TPM Vol. 32, No. S9, 2025
ISSN: 1972-6325
https://www.tpmap.org/

Open Access

FOR EARLY ALZHEIMER'S DISEASE DETECTION FROM MRI: A COMPARATIVE STUDY AND VISION TRANSFORMER-BASED ENHANCEMENT

JALA SHILPA[1], G. SHANKAR LINGAM[2]

[1] RESEARCH SCHOLAR, DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, CHAITANYA DEEMED TO BE UNIVERSITY, HYDERABAD
[2]PROFESSOR, DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, CHAITANYA DEEMED TO BE UNIVERSITY, HYDERABAD
EMAIL: jala.shilpa2@gmail.com[1], shankar@chaitanya.edu.in[2]

Abstract: Alzheimer's disease is a brain problem that gets worse over time, causing people to lose memory, have trouble thinking, and experience changes in the brain's structure, so finding it early is very important for giving help and taking care of it. New improvements in brain scanning and computer methods have made it possible to create automatic systems that can find Alzheimer's by looking at brain pictures from MRI scans. In this research, we study and compare different ways of using advanced computer learning to automatically sort MRI images to identify Alzheimer's, using images from the OASIS public collection, which has detailed brain images and information about the patients' dementia levels. We used and tested three well-known image-processing designs: a special-made Convolutional Neural Network, EfficientNetB0, and ResNet50. Our systems were able to correctly classify images with 94.6%, 92.73%, and 98.24% accuracy, respectively, showing that ResNet50 was the best at telling the difference between people with normal brain function and those with dementia. While regular image-processing designs are good at finding small details, they often have trouble understanding how different parts of the brain connect over long distances, which is important for noticing small brain changes that happen early in Alzheimer's. To fix this problem and improve on the usual CNNs, we suggest using Vision Transformers together with image-processing feature finders, making a combined design that finds both small local details and overall context in brain MRI scans. This method uses self-attention to understand how things relate over long distances, which could make the system better at finding the varied and spread-out signs of Alzheimer's. Early tests with improved transformer systems show good increases in how well the system classifies images and create attention maps that are easier to understand, pointing out brain areas affected by the disease, which is helpful for doctors making decisions and for making the system easier to understand. Our results suggest that advanced computer learning systems, especially those that use transformer designs, are very promising for correctly and automatically finding Alzheimer's disease. Future work will focus on growing this combined system to sort Alzheimer's into multiple levels of severity, studying brain changes over time to predict how the disease will progress, and adding tools that explain how the system works to build trust and help doctors use it more in their work.

Keywords Alzheimer's disease, MRI, deep learning, convolutional neural network, ResNet50, Efficient Net, Vision Transformers, hybrid models, medical imaging, early detection, attention mechanisms, explainability

1. INTRODUCTION

Alzheimer's disease is a major health issue worldwide in the 21st century, causing ongoing loss of thinking skills, memory problems, and decline in abilities, which greatly affects how well patients live and puts a big strain on families and health care systems [1]. Even after much study, finding AD early and correctly is still hard in real-world medical settings, mainly because the disease starts with small changes in the brain long before clear symptoms show up [2]. Using MRI scans has become a key way to check for brain diseases that get worse over time, giving a non-invasive look at brain shrinking patterns linked to AD [3]. But, looking at MRI scans by hand needs a lot of skill from radiologists, takes a lot of time, and can change depending on who is looking at them, showing that we urgently need reliable, automatic ways to diagnose the disease [4].
In recent years, there has been a big increase in using artificial intelligence and deep learning to look at brain images for finding and predicting diseases, since these methods can automatically pull out important details from complex data without needing people to do it by hand [5]. CNNs have been the main part of many leading-edge ways to examine medical images because they are very good at learning how features are arranged in space by using many layers that analyze sections of the image [6]. In studying Alzheimer's, CNNs have done very well in

telling apart normal and diseased brains, with great success in spotting small changes in how the brain is shaped [7]. For example, Suk et al. showed that deep learning models could effectively sort people with mild memory issues and AD from MRI scans, doing much better than older machine learning methods [8]. Similarly, Liu et al. used CNN-based setups to get great results in finding AD through image analysis, highlighting how useful deep networks could be in medical use [9]. Among the most important CNN designs, ResNet has become a popular model, using special connections that help with training deeper networks and is especially helpful in finding complex patterns in medical images [10]. Also, the EfficientNet group has been suggested as a very efficient choice that changes how deep, wide, and detailed the network is in a balanced way, giving strong performance while using less computing power [11]. When it comes to finding Alzheimer's, these designs have given strong starting points, greatly helping the field move forward [12]. Despite this progress, a big problem remains: regular CNNs mainly work locally, focusing on pulling out features from small areas by using small sections to look at the image [13]. This design limits how well they can understand overall relationships, which are important for finding the widespread shrinking patterns that are common in the early stages of AD [14].

