Computers in Biology and Medicine 179 (2024) 108874

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Ensemble approach of transfer learning and vision transformer leveraging explainable AI for disease diagnosis: An advancement towards smart healthcare 5.0

Ramesh Chandra Poonia [a,*], Halah A. Al-Alshaikh [b]

[a] Department of Computer Science, CHRIST (Deemed to be University), Delhi NCR, 201003, India
[b] Information Systems Department, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), 11432, Riyadh, Saudi Arabia

ARTICLE INFO

Keywords:
Smart healthcare
Disease diagnosis
Transfer learning
Vision transformer
Explainable AI

ABSTRACT

Smart healthcare has advanced the medical industry with the integration of data-driven approaches. Artificial intelligence and machine learning provided remarkable progress, but there is a lack of transparency and interpretability in such applications. To overcome such limitations, explainable AI (EXAI) provided a promising result. This paper applied the EXAI for disease diagnosis in the advancement of smart healthcare. The paper combined the approach of transfer learning, vision transformer, and explainable AI and designed an ensemble approach for prediction of disease and its severity. The result is evaluated on a dataset of Alzheimer's disease. The result analysis compared the performance of transfer learning models with the ensemble model of transfer learning and vision transformer. For training, InceptionV3, VGG19, Resnet50, and Densenet121 transfer learning models were selected for ensembling with vision transformer. The result compares the performance of two models: a transfer learning (TL) model and an ensemble transfer learning (Ensemble TL) model combined with vision transformer (ViT) on ADNI dataset. For the TL model, the accuracy is 58 %, precision is 52 %, recall is 42 %, and the F1-score is 44 %. Whereas, the Ensemble TL model with ViT shows significantly improved performance i.e., 96 % of accuracy, 94 % of precision, 90 % of recall and 92 % of F1-score on ADNI dataset. This shows the efficacy of the ensemble model over transfer learning models.

1. Introduction

In recent years, the healthcare industry has observed noteworthy advancements in technology, leading to the rise of smart healthcare systems. These systems influence various technologies, as well as artificial intelligence (AI), to improve healthcare delivery, improve patient-centric results, and rationalize medical procedures. One crucial feature of smart healthcare is disease diagnosis, which plays an important role in recognizing and treating illnesses promptly and precisely [1,2]. The start of smart healthcare systems, determined by technological advancements for example AI and machine learning (ML) that revolutionized the healthcare industry [3–5]. Disease diagnosis, a serious component of healthcare, has been meaningly impacted by these advancements, paving the way for more accurate, efficient, and patient-centric diagnostic procedures. The addition of smart technologies into healthcare settings offers huge potential for refining disease and eventually enhancing patient consequences. Researchers have recognized the status of patient's health in smart healthcare and have investigated and explored the domain. Chui et al. [1] highlighted the innovation, technologies, and applications relevant to disease prediction within smart healthcare. Similarly, Tian et al. [2] discuss the broader concept of smart healthcare, emphasizing the need to make medical care more intelligent. Although not explicitly focused on disease prediction, their work sets the stage by establishing the context of smart healthcare and its potential to transform healthcare services. To further delve into disease diagnosis in smart healthcare, the research emphasizes the importance of innovation, improvement, and skill development in this field, providing insights into advancements and strategies for enhancing investigative processes within smart healthcare systems [3]. The integration of AI and IoT for healthcare is explored, highlighting the benefits of combining these technologies to create intelligent models [4]. Disease diagnosis within the context of applying AI in healthcare is discussed, shedding light on the potential benefits and challenges associated with implementing AI-based systems, setting the stage for further

* Corresponding author.
E-mail addresses: rameshchandra.poonia@christuniversity.in (R.C. Poonia), hamshaikh@imamu.edu.sa (H.A. Al-Alshaikh).

https://doi.org/10.1016/j.compbiomed.2024.108874
Received 13 August 2023; Received in revised form 29 May 2024; Accepted 18 June 2024
Available online 15 July 2024

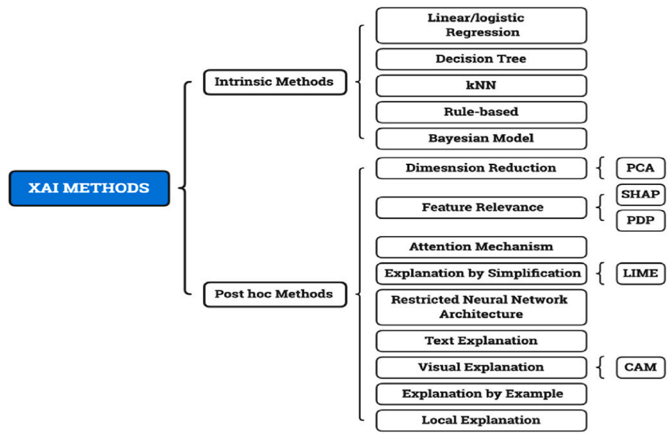

Fig. 1. Taxonomy of EXAI methods [15].

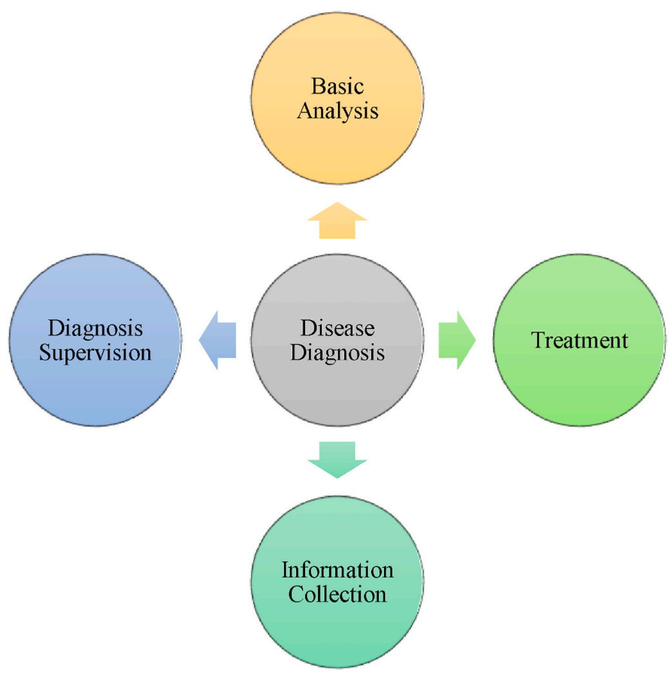

Fig. 2. Flowchart of smart healthcare.

advancements in this area [5–7]. These studies emphasize the vital role and transformative potential of AI/ML in smart healthcare for healthcare delivery. By leveraging smart technologies and incorporating AI-driven approaches, healthcare providers can enhance accuracy, and efficiency, and ultimately improve patient outcomes. Explainable AI (EXAI) is of paramount importance within smart healthcare systems due to the rising utilization of AI algorithms for disease prediction. Ensuring transparency and interpretability in the decision-making process is imperative [8–14]. EXAI techniques provide insights into how AI algorithms arrive at their decisions, fostering trust and acceptance among healthcare professionals, patients, and regulatory bodies. By understanding the reasoning behind AI-driven diagnoses, confidence in the technology is enhanced, leading to improved adoption and more reliable outcomes. EXAI has gained significant attention and has been applied in various healthcare domains, including disease diagnosis. Zhang et al. [15] discuss the applications of explainable AI in disease prediction and surgery. Their study explores how EXAI techniques can improve diagnostic accuracy and surgical decision-making by providing interpretable insights into AI algorithms' decision processes [7]. The taxonomy of EXAI is presented in Fig. 1.

In the context of smart healthcare 5.0, AI/EXAI/ML models have been employed to analyze MRI scans and showcase their potential in diverse medical applications. The flowchart in Fig. 2 illustrate the smart healthcare framework. One approach that has gained traction is the utilization of transfer learning models. MRI scans are commonly used to identify diseases. But the manual interpretation of these scans can be time-consuming and produce inaccurate results. Deep learning (DL) algorithms offer a promising solution by automating and improving the accuracy of tumor detection. In research studies, DL algorithms have been employed for tasks such as automatic segmentation and prediction of MRI scans. Techniques like Geometric Median Shift and convolutional neural networks (CNNs) have been used for automatic tumor detection and segmentation. This study aims to introduce an ensemble model that combines transfer learning (TL)-based models with a vision transformer model and EXAI. The ensemble model seeks to enhance the performance of tumor detection by leveraging the strengths of multiple TL models. Using a single model faces the challenges of overfitting and biased conclusions. By utilizing an ensemble model that combines the strengths of transfer learning (TL) models with the vision transformer model and EXAI for better feature extraction and improved performance in solving complex problems.

The paper discusses the remarkable advancements in smart healthcare, emphasizing the role of AI and explainable AI (EXAI) in disease diagnosis. Key achievements include technological innovations, the integration of AI and IoT, and the importance of transparent AI decision-making. Advanced achievements involve using transfer learning for MRI analysis, interdisciplinary collaboration, regulatory advancements, and the global impact of these technologies. These advancements aim to improve patient outcomes and make healthcare more efficient and accessible. It underscores the importance of Explainable AI (EXAI) for transparency and trust in AI-driven diagnoses.

