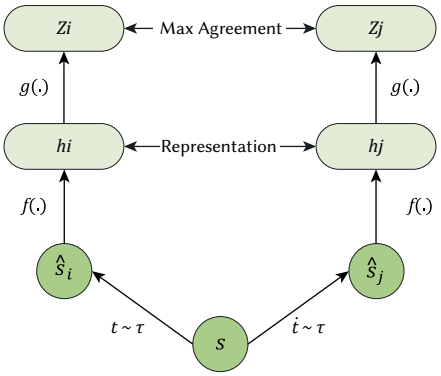

Fig. 2. The contrastive learning of visual representations.

where $\otimes$ represents element-wise multiplication and $\mathrm{AttG}(f)$ denotes the attention-weighted feature map created via attention gating. The effectiveness of the attention gating relies significantly on the complete feature map f. At any given position (i, j, k), Equation (5) can be expressed in scalar form as:

$$f'_{[i,j,k]} = \mathrm{AttG}(f)_{[i,j,k]} \cdot f_{[i,j,k]} = g_{[i,j,k]} \cdot f_{[i,j,k]} \tag{6}$$

where g performs complex gating by aggregating feature contexts for the position (i, j, k), yielding attention weights for $F_{[c,j,k]}$.

According to [31], activation can also be represented as a gating function:

$$f'_{[i,j,k]} = g(X_{[i,j,k]}) \cdot X_{[i,j,k]} \tag{7}$$

In a typical ReLU activation function, the scalar function g acts as an indicator by determining whether the input is greater or less than zero. The AttAct, aside from introducing non-linearity, enables ResNet to carry out context-aware, layer-wise, and adaptive feature refinement. In this study, AttAct is implemented as the activation function throughout the entire ResNet architecture.

AttAct utilizes point-by-point convolutions to incorporate local attention. These convolution operations allow for the exploration of cross-channel relationships in a point-by-point style while efficiently utilizing only a small set of training parameters. The attentional activation, as in Fig. 3, empowers ResNet to activate and refine selective features element-wise, leveraging point-by-point cross-channel correlations. The attentional weights $\mathrm{Att}(f)$ are calculated using a bottleneck configuration to reduce the number of parameters:

$$\mathrm{Att}(f) = \sigma(\mathrm{BN}(\mathrm{PWConv}_2(\delta(\mathrm{BN}(\mathrm{PWConv}_1(f)))))) \tag{8}$$

where δ and σ represent the ReLU and sigmoid activation functions, and BN refers to batch normalization. The kernel sizes are $\mathrm{PWConv}_1 = (C/r \times C \times 1 \times 1)$ and $\mathrm{PWConv}_2 = (C \times C/r \times 1 \times 1)$, where r is the channel reduction ratio.

The function $Att(f)$ has the same shape as the input feature map, and the resultant activated features are obtained via element-wise multiplication:

$$f' = \mathrm{Att}(f) \otimes f \tag{9}$$

Instead of using the traditional ReLU activation function, AttAct selectively activates and refines features in the ResNet-50 layers.

III. Experiments and Settings

A. Dataset and Data Splitting

This study uses the Kaggle Alzheimer's classification Dataset (KACD), which contains brain MRI scans and was obtained from a challenge on the Kaggle. The dataset consists of 6400 MRI samples

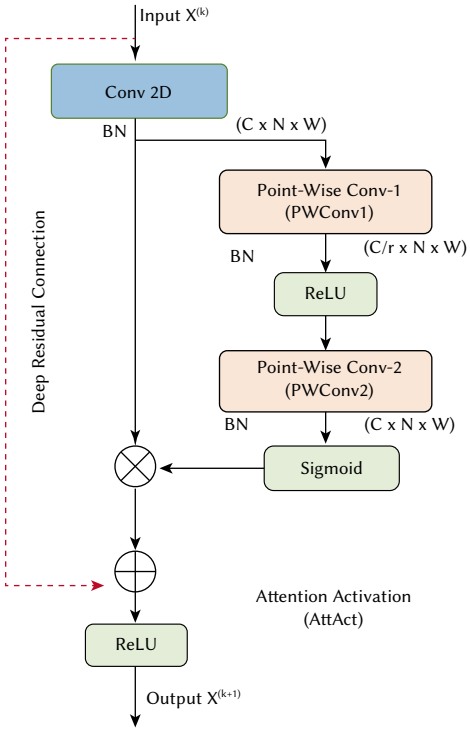

Fig. 3. Architecture of the attentional activation (AttAct).

(5121 training and 1279 testing samples). The dataset is divided into non-demented, very mild demented, mild demented, and moderate demented. The scans are collected in the axial plane with (128× 128) pixels dimensions, as shown in Fig. 4. The non-demented class indicates normal cognition, whereas the very mild demented class indicates an early stage of disease with fine symptoms. The mild demented describes the stage of Alzheimer's disease at which symptoms are noticeable and can be identified. In the moderate demented class, a patient encounters difficulty in thinking, reasoning, movement, and day-to-day tasks. Table I provides the details of the KACD. Cross-validation is a useful approach used to evaluate the performance of a model on new, unseen data, which helps generalize the model for independent datasets. This study utilizes a k-fold cross-validation approach with k set to 5. The process begins by splitting the dataset into a training set (80%) and a testing set (20%). Following the initial assessment, k-fold cross-validation is implemented by dividing the training set into k subsets. Finally, the performance of the model is evaluated on a 20% test set, which is not used for either training or cross-validation.