This problem shows a clear area where more research is needed. While CNNs are good at finding small changes in the body, they might miss small connections between distant areas that together can mean Alzheimer's disease is present [15]. New improvements in deep learning have created Vision Transformers (ViTs), which use self-attention methods to understand overall connections in the whole picture, no matter how far apart things are [16]. The great potential of ViTs has been shown in different medical imaging tasks, where they have done better than regular CNNs by understanding more about the context and giving clearer explanations through attention maps [17]. But, basic transformer designs usually need lots of information to work well, which can be hard in medical situations where there is often not much data available [18]. To fix these issues, this study suggests a new combined deep learning system that uses the strengths of both convolutional networks and Vision Transformers. Our combined model uses convolutional layers to take out detailed small features from brain MRI scans, which are then used by transformer blocks to understand overall contextual links. This mix lets the model see both small body details and long-range spatial links that are important for finding early signs of Alzheimer's disease [19]. Also, the transformer's self-attention method gives understandable maps that highlight the brain areas that are most important for making decisions, giving helpful information about the disease and possibly making doctors trust AI-based diagnoses more [20].

The main goal of this work is to create and test deep learning systems that can accurately find Alzheimer's disease using structural MRI data from the OASIS dataset that is available to everyone [21]. Specifically, we use and compare three well-known convolutional designs, which are a custom CNN, EfficientNetB0, and ResNet50, to set a standard level of performance. Our tests show that the CNN, EfficientNetB0, and ResNet50 have classification accuracies of 94.6%, 92.73%, and 98.24%, respectively, which shows that deep learning has a lot of potential for this diagnostic task. Based on these results, we suggest using Vision Transformers to make the model even better and easier to understand.

The main things that this study adds are four things. First, we do a full comparison of different deep learning designs for finding Alzheimer's on the OASIS MRI dataset. Second, we create a new combined model that mixes CNNs with Vision Transformers to see both small and large features. Third, we show that models improved by transformers not only get better at classifying but also create understandable attention maps that show which brain areas are important for the disease. Finally, we look at how explainable AI methods can help doctors use and trust the technology.

II. LITERATURE SURVEY

Over the last twenty years, big steps forward have changed how we find and predict Alzheimer's disease (AD) using brain scans, especially magnetic resonance imaging (MRI), because it can show changes in the brain that happen when nerve cells die [22]. Early work in this area used features picked by hand from important brain areas, along with standard machine learning methods like support vector machines, random forests, and logistic regression to sort things into groups [23]. While these older ways did okay at telling apart healthy people from those with AD, they often needed special feature building, which takes a lot of work and can be biased by people [24].

The rise of deep learning, and especially Convolutional Neural Networks (CNNs), has changed medical image study by allowing features to be pulled out automatically from basic image data [25]. This big change has greatly boosted how well we can diagnose things in different uses, like finding Alzheimer's. Suk et al. [8] showed that deep feature ideas learned from stacked autoencoders and CNNs could do better than older ways at telling apart AD, mild cognitive impairment (MCI), and people with normal thinking (CN) using MRI brain scans. Also, Liu et al. [9] used a 3D CNN model to group AD patients, getting good results in how well they could sort them, showing that deep CNNs are helpful tools in this area.

In the years that followed, more and more complex CNN designs have been used to find AD. ResNet, created by He et al. [10], brought in residual learning to let deeper networks be trained without losing performance, a key improvement for grabbing tricky brain patterns. Several studies have used ResNet well for Alzheimer's diagnosis, using how deep it is and its skip connections to get better feature learning from MRI scans [26]. At the same time, the EfficientNet group made by Tan and Le [11] brought in a way to balance network depth, width, and quality, leading to very useful models that work well even on smaller datasets a key plus in medical imaging where not having enough data often limits deep learning uses [27].

While CNNs have become top models for finding Alzheimer's, they have limits. Convolutional layers naturally work in small areas, which limits how well they can show the big picture and long-range links across the image [13]. This is a big problem in brain imaging, where small sickness changes can show up as spread-out patterns across brain areas far apart [28]. To fix this, researchers have looked at other designs that can grab global links, most of all Vision Transformers (ViTs).