Motivated by this, the paper contributes following:

- In this paper, transfer learning algorithms, namely VGG19, ResNet50, Densenet121, and InceptionV3 models are compared for Alzheimer's disease severity prediction using brain MRI images.
- Then ensemble of vision transformer models is implemented with transfer learning algorithms, namely VGG19, ResNet50, Densenet121, and InceptionV3.
- The paper presented the comparative analysis of single transfer learning models with ensemble models.
- The ensemble model outperformed the other CNN-based transfer learning algorithms.
- Additionally, the paper implemented the GradCAM model to provide the explainability of the model.

The remaining sections of the paper include: Section 2 provides a discussion of related work; a description of the proposed model's architecture and a description of models are illustrated in Section 3. Section 4 presents the performance evaluation and finally in section 5 concluding remarks with future research scopes are presented.

2. Related work

Dave et al. [8] conducted a study using explainable AI on a heart disease dataset. Their research highlights the importance of EXAI in healthcare, demonstrating how interpretability can enhance the understanding of AI models' decisions and improve the diagnosis of heart disease. Otaki et al. [10] explored EXAI in clinical applications such as single-photon emission computed tomography (SPECT) for coronary artery disease (CAD) diagnosis. In the area of cardiac disorders, Anand et al. [11] presented an explainable AI for designing a decision model for disease diagnosis. Author analyzed the ECG data using SHAP for decision making capability of deep CNN. Basheera et al. [16] proposed hybrid CNN model for Alzheimer's disease classification. Author extracted the gray matter from brain voxels. For image enhancement gaussian filter was used with independent component analysis (ICA) for segmentation process. Rosenson et al. [17] provided an approach of AI to identify dysfunctional high-density lipoprotein (HDL) and its association with atherosclerotic cardiovascular disease. Al'Aref et al. [18] explored the applications of machine learning for the identification of cardiovascular disease using different imaging modalities. Joo et al. [19] focused on the clinical inferences of machine learning in forecasting the incidence of cardiovascular diseases. Dey et al. [20] provided an inclusive review of the applications of AI using cardiovascular imaging and explored its efficacy to improve image analysis, risk prediction, and treatment response assessment, thereby enhancing cardiovascular disease management. Huang et al. [21] used AI applications for the diagnosis of different cancer types and underscored the importance of AI-driven solutions in personalized medicine. Xu et al. [22] focused on the conversion of cancer genomics for precision medicine with the application of AI. The authors discuss the challenges and potential of integrating AI with cancer genomics data, emphasizing its impact on treatment strategies, patient outcomes, and precision medicine. They highlight role of AI in identifying therapeutic targets, predicting treatment responses, and facilitating personalized approaches. Bi et al. [23] examined the role of AI in analyzing radiological images with pathological data. Adir et al. [24] investigated the integration of AI and nanotechnology in precision cancer medicine. They emphasized the potential of AI to analyze large datasets from nanotechnology-based imaging, enabling improved diagnosis, targeted therapy, and treatment response monitoring. Multidisciplinary collaborations are crucial to fully leverage AI and nanotechnology in cancer care. Coccia [25] used the advancement of AI tools for cancer detection with the integration of imaging applications. Integrating AI-driven image analysis with clinical practices has the potential to enhance early detection, treatment planning, and patient outcomes in cancer care. Vieira et al. [26] presented the correlation between psychiatric and neurological disorders with the application of convolutional neural networks (CNNs). Zhang et al. [27] proposed a machine-learning approach for identification of the neurological cerebrovascular disease using MRI data. Surianarayanan et al. [28] applied machine learning for diagnosing neurological conditions with radiological images. Sappagh et al. [29] developed a multimodal system for the diagnosis of Alzheimer's disease with the incorporation of explainable AI. Sudar et al. [30] used the EXAI model with VGG for the identification of Alzheimer's disease and its severity stage. Lombardi et al. [31] focused on the development of a robust framework for assessing and predicting cognitive decline in patients with neurodegenerative diseases like Alzheimer's. Jain et al. [32] presented a Visual Explainable AI model for identifying and classifying dementia using MRI scans and achieved an accuracy of 74 % only. Whereas, Tuvshinjargal et al. [33] classified the diseases with a quantization technique that was applied over MRI data and used the VGG-C Transform model. Mohi et al. [34] employed Convolutional Neural Network (CNN) model for Alzheimer's disease stages detection. Nijaguna et al. [35] explored the use of the Quantum Fruit Fly Algorithm (QFFA) in combination with ResNet50 and VGG16 deep learning models for medical diagnosis. This approach offers a promising solution for improving medical diagnosis accuracy in handling large datasets. Chhabra and Kumar [36] presented a smart healthcare system that employs the DenseNet 121 deep learning model to simultaneously detect multiple diseases from chest X-ray images. The research addresses the challenge of identifying various chest conditions. This approach holds promise for enhancing healthcare diagnostics and decision-making. Odusami et al. [37] addressed the challenge of early Alzheimer's disease (AD) detection using multimodal neuroimaging data from MRI and PET scans. The proposed approach combines these data sources using discrete wavelet transform (DWT) and transfer learning with VGG16. The fused images are classified using a vision transformer. Testing on ADNI dataset showed high accuracy, with 93.75 % accuracy for PET data, surpassing previous studies and highlighting its potential for improving AD diagnosis accuracy. Miltiadous et al. [38] introduced a novel method called DICE-net for classifying Alzheimer's disease (AD) using EEG data. DICE-net outperforms baseline models, achieving an 83.28 % accuracy in distinguishing AD patients from healthy individuals, with potential applications in early AD diagnosis and interventions. Qiu et al. [41] presented an interpretable deep learning strategy for identifying Alzheimer's disease signatures. It utilizes multimodal inputs including MRI, age, gender, and Mini-Mental State Examination scores. The framework combines a fully convolutional network, which creates high-resolution maps of disease probability from brain structure, with a multilayer perceptron. This approach allows for precise and intuitive visualization of individual Alzheimer's disease risk, aiding in accurate diagnosis. Zhu et al. [42] introduced a dual attention multi-instance deep learning network (DA-MIDL) for early diagnosis of Alzheimer's disease (AD) and its prodromal stage, mild cognitive impairment (MCI). DA-MIDL has three main components: Patch-Nets with spatial attention blocks for feature extraction, an attention multi-instance learning (MIL) pooling operation for balanced patch contribution, and an attention-aware global classifier for AD-related classification. Tested on 1689 subjects from ADNI and AIBL datasets, the DA-MIDL model shows improved accuracy and generalizability in identifying pathological locations and classification performance. Xia et al. [43] proposed a deformable self-attention module, which selects the positions of key and value pairs in a data-dependent manner. This method allows the self-attention module to focus on relevant regions and capture more informative features. Based on this, they present the Deformable Attention Transformer, a general backbone model with deformable attention for image classification and dense prediction tasks. Their experiments demonstrate consistently improved results across various benchmarks. Hu et al. [44] presented the Conv-Swinformer for Alzheimer's disease prediction from brain MRI images by hybridization of CNN and Transformer

Table 1
Recent research contributions.

Ref	Technique Used	Disease Type	Input Images
Otaki et al. [10]	EXAI	Coronary Artery Disease (CAD)	SPECT
Basheera et al. [16]	Hybrid CNN	Alzheimer's Disease	MRI
Zhang et al. [27]	Machine Learning	Cerebrovascular Disease	MRI
Surianarayanan et al. [28]	Machine Learning	Neurological Conditions	Radiological Images
Sappagh et al. [29]	EXAI	Alzheimer's Disease	MRI
Sudar et al. [30]	EXAI with VGG	Alzheimer's Disease	MRI
Lombardi et al. [31]	Machine Learning	Neurodegenerative Diseases (Alzheimer's)	Neuropsychological Test Results
Jain et al. [32]	Visual EXAI	Dementia	MRI Scans
Tuvshinjargal et al. [33]	VGG-C Transform with Quantization	Various Diseases	MRI
Mohi et al. [34]	CNN	Alzheimer's Disease Stages	MRI
Odusami et al. [37]	DWT and Transfer Learning with VGG16	Alzheimer's Disease (AD)	MRI and PET Scans
Miltiadous et al. [38]	DICE-net	Alzheimer's Disease (AD)	EEG Data
Qiu et al. [41]	Interpretable Deep Learning	Alzheimer's Disease	MRI, Age, Gender, MMSE Scores
Zhu et al. [42]	DA-MIDL	Alzheimer's Disease (AD) and Mild Cognitive Impairment (MCI)	MRI
Hu et al. [44]	Conv-Swinformer	Alzheimer's Disease	Brain MRI Images