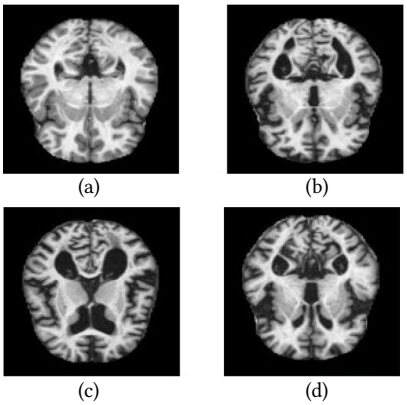

Fig. 4. MRI images. (A) Non-demented, (B) Very Mild, (C) Mild, and (D) Moderate.

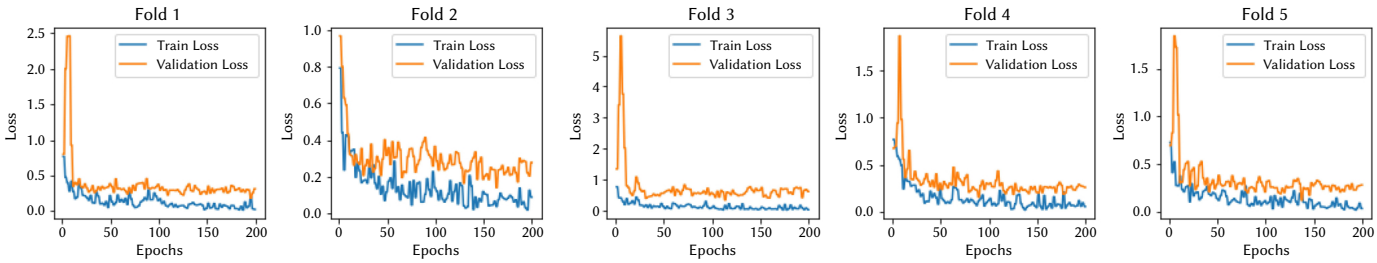

Fig. 5. Learning curves on KACD dataset. The training occurs to be stable even on limited MRI data.

TABLE I. Details of the KACD With MRI Brain Images

Class Name	Train Set	Test Set	Total
Non-Demented	2560	640	3200
Very Mild Demented	1792	448	2240
Mild Demented	717	179	896
Moderate Demented	52	12	64
Total Class	5121	1279	6400

B. Network and Performance Metrics

ResNet-50 is a deep convolutional neural network (DCNN), and the architecture details are given in Table II. In the network configuration, the training is set to run for 200 epochs utilizing Stochastic Gradient Descent (SGD) with Sharpness-Aware Minimization (SAM) as the optimization strategy. The loss function employed is the normalized temperature-scaled cross-entropy, which is designed to improve model performance by addressing the issue of overconfidence in class predictions.

TABLE II. The ResNet-50 Backbone Network Architecture Where ReLU Is Replaced With Attentional Activation (AttAct)

Layer	Output	50-Layers Description
Conv1	112 x 112	7 x 7Conv, 64
Conv2	56 x 56	[(1 x 1, 64), (3 x 3, 64), (1 x 1, 256)] x 3
Conv3	28 x 28	[(1 x 1, 128), (3 x 3, 128), (1 x 1, 512)] x 4
Conv4	14 x 14	[(1 x 1, 256), (3 x 3, 256), (1 x 1, 1024)] x 6
Conv5	7 x 7	[(1 x 1, 512), (3 x 3, 512), (1 x 1, 2048)] x 3
–	1 x 1	Avg Pool, 1000d Fully Connected, Softmax
FLOPs		3.8×10^9

The learning rate is initialized at 0.001, with a momentum of 0.9 to help accelerate the training process and improve convergence. The ρ parameter is set to 2, controlling the sharpness penalty in the SAM optimizer. Additionally, weight decay is applied with a value of 1×10^{-8} to regularize the model and prevent overfitting. The optimizer used for training the network is Adam, known for its adaptive learning rate and weight decay, which further enhances the training stability and performance of the ResNet-50 model. β is the temperature parameter which controls the sharpness of the similarity distribution, with a value fixed to 0.1. The models are implemented using Python and PyTorch, running on an Intel(R) Core(TM) i5-1135G7 CPU @ 2.40GHz (8 CPUs) processor with 8GB RAM and an NVIDIA GeForce GTX 960. Fig. 5 displays the learning curves for the AD stage classification task using the KACD dataset, showing no signs of overfitting throughout the training process.

Key metrics for the performance evaluation include Accuracy (ACC), Sensitivity (Recall), Precision, F1-score, and Area Under Curve (AUC). These metrics present comprehensive details of a model's performance. The ACC, Recall, and Precision metrics are defined as:

$$\text{Accuracy (ACC)} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$\text{Recall (Sensitivity)} = \frac{TP}{TP + FN} \tag{11}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

where *TP*, *TN*, *FP*, and *FN* represent the true positives, true negatives, false positives, and false negatives, respectively.

