frontiers | Frontiers in Artificial Intelligence

TYPE Original Research
PUBLISHED 12 March 2025
DOI 10.3389/frai.2025.1540646

OPEN ACCESS

EDITED BY
Tony J. Prescott,
The University of Sheffield, United Kingdom

REVIEWED BY
Takashi Kuremoto,
Nippon Institute of Technology, Japan
Sid Ahmed Benabderrahmane,
New York University, United States

*CORRESPONDENCE
Faisal Binzagr
✉ fbinzagr@kau.edu.sa

RECEIVED 06 December 2024
ACCEPTED 11 February 2025
PUBLISHED 12 March 2025

CITATION
Binzagr F and Abulfaraj AW (2025) InGSA: integrating generalized self-attention in CNN for Alzheimer's disease classification.
Front. Artif. Intell. 8:1540646.
doi: 10.3389/frai.2025.1540646

InGSA: integrating generalized self-attention in CNN for Alzheimer's disease classification

Faisal Binzagr[1]* and Anas W. Abulfaraj[2]

[1]Department of Computer Science, King Abdulaziz University, Rabigh, Saudi Arabia, [2]Department of Information Systems, King Abdulaziz University, Rabigh, Saudi Arabia

Alzheimer's disease (AD) is an incurable neurodegenerative disorder that slowly impair the mental abilities. Early diagnosis, nevertheless, can greatly reduce the symptoms that are associated with the condition. Earlier techniques of diagnosing the AD from the MRI scans have been adopted by traditional machine learning technologies. However, such traditional methods involve depending on feature extraction that is usually complex, time-consuming, and requiring substantial effort from the medical personnel. Furthermore, these methods are usually not very specific as far as diagnosis is concerned. In general, traditional convolutional neural network (CNN) architectures have a problem with identifying AD. To this end, the developed framework consists of a new contrast enhancement approach, named haze-reduced local-global (HRLG). For multiclass AD classification, we introduce a global CNN-transformer model InGSA. The proposed InGSA is based on the InceptionV3 model which is pre-trained, and it encompasses an additional generalized self-attention (GSA) block at top of the network. This GSA module is capable of capturing the interaction not only in terms of the spatial relations within the feature space but also over the channel dimension it is capable of picking up fine detailing of the AD information while suppressing the noise. Furthermore, several GSA heads are used to exploit other dependency structures of global features as well. Our evaluation of InGSA on a two benchmark dataset, using various pre-trained networks, demonstrates the GSA's superior performance.

KEYWORDS

Alzheimer's disease classification, generalized self-attention, CNN, transfer learning, transformer

1 Introduction

Alzheimer's disease (AD) is a type of dementia that is not curable, which becomes worse over years as it affects the human brain, but early diagnosis helps to minimize the symptoms and the management of the patient (McKhann et al., 1984). Its manifestation involves impaired memory because patients cannot organize or recall information properly, and poor judgment that renders the affected persons completely helpless and in need of care as the disease develops (Choi et al., 2020). The probability raised from 2% at 65 years to 35% at 85 years for AD. Approximately 26.6 million people had it in 2006; the figure rose to over 55 million in 2020 and is expected to reach 152 million by 2050 (Gunawardena et al., 2017). Neuronal loss and synaptic impairment can occur at least one or two decades before disease onset (Böhle et al., 2019). It is essential to detect AD in the prodromal stage, which is characterized by moderate cognitive impairment (MCI), as there is currently no cure. Early MCI (EMCI) is a cognitive impairment stage that precedes MCI (Kang et al., 2020). The early detection of EMCI has the potential to

prevent the progression of EMCI to AD. The importance of diagnosing MCI patients has been emphasized by studies that has identified the distinctions between early MCI (EMCI) and late MCI (LMCI) groups (Nozadi et al., 2018; Edmonds et al., 2019; Zhang T. et al., 2019). MCI has a symptom profile that is similar, but less severe, to AD (Varatharajah et al., 2019). Nowadays, this disease is also defined as mild cognitive impairment associated with the existence of Alzheimer's disease; according to recent investigations, ∼80% of patients diagnosed with MCI develop AD in 7 years. For monitoring of variations in the densities of the brain tissues, magnetic resonance imaging (MRI) and positron emission tomography (PET) are frequently used since they do not include the invasion of the tissues (Ramzan et al., 2020; Gao, 2021). Neuroimaging, especially using MRI, is crucial for the study of the nervous system structures more closely (Tuvshinjargal and Hwang, 2022); this test helps in diagnosis of certain diseases such as tumors and cancer (Tehsin et al., 2024). MRI does work in the case of Alzheimer's; it allows capturing structural changes in the brain, for instance, the reduction of certain regions and the appearance of new formations, heterogeneous density, and the presence of abnormal substances typical of the disease (Simic et al., 2009).

In recent years, medical imagery such as MRI has been used with machine learning (ML) and deep learning (DL). These methods are used in health checks and early AD diagnosis. They also excel at categorizing images in health and computer vision (Nasir et al., 2021, 2020, 2022; Yousafzai et al., 2024; Nasir et al., 2023). In recent decades, neuroimaging data have grown, allowing ML and DL algorithms to better characterize AD. The authors used such methodologies to offer prospective AD diagnosis and prognostic outcomes (Nagarajan et al., 2021). These works executed features from several image processing pipeline streams using random forest classifier, decision tree, or support vector machine (SVM). Lately, DL techniques have showed potential in medical imaging with good picture classification accuracy (Ajagbe et al., 2021). Automatic feature extraction from images using CNNs and transfer learning (TL) is more efficient than typical ML methods (Raju et al., 2021). However, working with medical data is problematic due to imbalanced dataset, including AD. In this strategy, various sample sizes are used for different classes, the model is always biased, and it cannot generalize beyond the training dataset. DL models can process raw data better than simple feed forward, but they can overfit when solving complicated problems such as class imbalance. In real-world circumstances, such models perform poorly in generalization, efficacy, and reliability. The main contribution of this study is as follows:

- We introduce a contrast enhancement method called haze-reduced local-global, inspired by the haze reduction principle.
- We suggest a new global CNN-transformer architecture, InGSA, for the classification of multiclass AD. A pre-trained CNN is integrated with a specialized transformer module in InGSA network.
- This network, comprised of several generalized self-attention module (GSA), is designed to effectively capture extensive feature dependencies across different brain regions by establishing global connections along both the channel and spatial dimensions.
- The InGSA model is tested on a two publicly available dataset, where we also use various pre-trained CNN models to demonstrate its effectiveness. Furthermore, we perform a comparative analysis between InGSA and modern attention mechanisms, as well as the latest approaches in multiclass AD classification.

The structure of this research is comprehensively examined in the following manner: Related works are detailed in Section 2. Section 3 delineates the fundamental concepts and proposed methodology. The experimental results are the subject of Section 4. The study is concluded in Section 5.

2 Related work

Over the past few years, the usage of DL methods for the identification of AD has received much attention (Mohammed et al., 2021; Ahmed et al., 2022; Menagadevi et al., 2023). For instance, a study employed DL with stacked auto-encoders and uses the softmax function in the final layer to address problem of bottlenecks. Their approach needed far less training data compared to their peers, as well as very small input to classify several groups with ∼87.70% accuracy. One of the observations from the current study was that the use of several features improves classification (Frizzell et al., 2022). Furthermore, a classification framework was built, based on the use of multiple different input databases since it is complementary. To combine features from different modalities, they used a process known as non-linear graph mixture model. Using this method, the areas under the curve were calculated with 98.1% accuracy when differentiating between AD and CN images, 82.40% between NC and MCI images, with the overall classification performance being 77.90% (Guo and Zhang, 2020).