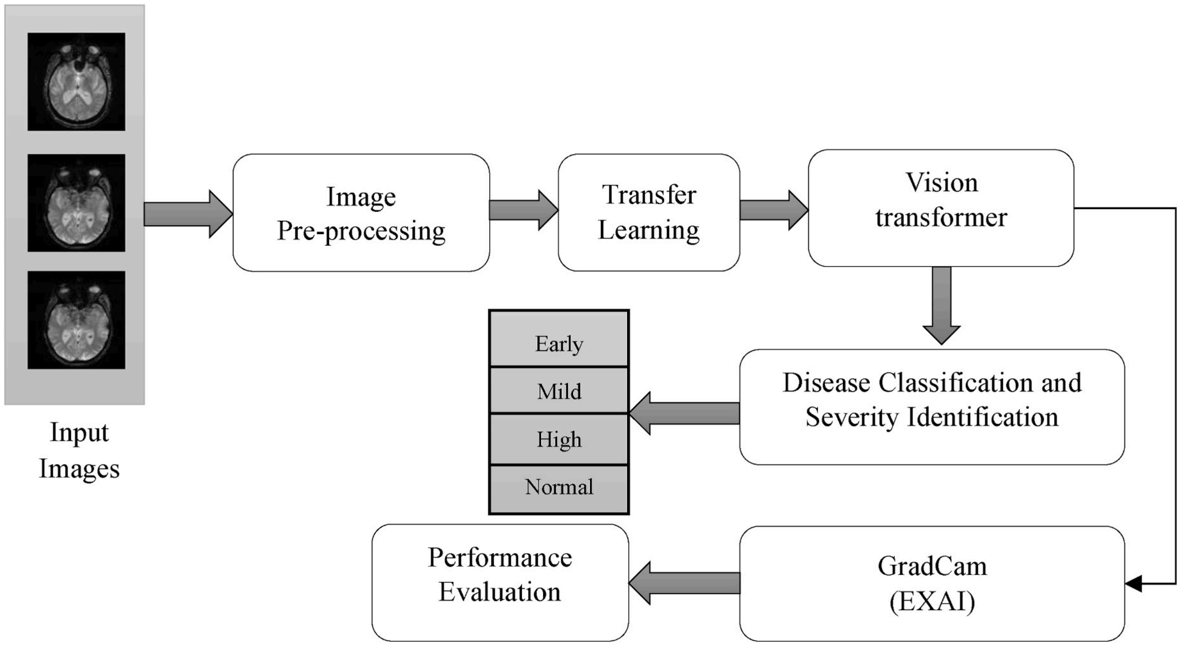

Fig. 3. Flowchart of methodology.

technologies. Some of the critical literature used for medical image processing are presented in Table 1.

The presented technologies offer advantages such as improved diagnostic accuracy, potential for early disease detection, and personalized treatment options. However, they also come with challenges including the need for large datasets, data preprocessing complexity, and specialized equipment requirements. Overall, these innovative techniques hold great promise in revolutionizing healthcare by providing more accurate and timely diagnoses, ultimately leading to improved patient outcomes and interventions.

3. Methodology used

In this paper, a smart healthcare model is presented for disease diagnosis using an ensemble approach of transfer learning and vision transformer leveraging Explainable AI. For experimental evaluation, Alzheimer's disease dataset [39] is considered. The dataset consists of 819 subjects. The dataset contains 3 groups of subjects: AD, MCI, and CN. Here the model is designed to predict the disease and its severity level as "Early, Mild, High, and Normal". The detection and classification followed several key steps, as presented in Fig. 3. Algorithm for the proposed work is presented below:

1: Start
2: Input MRI brain images.
3: Preprocess the images using image enhancement and augmentation techniques.
4: Extract features using an ensemble model of transfer learning InceptionV3 and ViT.
5: Build a classification model with a softmax layer.
6: Train the model.
7: End training.
8: Input new MRI brain images for Alzheimer severity detection.
9: Feed the images to the trained model.
10: Apply the softmax function to obtain probability scores for each severity class.
11: Determine the detected Alzheimer's severity based on the probability scores.
12: End

In this work, Alzheimer's Disease Neuroimaging Initiative (ADNI) [39] is used. The ADNI is a significant, multi-site study focused on improving clinical trials for Alzheimer's disease (AD) prevention and treatment. The data set comprises MRI images, divided into four classes in both training and testing sets: Mild Demented, Moderate Demented, Non-Demented, and Very Mild Demented.

Image Preprocessing is applied to distinguishing healthy areas from anomalous regions in medical images is challenging. Therefore, preprocessing is essential. Medical images often have a limited dynamic range, making it difficult to adequately expose all pixels. To address this, increasing the exposure in some under-exposed regions can enhance the visibility and clarity of the image. This is mathematically represented as:

$$I_e = \sum_{m=1}^{M} I_m * q_i \tag{1}$$

Where, exposure or enhancement is represented as q_i in image I_m with color channels M and result in improved exposure image.

Then augmentation is applied such as flip and rotation, translation, scaling and cropping, and shearing for medical imaging learning.

The transfer learning (TL) models are ensembled together with ViT-based models to enhance their performance. The proposed TL methods integrated with ViT exhibit superiority in recognizing complex patterns and adapting to diverse healthcare datasets, thereby providing enhanced performance in smart healthcare data analysis. The advanced combination allows for efficient learning with less labelled data, making it a valuable approach in the data-scarce healthcare field. While introducing complexity due to the need for careful tuning and the integration of VT, the significant improvement in performance metrics justifies this complexity. Other machine learning methods were not considered in the study to focus on enhancing and optimizing existing, proven architectures, thereby ensuring that the research findings could be readily applied and generate immediate value in the field of smart healthcare.

Here, in this paper, we have analyzed different transfer learning models ensembled together with ViT-based models. After running these models, we analyzed and compared their results. Furthermore, we assessed the classification accuracy of ensemble models using GradCam, which is an Explainable AI tool. The input image dimensions were set to 128 × 128, and we used weights from the ImageNet dataset. A batch size of 64 was used throughout. This model converts the images into 1Darray feature maps. The model's final layer uses softmax activation for multiclass classification. Adam optimizer with a learning rate of 1e-5 is used. Adam adjusts the learning rates for specific parameters using the adaptive learning rate technique as stated below:

$$l = -\sum_{c=1}^{M} y_{o,c} \log\left(p_{o,c}\right) \tag{2}$$

$$\sigma(\overrightarrow{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \tag{3}$$

Where σ is termed as softmax function. Here input vector is represented as z, and the exponential function as $z(i)$ represents the standard exponential function for the ith feature vector. The output vector consists of C classes, where the exponential function of $z(j)$ represents the standard exponential function. For training, InceptionV3, VGG19, Resnet50, and Densenet121 transfer learning models were selected for ensembling with vision transformer. The ensemble model with transfer learning and vision transformer has several benefits: improved performance by combining multiple models, enhanced robustness and generalization by reducing errors and capturing diverse patterns, utilization of complementary information through different transformations, leverage of pre-trained models for faster initialization, and overcoming data scarcity, and enhanced resistance to perturbations through data augmentation. This approach is effective for vision-based tasks. These are discussed below:

3.1. VGG19

One of the popular transfer learning model is VGG19 which is a type of convolutional neural network model with 19 layers. Out of 19 layers, 16 layers are of convolution layer, and 3 fully-connected layers with 5 max-pooling layers, and 1 softmax layer. The kernel size is of 3× 3 with stride value of 1. Alongwith that spatial padding is also added to preserve the resolution of image. Activation function used in VGG19 is the ReLU activation after each convolution layer. This preserves the non-linearity of the model. By default input image size of the VGG19 is 224 × 224 pixels with 3 RGB channels. Spatial dimension of the processed features in VGG19 is reduced by applying Maxpooling layer. This layer use kernel size of 2 × 2 with stride value of 2 is used. The last fully-connected layer consists of 4096 neurons. Finally, the fully-connected layer is connected with softmax layer. The learning parameters for VGG19 is 19.6 billion [40].

3.2. ResNet50

ResNet is a deep neural network with nearly 50 layers that uses residual blocks. It achieved high performance in various image recognition competitions. The concept of adding more layers to learn complex features has limitations in traditional CNNs. ResNet addresses this issue by introducing shortcut connections that allow direct mapping from input to output. This architecture is widely used for image classification, object localization, and detection. It can also be extended to other tasks, providing depth advantages while reducing processing costs. ResNet models like ResNet101 and ResNet152 have fewer filters and a simpler structure compared to VGG networks [41]. The authors of ResNet introduced identity connections that directly connect the input of a layer to its output. This enables the layers to learn a residual mapping, denoted as H(x), where x is the input and F is the output from the identity connections. They also allowed the nonlinear layers to learn a different mapping. The ResNet architecture includes skip connections that connect the input directly to the output, allowing for the learning of residual mappings. This helps alleviate the issue of vanishing gradients and enables better performance of deeper layers. The ResNet50 model, in particular, uses a bottleneck architecture and consists of convolutional and pooling layers, repeated several times with varying kernel sizes. The ResNet50 architecture is a deep neural network consisting of several convolutional and pooling layers. Here is a summary of its structure:

- The network starts with a convolutional layer with a kernel size of 7 × 7 and 64 distinct kernels. Each kernel has a stride size of 2, resulting in one layer.
- Then it is followed by a max-pooling layer that contains a stride value of 2.
- Next, there are three sets of convolutional layers:
    - The first set includes three consecutive layers with kernel sizes of 1 × 1, 64; 3 × 3, 64; and 1 × 1, 256. This set is repeated three times, resulting in nine layers.
    - The second set consists of four layers with kernel sizes of 1 × 1, 128; 3 × 3, 128; and 1 × 1, 512. This set is repeated four times, total of 12 layers.
    - The third set includes six layers with kernel sizes of 1 × 1, 256; 3 × 3, 256; and 1 × 1, 1024. This set is repeated six times, resulting in 18 layers.
- Following the third set, there is another set of three layers with kernel sizes of 1 × 1, 512; 3 × 3, 512; and 1 × 1, 2048. This set is repeated thrice and results in nine layers.
- After the convolutional layers, an average pooling layer is applied.
- Finally, there is a fully connected layer with 1000 nodes, followed by a softmax function, resulting in one layer.