IV. Results and Discussions

We present the results of the proposed Alzheimer's Disease classification model, evaluated using Accuracy, Recall, F1-score, and Precision. The performance of our model is compared to other DNN-based models for Alzheimer's Disease classification. To determine the best accuracy, we employ 5-fold cross-validation.

A. Overall Performance

Table III demonstrates the results of the proposed SSA-Net. This model involves training ResNet-50 using the SimCLR learning framework and replacing the standard ReLU with attentional activation. The model's performance is evaluated based on Accuracy, Precision, F1-score, Recall, Error rate, and False-Positive Rate (FPR) across non-demented (ND), very mild demented (VMD), mild demented (MD), and moderate demented (MOD).

TABLE III. Performance of the Proposed SSA-Net on Alzheimer's Disease Detection and Classification

Metric	ND	VMD	MD	MOD
Accuracy	99.77%	99.77%	99.84%	99.97%
Precision	99.69%	99.78%	99.44%	99.97%
F1-Score	99.76%	99.69%	99.44%	99.97%
Recall	99.84%	99.60%	99.44%	99.97%
Error Rate	0.0023	0.0023	0.0016	0.00
FPR	0.0031	0.0012	0.0009	0.00

The results indicate highly accurate Alzheimer's Disease detection and classification across all classes: mild demented (99.84% accuracy, 99.44% precision, 99.44% recall, and 99.44% F1 score), moderate demented (100% accuracy, precision, recall, and F1 score), non-demented (99.77% accuracy, 99.69% precision, 99.84% recall, and 99.76% F1 score), and very mild demented (99.77% accuracy, 99.78% precision, 99.56% recall, and 99.67% F1 score). These metrics demonstrate minimal error rates (ranging from 0% to 0.23%) and low false positive rates (ranging from 0% to 0.31%), underscoring the model's robustness in accurately distinguishing between different dementia stages. These findings indicate the critical importance of early detection and effective procedures in clinical settings, supporting the potential of deep learning models in improving the detection of AD accuracy.

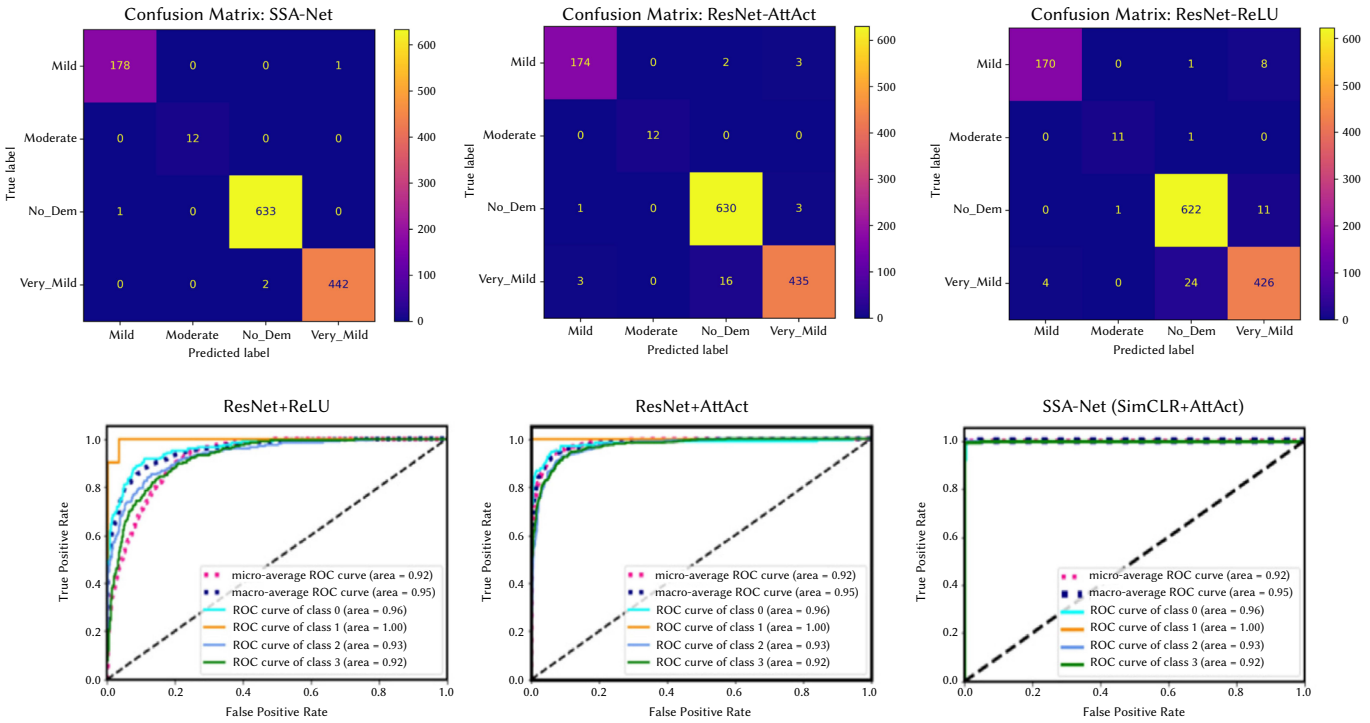

Fig. 6. Confusion Matrix Analysis. ResNet-ReLU (with ReLU activation in the encoder); ResNet-AttAct (with attentional activation in the encoder); and SSA-Net (trained with SimCLR and attentional activation in the encoder). Receiver Operating Characteristic Curve (ROC). ResNet+ReLU; ResNet+AttAct, and SSA-Net (SimCLR+AttAct).