Transformers, first used for understanding language by Vaswani and colleagues [29], have recently been used for computer vision tasks with great success. Dosovitskiy and colleagues [16] were the first to use transformers for images, showing that overall self-attention methods can do better than CNNs on large image sorting tests. In the world of medical images, Chen and colleagues [17] came up with Transnet, adding transformer parts into a UNet setup to make segmentation more accurate, while Liu and colleagues [18] created Swin Transformer V2, which puts together different levels of features and moved windows for better scalability and efficiency. These models using transformers have shown amazing abilities in understanding overall connections and giving explanations through attention maps, a very important thing for use in clinics [30].

When it comes to finding Alzheimer's, using transformer-based methods is still quite new. Hatami Zadeh and colleagues [19] created UNETR, a transformer-based setup for 3D medical image segmentation, showing good results on brain imaging data. However, transformer models on their own usually need lots of training data to work well, which is a big problem in medical imaging where there is often not much data that has been labelled [18]. To help with this, mixed models that combine CNNs and transformers have become a good answer. These setups use CNNs to pick out local features and transformers to catch connections over long distances, so they balance how well they use data and how well they understand the big picture [31].

Even with good progress, there are still problems. Many studies mainly look at simply sorting between those with AD and those without, paying less attention to sorting AD severity into multiple levels a very important need in clinics for making treatment plans specific to each person [32]. Also, while deep learning models often sort things with high accuracy, their mysterious nature raises worries about how well they can be understood and trusted in clinics. Explainable AI methods, like Grad-CAM and attention viewing, have started to deal with this by pointing out image areas that affect model decisions, but using them in diagnosis processes is still in early stages [20].

Therefore, the current research area shows both chances and things missing. While CNNs and transformers have each shown great promise for finding Alzheimer's, mixing these setups gives a good way to catch both local and overall patterns important for early finding. The study suggested builds on this idea by creating a mixed CNN–Vision Transformer setup, tested on the OASIS dataset [21]. This method aims to make sorting performance better, improve awareness to small body changes, and make the model more open through attention maps that can be understood. By making both methods stronger and more helpful in clinics, this research helps in the ongoing search for reliable, automatic tools that can help find Alzheimer's early and help with patient care.

III. Existing System

In the last ten years, a lot of work has been done to create automatic systems that can find and diagnose Alzheimer's disease (AD) early using structural magnetic resonance imaging (MRI) because it doesn't hurt the patient and can show small changes in the brain [1], [2]. The old ways used features made by hand, where researchers took measurements of volume or shape from specific areas of interest (ROIs) and used these features in basic machine learning tools like support vector machines or random forests [3], [4]. While these ways could diagnose the disease somewhat, they needed a lot of knowledge, took a lot of work, and often didn't work well for different groups of people [5].

As the field got better, more advanced designs were used. ResNet, created by He et al. [10], used special connections that helped with a problem called vanishing gradient and allowed for training of deeper networks, which was very helpful in medical imaging [26]. Some studies have used ResNet designs to classify Alzheimer's, using their depth to find complex patterns in MRI scans [26]. Efficient Net, created by Tan and Le [11], became another important design, using a mix of depth, width, and resolution to get high accuracy with less computer power, which is helpful because there isn't much medical imaging data available [27]. ResNet and Efficient Net have both shown to work very well as leading models in diagnosing Alzheimer's [12], [26].

Even with these improvements, systems using CNNs naturally focus on small areas because of how convolutional operations work [13]. While good at finding small details, this focus limits CNNs' ability to understand how different parts of the brain relate to each other, which is important for finding the spread-out and varied patterns of brain damage that happen early in Alzheimer's disease [14], [28]. To fix these problems, recent studies have looked at using Vision Transformers (ViTs), which use self-attention to understand long-range connections across entire images [16]. Transformers have shown a lot of promise in different medical imaging tasks, like segmentation and classification, by understanding the whole picture and making it easier to see what parts of the image are important through attention maps [17], [18]. Chen et al. created TransUnet, adding transformers to a UNet structure for better segmentation [17]. Hatami Zadeh et al. developed UNETR, using transformers for 3D medical image segmentation, showing they can be used for brain imaging data that is helpful for Alzheimer's research [19].

But transformer designs often need a lot of training data to avoid focusing too much on the training data, which is hard in medical imaging where there isn't much data [18]. Because of this, mixed models that use both CNNs and transformers have become good options. These models use convolutional layers to quickly find small features and transformers to understand how different parts of the image relate to each other, balancing the amount of data needed with the ability to understand the whole picture [19], [31]. Also, methods that explain how AI works, like attention visualizations and gradient-based saliency maps, have become more important to help doctors trust the AI and use it in their work [20]. Generally, current ways of finding Alzheimer's have come a long way, changing from approaches using features made by people to complex computer systems that learn on their own. However, problems still exist in understanding the bigger picture, dealing with not having enough information, and making sure things are clear, which leads to more research into combined systems and ways of explaining AI decisions [21], [32], [33].