A novel rapid, low-cost, and efficient diagnostic model was implemented using brain MRI scans. They used DenseNet121 model which is a computationally heavy model, and to this model, they achieved an accuracy of 87% in detecting the disease. To rectify this, the authors employed an idea of fine-tuning two models of AlexNet and LeNet models where features were extracted in three ways through parallel filters. The new model they came up with was able to predict the disease with an accuracy of 93% (Hazarika et al., 2023). In the same manner, the researchers in Acharya et al. (2021) used VGG-16 based CNN transfer learning to diagnose AD with an overall accuracy of 95.7%. Another study used DL for distinguishing dementia and Alzheimer's from the MRI images (Murugan et al., 2021).

Abbreviations: CNN, convolutional neural network; HRLG, haze-reduced local-global; GSA, generalized self-attention; AD, Alzheimer's disease; MCI, moderate cognitive impairment; EMCI, Early MCI; LMCI, late MCI; MRI, magnetic resonance imaging; PET, positron emission tomography; ML, machine learning; DL, deep learning; SVM, support vector machine; TL, transfer learning; ELM, extreme learning machine; DAG, directed acyclic graph; Mob, MobileNet; Den, DenseNet201; Res, ResNet50; Sq, SqueezeNet; InV3, inceptionV3; CBAM, convolutional block attention module; CSDAB, channel split dual attention block; ViT, vision transformer; DEiT, data efficient image transformer; PVT, pyramid vision transformer.

The approach used in Murugan et al. (2021) learns individual Alzheimer's likelihood using multilayer perceptron representations and also generates disease probability heat maps from brain region activity. To overcome the problem of class imbalance, the samples are divided in equal proportion. The five ADNI subtypes consist of 1,296 images comprising of AD, MCI, EMCI, LMCI, and CN images processing the DEMNET model by resizing the images to 176 × 176 and obtained an accuracy of 84.83%. In the same way, Oktavian et al. (2022) presented the fine-tuned ResNet18 model for distinguishing between MCI, AD, and CN using MRI and PET datasets. This model incorporated transfer learning and used the technologies such as weighted loss function for ascending the class imbalance, and mish activation function to augment its accuracy, and it obtained 88.3% overall classification. On the other hand, the authors in Dyrba et al. (2021) adopted a CNN with 663 T1-weighted MRI scans belonging to dementia and amnestic MCI patients. To confirm their model, they performed cross-validation and used an additional three datasets that included an overall of 1,655 cases. To further provide the clinical relevance of the method, they correlated the relevance scores to the hippocampal volume. A friendly model assessment tool was created through importance maps of 3D CNN, achieving accuracy of 94.9% of AD vs. CN. A particular drawback of many papers on the detection of Alzheimer's is related to the imbalance of classes, which creates problems of overfitting and lowering predictive ability in almost all existing deep learning models. The yield is further magnified by the fact that realistic training data are also scarce. To overcome this, we utilized the data augmentation approach to balance datasets and improve DL results since the technique synthesizes new data samples.

3 Proposed methodology

The configuration of the proposed InGSA is illustrated in Figure 1, comprising a fine-tuned CNN model, a generalized self-attention (GSA), and a classifier. The fine-tuned CNN models aid in extracting abstract feature representations from the input MR images. The GSA block has various components to comprehend global interdependence across spatial and channel dimensions, facilitating the extraction of more nuanced and category-specific information. The extreme learning machine (ELM) classifier is employed to categorize AD. This section offers a comprehensive overview of the InGSA architecture and its fundamental components.

3.1 Haze reduced local global image enhancement

Traditional haze elimination procedures are developed for improving the visual distinctiveness of scenes by increasing the contrast and color saturation. By applying these techniques, the total clarity of the scene which is captured in the given image is likely to be enhanced. In this research, we formally propose a new type of contrast enhancement method that adopts both haze removal and local-global transformation techniques.

Let D denotes a complex image database that is composed of N images. While the original image is represented by the dimensions of $N \times M \times 3$ as $I(x, y)$, the $Y(x, y)$ denotes the improved image. First, a haze reduction method utilizing the dark channel prior is employed on the first image. This process of haze reduction can be mathematically expressed as follows:

$$C(x) = \gamma(x)j(x) + l(1 - t(x)) \tag{1}$$

where C denotes the measured intensity values, γ represents the scene radiance, $j(x)$ designates the transmission map, and l denotes the atmospheric light intensity. The dehazing algorithm utilized aims to restore the scene radiance γ based on the estimations of both the transmission map and the atmospheric light, as expressed in the following manner:

$$\gamma(x) = \frac{C(x) - \alpha}{\max(t(x), t_0)} + \alpha \tag{2}$$

The resulting $\gamma(x)$ is subsequently employed to calculate the global contrast of an image using the following equation:

$$G_0 = (1 + g_k) \times (G_i - k_{\text{mean}}) + \sigma \tag{3}$$

In this regard, G_0 stands for the global contrast image of the original image while g_k represents gain factor of global contrast, G_i for the value of pixel $\gamma(x)$, k_{mean} for the overall average pixel value of $\gamma(x)$, and σ for the standard deviation of $\gamma(x)$. In the subsequent step, we assessed the local contrast of the haze-reduced image using the following mathematical expression:

$$H(x, y) = \frac{LC}{\sigma(i, j) + \varphi} \times \mu(x, y) \tag{4}$$

where LC for local contrast, φ for a small constant, and $\mu(x, y)$ means the mean value of the dehazed image. Finally, these two resultant images of local and global contrast were incorporated toward a single image in this way that we adopt the following mathematical formula to produce the final enhancement output.

$$Y(x, y) = [G(x, y) + H(x, y)] - I(x, y)] \tag{5}$$

3.2 Deep transfer learning

InceptionV3 is a directed acyclic graph (DAG) network that has 316 layers and 350 links that include 94 as convolutional layers (Szegedy et al., 2016). Such a structure facilitates the provision of adequate employment of complicated dependency relations in the network, having many inputs and outputs at different layers. Differently from the standard CNN model in which the filter size is fixed throughout the layers of that model, InceptionV3 has different filter sizes within the same layer, which increases its capability of feature extraction on the data. Originally trained on ImageNet (Deng et al., 2009) on which includes over one million images split into one thousand categories, the InceptionV3 has the ability to read most features. The model takes images of input size 299 × 299 × 3. In this study, the model has been adapted to classify various stages of AD for which transfer learning from the ImageNet

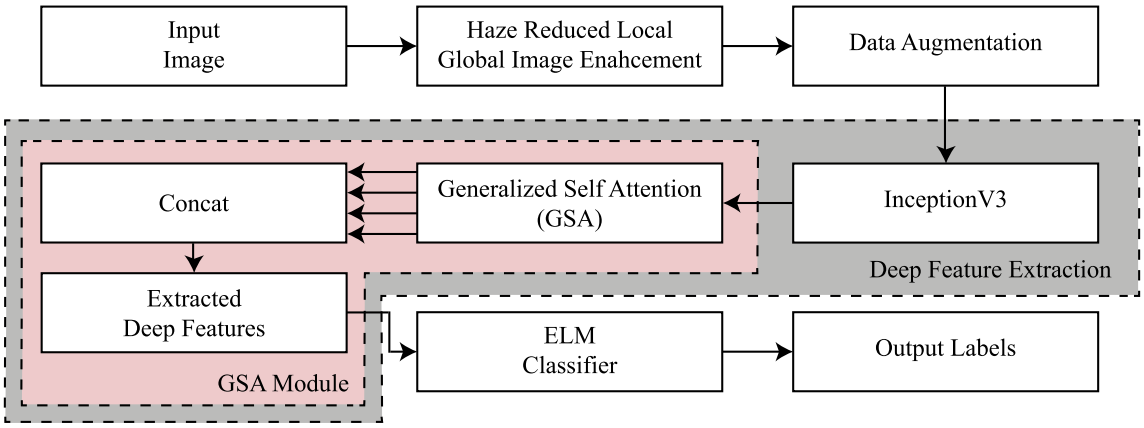

FIGURE 1
General flow of the proposed framework.