To demonstrate the influence of attentional activation in the proposed model, we trained the ResNet-50 model using the SimCLR learning framework and the standard ReLU activation. The results indicated that with ReLU activation, the model's performance declined across all metrics and classes, except for the moderate dementia class. Table IV presents the results of the model without attentional activation. The accuracy of SSA-Net is increased by 1.26% (non-demented), 1.54% (very mild demented), 0.39% (mild demented), and both models perform equally well for the moderate demented category. The F1-score of SSA-Net has seen an increase of 1.23% for the non-demented category, 2.13% for very mild demented, and 1.41% for mild demented. Both models perform equally well for the moderate demented category.

TABLE IV. Performance of ResNet-50 With Attentive Activation Replacing ReLU on Alzheimer's Disease Detection and Classification

Metric	ND	VMD	MD	MOD
Accuracy	98.51%	98.28%	99.45%	99.96%
Precision	97.41%	98.90%	98.86%	99.96%
F1-Score	98.53%	97.56%	98.03%	99.96%
Recall	99.68%	96.25%	97.21%	99.96%
Error Rate	0.0148	0.0172	0.0055	0.00
FPR	0.0266	0.0060	0.0018	0.00

Fig. 6 shows an analysis of the confusion matrix for three deep learning models: SSA-Net, ResNet trained by SimCLR, and ResNet-50 with ReLU activation. The confusion matrices confirm the results and show the success of the SSA-Net. The Receiver Operating Characteristic Curve (ROC) of the ResNet+ReLU, ResNet+AttAct, and SSA-Net (SimCLR+AttAct) are also plotted in Fig. 6. Fig. 7 visualizes the heatmaps of ResNet-50 with SimCLR and attentional activation on the disease classes.

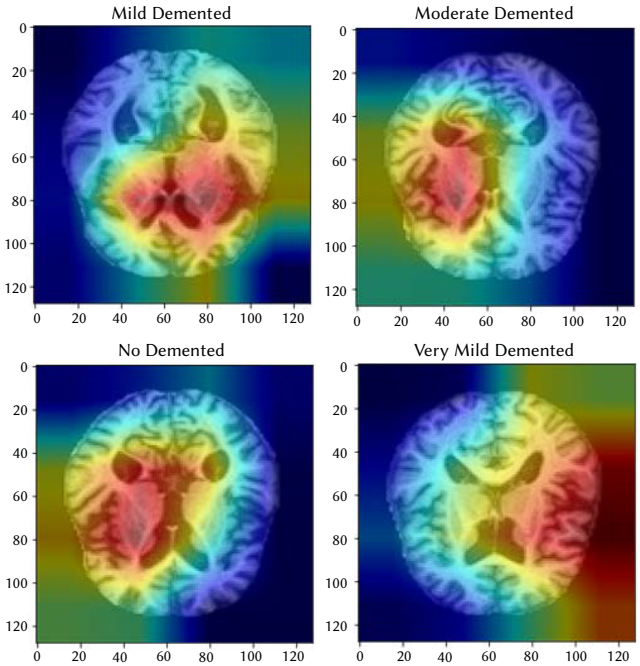

Fig. 7. Visualization of the heatmaps on the disease classes.

Table V provides experimental results for 5-fold cross-validation on the KACD dataset. The proposed SSA-Net model demonstrated outstanding accuracy, achieving a mean precision score of 99.72%. This reflects its capability to correctly identify positive cases within the dataset. Additionally, it achieved a mean recall rate of 99.75%, indicating its effectiveness in capturing actual positive instances and minimizing false negatives. The F1-score, which balances precision and recall, was an outstanding 99.72%, highlighting the robustness of our model's performance. Furthermore, the model excelled in precision,

with a mean score of 99.7%, showcasing a low number of false positive predictions and demonstrating strong capability in predicting positive cases correctly.

TABLE V. Experimental Results of 5-Fold Cross-Validation

Fold No.	Accuracy	Precision	F1-Score	Recall
1	99.76%	99.71%	99.76%	99.84%
2	99.68%	99.68%	99.68%	99.68%
3	99.75%	99.74%	99.75%	99.75%
4	99.65%	99.65%	99.65%	99.65%
5	99.77%	99.71%	99.77%	99.85%

B. Comparison With Other Models

To assess the performance of the proposed SSA-Net, we compare the results with other state-of-the-art (SOA) neural models. The SOA models include SADNN [32], DCNN [33], DenseNet [34], MSD-Net [35], MS-RLM [36], MM3D-DL [37], ensemble CNN-LSTM [38], ADG-Net [39], PDL [40], and PL-Net [41].