3.3. Densenet121

DenseNet is a type of fully connected convolutional neural network (CNN) that employs dense connections between layers. Unlike traditional CNNs, where layers are sequentially connected, DenseNet connects each layer to all preceding and subsequent layers, forming dense connections. This design addresses the vanishing gradient problem and improves accuracy in high-level neural networks. DenseNet is a convolutional neural network architecture that shares similarities with ResNet but also incorporates some distinctive features [42]. One of its key advantages is parameter efficiency, achieved by utilizing a small number of parameters per layer. Moreover, DenseNet benefits from deep implicit supervision, which enhances the flow of gradients throughout the network. This architecture employs dense connections, ensuring that information transmitted across multiple levels is not lost or diminished as it traverses the network. Unlike traditional CNNs that sum feature maps, DenseNet concatenates them, enabling feature reuse and reducing the overall number of parameters. The fundamental structure of DenseNet includes batch normalization layers, ReLU activation functions, and 3x3 convolutions at each step. Each dense block within the architecture comprises a varying number of layers (repeats), with each layer consisting of two convolutional layers: a bottleneck layer with a 1x1 size then a kernel of size 3x3 is used for convolution operation and a second layer is convolutional layer with 1x1 size with average pooling layer with stride 2.

3.4. InceptionV3

Inception V3 is an enhanced version of Inception designs that aim to reduce computational resources. Compared to VGGNet, Inception Networks are more computationally efficient with fewer parameters and lower costs. To maintain these advantages while modifying an Inception Network, Inception v3 suggests optimization strategies such as factorized convolutions, regularization, dimension reduction, and parallelized calculations. These strategies facilitate easy adaptation of the model for different use cases while preserving computational efficiency. Inception v3 is constructed step-by-step, incorporating factorized convolutions, smaller convolutions, asymmetric convolutions, an auxiliary classifier, and an efficient technique for grid size reduction. The final architecture

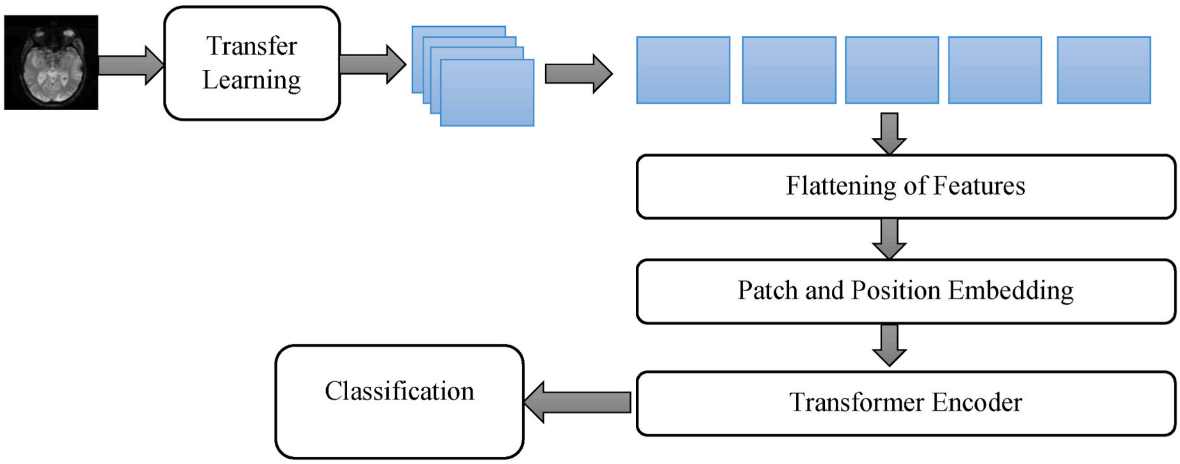

Fig. 4. Ensemble Model of Transfer Learning and Vision transformer.

of Inception v3 optimizes both computational efficiency and performance [36].

3.5. Vision transformer

Vision Transformer (ViT) is a type of deep learning model that is based on embedding layered architecture that is used to process natural language requirements. Based on that concept the image patches are flattened and fed as sequences to the model for learning. These patches are flattened and fed into a Transformer encoder, which includes self-attention layers and feed-forward networks. ViT models are pre-trained on large datasets and then fine-tuned for specific tasks, achieving competitive performance. However, they require more computational resources compared to convolutional neural networks (CNNs) [43]. Researchers are exploring variations and improvements, such as hybrid models and task-specific approaches. To adapt the Transformer model for images, the image is reshaped into patches and represented as token embeddings. Position embeddings are added to preserve positional information. The sequence of patch embeddings, along with position embeddings, is inputted to the Transformer encoder, which consists of multiple layers of self-attention and MLP blocks. Layer normalization and residual connections are applied for improved gradient flow. A learnable embedding is added at the beginning of the sequence, serving as the image representation. A classification head, typically an MLP or linear layer, is attached to this representation for tasks like pre-training or fine-tuning.

An alternative approach to forming the input sequence in the Transformer model is to utilize feature maps obtained from a convolutional neural network (CNN), in addition to raw image patches. In the ensemble model depicted, the patch embedding projection is applied to patches that are extracted from the CNN feature map, as illustrated in Fig. 4. One option is to utilize patches with a spatial size of 1x1, effectively flattening the spatial dimensions of the feature map and projecting it to the appropriate dimension required by the Transformer. To supplement the patch embeddings and preserve positional information within the input sequence, the classification input embedding and position embeddings are incorporated. Detailed steps are provided as below:

The ViT is a deep learning architecture that utilizes transformer models for visual tasks. The mathematical expressions involved in the ViT working flow can be represented as follows:

1: Patch Extraction

$$Patches\ (P,P,C) \leftarrow Input\ image(H,W,C)\ \{Where, H = height, W = width, P = patch\ sze, C = No.\ of\ channel\}\ .$$

2: Patch Embedding

$$LD\ Embedding\ Space\ (N,D) \xleftarrow{linear\ projection} Patches(P,P,C)\ \{Where, D = Embedding\ Size, N = No.\ of\ patches\}\ .$$

3: Positional Encoding

$$Positional\ Encoded\ Patches \xleftarrow{Sine\ and\ Cosine\ Functions} LD\ Embedding\ Space\ (N,D)\ .$$

4: Transformer Encoder

$$Feature\ Maps \xleftarrow{attention\ network} Positional\ Encoded\ Patches\ .$$

5: Classification

$$Severity\ Probablity \xleftarrow{softmax} Feature\ Maps.$$

3.6. GradCAM

For designing deep learning more understanding and interpretable for medical imaging applications GradCAM was introduced. This is class activation map that have ability to visualize and explain the decisions. During detection and prediction of task it is required to visualize the internal feature maps generated and provide the visual explanation. In this approach a localization map [46] is created as

$$L_{GRADCAM} = RELU\left(\sum_k \alpha_k A_k\right) \tag{4}$$

Where, α_k feature weight and A_k is the feature map.

4. Results and discussions

This section presented the results and discussions for Alzheimer's disease diagnosis and its severity analysis for the deployment of smart healthcare. The dataset consists of ~5000 images [39]. Among these 70 % of images are used for training and 30 % for testing or validation. The entire model is simulated on python platform over google colab with facility of Tesla P100-PCIE GPU. The performance evaluation included accuracy, precision, recall, and F1_score, as stated below:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{5}$$