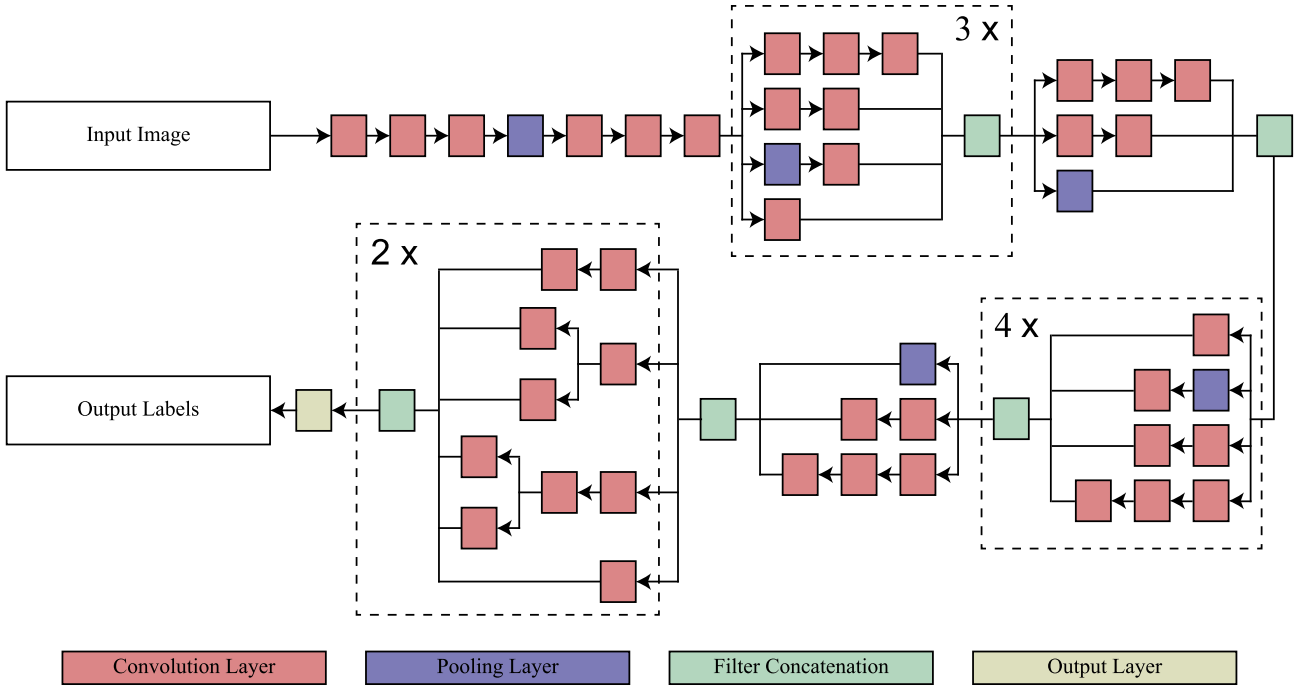

FIGURE 2
Architecture of InceptionV3 model.

training phase has been applied to achieve efficient medical image classification. The architecture of InceptionV3 model is shown in Figure 2.

Transfer learning (Pan and Yang, 2009) is a popular method in recognition and detection tasks, allowing for improved model performance by leveraging pre-trained models. In this context, the domain D consists of a feature vector $Y = y_1, y_2, \cdots, y_n$ with a corresponding probabilistic distribution $P(Y)$, forming $B = Y, P(Y)$. The task, denoted as T, consists of the ground truth $Z = z_1, z_2, \cdots, z_n$. The function can be expressed in probabilistic form as $P(z|y)$. In the context of transfer learning, this can be represented concerning the source domain as $B_T = (x(T_1), x(T_2)), (x(T_2), x(T_2)), \cdots, (x(T_n), x(T_n))$ along with the learning rate S_T. The target output is denoted as $B_S = (x(S_1), x(S_2)), (x(S_2), x(S_2)) \cdots, (x(S_n), x(S_n))$, and the associated function for the targeted neural network is represented as S_S. The primary objective of transfer learning is to improve the learning rate for predicting the target object by utilizing the recognition function $F_S(.)$, which is informed by training on both B_T and B_S, where $B_T \neq B_S$ and $S_T \neq S_S$. Inductive transfer learning proves to be effective in pattern recognition tasks. An annotated dataset is essential for efficient training and evaluation when implementing inductive transfer learning. This process can involve distinct class labels $Z_T \neq Z_S$ and differing distributions $P(Z_T|Y_T) \neq (Z_S|Y_S)$.

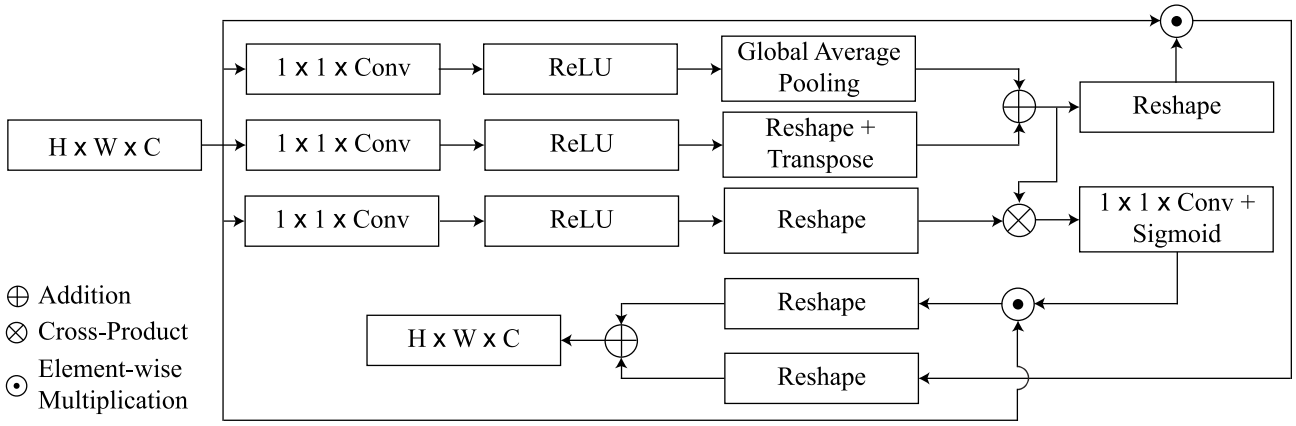

FIGURE 3
Overview of the GSA block.