SADNN is a model for early AD detection using a 3D-residual attention DNN for end-to-end learning from MRI scans. The model introduces a residual self-attention DNN to capture local, global, and spatial information, improving detection performance. DCNN uses a convolutional neural network for AD classification from MRI scans. DenseNet proposes a hippocampus analysis model that combines global and local features using densely connected convolutional networks and shape analysis for AD detection. MSD-Net proposes a multi-view separable pyramid network, which learns representations from axial, coronal, and sagittal scan views for AD classification. MS-RLM proposes a multi-scale region classification block which adaptively combines multi-scale regions and drives decision fusion. MM3D-DL proposes a multimodal approach that extracts features without loss using a depthwise separable convolution block without an activation function.

Table VI provides a comprehensive performance overview of various deep learning models in classifying neurodegenerative diseases across different tasks and modalities (MRI and PET). Each model's effectiveness is evaluated through metrics including accuracy, recall, and F1-score. The proposed SSA-Net in this study stands out with excellent performance metrics (99.70% accuracy, 99.80% recall, and 99.70% F1-score) when utilizing MRI scans from regions of interest (ROI) to distinguish between non-demented (ND) and Alzheimer's Disease (AD). This underlines the robust ability of the proposed model to accurately classify disease states, important for the early detection of AD.

On the other hand, models such as MS-RLM and MM3D-DL exhibit varying performance depending on the task and modality used. MS-RLM achieves high accuracy (98%), balanced recall (96%), and F1-score (96%) when using PET data across all regions to classify ND vs AD. This indicates PET imaging's effectiveness in capturing disease markers across brain areas, contributing to accurate disease classification. MM3D-DL, employing both MRI and PET modalities together (MRI+PET), shows mixed results across tasks. For instance, it achieves 84% accuracy, 76%recall, and 81% F1-score in distinguishing Moderate (MOD) vs Mild Dementia (MD) using combined imaging data from all brain regions. This reflects the complexity of integrating multimodal data for disease classification, where combining MRI and PET provides complementary information but also introduces challenges in data integration and model training.

To assess if there are statistically significant differences in the SSA-Net performance between different models, we conducted ANOVA tests. In Table VII, the symbol (+) signifies a significant difference with a 95% confidence level. The statistical analysis revealed that the results of the proposed SSA-Net model are indeed significant ($p < 0.0001$), supporting the null hypothesis (+). Similarly, the benchmark models also yielded significant results ($p < 0.0001$) in support of the null hypothesis. Table VIII further examines the

TABLE VI. Comparison Against State-of-the-art (SOA) Models. "—" Indicates No Results in the Original Study

Model	Optimizer	Architecture	Modality	MRI Region	Accuracy	Recall	F1-Score	Dataset
SADNN	AdamW	3D Residual CNN	MRI	All	91.00	91.00	—	ANDI
DCNN	Adam	3D Deep CNN	MRI	ROI	—	88.00	—	ANDI
DenseNet	Adam	Dense CNN	MRI	ROI	92.00	94.00	—	ANDI
MSD-Net	SGD	Multi-View CNN	PET	ROI	93.00	91.00	—	ANDI
MS-RLM	Adam	Multi-Scale CNN	PET	All	98.00	96.00	96.00	ADNI
MM3D-DL	Adam	Multimodal CNN	MRI+PET	All	94.00	99.00	93.00	ANDI
CNN-LSTM	Adam	Conv1CNN+LSTM	MRI	ROI	98.75	98.79	98.81	KACD
ADG-Net	Adam	Deep CNN	MRI	All	99.61	99.53	99.61	KACD
PDL	Adam	Pre-Train CNN	MRI	All	99.30	99.01	99.01	KACD
PL-Net	Adam	CNN Backbone	MRI	All	99.50	99.80	—	KACD
SSA-Net	(Ours)	AdamW	ResNet50+SimCLR	MRI	All	99.70	99.80	99.70

TABLE VII. ANOVA Statistical Analysis With A 95% Confidence Interval

Model	Accuracy		Recall		F1-Score	
	p-value	H0	p-value	H0	p-value	H0
SADNN	<0.0024	(+)	<0.0011	(+)	<0.0019	(+)
DCNN	<0.0028	(+)	<0.0015	(+)	<0.0020	(+)
DenseNet	<0.0010	(+)	<0.0012	(+)	<0.0012	(+)
MSD-Net	<0.0015	(+)	<0.0011	(+)	<0.0012	(+)
MS-RLM	<0.0001	(+)	<0.0001	(+)	<0.0001	(+)
MM3D-DL	<0.0002	(+)	<0.0009	(+)	<0.0009	(+)
SSA-Net	<0.0001	(+)	<0.0001	(+)	<0.0001	(+)

performance of SSA-Net against other CNN models on the accuracy, F1-score, recall, AUC, and parameter count. The parameter count provides the computational complexity of the SSA-Net.