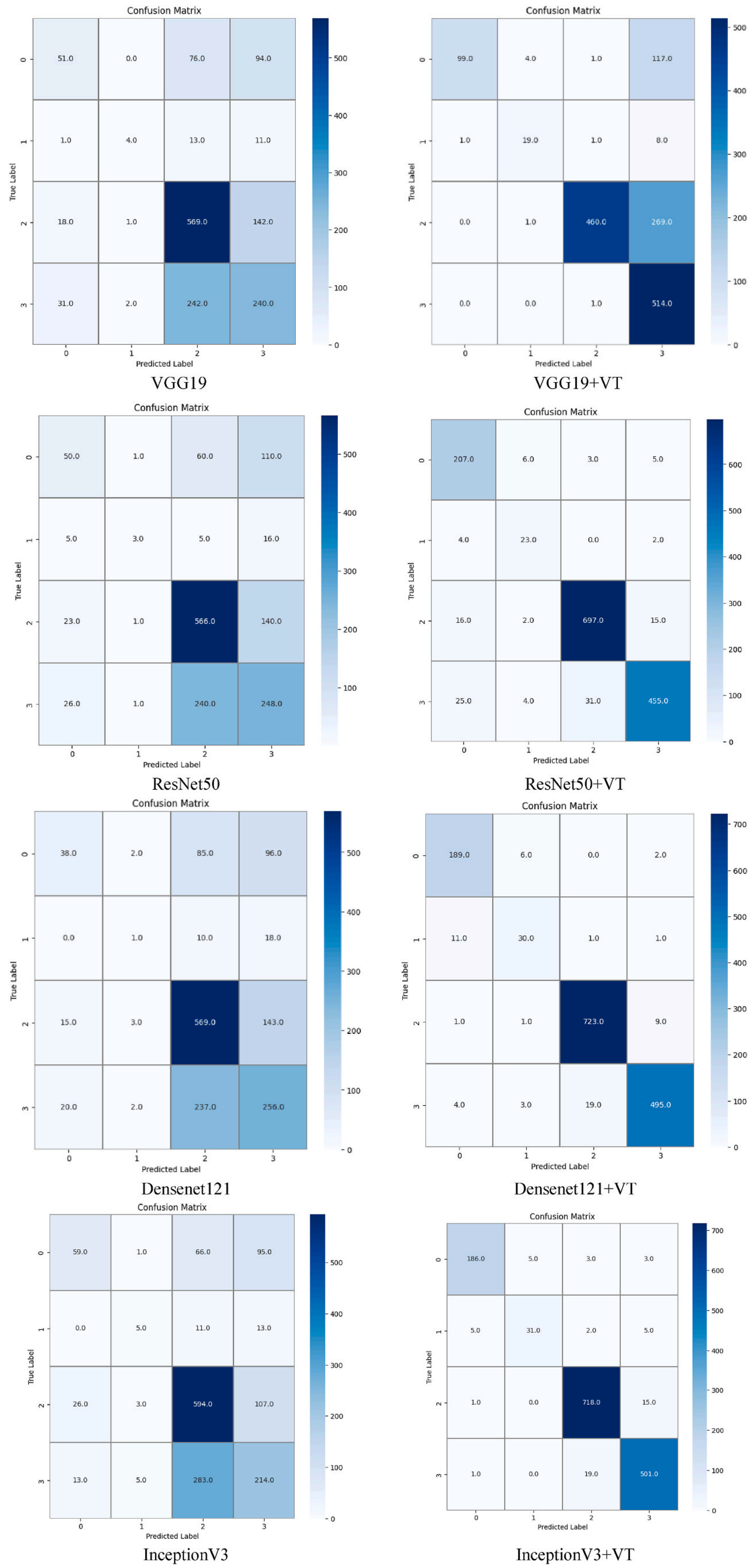

Fig. 5. Comparison of Confusion Matrix of Transfer Learning Model with Ensemble Transfer Learning with Vision transformer Model.

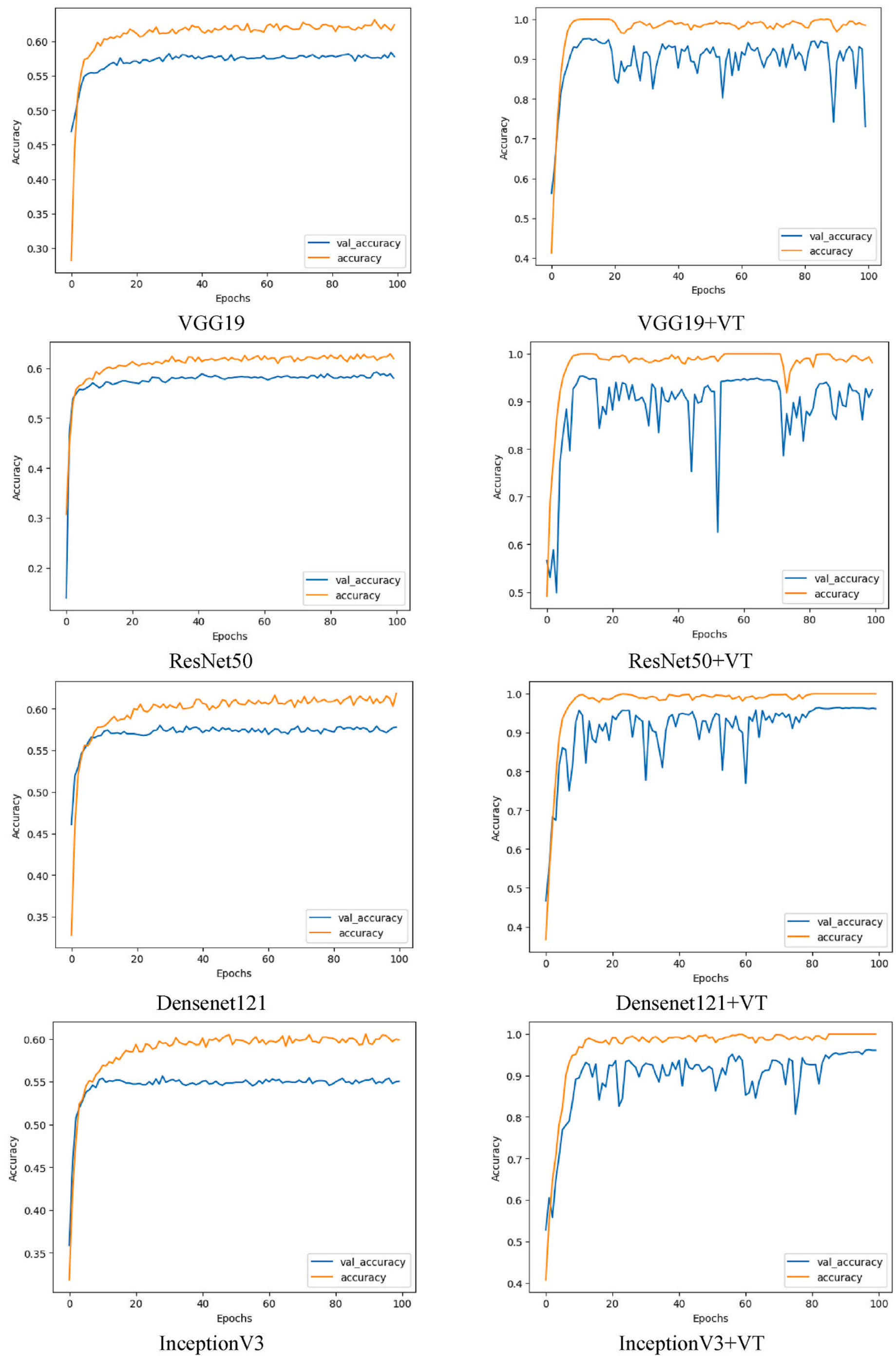

Fig. 6. Comparison of Training and Validation Accuracy of Transfer Learning Model with Ensemble Transfer Learning with Vision transformer Model.

$$Precision = \frac{(TP)}{(TP + FP)} \tag{6}$$

$$Recall = \frac{(TP)}{(TP + FN)} \tag{7}$$

$$F1\_score = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} \tag{8}$$

Fig. 5 shows a comparison of the confusion matrix of the transfer learning model with ensemble transfer learning with the vision transformer model for VGG19, VGG19+VT, ResNet50, ResNet50+VT, Densenet121, Densenet121+VT, InceptionV3, and InceptionV3+VT methods. The confusion matrix of VGG19 shows more missclassification as compared to VGG19+VT. Whereas ResNet50 shows high missclassification as compared to ResNet50+VT, Densenet121+VT shows better classification result as compared to Densenet121. Similar pattern was also observed in InceptionV3, and InceptionV3+VT.

Fig. 6 shows a comparison of training and validation accuracy of the transfer learning model with ensemble transfer learning with vision transformer model in which graph is plotted between accuracy and epochs. For VGG19 model training accuracy is approx. 60 % whereas the validation accuracy was approx. 55 % but VGG19+VT model achieved training and validation accuracy of more than 90 %. Similar pattern was observed in ResNet50 as well as ResNet50+VT model. For Densenet121 model accuracy is also ranging between 55 and 60 % and