3.3 Generalized self-attention module

The proposed GSA module is aimed at achieving detailed description of AD characteristics while avoiding irrelevant features. Its architecture was influenced by the self-attention mechanisms employed in GCNet (Cao et al., 2019; Zhang et al., 2019), as illustrated in Figure 3. However, unlike these methods, it positions global dependency across both spatial and channel dimensions at the same time. Spatial attention worked for the relationships of the global features in the spatial location, while channel attention worked on the importance of a point channel out of all the channels.

As initial input for the GSA module, we utilize the high level activation maps $Z \in \mathbb{R}^{H\times W\times C}$, whereas the GSA module returns the refined feature maps $Z_{gs} \in \mathbb{R}^{H\times W\times C}$. The feature map is divided into the keys, queries, and values, similar to a transformer architecture, which is supported by three attributes Q, K, and V.

The query function $q(Z)$ is defined by a convolution of 1×1 consisting of $C' = C/8$ channels and global average pooling to attain the vector $Q(Z) \in \mathbb{R}^{1\times C'}$. On the other hand, the key and value functions are carried out by 1×1 convolution followed by reshape operations but without global average pooling and the outputs are maps $K(Z) \in \mathbb{R}^{HW\times C'}$ and $V(Z) \in \mathbb{R}^{HW\times C'}$). Next, The spatial attention weights are generated by calculating the matrix product between Q and K and applying a softmax activation function given as

$$Z' = \phi(q(Z) \otimes k(Z)^T) \tag{6}$$

With regard to the abbreviations used here, we have $\otimes$ indicating the cross-product of the matrix, ϕ which stands for the softmax activation function of the formula and the double dagger T showing the operation of matrix transposition. Following this, the spatial attention feature map $Z_{sp} \in \mathbb{R}^{H\times W\times C}$ derived by performing element-wise multiplication among Z' and Z is as follows:

$$Z_{sp} = \text{reshape}(Z') \odot Z \tag{7}$$

Similarly, a matrix cross-product of Z' with $v(Z)$ leads to the channel attention weights which are passed through a 1×1 convolution layer and a sigmoid non-linearity. It also increases the channels from C' to C. This process is also called linear embedding. The mathematical formulation for this global transformation is given by

$$Z'' = \sigma(\text{conv}(Z' \otimes v(Z))) \tag{8}$$

Next, the channel-wise attention maps $Z_{ch} \in \mathbb{R}^{H\times W\times C}$ are calculated as

$$Z_{ch} = Z'' \odot Z \tag{9}$$

Finally, we integrate the spatial attention feature maps Z_{sp} and the channel attention maps Z_{ch} by taking their weighted sum, producing the refined attention feature map $Z_{gs} \in \mathbb{R}^{H\times W\times C}$, defined as

$$Z_{gs} = W_1 Z_{sp} + W_2 Z_{ch} \tag{10}$$

where W_1 and W_2 are two trainable scalar weights. In summary, GSA obtains the channel-wise and spatial dependencies concurrently from MR images and then improves the features representation. We merge the feature attention maps produced by the GSA heads through concatenation, preceding 1×1 convolution to generate the final output of the proposed GSA, denoted as $Z_{tm} \in \mathbb{R}^{H\times W\times C}$. Mathematically, $Z_t m$ can be represented as follows:

$$Z_{tm} = \text{concat}(Z^1_{gs}, Z^2_{gs}, \ldots, Z^h_{gs}) \tag{11}$$

In this study, h, representing the number of GSA heads, is set empirically to a value of 4. This specific choice of $h = 4$ was determined empirically.

3.4 Classification

The ELM [42] was used as a classifier to differentiate AD stages. Given z sample (Z, o), the ELM's output with no errors can be mathematically expressed as follows:

$$o = \sqrt{\sum \alpha t(w_i + b)} \tag{12}$$

In this instance, the activation function is denoted by $t(\cdot)$, and the input and output samples are Z and o, respectively. The variables w and b are weights and bias, respectively, and α is the weight coefficient. The output O is provided as $O = H\alpha$ whereby $O = (o_1, o_2, \cdots, o_n)$ symbolizes the output vector and $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ denotes the weight vector. The hidden layers can be expressed as

$$H = \begin{bmatrix} t(w_1 i_1 + b_1) & \cdots & t(w_n i_1 + b_n) \\ \vdots & \ddots & \vdots \\ t(w_1 i_m + b_1) & \cdots & t(w_n i_m + b_n) \end{bmatrix} \tag{13}$$

The number of nodes in the hidden layer needs to be below the total amount of samples. Description of the structured model of a single hidden-layer ELM neural network utilized for AD classification is provided in Equation 13. The hidden layer, denoted by H, is composed of nodes and activation functions $t(\cdot)$. Weights w_i and biases b are connected to each hidden layer node, with i ranging from 1 to m and representing input variables. The formula used in the production of the output of the hidden layer O is the summation of the product between each of the node's activation function and weights then passed through $t(\cdot)$. The mechanism can be mathematically represented as $O = H\alpha$. Equation 13 defines the structure of the hidden layer, which further elucidates that the H is a concatenation of n nodes. The weighted sum of input features $i = (i_1, i_2, \cdots, i_m)$ is computed for each node's activation by utilizing weights w_i and biases b for each node. The function $t(\cdot)$ adds non-linearity and provides the network with ability to learn more complex input data patterns. Determining the quantity of nodes in the hidden layers is essential; they should be fewer than the amount of samples to avert overfitting. During training, we obtain the weights and biases to minimize the mapping function between the input features and the associated output for AD classification.

4 Experimental results

The analysis and experimental results of the proposed models are detailed in this section. The presentation includes information regarding the dataset, implementation characteristics, and comparison analysis.

4.1 Experimental setup and dataset

The model was trained on a high-performance machine equipped with an Intel Core i9-14900HX processor and an NVIDIA RTX 4090 GPU, providing substantial computational power for deep learning tasks. The system included 64GB of DDR5 RAM operating at 5,600MT/s, ensuring efficient handling of large-scale data. CUDA 12.6 was utilized to enable GPU-accelerated training. The model was trained with a learning rate of 0.0001, a value chosen to balance the stability and convergence speed of the training process.

TABLE 1 ADNI dataset image count before and after augmentation.

Classes	No. of images	Augmented	No. of utilized images
AD	8,346	n/a	500
CN	8,650	n/a	500
EMCI	480	n/a	480
LMCI	144	432	432
MCI	1,155	n/a	500

TABLE 2 Total number of images and number of images utilized from OASIS dataset.

Classes	No. of images	No. of utilized images
Mild dementia	5,002	500
Moderate dementia	488	488
Non-demented	67,200	500
Very mild dementia	13,700	500

In this experiment, ADNI dataset was used which consisted of five classes: AD, CN, EMCI, LMCI, and MCI. The original number of images varied significantly across classes, with AD, CN, and MCI having thousands of images, while EMCI and LMCI had considerably fewer as shown in Table 1. To address this class imbalance, data augmentation was applied exclusively to the LMCI class, which originally had only 144 images. Through augmentation, the LMCI class was expanded to 432 images, increasing the total number of samples used in the training process. For the other classes, 500 images were randomly selected from AD and CN, while all available images were used for EMCI and MCI.

The augmentation methods used to enhance the LMCI class included rotation, scaling, and flipping. Rotation involved rotating images by various angles to introduce diversity without altering key signal features. Scaling was applied to adjust the size of images while maintaining their aspect ratio, simulating variability in data capture. Flipping, both horizontally and vertically, was also used to further diversify the dataset, making the model more robust to orientation changes. These augmentation techniques were critical in improving class balance and ensuring better generalization during model training.