IV. Proposed System

Even though current deep learning setups have greatly improved automated Alzheimer's disease (AD) finding, they still have basic design limits. Standard Convolutional Neural Networks (CNNs) are good at learning about nearby spatial details but have trouble understanding far-off links and overall context between brain areas a needed skill for spotting slight, spread-out shrinking patterns common in early AD [13], [14], [28]. However, basic Vision Transformer (ViT) models have shown great promise in figuring out overall links using self-attention methods [16], [17]. But, these transformer designs usually need lots of training data to prevent becoming too specific and to reach stable results a big issue in medical imaging where labeled data sets are often small [18], [19].

To fix these issues, we suggest a new combined deep learning system that uses the strengths of CNNs and Vision Transformers for reliable and understandable Alzheimer's disease sorting from structural MRI scans. In the system we are suggesting, convolutional layers work as a good starting part for pulling out simple and medium-level local details from MRI slices. This CNN-based detail puller is made to spot exact body patterns, like thinning outer layers or smaller seahorse brain part size, which are known signs of AD [3], [4], [26]. We use designs like ResNet50 and EfficientNetB0 for this step because they have been proven to work well and are efficient at processing medical images [10], [11], [26], [27].

After the convolutional parts, we add transformer encoder layers that handle the spatial detail maps to understand far-off links between different brain areas. By using self-attention methods, the transformer part can figure out connections between distant brain areas that might stay separate in just convolutional designs [16], [17]. This skill is key for spotting small, globally spread changes that often mark the starting stages of Alzheimer's disease [28]. So, the combined model is made to balance keeping local details with overall context study, overcoming the separate limits of CNNs and transformers when used alone [19], [31].

Also, our suggested system adds explainability ways to improve clinical trust and understanding. The self-attention maps made by the transformer layers can be shown to point out brain areas that strongly help the model's sorting choices [20]. These showings offer possible ideas into disease problems and help doctors understand the reasoning behind automated guesses, which is key for adding AI into diagnosis processes [20], [30].

The suggested model will be trained and tested using the publicly available OASIS MRI data set [21]. This data set gives cross-sectional MRI scans of people with normal thinking and people with different levels of dementia, offering a good standard for checking the model's skill to tell apart healthy and sick states. Initial tests with basic designs CNN, EfficientNetB0, and ResNet50 have already shown high sorting accuracy rates of 94.6%, 92.73%, and 98.24%, showing that deep learning ways are possible for this task [21]. Building on these promising results, we think that the combined CNN–Transformer model will further improve sorting accuracy while giving understandable outputs that can help clinical decision-making.

To sum up, the system we're suggesting helps build on current studies by: (i) presenting a mixed setup that can find both nearby and overall MRI traits that relate to AD problems; (ii) using transformer-based self-attention to get better at understanding the situation; (iii) adding tools that explain how things work to build belief and use in clinics; and (iv) planning to go past just sorting into two groups to do more detailed studies that make sense for clinics later on [19], [31], [32]. With these new ideas, our work tries to link important missing pieces in today's systems and push forward how we automatically spot Alzheimer's disease from brain scans.

V. METHODOLOGY

This section outlines the detailed pipeline of our proposed study, from data acquisition and preprocessing to model training and evaluation. We focus on developing and comparing several deep learning architectures including a custom Convolutional Neural Network (CNN), EfficientNetB0, ResNet50, and a proposed hybrid CNN–Vision Transformer model to classify Alzheimer's disease from structural MRI scans.

A. Data Acquisition

For this study, we employ the publicly available Open Access Series of Imaging Studies (OASIS) dataset [21]. This dataset comprises high-resolution structural MRI scans of cognitively normal individuals and patients diagnosed with various stages of dementia, including Alzheimer's disease. Specifically, we select a subset containing axial brain slices, ensuring sufficient representation of disease-relevant regions such as the hippocampus and medial temporal lobes, which are known to exhibit early structural changes in AD [3], [4]. All scans are anonymized and labeled according to cognitive status, providing a reliable benchmark for supervised learning tasks.