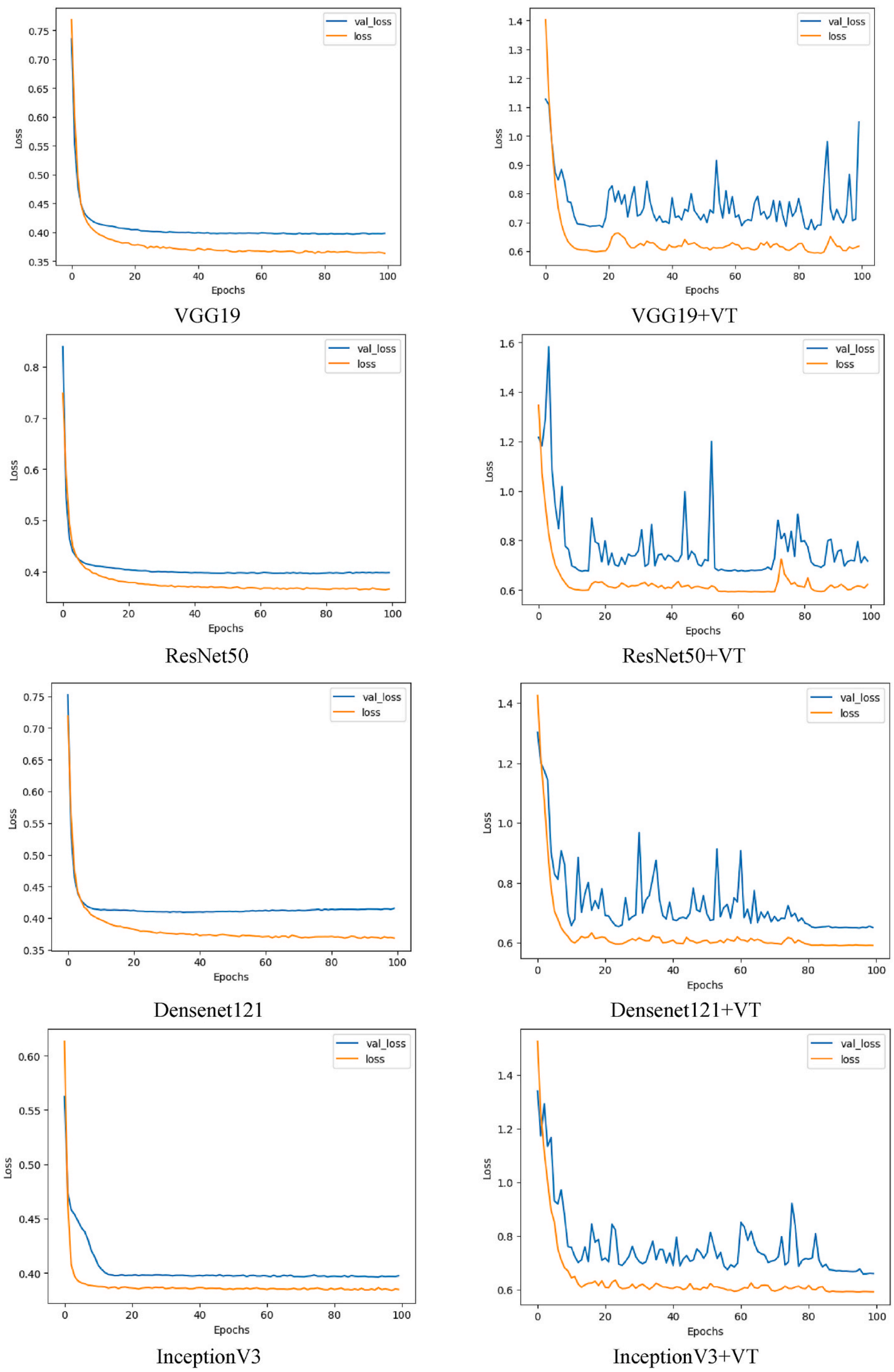

Fig. 7. Comparison of Training and Validation Loss of Transfer Learning Model with Ensemble Transfer Learning with Vision transformer Model.

Table 2
Performance of transfer learning models.

	"Accuracy"	"Precision"	"Recall"	"F1-score"
Vgg19	0.58	0.55	0.40	0.43
ResNet50	0.58	0.53	0.40	0.42
DenseNet121	0.58	0.56	0.41	0.41
InceptionV3	0.58	0.52	0.42	0.44

Densenet121+VT model achieved an accuracy of approx. 95 %. Similar pattern was also observed in InceptionV3 as well as InceptionV3+VT. Fig. 7 shows a comparison of training and validation loss of the transfer learning model with ensemble transfer learning with vision transformer model in which graph is plotted between loss and epoch for VGG19 model loss is near about 0.4 and for VGG19+VT model loss is near about 0.6. Similar pattern was observed in all other models i.e., ResNet50, ResNet50+VT, Densenet121, Densenet121+VT, InceptionV3 and InceptionV3+VT.

Table 2 shows that the four models (Vgg19, ResNet50, DenseNet121, and InceptionV3) have similar performance with an accuracy of 0.58. They also exhibit comparable precision and recall values, ranging from 0.52 to 0.56 and 0.40 to 0.42, respectively. InceptionV3 achieves the highest F1-score at 0.44, while the other models have slightly lower scores ranging from 0.42 to 0.43. Overall, the models' performance metrics indicate limited discrimination and suggest that they may

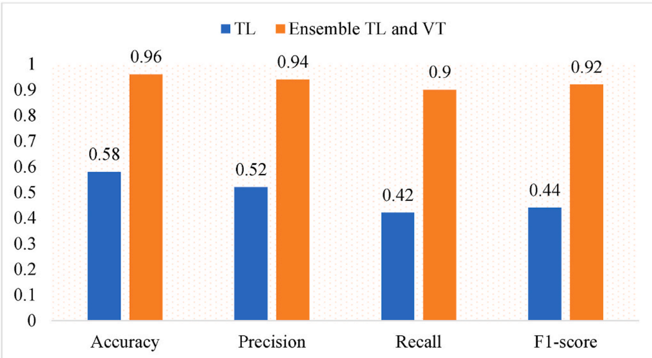

Fig. 8. Comparison of Transfer Learning Model with Ensemble Transfer Learning with Vision transformer Model.

Table 3
Performance of Ensemble Transfer Learning with Vision transformer Model.

	"Accuracy"	"Precision"	"Recall"	"F1-score"
Vgg19+VT	0.73	0.84	0.68	0.71
ResNet50+VT	0.92	0.85	0.89	0.87
DenseNet121+VT	0.96	0.91	0.90	0.90
InceptionV3+VT	0.96	0.94	0.90	0.92

Table 3 compares the performance of four different ensemble transfer learning models with vision transformer: Vgg19+VT, ResNet50+VT, DenseNet121+VT, and InceptionV3+VT. The table shows that DenseNet121+VT, and InceptionV3+VT achieved highest accuracy, InceptionV3+VT achieved highest precision, recall and f1-score.

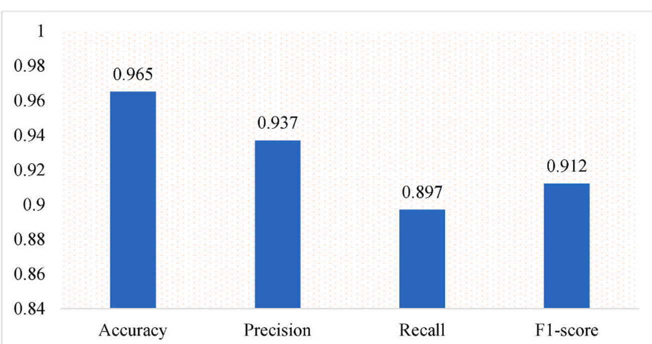

Fig. 9. Cross-validation result.

struggle to capture complex patterns in the data.

Fig. 8 compares the performance of two models: a transfer learning (TL) model and an ensemble transfer learning (Ensemble TL) model combined with vision transformer (VT). The ensemble transfer learning models with vision transformer show improved performance compared to the individual models. They achieve higher accuracy, precision, recall, and balanced F1-scores, indicating an overall boost in classification performance. This suggests that combining multiple models and incorporating vision transformers enhances the models' ability to correctly predict positive instances, capture actual positives, and achieve a good trade-off between precision and recall.

Fig. 9 presents the 10-fold cross validation for the ensemble transfer learning with vision transformer and achieved an average accuracy of 96.5 %, average precision of 93.7 %, average recall of 89.7 %, and an average F1-score of 91.2 %. The results shows the model's robustness and reliability, making it a potent tool for applications demanding high predictive accuracy and efficiency.

Fig. 10 represents the GradCAM representation of classified outcome. GradCAM is an EXAI approach in computer vision that highlights critical regions in an image to help comprehend and interpret a model's judgements. The image appears to be a composite of brain scans displayed with a color map overlay, which is commonly used in medical imaging to enhance visual analysis using GradCAM [45]. The spectrum of colors, ranging from cool to warm (blue to red) that represents the gradient from low to high in white matter of brain that can identify the region for analysis of Alzheimer's disease. The color variations help in analysis of the critical region for decision-making for medical professionals. This will enhance the trust on AI-assisted diagnosis.

In Fig. 11 comparative evaluation of various machine learning models, their respective performance scores were reported for a specific task. VGG16 was used in Ref. [37] and achieved performance of approx. 81 %. DICE-Net [38] achieved an accuracy score of 83.28 %, while MLP [41] slightly outperformed it with an accuracy score of 83.40 %. However, AttentionCNN [42] achieved an accuracy of 92.40 %. Conv-Swinformer [44] achieved an accuracy of 93.56 % The highest-performing model in this comparison was InceptionV3+VT, which reached an impressive accuracy score of 96 %.

5. Discussion

The initial transfer learning models—VGG19, ResNet50, DenseNet121, and InceptionV3—demonstrated limited data discrimination and complex pattern recognition performance, yielding approximately 58 % accuracy and F1-scores below 44 %. However, upon integrating the Vision Transformer (VT), the models exhibited significant enhancements, notably improving accuracy, precision, recall, and F1-score metrics.

This improvement carries important managerial implications. Firstly, the enhanced models support informed and reliable decision-making, contributing to better healthcare diagnosis, treatment planning, and outcome prediction. Secondly, the efficient models enable optimal resource utilization, potentially reducing operational costs. Moreover, the heightened accuracy and reliability in data analysis mitigate risks associated with misdiagnosis and improper treatment planning. The models further provide valuable insights for effective strategic planning and policy-making in healthcare.

In the context of managerial implications, it is crucial to emphasize that investing in training for staff to use and interpret data generated through these advanced models effectively is essential. This ensures that the benefits of the improved models are fully realized and integrated into healthcare practices.

Key takeaways from the research underscore the significance of the VT integration with transfer learning models in addressing performance gaps of traditional models. This integration offers substantial advantages for healthcare practitioners and administrators, supporting decision-making, resource optimization, risk mitigation, and strategic planning. Continuous improvement, staff training, and fostering collaboration with technology experts emerge as pivotal elements for leveraging these advanced models effectively in the delivery of smart healthcare.

The integration of vision transformer with traditional models like VGG19, ResNet50, DenseNet121, and InceptionV3 is examined for its significant improvements in accuracy, precision, recall, and F1-scores. This enhancement has profound implications in healthcare, aiding in informed decision-making, resource optimization, risk mitigation, and strategic planning. While acknowledging the limitations of the current research, the section also suggests future directions, emphasizing the need for continuous improvement and the application of these advanced models in healthcare. This comprehensive analysis not only highlights the technical advancements but also situates them within the broader context of their practical impact and future potential.