Another dataset used in this experiment is the Open Access Series of Imaging Studies (OASIS), a widely utilized resource for neuroimaging research, particularly in the study of brain health and dementia. Table 2 presents the distribution of images from the OASIS dataset across four classes: Mild Dementia, Moderate Dementia, Non-Demented, and Very Mild Dementia.

4.2 ADNI results

Table 3 indicates the performances of the proposed model in terms of classification for Alzheimer's detection under the different cognitive conditions. The network can accurately predict an image with an average of 96.67% and high values of precision, recall, and F1-scores for all classes. For the LMCI class, the study realized a precision of 98.80%, while at the same time, AD has the highest recall of 98.00% to show effective detection. All classes have an F1-score of more than 96%, and it can be seen that the approach balanced precision and recall. The AUC values are also high, and AD reached 98.21%. Confusion matrix for proposed model is given in Figure 4.

Table 4 gives a quantitative analysis of the number of correct detections of the various models augmented with different

TABLE 3 Classification performance of InGSA on ADNI dataset.

Class	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
AD	94.23	98.00	96.08	98.21
CN	97.92	95.92	96.91	97.70
EMCI	95.88	96.88	96.37	97.92
LMCI	98.80	95.35	97.04	97.55
MCI	97.00	97.00	97.00	98.11
Accuracy	96.67			
Macro average	97.76	96.63	96.68	97.90
Weighted average	96.71	96.67	96.67	97.91

TABLE 4 Performance comparison of different attention mechanisms.

Attention mechanism	Accuracy (%)				
	Mob	Den	Res	Sq	InV3
None	82.76	76.65	71.27	86.32	90.17
Self attention	85.43	79.89	76.39	87.36	93.44
CBAM	83.70	78.72	71.89	89.78	91.23
CSDAB	83.00	78.21	73.47	88.24	90.87
Proposed (GSA)	84.54	82.87	79.98	91.43	96.67

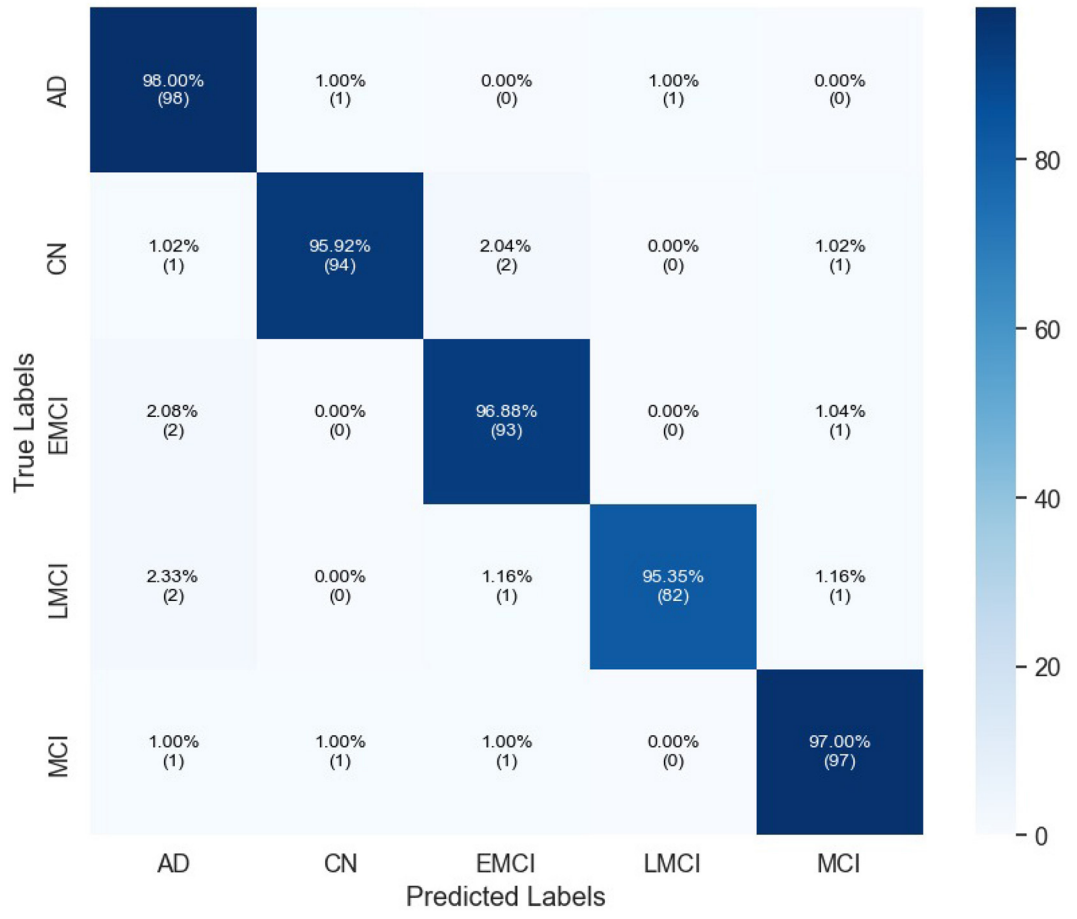

FIGURE 4
Confusion matrix of proposed InGSA on ADNI Dataset.

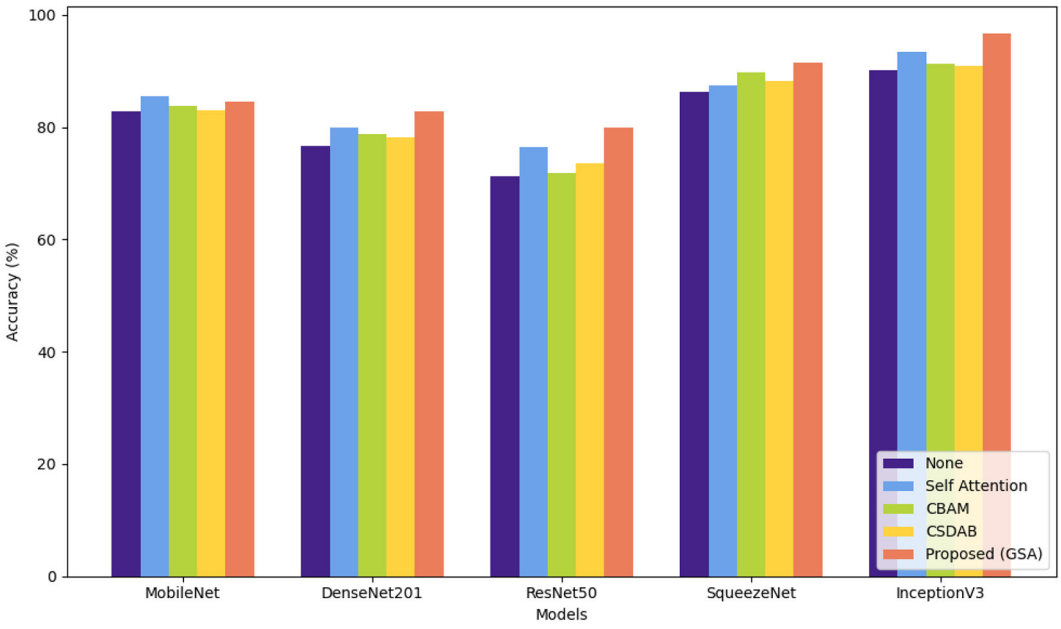

FIGURE 5
Visual comparison of various attention mechanisms on ADNI dataset.

attention mechanisms: MobileNet (Mob), DenseNet201 (Den), ResNet50 (Res), SqueezeNet (Sq), and InceptionV3 (InV3). With the proposed GSA, the largest performance improvements were achieved with InceptionV3, from 90.17% to 96.67% (with attention) and with SqueezeNet from 86.32% to 91.43%. Here, DenseNet201 shows improvement of 6.22% from 76.65 to 82.87, while ResNet50 goes from 71.27% to 79.98%. Self-Attention (Zhang et al., 2019) also presented substantial enhancements for InceptionV3 from 93.15% to 93.44%, as well as for SqueezeNet from 86.59% to 87.36%. Both convolutional block attention module (CBAM) (Woo et al., 2018) and channel split dual attention block (CSDAB) (Dutta and Nayak, 2022) result in moderate accuracy increases, with CBAM improving SqueezeNet to 89.78% and CSDAB raising it to 88.24%. In general, GSA consistently improves accuracy in all models being tested. Visual analysis of attention mechanisms with pre-trained model on ADNI dataset is shown in Figure 5.