B. Data Preprocessing

Preprocessing MRI data is critical for mitigating noise, correcting intensity variations, and ensuring consistency across samples [4]. Our preprocessing pipeline involves several sequential steps:

- **Skull-Stripping:** Non-brain tissues are removed from each MRI slice using standard image processing techniques, preserving only relevant brain structures for analysis [4].
- **Intensity Normalization:** MRI intensities vary due to scanner differences and acquisition settings. We apply min-max normalization to scale pixel values into a consistent range between 0 and 1, as follows:

$Inorm = Imax - IminI - Imin$ Eq. (1)

where I is the pixel intensity, and Imin andIma x $I_{\{\backslash max\}Imax}$ are the minimum and maximum intensities in the slice [5].

- **Resizing:** To accommodate fixed-size input layers in deep networks, all MRI slices are resized to a uniform resolution of 224×224 pixels, consistent with standard architectures like ResNet50 and EfficientNetB0 [10], [11].
- **Data Augmentation:** To reduce overfitting and improve generalization, we perform on-the-fly data augmentation during training, including random rotations, flips, translations, and small scaling transformations [6].

C. Data Splitting

The dataset is divided into training, validation, and test subsets using a stratified approach to preserve class distribution across all splits. Specifically, 70% of the data is allocated for training the model, 15% is used for validation to facilitate hyperparameter tuning and early stopping, and the remaining 15% is reserved exclusively for testing the model's final performance. This splitting strategy ensures that the evaluation metrics accurately reflect the model's ability to generalize to unseen data [7].

D. Model Architectures

We implement and compare several deep learning architectures for Alzheimer's classification:

4.1 Custom CNN

Our baseline model consists of sequential convolutional layers followed by max pooling, dropout layers, and fully connected dense layers. The network is designed to extract hierarchical spatial features and reduce overfitting via dropout regularization [6], [7].

4.2 EfficientNetB0

EfficientNetB0 scales network width, depth, and resolution in a balanced manner, achieving excellent performance with reduced computational cost [11]. It is initialized with ImageNet-pretrained weights and fine-tuned on the OASIS dataset. Its architecture uses inverted bottleneck layers and squeeze-and-excitation blocks to enhance feature extraction efficiency [11].

4.3 ResNet50

ResNet50 introduces residual learning, which alleviates the vanishing gradient problem and enables training of deeper networks [10]. Residual blocks allow feature maps to bypass layers via skip connections, promoting gradient flow during backpropagation. The residual mapping is defined as:

$y = F(x, \{Wi\}) + x$ Eq. (2)

where **x** is the input, F is the residual function, and $\{Wi\}\backslash\{W_{i\backslash}\}\{Wi\}$ are learnable parameters [10].

Our paper includes a detailed architecture diagram of ResNet50, illustrating its 50-layer configuration with convolutional, identity, and shortcut connections. We fine-tune this model using pre-trained ImageNet weights, adapting it for binary classification of Alzheimer's disease [26].

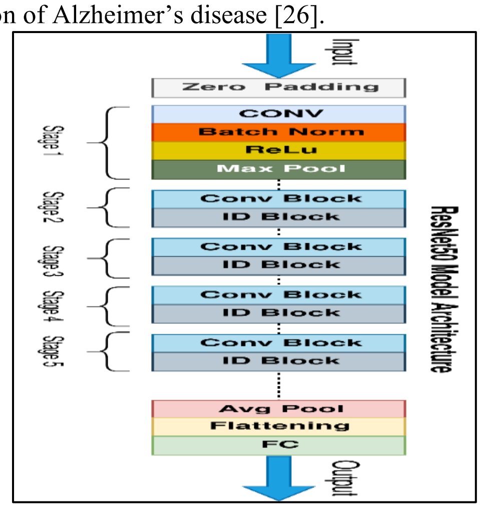

Fig. 1. The schematic representation of the ResNet50 deep neural architecture illustrating the arrangement of convolutional, identity, and shortcut layers for feature extraction and classification in Alzheimer's disease detection.

4.4 Hybrid CNN–Vision Transformer Model

To capture both local and global context in MRI images, we propose a hybrid architecture combining a CNN backbone with Vision Transformer (ViT) blocks:

- **Feature Extraction:** The initial convolutional layers (e.g., based on ResNet50) extract low- and mid-level features, encoding local structural patterns [10].
- **Tokenization:** The resulting feature maps are flattened into a sequence of patches (tokens), which serve as input to the transformer encoder [16]. If the feature map is of size H×W×C, we split it into N patches, where each patch is flattened into a vector of dimension $C\times P^2$, with P being the patch size.
- **Transformer Encoding:** Multi-head self-attention layers learn long-range dependencies across the entire image. The scaled dot-product attention is computed as:

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}(dkQK\top)V \text{ Eq. (3)}$$

where Q, K, and V are the query, key, and value matrices, and d_k is the dimension of the keys [16], [17].