Incorporating Vision Transformers (VT) with traditional machine learning models significantly enhances healthcare diagnostics and decision-making. However, a crucial gap in the discussion is the

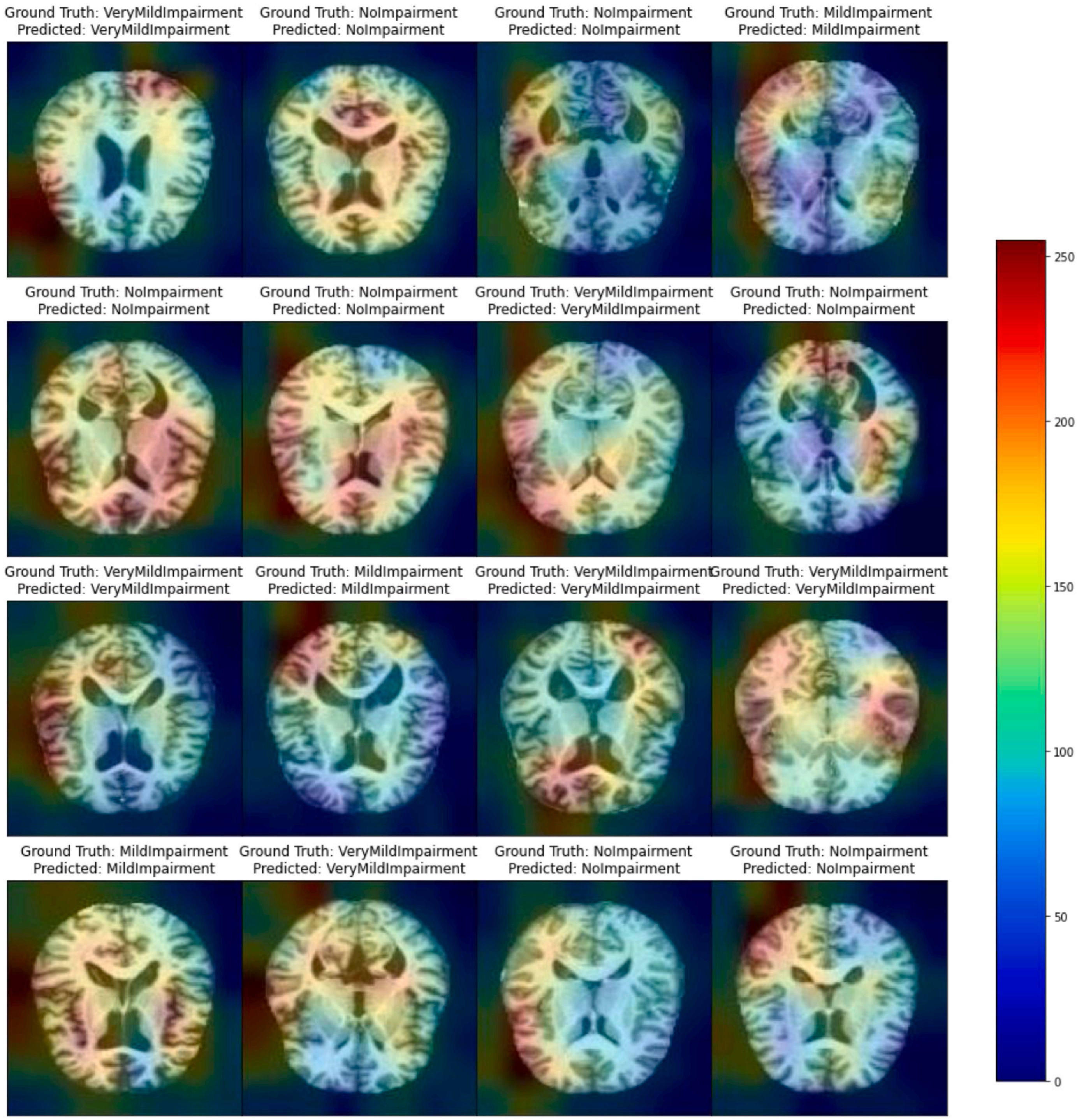

Fig. 10. Comparison of Transfer Learning Model with Ensemble Transfer Learning with Vision transformer Model.

interpretability and explainability of these models. For addressing this, Explainable AI (XAI) techniques can be utilized feature visualization that will enhance the interpretable of the layers of model and focusing on user-centric explainability that can bridge this gap. In future, this work will incorporate a detailed analysis of explainable AI (XAI) techniques, such as Layer-wise Relevance Propagation (LRP), SHAP, and LIME, could provide insights into the decision-making processes of these models.

6. Conclusion

The paper proposes a combined approach involving transfer learning, transformation techniques, and EXAI for diagnosing disease severity. The evaluation is conducted using an Alzheimer's disease dataset, where pre-trained models and brain imaging data transformation improve diagnostic accuracy. EXAI techniques facilitate clinicians in understanding the factors contributing to the diagnosis, leading to better patient care. The proposed approach has the potential to enhance smart healthcare systems by enabling early detection and personalized treatment plans for Alzheimer's disease, resulting in

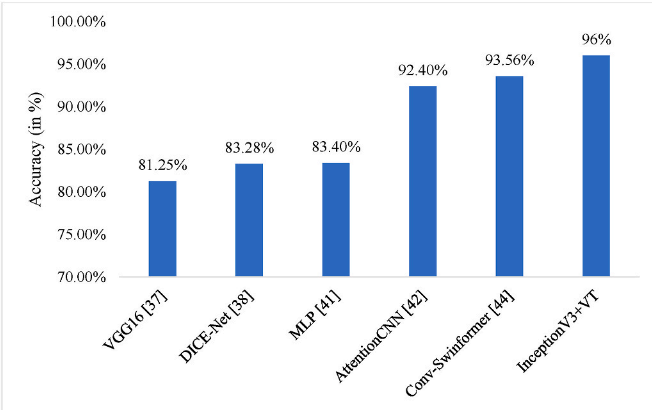

Fig. 11. A comparative state-of-art of diagnostic methodologies.

improved patient outcomes and reduced healthcare costs. Comparative analysis shows that ensemble transfer learning models with vision transformer significantly improve classification performance in terms of accuracy, precision, recall, and F1-score. In summary, the individual TL models have limited discrimination and struggle to capture complex patterns. However, the ensemble transfer learning models with vision transformer show significant improvements as compared to individual models. This indicates that combining multiple models and incorporating vision transformers enhances classification performance. Therefore, it can be inferred that using ensemble transfer learning models with vision transformer can lead to improved classification performance compared to individual models or traditional transfer learning approaches. In future, this work will be extended with other cyber-physical systems for smart healthcare applications. Future work will combine multiple disease diagnosis in single platform for real-time investigation.

Funding

Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-RP23114).

Data availability statement

The data presented in this study are available upon request from the corresponding author.

CRediT authorship contribution statement

Ramesh Chandra Poonia: Writing – review & editing, Writing – original draft, Validation, Methodology, Conceptualization. **Halah A. Al-Alshaikh:** Writing – review & editing, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (grant number IMSIU-RP23114).

References

[1] K.T. Chui, W. Alhalabi, S.S.H. Pang, P.O. de Pablos, R.W. Liu, M. Zhao, Disease diagnosis in smart healthcare: innovation, technologies and applications, Sustain. Times 9 (12) (Dec. 2017) 2309, https://doi.org/10.3390/SU9122309.

[2] S. Tian, W. Yang, J.M. Le Grange, P. Wang, W. Huang, Z. Ye, Smart healthcare: making medical care more intelligent, Glob. Health J. 3 (3) (Sep. 2019) 62–65, https://doi.org/10.1016/J.GLOHJ.2019.07.001.

[3] A. Ray, A.K. Chaudhuri, Smart healthcare disease diagnosis and patient management: innovation, improvement and skill development, Mach. Learn. with Appl. 3 (Mar. 2021) 100011, https://doi.org/10.1016/J.MLWA.2020.100011.

[4] R.F. Mansour, A. El Amraoui, I. Nouaouri, V.G. Diaz, D. Gupta, S. Kumar, Artificial intelligence and internet of things enabled disease diagnosis model for smart healthcare systems, IEEE Access 9 (2021) 45137–45146, https://doi.org/10.1109/ACCESS.2021.3066365.

[5] M. Nasr, M.M. Islam, S. Shehata, F. Karray, Y. Quintana, Smart healthcare in the age of AI: recent advances, challenges, and future prospects, IEEE Access 9 (2021) 145248–145270, https://doi.org/10.1109/ACCESS.2021.3118960.

[6] M. Sony, J. Antony, G.L. Tortorella, Critical success factors for successful implementation of healthcare 4.0: a literature review and future research agenda, Int. J. Environ. Res. Publ. Health 20 (5) (Mar. 2023), https://doi.org/10.3390/IJERPH20054669.

[7] Y. Zhang, Y. Weng, J. Lund, Applications of explainable artificial intelligence in diagnosis and surgery, Diagnostics 12 (2) (Feb. 2022), https://doi.org/10.3390/DIAGNOSTICS12020237.

[8] D. Dave, H. Naik, S. Singhal, P. Patel, Explainable AI meets healthcare: a study on heart disease dataset [Online]. Available: https://arxiv.org/abs/2011.03195v1, Nov. 2020. (Accessed 22 May 2023).

[9] Q. Ye, J. Xia, G. Yang, Explainable AI for COVID-19 CT classifiers: an initial comparison study, Proc. - IEEE Symp. Comput. Med. Syst. (Apr. 2021) 521–526, https://doi.org/10.1109/CBMS52027.2021.00103, 2021-June.