Table 5 shows accuracy and F1-score when comparing the current existing models, namely, vision transformer (ViT), data efficient image transformer (DEiT), and pyramid vision transformer (PVT) with the InGSA model proposed in this study. Out of all the models, the DEiT comes with the highest accuracy and F1-score with accuracies of 89.44%, and F1-score of 88.12%, with PVT coming second with accuracies of 88.48% and F1-scores of 85.98%. ViT has the worst performance with an accuracy of 86.24% and the F1-score at 84.67%. Comparing the proposed model InGSA with others, it is clear that the proposed model InGSA outperforms the other models with accuracy of 96.67% and F1-score of 96.68% that indicate effectiveness of the proposed model InGSA.

TABLE 5 Comparison of proposed InGSA with transformer-based models on ADNI dataset.

Model	Accuracy (%)	F1-score (%)
ViT	86.24	84.67
DEiT	89.44	88.12
PVT	88.48	85.98
InGSA	96.67	96.68

TABLE 6 Classification performance of InGSA on OASIS dataset.

Class	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
Mild demented	98.02	99.00	98.51	99.16
Moderate demented	98.97	97.96	98.46	98.81
Non-demented	97.94	95.00	96.45	97.16
Very mild demented	97.09	100.00	98.52	99.50
Accuracy	97.99			
Macro average	98.00	97.99	97.98	98.66
Weighted average	98.00	97.99	97.98	98.66

4.3 OASIS results

Table 6 presents the classification performance for different categories of dementia using OASIS dataset. High precision of

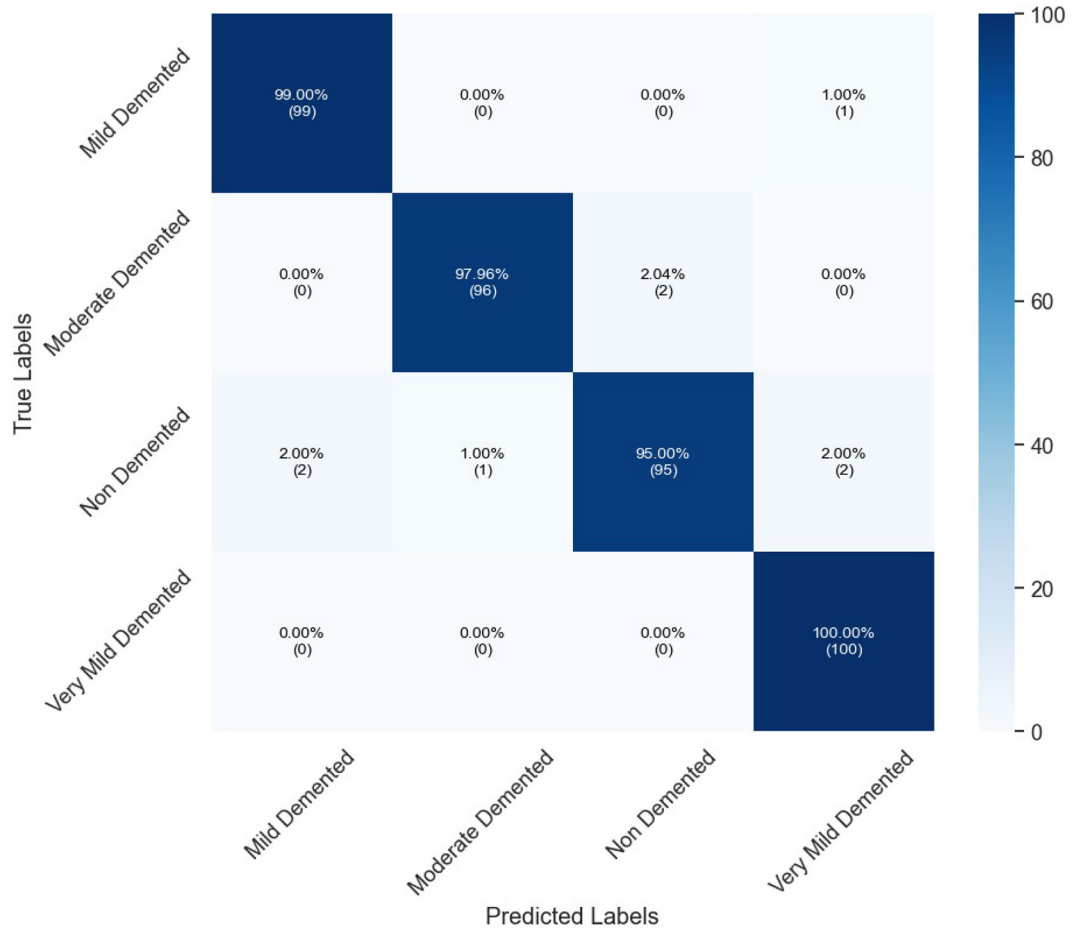

FIGURE 6
Confusion matrix of proposed InGSA on OASIS dataset.

TABLE 7 Performance comparison of different attention mechanisms.

Attention mechanism	Accuracy (%)				
	Mob	Den	Res	Sq	InV3
None	87.35	80.78	72.34	87.54	92.87
Self attention	88.45	84.67	73.67	89.79	94.76
CBAM	88.89	84.32	77.93	88.38	93.62
CSDAB	88.95	82.19	74.05	88.75	93.90
Proposed (GSA)	90.74	87.64	79.85	91.02	97.99

all classes set by the model, specifically Moderate Demented presenting 98.97% and Mild Demented 98.02%. The Recall is exceptional for Very Mild Demented at 100% which means that all the cases belonging to this class are identified rightly. The F1-score, therefore, averaged over all classes is unbiased, being 96.45% for Non-Demented and 98.52% for both Mild and Very Mild Demented classes. AUC's are high, notably Very Mild Demented with a highest of 99.50%. In general, the proposed model renders high performance in the presented study with the overall accuracy of 97.99%. Figure 6 depicts confusion matrix of proposed model for OASIS dataset.

Table 7 shows the percentage of classification accuracy of multiple models when using the OASIS dataset, with different attention mechanisms. All the attention approaches improve on the baseline accuracy of all the models including the proposed GSA model. For instance, they achieve 90.74% accuracy with MobileNet and an outstanding 97.99% with InceptionV3 confirming how efficient the proposed approach is in enhancing model accuracy. Self-attention mechanism also plays a useful role, especially in MobileNet and InceptionV3 models and in this experiment reached a throughput of 88.45 and 94.76%, correspondingly. While structure imported with CBAM and CSDAB mechanisms may be less fortunate than the GSA model, it has revealed improvements. Figure 7 illustrates the visual analysis of attention mechanisms using a pre-trained model on the OASIS dataset.