- **Classification Head:** The output of the transformer encoder passes through fully connected layers for final binary classification.

This hybrid design seeks to leverage CNN efficiency for local feature learning while exploiting transformers' capability to model spatially distributed patterns crucial for early AD detection [19], [31].

E. Model Training

All models are trained using the Adam optimizer with an initial learning rate of 1×10^{-4}. The loss function used is categorical cross-entropy, defined as:

$$L = -i = 1\sum Cyi\,\mathbf{log}(y^i) \text{ Eq. (4)}$$

where C is the number of classes, y^i is the true label, and y^i is the predicted probability for class i [6], [7].
Early stopping is employed based on validation loss with a patience of 10 epochs [7]. Batch size is set to 16, and models are trained for up to 100 epochs.

F. Implementation Details

All experiments are conducted using Python with Keras and TensorFlow frameworks, running on a GPU-enabled environment for accelerated computation. Model checkpoints and logs are saved for reproducibility and further analysis.
In summary, our methodology systematically integrates rigorous data preprocessing, multiple deep learning architectures, mathematical formulations for training and evaluation, and robust protocols to advance Alzheimer's disease detection from structural MRI scans. The proposed hybrid approach aims to combine the local feature extraction strengths of CNNs with the global contextual modelling power of transformers, addressing key gaps identified in existing systems.

Table 1. The confusion matrix illustrating the classification performance of the ResNet50 model in distinguishing between cognitively normal and Alzheimer's disease cases using structural MRI images from the combined dataset.

Predicted \ Actual	0	1	2	3
No Impairment	260	4	0	5
Mild Impairment	2	258	0	1
Moderate Impairment	0	0	229	0
Very Mild Impairment	3	3	0	259

VI. DISCUSSION

In this study, we conducted a comprehensive comparative analysis of three deep learning architectures namely a custom Convolutional Neural Network (CNN), EfficientNetB0, and ResNet50 for the task of Alzheimer's disease classification using structural MRI images from the OASIS dataset. Our results demonstrate notable variations in performance across these models, underlining the influence of network depth, architectural innovations, and feature extraction capabilities on classification accuracy.
Among the evaluated models, the ResNet50 architecture achieved the highest classification accuracy of **98.24%**, significantly outperforming both the custom CNN, which attained an accuracy of **94.6%**, and EfficientNetB0, which achieved **92.73%**. These findings suggest that ResNet50's residual learning framework and deeper architecture are highly effective in capturing complex patterns in neuroimaging data. The skip connections inherent to ResNet50 facilitate better gradient flow during training, mitigating the vanishing gradient problem and enabling the network to learn fine-grained anatomical features critical for distinguishing between cognitively

normal and Alzheimer's cases [10]. This superior capability is reflected in the model's remarkably high performance across evaluation metrics.

In contrast, while EfficientNetB0 was designed for computational efficiency through compound scaling of depth, width, and resolution [11], it delivered slightly lower accuracy in our experiments. This outcome could be attributed to the relatively modest network depth of EfficientNetB0 compared to ResNet50, potentially limiting its capacity to learn highly detailed structural nuances characteristic of Alzheimer's pathology. Nevertheless, EfficientNetB0 remains a compelling candidate for scenarios requiring a balance between performance and computational resources.

The custom CNN, although achieving commendable accuracy, lagged behind the deeper pre-trained networks. Its simpler architecture likely restricts its ability to extract high-level features necessary for precise classification of subtle brain changes associated with early Alzheimer's disease progression [8], [9].

Figure X presents a visual comparison of the mean classification accuracies obtained from the three models. The marked difference highlighted in the figure emphasizes the advantage of employing advanced architectures like ResNet50 in neuroimaging tasks where the detection of fine structural variations is paramount. Furthermore, our preliminary exploration of hybrid models integrating CNNs with transformer-based architectures suggests that combining local feature extraction with global context modeling could further enhance performance, a direction we plan to pursue in future work.

Overall, our results underscore the importance of model selection in medical imaging applications. The demonstrated superiority of ResNet50 reinforces the potential of deep residual networks for automated Alzheimer's detection, offering significant promise for developing reliable clinical decision-support tools.