TABLE VIII. Performance of SSA-Net Against Other CNN Models

Model	Para (M)	Accuracy	Precision	Recall	AUC
MobileNetV2	224M	98.74	99.39	99.37	99.95
DenseNet121	10M	84.39	64.98	53.14	80.83
AlexNet	60M	83.75	62.94	60.01	77.64
InceptionV2	154M	94.39	94.49	94.49	99.49
ResNet50	25M	84.88	77.84	70.74	80.64
SSA-Net	25.5M	99.72	99.70	99.75	99.98

V. Conclusion

This study introduces a self-supervised attentive feature learning network (SSA-Net) for classifying Alzheimer's disease. The method enhances model accuracy and reliability by combining self-supervised learning and attention mechanisms. The backbone ResNet-50 model with attentive activation in the encoder is used instead of ReLU, improving focus on relevant features. SimCLR, a self-supervised learning framework, is employed to learn high-quality visual representations in brain MRI scans without labels. On the Kaggle Alzheimer's classification dataset (KACD) for training and testing, the proposed SSA-Net achieved 99.7% classification accuracy, surpassing the accuracy (98.1%) of ResNet-50. The experiments are conducted to examine the proposed SSA-Net using Accuracy, Precision, F1-score, Recall, Error rate, and False-Positive Rate (FPR) across non-demented (ND), very mild demented (VMD), mild demented (MD), and moderate demented (MOD) categories. The results conclude highly accurate Alzheimer's Disease detection and classification across all classes: mild demented (99.84%accuracy), moderate demented (100% accuracy), non-demented (99.77% accuracy), and very mild demented (99.77% accuracy). They also demonstrate minimal error rates (ranging from 0% to 0.23%) and low false positive rates (ranging from 0% to 0.31%), underscoring the model's robustness in accurately distinguishing between different dementia stages. These findings emphasize the importance of early detection and effective procedures, supporting the potential of deep learning models in improving the detection of AD. The results concluded that with ReLU activation, the model's performance declined across all metrics and classes, except for the moderate dementia class. The accuracies of SSA-Net increased by 1.26% (non-demented), 1.54% (very mild demented), and 0.39% (mild demented) over ResNet+ReLU. The confusion matrices and ROC curves further confirm the excellent performance of SSA-Net. On 5-fold cross-validation on the KACD dataset, SSA-Net achieved a mean precision score of 99.72%, a mean recall rate of 99.75%, and a mean F1-score of 99.72%. Overall, SSA-Net shows better performance, emphasizing the potential of deep learning models in enhancing the accuracy of AD detection and classification, supporting their application in clinical settings for early detection and effective treatment.

In future studies, the investigation can combine SSA-Net with clinical data, such as demographics of patients, genetic information, and cognitive tests to improve the prediction of the model. Additionally, we can validate the performance of SSA-Net on additional diverse datasets to ensure its generalization and robustness across different imaging conditions.

Acknowledgment

This research project was funded by the Deanship of Scientific Research and Libraries, Princess Nourah bint Abdulrahman University, through the Program of Research Project Funding After Publication, grant No (RPFAP-75-1445).

References

[1] S. Afzal, M. Maqsood, U. Khan, I. Mehmood, H. Nawaz, F. Aadil, O. Y. Song, Y. Nam, "Alzheimer disease detection techniques and methods: A review," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 7, pp. 26–38, 2021.

[2] M. H. Al-Adhaileh, "Diagnosis and classification of alzheimer's disease by using a convolution neural network algorithm," *Soft Computing*, vol. 26, no. 16, pp. 7751–7762, 2022.

[3] M. U. Ali, K. S. Kim, M. Khalid, M. Farrash, A. Zafar, S. W. Lee, "Enhancing alzheimer's disease diagnosis and staging: a multistage cnn framework using mri," *Frontiers in Psychiatry*, vol. 15, p. 1395563, 2024.

[4] I. Ali, N. Saleem, M. Alhussein, B. Zohra, K. Aurangzeb, Q. M. u. Haq, "Deepcgan: early alzheimer's detection with deep convolutional generative adversarial networks," *Frontiers in Medicine*, vol. 11, p. 1443151, 2024.

[5] M. A. Kumaar, D. Samiayya, V. Rajinikanth, P. D. R. Vincent, S. Kadry, "Brain tumor classification using a pre-trained auxiliary classifying style-based generative adversarial network," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 6, pp. 101–111, 2024.

[6] M. Irfan, S. Shahrestani, M. ElKhodr, "The application of deep learning for classification of alzheimer's disease stages by magnetic resonance imaging data.," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 9, no. 2, 2025.

[7] I. Ayus, D. Gupta, "A novel hybrid ensemble based alzheimer's identification system using deep learning technique," *Biomedical Signal Processing and Control*, vol. 92, p. 106079, 2024.

[8] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proceedings of the 37th International Conference on Machine Learning (ICML)*, 2020, pp. 1597– 1607.

[9] V. A. Ciurea, R.-A. Covache-Busuioc, A. G. Mohan, H. P. Costin, V. Voicu, "Alzheimer's disease: 120 years of research and progress," *Journal of Medicine and Life*, vol. 16, no. 2, p. 173, 2023.

[10] R. Cui, M. Liu, "Hippocampus analysis by combination of 3-d densenet and shapes for alzheimer's disease diagnosis," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 5, pp. 2099–2107, 2018.