The findings in Table 8 reveal that all algorithms provide high accuracy, where proposed InGSA model performs the best with accuracy of 97.99% and F1-score of 97.88%. This implies that InGSA is proud not only to classify instances accurately but also to have a low percentage of false positive and false

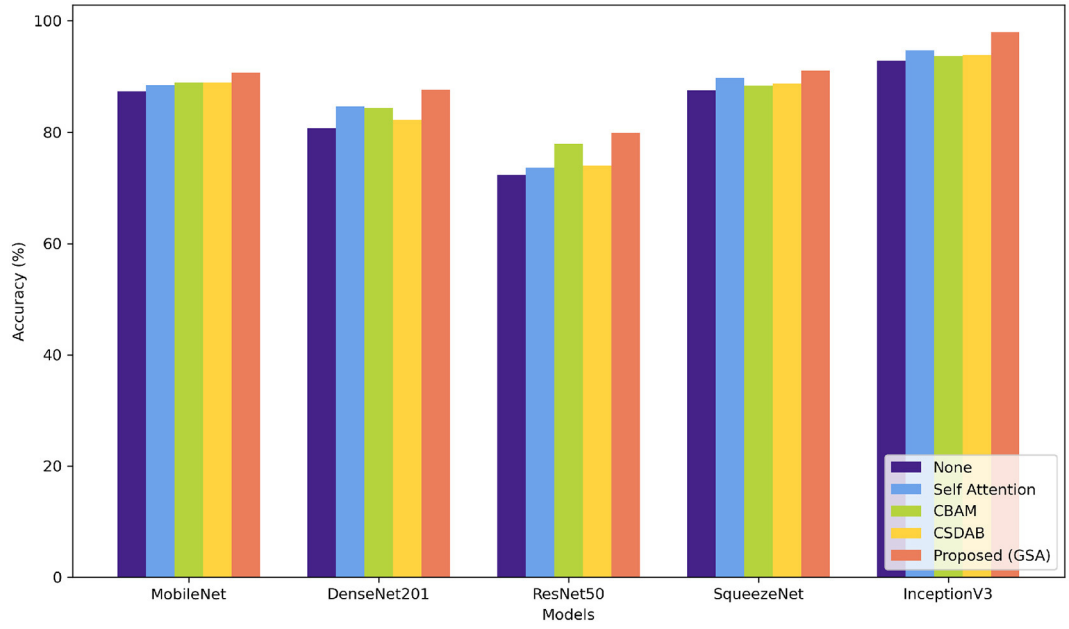

FIGURE 7
Visual comparison of various attention mechanisms on OASIS dataset.

TABLE 8 Comparison of proposed InGSA with transformer-based models on OASIS dataset.

Model	Accuracy (%)	F1-score (%)
ViT	91.65	86.78
DEiT	89.43	89.12
PVT	93.79	92.61
InGSA	97.99	97.88

negative. Next is the PVT model which gives classification accuracy of 93.79% and F1-score of 92.61% demonstrating the good classification prowess of the model. The performance of both ViT and DEiT models is reasonable, with accuracies of 91.65% and 89.43%, respectively.

5 Conclusion

Alzheimer's disease, diagnosed and classified with multiclass datasets in the early stage, needed a proficient automatic system identification. This study puts forward a CNN-Transformer model to diagnose Alzheimer's cases from multiclass datasets using transfer learning. First, a method of contrast enhancement is utilized to help better visualize important features. Furthermore, we introduce a new global CNN-transformer network known as InGSA for multiclass AD classification to facilitate end-to-end training. The InGSA architecture is based on the CNN and transformer, and GSA blocks are placed on top of pre-trained InceptionV3 model. GSA blocks are important for expression subscalar detection of global dependencies of features. The GSA component improves the extraction of detailed information by learning channel-wise and spatial-wise attention weights at the same time. In-depth experiments on two benchmark datasets demonstrate that our proposed InGSA achieves superior performance compared to the state-of-the-art techniques. Furthermore, GSA yields better results than other traditional attention methods. For the future works, we aim to test GSA on more extensive set and diverse dataset, and we also want to apply our proposed method in the other vision-related tasks.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://adni.loni.usc.edu/data-samples/adni-data/.

Author contributions

FB: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing – original draft. AA: Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The authors would like to express sincere gratitude to the Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Saudi Arabia, for their invaluable support and guidance.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Acharya, H., Mehta, R., and Singh, D. K. (2021). "Alzheimer disease classification using transfer learning," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (Erode: IEEE), 1503–1508. doi: 10.1109/ICCMC51019.2021.9418294

Ahmed, G., Er, M. J., Fareed, M. M. S., Zikria, S., Mahmood, S., He, J., et al. (2022). Dad-net: classification of Alzheimer's disease using ADASYN oversampling technique and optimized neural network. *Molecules* 27:7085. doi: 10.3390/molecules27207085

Ajagbe, S. A., Amuda, K. A., Oladipupo, M. A., Oluwaseyi, F. A., and Okesola, K. I. (2021). Multi-classification of alzheimer disease on magnetic resonance images (MRI) using deep convolutional neural network (DCNN) approaches. *Int. J. Adv. Comput. Res.* 11:51. doi: 10.19101/IJACR.2021.1152001

Böhle, M., Eitel, F., Weygandt, M., and Ritter, K. (2019). Layer-wise relevance propagation for explaining deep neural network decisions in MRI-based Alzheimer's disease classification. *Front. Aging Neurosci.* 11:456892. doi: 10.3389/fnagi.2019.00194

Cao, Y., Xu, J., Lin, S., Wei, F., and Hu, H. (2019). "GCNET: non-local networks meet squeeze-excitation networks and beyond," in *Proceedings of the IEEE/CVF international conference on computer vision workshops* (Seoul: IEEE). doi: 10.1109/ICCVW.2019.00246

Choi, B.-K., Madusanka, N., Choi, H.-K., So, J.-H., Kim, C.-H., Park, H.-G., et al. (2020). Convolutional neural network-based MR image analysis for Alzheimer's disease classification. *Curr. Med. Imaging* 16, 27–35. doi: 10.2174/1573405615666191021123854

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., et al. (2009). "Imagenet: a large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition* (Miami, FL: IEEE), 248–255. doi: 10.1109/CVPR.2009.5206848

Dutta, T. K., and Nayak, D. R. (2022). "CDANET: channel split dual attention based CNN for brain tumor classification in MR images," in *2022 IEEE international conference on image processing (ICIP)* (Bordeaux: IEEE), 4208–4212. doi: 10.1109/ICIP46576.2022.9897799

Dyrba, M., Hanzig, M., Altenstein, S., Bader, S., Ballarini, T., Brosseron, F., et al. (2021). Improving 3d convolutional neural network comprehensibility via interactive visualization of relevance maps: evaluation in Alzheimer's disease. *Alzheimers Res. Ther.* 13, 1–18. doi: 10.1186/s13195-021-00924-2

Edmonds, E. C., McDonald, C. R., Marshall, A., Thomas, K. R., Eppig, J., Weigand, A. J., et al. (2019). Early versus late MCI: improved mci staging using a neuropsychological approach. *Alzheimers Dement.* 15, 699–708. doi: 10.1016/j.jalz.2018.12.009