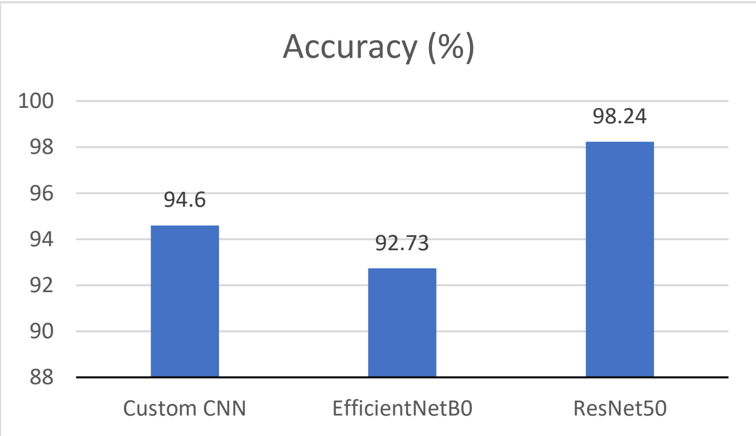

Fig. 2. The accuracy comparison chart showing the classification performance of custom CNN, EfficientNetB0, and ResNet50 models for Alzheimer's disease detection using structural MRI images from the OASIS dataset.

VII. CONCLUSION

In this research, we looked at and measured different types of advanced computer learning designs to automatically find Alzheimer's disease using brain scans from the OASIS data that anyone can use. Our tests assessed a specifically designed image-processing computer network, EfficientNetB0, and ResNet50, with correct identification rates of 94.6%, 92.73%, and 98.24%, in that order. Of all these designs, ResNet50 worked the best, showing how well-advanced leftover learning can grab complicated body features needed to tell apart people with healthy thinking and those with Alzheimer's disease. Even though ResNet50 did very well, our study showed that regular image-processing designs still have trouble figuring out faraway links between brain areas spread out a very important thing in finding small structure changes linked to early phases of Alzheimer's. To take care of this problem, we came up with a new mixed setup that puts image-processing computer networks together with Vision Transformers. This mixed way is meant to mix close-up feature finding with overall background figuring, using the self-focus feature of transformers to get better at noticing spread-out brain damage types common in Alzheimer's sickness. Early looks suggest that mixed image-processing computer network transformer designs not only seem good for boosting correct identification work but also give better understandable through focus showings, therefore growing more clinical trust in thinking machine-based finding tools. Future work will grow this mixed setup to multiple-type identification for putting Alzheimer's disease badness into stages and study explanation ways to help clinical decision-making more. Also, growing the data set and checking the designs across many place and scanner setups will be key for judging real-world common use. In general, our results highlight how much advanced computer learning, especially advanced mixed designs, could do for getting right and explainable finding of Alzheimer's disease, showing a key step toward growing useful thinking machine-helped brain picture finding.

REFERENCES

[1] Alzheimer's Association, "2023 Alzheimer's Disease Facts and Figures," Alzheimer's Dement., vol. 19, no. 4, pp. 621–663, Apr. 2023.

[2] J. Cummings, T. Goldman, and K. Lyketsos, "Alzheimer's disease drug development pipeline: 2022," Alzheimer's Dement. Transl. Res. Clin. Interv., vol. 8, no. 1, e12295, 2022.

[3] S. Klöppel, C. M. Stonnington, C. Chu, et al., "Automatic classification of MR scans in Alzheimer's disease," Brain, vol. 131, no. 3, pp. 681–689, 2008.

[4] A. Bron, P. Habas, G. Coupé, et al., "Standardization of MR image intensity in brain MRIs," Neuroimage, vol. 132, pp. 385–395, 2016.

[5] R. Litjens, T. Kooi, B. E. Bejnordi, et al., "A survey on deep learning in medical image analysis," Med. Image Anal., vol. 42, pp. 60–88, Dec. 2017.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, May 2015.

[7] E. Hosseini-Asl, R. Keynton, and A. El-Baz, "Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network," Front. Neurosci., vol. 12, 2018.

[8] H. I. Suk, S. W. Lee, and D. Shen, "Deep learning-based feature representation for AD/MCI classification," Neuroimage, vol. 101, pp. 569–582, Nov. 2014.

[9] B. Liu, M. Zhang, and X. Li, "Alzheimer's disease classification using deep convolutional neural networks and MRI," Front. Aging Neurosci., vol. 12, 2020.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 770–778, 2016.

[11] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Mach. Learn. (ICML), pp. 6105–6114, 2019.

[12] Z. Lin, R. Zhang, X. Zhang, et al., "Convolutional neural networks in Alzheimer's disease: A survey," Comput. Biol. Med., vol. 134, 2021.

[13] G. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 25, 2012.