[10] Y. Otaki, et al., Clinical deployment of explainable artificial intelligence of SPECT for diagnosis of coronary artery disease, JACC Cardiovasc. Imaging 15 (6) (Jun. 2022) 1091–1102, https://doi.org/10.1016/J.JCMG.2021.04.030.

[11] A. Anand, T. Kadian, M.K. Shetty, A. Gupta, Explainable AI decision model for ECG data of cardiac disorders, Biomed. Signal Process Control 75 (May 2022) 103584, https://doi.org/10.1016/J.BSPC.2022.103584.

[12] P. Bourdon, O. Ben Ahmed, T. Urruty, K. Djemal, C. Fernandez-Maloigne, Explainable AI for Medical Imaging: Knowledge Matters, Multi-faceted Deep Learn, Feb. 2021, pp. 267–292, https://doi.org/10.1007/978-3-030-74478-6_11.

[13] G. Antoniou, E. Papadakis, G. Baryannis, Mental health diagnosis: a case for explainable artificial intelligence 31 (3) (May 2022), https://doi.org/10.1142/S0218213022410032.

[14] J. Peng, et al., An explainable artificial intelligence framework for the deterioration risk prediction of hepatitis patients, J. Med. Syst. 45 (5) (May 2021), https://doi.org/10.1007/S10916-021-01736-5.

[15] Y. Zhang, Y. Weng, J. Lund, Applications of explainable artificial intelligence in diagnosis and surgery, Diagnostics 12 (2) (Feb. 2022), https://doi.org/10.3390/DIAGNOSTICS12020237.

[16] Shaik Basheera, M. Satya Sai Ram, Convolution neural network–based Alzheimer's disease classification using hybrid enhanced independent component analysis based segmented gray matter of T2 weighted magnetic resonance imaging with clinical valuation, Alzheimer's Dementia: Translational Research & Clinical Interventions 5 (2019) 974–986.

[17] R.S. Rosenson, et al., Dysfunctional HDL and atherosclerotic cardiovascular disease, Nat. Rev. Cardiol. 13 (1) (Jan. 2016) 48–60, https://doi.org/10.1038/NRCARDIO.2015.124.

[18] S.J. Al'Aref, et al., Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging, Eur. Heart J. 40 (24) (Jun. 2019) 1975–1986, https://doi.org/10.1093/EURHEARTJ/EHY404.

[19] G. Joo, Y. Song, H. Im, J. Park, Clinical implication of machine learning in predicting the occurrence of cardiovascular disease using big data (Nationwide Cohort Data in Korea), IEEE Access 8 (2020) 157643–157653, https://doi.org/10.1109/ACCESS.2020.3015757.

[20] D. Dey, et al., Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review, J. Am. Coll. Cardiol. 73 (11) (Mar. 2019) 1317–1335, https://doi.org/10.1016/J.JACC.2018.12.054.

[21] S. Huang, J. Yang, S. Fong, Q. Zhao, Artificial intelligence in cancer diagnosis and prognosis: opportunities and challenges, Cancer Lett. 471 (Feb. 2020) 61–71, https://doi.org/10.1016/J.CANLET.2019.12.007.

[22] J. Xu, et al., Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives, Hum. Genet. 138 (2) (Feb. 2019) 109–124, https://doi.org/10.1007/S00439-019-01970-5.

[23] W.L. Bi, et al., Artificial intelligence in cancer imaging: clinical challenges and applications, CA A Cancer J. Clin. 69 (2) (Feb. 2019), https://doi.org/10.3322/CAAC.21552.

[24] O. Adir, et al., Integrating artificial intelligence and nanotechnology for precision cancer medicine, Adv. Mater. 32 (13) (Apr. 2020), https://doi.org/10.1002/ADMA.201901989.

[25] M. Coccia, Deep learning technology for improving cancer care in society: new directions in cancer imaging driven by artificial intelligence, Technol. Soc. 60 (Feb. 2020) 101198, https://doi.org/10.1016/J.TECHSOC.2019.101198.

[26] S. Vieira, W.H.L. Pinaya, A. Mechelli, Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications, Neurosci. Biobehav. Rev. 74 (Pt A) (Mar. 2017) 58–75, https://doi.org/10.1016/J.NEUBIOREV.2017.01.002.

[27] L. Zhang, Y. Li, L. Bian, Q. Luo, X. Zhang, B. Zhao, Cognitive impairment of patient with neurological cerebrovascular disease using the artificial intelligence technology guided by MRI, Front. Public Health 9 (Mar. 2022) 2175, https://doi.org/10.3389/FPUBH.2021.813641/BIBTEX.

[28] C. Surianarayanan, J.J. Lawrence, P.R. Chelliah, E. Prakash, C. Hewage, Convergence of artificial intelligence and neuroscience towards the diagnosis of neurological disorders—a scoping review, Sensors 23 (6) (2023) 3062, https://doi.org/10.3390/S23063062.

[29] S. El-Sappagh, J.M. Alonso, S.M.R. Islam, A.M. Sultan, K.S. Kwak, A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease, 111, Sci. Rep. 11 (1) (2021) 1–26, https://doi.org/10.1038/s41598-021-82098-3.

[30] K.M. Sudar, P. Nagaraj, S. Nithisaa, R. Aishwarya, M. Aakash, S.I. Lakshmi, Alzheimer's disease analysis using explainable artificial intelligence (XAI), Int. Conf. Sustain. Comput. Data Commun. Syst. ICSCDS 2022 - Proc. (2022) 419–423, https://doi.org/10.1109/ICSCDS53736.2022.9760858.

[31] A. Lombardi, et al., A robust framework to investigate the reliability and stability of explainable artificial intelligence markers of Mild Cognitive Impairment and Alzheimer's Disease, Brain informatics 9 (1) (Dec. 2022), https://doi.org/10.1186/S40708-022-00165-5.

[32] V. Jain, O. Nankar, D.J. Jerrish, S. Gite, S. Patil, K. Kotecha, A novel AI-based system for detection and severity prediction of dementia using MRI, IEEE Access 9 (2021) 154324–154346, https://doi.org/10.1109/ACCESS.2021.3127394.

[33] B. Tuvshinjargal, H. Hwang, VGG-C transform model with batch normalization to predict Alzheimer's disease through MRI dataset, Electron 11 (16) (2022) 2601, https://doi.org/10.3390/ELECTRONICS11162601.

[34] Mohi ud din dar, Gowhar, et al., A novel framework for classification of different Alzheimer's disease stages using CNN model, Electronics 12 (2) (2023) 469.

[35] G.S. Nijaguna, J.A. Babu, B.D. Parameshachari, R.P. de Prado, J. Frnda, Quantum Fruit fly algorithm and ResNet50-VGG16 for medical diagnosis, Appl. Soft Comput. 136 (Mar. 2023) 110055, https://doi.org/10.1016/J.ASOC.2023.110055.

[36] M. Chhabra, R. Kumar, A smart healthcare system based on classifier DenseNet 121 model to detect multiple diseases, Lect. Notes Networks Syst. 339 (2022) 297–312, https://doi.org/10.1007/978-981-16-7018-3_23/COVER.

[37] Modupe Odusami, Rytis Maskeliūnas, Robertas Damaševičius, Pixel-level fusion approach with vision transformer for early detection of Alzheimer's disease, Electronics 12 (5) (2023) 1218.

[38] A. Miltiadous, E. Gionanidis, K.D. Tzimourta, N. Giannakeas, A.T. Tzallas, DICE-net: a novel convolution-transformer architecture for alzheimer detection in EEG signals, IEEE Access 11 (2023) 71840–71858, https://doi.org/10.1109/ACCESS.2023.3294618.

[39] R.C. Petersen, et al., Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization, Neurology 74 (3) (2010) 201–209, https://doi.org/10.1212/WNL.0B013E3181CB3E25.

[40] Z. Zhou, X. Yang, J. Ji, Y. Wang, Z. Zhu, Classifying fabric defects with evolving Inception v3 by improved L2,1-norm regularized extreme learning machine 93 (3–4) (Jul. 2022) 936–956, https://doi.org/10.1177/00405175221114633.

[41] S. Qiu, P.S. Joshi, M.I. Miller, C. Xue, X. Zhou, C. Karjadi, G.H. Chang, A.S. Joshi, B. Dwyer, S. Zhu, M. Kaku, Development and validation of an interpretable deep learning framework for Alzheimer's disease classification, Brain 143 (6) (2020) 1920–1933.

[42] Wenyong Zhu, et al., Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI, IEEE Trans. Med. Imag. 40 (9) (2021) 2354–2366.

[43] Z. Xia, X. Pan, S. Song, L.E. Li, G. Huang, Vision transformer with deformable attention, IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn. (Jan. 2022) 4784–4793, https://doi.org/10.1109/CVPR52688.2022.00475, 2022-June.

[44] Zhentao Hu, et al., Conv-Swinformer: integration of CNN and shift window attention for Alzheimer's disease classification, Comput. Biol. Med. 164 (2023) 107304.

[45] Ramprasaath R. Selvaraju, et al., Grad-cam: visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE international conference on computer vision, 128, 2017, pp. 336–359.

[46] Jaekeol Choi, Jungin Choi, Wonjong Rhee, Interpreting neural ranking models using grad-cam, arXiv preprint arXiv:2005.05768 (2020).