Frizzell, T. O., Glashutter, M., Liu, C. C., Zeng, A., Pan, D., Hajra, S. G., et al. (2022). Artificial intelligence in brain MRI analysis of Alzheimer's disease over the past 12 years: a systematic review. *Ageing Res. Rev.* 77:101614. doi: 10.1016/j.arr.2022.101614

Gao, F. (2021). Integrated positron emission tomography/magnetic resonance imaging in clinical diagnosis of Alzheimer's disease. *Eur. J. Radiol.* 145:110017. doi: 10.1016/j.ejrad.2021.110017

Gunawardena, K., Rajapakse, R., and Kodikara, N. D. (2017). "Applying convolutional neural networks for pre-detection of Alzheimer's disease from structural MRI data," in *2017 24th international conference on mechatronics and machine vision in practice (M2VIP)* (Auckland: IEEE), 1–7. doi: 10.1109/M2VIP.2017.8211486

Guo, H., and Zhang, Y. (2020). Resting state fMRI and improved deep learning algorithm for earlier detection of Alzheimer's disease. *IEEE Access* 8, 115383–115392. doi: 10.1109/ACCESS.2020.3003424

Hazarika, R. A., Maji, A. K., Kandar, D., Jasinska, E., Krejci, P., Leonowicz, Z., et al. (2023). An approach for classification of Alzheimer's disease using deep neural network and brain magnetic resonance imaging (MRI). *Electronics* 12:676. doi: 10.3390/electronics12030676

Kang, L., Jiang, J., Huang, J., and Zhang, T. (2020). Identifying early mild cognitive impairment by multi-modality MRI-based deep learning. *Front. Aging Neurosci.* 12:206. doi: 10.3389/fnagi.2020.00206

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E. M., et al. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group* under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology* 34, 939–939. doi: 10.1212/WNL.34.7.939

Menagadevi, M., Mangai, S., Madian, N., and Thiyagarajan, D. (2023). Automated prediction system for Alzheimer detection based on deep residual autoencoder and support vector machine. *Optik* 272:170212. doi: 10.1016/j.ijleo.2022.170212

Mohammed, B. A., Senan, E. M., Rassem, T. H., Makbol, N. M., Alanazi, A. A., Al-Mekhlafi, Z. G., et al. (2021). Multi-method analysis of medical records and MRI images for early diagnosis of dementia and Alzheimer's disease based on deep learning and hybrid methods. *Electronics* 10:2860. doi: 10.3390/electronics10222860

Murugan, S., Venkatesan, C., Sumithra, M., Gao, X.-Z., Elakkiya, B., Akila, M., et al. (2021). Demnet: a deep learning model for early diagnosis of Alzheimer diseases and dementia from MR images. *IEEE access* 9, 90319–90329. doi: 10.1109/ACCESS.2021.3090474

Nagarajan, S. M., Muthukumaran, V., Murugesan, R., Joseph, R. B., and Munirathanam, M. (2021). Feature selection model for healthcare analysis and classification using classifier ensemble technique. *Int. J. Syst. Assur. Eng. Manag.* 1–12. doi: 10.1007/s13198-021-01126-7

Nasir, I. M., Bibi, A., Shah, J. H., Khan, M. A., Sharif, M., Iqbal, K., et al. (2021). Deep learning-based classification of fruit diseases: an application for precision agriculture. *Comput. Mater. Contin.* 66, 1949–1962. doi: 10.32604/cmc.2020.012945

Nasir, I. M., Khan, M. A., Yasmin, M., Shah, J. H., Gabryel, M., Scherer, R., et al. (2020). Pearson correlation-based feature selection for document classification using balanced training. *Sensors* 20:6793. doi: 10.3390/s20236793

Nasir, I. M., Raza, M., Shah, J. H., Khan, M. A., Nam, Y.-C., Nam, Y., et al. (2023). Improved shark smell optimization algorithm for human action recognition. *Comput. Mater. Contin.* 76, 2667–2684. doi: 10.32604/cmc.2023.035214

Nasir, I. M., Raza, M., Shah, J. H., Wang, S.-H., Tariq, U., Khan, M. A., et al. (2022). Harednet: A deep learning based architecture for autonomous video surveillance by recognizing human actions. *Comput. Electr. Eng.* 99:107805. doi: 10.1016/j.compeleceng.2022.107805

Nozadi, S. H., Kadoury, S., and The Alzheimer's Disease Neuroimaging Initiative (2018). Classification of Alzheimer's and mci patients from semantically parcelled pet images: a comparison between av45 and FDG-pet. *Int. J. Biomed. Imaging* 2018:1247430. doi: 10.1155/2018/1247430

Oktavian, M. W., Yudistira, N., and Ridok, A. (2022). Classification of Alzheimer's disease using the convolutional neural network (CNN) with transfer learning and weighted loss. *arXiv* [Preprint]. arXiv:2207.01584. doi: 10.48850/arXiv.2207.01584

Pan, S. J., and Yang, Q. (2009). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191

Raju, M., Gopi, V. P., and Anitha, V. (2021). "Multi-class classification of Alzheimer's disease using 3dcnn features and multilayer perceptron," in *2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)* (Chennai: IEEE), 368–373. doi: 10.1109/WiSPNET51692.2021.9419393

Ramzan, F., Khan, M. U. G., Rehmat, A., Iqbal, S., Saba, T., Rehman, A., et al. (2020). A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks. *J. Med. Syst.* 44, 1–16. doi: 10.1007/s10916-019-1475-2

Simic, G., Stanic, G., Mladinov, M., Jovanov-Milosevic, N., Kostovic, I., Hof, P. R., et al. (2009). Does Alzheimer's disease begin in the brainstem? *Neuropathol. Appl. Neurobiol.* 35, 532–554. doi: 10.1111/j.1365-2990.2009.01038.x

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Las Vegas, NV: IEEE), 2818–2826. doi: 10.1109/CVPR.2016.308

Tehsin, S., Nasir, I. M., Damaševičius, R., and Maskeliūnas, R. (2024). DASAM: Disease and spatial attention module-based explainable model for brain tumor detection. *Big Data Cogn. Comput.* 8:97. doi: 10.3390/bdcc8090097

Tuvshinjargal, B., and Hwang, H. (2022). VGG-C transform model with batch normalization to predict Alzheimer's disease through MRI dataset. *Electronics* 11:2601. doi: 10.3390/electronics11162601

Varatharajah, Y., Ramanan, V. K., Iyer, R., and Vemuri, P. (2019). Predicting short-term mci-to-ad progression using imaging, CSF, genetic factors, cognitive resilience, and demographics. *Sci. Rep.* 9:2235. doi: 10.1038/s41598-019-38793-3

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). "CBAM: convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, eds. V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss (Cham: Springer), 3–19. doi: 10.1007/978-3-030-01234-2_1

Yousafzai, S. N., Shahbaz, H., Ali, A., Qamar, A., Nasir, I. M., Tehsin, S., et al. (2024). X-news dataset for online news categorization. *Int. J. Intell. Comput. Cybern.* 17, 737–758. doi: 10.1108/IJICC-04-2024-0184

Zhang, H., Goodfellow, I., Metaxas, D., and Odena, A. (2019). "Self-attention generative adversarial networks," in *International conference on machine learning* (Long Beach, CA: PMLR), 7354–7363.

Zhang, T., Zhao, Z., Zhang, C., Zhang, J., Jin, Z., Li, L., et al. (2019). Classification of early and late mild cognitive impairment using functional brain network of resting-state fMRI. *Front. Psychiatry* 10:572. doi: 10.3389/fpsyt.2019.00572