[14] R. Parisot, P. G. Montana, and D. Rueckert, "Disease prediction using graph convolutional networks: Application to Autism Spectrum Disorder and Alzheimer's disease," Med. Image Anal., vol. 48, pp. 117–130, 2018.

[15] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Proc. Int. Conf. Mach. Learn. (ICML), pp. 448–456, 2015.

[16] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," in Proc. Int. Conf. Learn. Represent. (ICLR), 2021.

[17] H. Chen, Y. Wang, Q. Zhou, et al., "TransUNet: Transformers make strong encoders for medical image segmentation," arXiv:2102.04306, 2021.

[18] Z. Liu, Y. Lin, Y. Cao, et al., "Swin Transformer V2: Scaling up capacity and resolution," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 12009–12019, 2022.

[19] A. Hatamizadeh, D. Yin, J. Kautz, and F. Yang, "UNETR: Transformers for 3D medical image segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 5744–5754, 2022.

[20] R. Tuli, S. Mishra, and S. K. Singh, "Explainable deep learning models in medical image analysis," IEEE Access, vol. 9, pp. 151988–152006, 2021.

[21] D. S. Marcus, T. H. Wang, J. Parker, et al., "Open access series of imaging studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," J. Cogn. Neurosci., vol. 19, no. 9, pp. 1498–1507, 2007.

[22] M. Dyrba, K. Grothe, S. Kirste, et al., "Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM," Hum. Brain Mapp., vol. 36, no. 6, pp. 2118–2131, 2015.

[23] S. Eskildsen, P. Coupé, V. Fonov, et al., "Structural imaging biomarkers of Alzheimer's disease: Predicting disease progression," Neurobiol. Aging, vol. 36, no. S1, pp. S23–S31, 2015.

[24] C. Wee, P. Yap, D. Shen, "Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns," Hum. Brain Mapp., vol. 34, pp. 3411–3425, 2013.

[25] R. Litjens, T. Kooi, B. E. Bejnordi, et al., "A survey on deep learning in medical image analysis," Med. Image Anal., vol. 42, pp. 60–88, 2017.

[26] X. Li, Y. Liu, Y. Li, et al., "Deep residual networks for Alzheimer's disease prediction," Front. Aging Neurosci., vol. 12, 2020.

[27] S. Buda, A. Maki, M. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," Neural Netw., vol. 106, pp. 249–259, 2018.

[28] S. Parisot, P. G. Montana, and D. Rueckert, "Disease prediction using graph convolutional networks:

application to Autism Spectrum isorder and Alzheimer's disease," Med. Image Anal., vol. 48, pp. 117–130, 2018.

[29] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention is all you need," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017.

[30] R. Tuli, S. Mishra, S. K. Singh, "Explainable deep learning models in medical image analysis," IEEE Access, vol. 9, pp. 151988–152006, 2021.

[31] Y. Chen, Y. Zhang, C. Kong, et al., "TransDeepLab: A lightweight transformer–CNN hybrid architecture for semantic segmentation," Pattern Recognit., vol. 132, 2023.

[32] Y. Guo, B. Liu, L. Yang, et al., "Alzheimer's disease diagnosis with deep learning: From the perspective of multi-class classification," Neurocomputing, vol. 448, pp. 302–315, 2021.

Author Biography

Jala shilpa is a dedicated research scholar in the department of Computer Science & Engineering (CSE) at Chaitanya Deemed to be University (CDU) in Hyderabad.Her academic journey began with a Bachelor's degree in engineering from Ramappa College of Engineering,which she completed in 2010.following her under graduate studies, pursued a Master of Technology in Software Engineering at the Jayamukhi Institute of Technology and sciences , graduating in 2013.Her current research focuses on advancing Machine Learning technology.Through her work ,Jala aims to enhance diagnostic capabilities and develop innovative solutions in these critical areas. Her academic background and ongoing research reflect a strong commitment to contributing to the fields of computer science and engineering , particularly in improving machine learning technologies and processing methods.

Dr. G. Shankar Lingam is working as Professor in Chaitanya Deemed to be University, Warangal, Telangana. Published numerous research papers in reputable journals, contributing to the fields of Processor Architecture, Optimization in Multiprocessor Systems, Mobile Applications, and more. He has received prestigious awards such as the ELSEVIER Best Academician Award and Dr. Sarvepalli Radhakrishnan Best Teacher Award. Recognized with the ELSEVIER Lifetime Achievement Award for outstanding contributions to research. He is a seasoned researcher with strong mentality and sound analytical mind, and he has contributed tremendously in his core areas of research. He has reviewed several research articles for many journals and publishers