[11] Y. Dai, S. Oehmcke, F. Gieseke, Y. Wu, K. Barnard, "Attention as activation," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 9156–9163, IEEE.

[12] A. El-Assy, H. M. Amer, H. Ibrahim, M. Mohamed, "A novel cnn architecture for accurate early detection and classification of alzheimer's disease using mri data," *Scientific Reports*, vol. 14, no. 1, p. 3463, 2024.

[13] X. Fan, H. Li, L. Liu, K. Zhang, Z. Zhang, Y. Chen, Z. Wang, X. He, J. Xu, Q. Hu, "Early diagnosing and transformation prediction of alzheimer's disease using multi-scaled self-attention network on structural mri images with occlusion sensitivity analysis," *Journal of Alzheimer's Disease*, vol. 97, no. 2, pp. 909-926, 2024.

[14] P. Goyal, R. Rani, K. Singh, "A multilayered framework for diagnosis and classification of alzheimer's disease using transfer learned alexnet and lstm," *Neural Computing and Applications*, vol. 36, no. 7, pp. 3777–3801, 2024.

[15] M. E. Hasan, A. Wagler, "New convolutional neural network and graph convolutional network-based architecture for ai applications in alzheimer's disease and dementia-stage classification," *AI*, vol. 5, no. 1, pp. 342–363, 2024.

[16] M. Hatami, F. Yaghmaee, R. Ebrahimpour, "Investigating the potential of reinforcement learning and deep learning in improving alzheimer's disease classification," *Neurocomputing*, vol. 597, p. 128119, 2024.

[17] M. Kaya, Y. Çetin Kaya, "A novel deep learning architecture optimization for multiclass classification of alzheimer's disease level," *IEEE Access*, vol. 12, pp. 46562-46581, 2024.

[18] S. K. Kim, Q. A. Duong, J. K. Gahm, "Multimodal 3d deep learning for early diagnosis of alzheimer's disease," *IEEE Access*, vol. 12, pp. 46278-46289, 2024.

[19] N. Kishore, N. Goel, "Deep learning based diagnosis of alzheimer's disease using fdg-pet images," *Neuroscience Letters*, vol. 817, p. 137530, 2023.

[20] Q. Li, M. Q. Yang, "Comparison of machine learning approaches for enhancing alzheimer's disease classification," *PeerJ*, vol. 9, p. e10549, 2021.

[21] S. Liang, Y. Gu, "Computer-aided diagnosis of alzheimer's disease through weak supervision deep learning framework with attention mechanism," *Sensors*, vol. 21, no. 1, p. 220, 2020.

[22] J. Liu, M. Li, Y. Luo, S. Yang, W. Li, Y. Bi, "Alzheimer's disease detection using depthwise separable convolutional neural networks," *Computer Methods and Programs in Biomedicine*, vol. 203, p. 106032, 2021.

[23] M. Manjavong, J. Kang, A. Diaz, M. Ashford, J. Eichenbaum, A. Aaronson, M. Miller, S. Mackin, R. Tank, M. Weiner, *et al.*, "Performance of plasma biomarkers combined with structural mri to identify candidate participants for alzheimer's disease- modifying therapy," *The Journal of Prevention of Alzheimer's Disease*, vol. 11, no. 5, pp. 1198-1205, 2024.

[24] T. Prasath, V. Sumathi, "Pipelined deep learning architecture for the detection of alzheimer's disease," *Biomedical Signal Processing and Control*, vol. 87, p. 105442, 2024.

[25] G. Mohi ud din dar, A. Bhagat, S. I. Ansarullah, M. T. B. Othman, Y. Hamid, H. K. Alkahtani, I. Ullah, H. Hamam, "A novel framework for classification of different alzheimer's disease stages using cnn model," *Electronics*, vol. 12, no. 2, p. 469, 2023.

[26] S. Murugan, C. Venkatesan, M. Sumithra, X.-Z. Gao, B. Elakkiya, M. Akila, S. Manoharan, "Demnet: A deep learning model for early diagnosis of alzheimer diseases and dementia from mr images," *IEEE Access*, vol. 9, pp. 90319–90329, 2021.

[27] H. Nawaz, M. Maqsood, S. Afzal, F. Aadil, I. Mehmood, S. Rho, "A deep feature-based real-time system for alzheimer disease stage detection," *Multimedia Tools and Applications*, vol. 80, pp. 35789–35807, 2021.

[28] E. Nichols, J. D. Steinmetz, S. E. Vollset, K. Fukutaki, J. Chalek, F. Abd-Allah, A. Abdoli, A. Abualhasan, E. Abu-Gharbieh, T. T. Akram, *et al.*, "Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the global burden of disease study 2019," *The Lancet Public Health*, vol. 7, no. 2, pp. e105–e125, 2022.

[29] X. Pan, T.-L. Phan, M. Adel, C. Fossati, T. Gaidon, J. Wojak, E. Guedj, "Multi-view separable pyramid network for ad prediction at mci stage by 18 f-fdg brain pet imaging," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 81–92, 2020.

[30] N. Saleem, S. Bourouis, H. Elmannai, A. D. Algarni, "Dpht-anet: Dual-path high-order transformer- style fully attentional network for monaural speech enhancement," *Applied Acoustics*, vol. 224, p. 110131, 2024.

[31] A. Puente-Castro, E. Fernandez-Blanco, A. Pazos, C. R. Munteanu, "Automatic assessment of alzheimer's disease diagnosis based on deep learning techniques," *Computers in Biology and Medicine*, vol. 120, p. 103764, 2020.

[32] Y. Sang, W. Li, "Classification study of alzheimer's disease based on self-attention mechanism and dti imaging using gcn," *IEEE Access*, vol. 12, pp. 24387-24395, 2024.

[33] J. V. Shanmugam, B. Duraisamy, B. C. Simon, P. Bhaskaran, "Alzheimer's disease classification using pre-trained deep networks," *Biomedical Signal Processing and Control*, vol. 71, p. 103217, 2022.

[34] R. Suganthe, M. Geetha, G. Sreekanth, K. Gowtham, S. Deepakkumar, R. Elango, "Multiclass classification of alzheimer's disease using hybrid deep convolutional neural network," *NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal*, pp. 145–153, 2021.

[35] P. Deshpande, R. Dhabliya, D. Khubalkar, P. A. Upadhye, K. A. Wagh, V. Khetani, "Alzheimer disease progression forecasting: Empowering models through hybrid of cnn and lstm with pso op-timization," in *2024 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 2024, pp. 1–5, IEEE.

[36] R. Venkatesh, S. Anantharajan, S. Gunasekaran, C. Yogaraja, I. Gethzi Ahila Poornima, "Prediction of alzheimer's disease using adaptive fine-tuned deep resnet-50 with attention mechanism," *SN Computer Science*, vol. 5, no. 4, p. 392, 2024.

[37] J. Wen, E. Thibeau-Sutre, M. Diaz-Melo, J. Samper- González, A. Routier, S. Bottani, D. Dormont, S. Durrleman, N. Burgos, O. Colliot, *et al.*, "Convolutional neural networks for classification of alzheimer's disease: Overview and reproducible evaluation," *Medical Image Analysis*, vol. 63, p. 101694, 2020.

[38] N. Yaqoob, M. A. Khan, S. Masood, H. M. Albarakati, Hamza, F. Alhayan, L. Jamel, A. Masood, "Prediction of alzheimer's disease stages based on resnet-self-attention architecture with bayesian optimization and best features selection," *Frontiers in Computational Neuroscience*, vol. 18, p. 1393849, 2024.

[39] M. Yildirim, A. Cinar, "Classification of alzheimer's disease mri images with cnn based hybrid method," *Ingénierie Des Systèmes d Inf.*, vol. 25, no. 4, pp. 413–418, 2020.

[40] M. A. Zayene, H. Basly, F. E. Sayadi, "Multi-view separable residual convolution neural network for detecting alzheimer's disease progression," *Biomedical Signal Processing and Control*, vol. 95, p. 106375, 2024.

[41] J. Zhang, X. He, L. Qing, Y. Xu, Y. Liu, H. Chen, "Multi-scale discriminative regions analysis in fdg-pet imaging for early diagnosis of alzheimer's disease," *Journal of Neural Engineering*, vol. 19, no. 4, p. 046030, 2022.

Hela Elmannai

Hela Elmannai received the Ph.D. degree in information technology from SUPCOM, Aryanah, Tunisia, in 2017. She is currently an Associate Professor with the Department of Information Technology, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia. Her research interests include artificial intelligence, networking, blockchain, and engineering applications.

Nasir Saleem

Nasir Saleem received Ph.D. in Digital Speech Processing and Deep Learning from UET Peshawar in 2021. He is currently a Research Fellow at ENU, UK and was a postdoctoral Fellow at IIUM, working on AI-based speech processing algorithms. Previously, he served as a Senior Lecturer (2008–2012) and is now an Assistant Professor at Gomal University. His research interests include Human–Machine Interaction, Speech Enhancement, AV Processing, and Machine Learning. He has published in leading venues (Elsevier, Springer, IEEE) and actively serves as a reviewer for journals.

Sami Bourouis

Sami Bourouis received the Engineer, M.Sc., and Ph.D. degrees in computer science from the University of Tunis, Tunisia, in 2003, 2005, and 2011, respectively. He is currently a Professor at the College of Computers and Information Technology, Taif University, Saudi Arabia. His research interests include data mining, image processing, AI, machine learning, cyber security, and pattern recognition applied to several real-life applications.

Reem Ibrahim Alkanhel

Reem Ibrahim Alkanhel received the B.S. degree in computer sciences from King Saud University, Riyadh, Saudi Arabia, in 1996, the M.S. degree in information technology (computer networks and information security) from the Queensland University of Technology, Brisbane, Australia, in 2007, and the Ph.D. degree in information technology (networks and communication systems) from Plymouth University, Plymouth, U.K., in 2019. She has been with Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, since 1997. She is currently a Teaching Assistant with the College of Computer and Information Sciences. Her current research interests include communication systems, networking, the Internet of Things, information security, information technology, quality of service and experience, software-defined networks, and deep reinforcement learning.