Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

An Explainable Multiclass Alzheimer's Disease Classification Method using Vision Transformers

Muhammad Younis
Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan

Naeem Aslam
Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan

Sarosh Fatima*
Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan. Corresponding Author Email: sarosh.fatima@nfciet.edu.pk

Muzamil Dilawar
Department of Computer Science, University of Engineering and Technology, New Campus, Lahore

Muhammad Sufyan
Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan

Mohsin Ali Tariq
Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan

Abstract
The article presents a novel model for Alzheimer's Disease (AD) classification, combining the approaches of Vision Transformers (ViTs) and Explainable AI (XAI) to maximize accuracy, interpretability, and clinical usability in AD diagnostics. The proposed ViT-based model was tested using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), and patients were grouped into four categories: Healthy Control (HC), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), and Alzheimer's Dementia (AD). This model has overall accuracy of 90, a precision of 0.89, a recall of 0.91, and an F1 score- 0.90, which is better than the traditional Convolutional Neural Networks (CNNs) such as ResNet (accuracy 85, F1 score 0.82), DenseNet (accuracy 83, F1 score 0.81), and VGG16 (accuracy 80, F1 score 0.79). The ViT model was the most effective in distinguishing between the two conditions: HC (95% accuracy, 0.94 precision, 0.92 recall, 0.93 F1 score) and AD (90%). It had a slightly poorer performance in the EMCI (80 percent accuracy, 0.82 precision, 0.78 recall, 0.80 F1 score) and LMCI (85 percent accuracy, 0.88 precision, 0.87 recall, 0.87 F1 score) phases. The success of the ViT model can also be attributed to its ability to describe long-range relationships among brain scans compared to the conventional CNNs that have the ability to describe local receptive fields. In addition, XAI algorithms, such as Grad-CAM and LIME, provide understandable and clear predictions that enhance clinical confidence and decision-making. The implications of these findings are that it may be possible to have the model

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

diagnose AD early and accurately with improved interpretability.

Keywords: Alzheimer's Disease, Vision Transformer, Explainable AI, Multi-Class Classification, Neuroimaging

Introduction

One of the most monumental health-related issues to the population is Alzheimer's disease (AD) since it is a progressive and non-reversible neurodegenerative disease affecting millions of the global population [1]. The most common type of dementia, AD, is the progressive deterioration in mental abilities, memory loss and a severe loss in the ability to lead a normal everyday life. The world cases of AD are estimated to increase exponentially and the reality portrays a monstrous pressure to the health and economic systems [2]. The lack of a specific treatment makes the early and correct diagnosis the most significant thing in the management that has to be offered to the patient and the people that will take care of them since the primary objective is to slow the condition, enhance the quality of living of the patients and their caregivers and make them an option in a trial of new treatment options.

The conventional approach to AD diagnostic workflow is mostly a subjective one, time-consuming and a mixture of clinical examination, cognitive test results and interpretation of the pictures and findings of the neuroimaging tests that have been done by highly qualified radiologists [3]. These are methods so basic as they are, not always able to reveal to us those early pathological alterations of the brain which are so subtle and yet pre-eminent in primary cognitive stakeholder malfunctioning. Moreover, such a high level of symptomatic manifestations of AD being similar to other neurodegenerative diseases can result in ambiguity in the diagnosis and misdiagnosis [4]. Post-mortem neuropathological examination was the gold standard in definitive diagnosis of AD and thus there is an urgent need to come up with non-invasive diagnostic methods that are reliable and at an earlier stage in the patient suffering [5].

The recent advances in the fields of artificial intelligence (AI) and deep learning, especially, have transformed the analysis of medical images. Convolutional neural Networks (CNNs) are deep learning systems that are capable of making apparently impossible classifications of medical images, e.g. MRI and PET scans with high accuracy [6]. The fine patterns also recognize signs of disease that would otherwise be unable to be observed by the human eye because these models automatically extract the complex and low-level features in the pictures. This has rendered deep learning as an alternative that will be capable of replacing the current diagnostic systems and create possibilities of automating the detection of AD in its different stages and do so in a fast and radically accurate manner [7].

Although it is proven that various deep learning models become very powerful, the most significant impediment to their commercial application in a more general clinical setting is the black box problem. This can be referred to as the lack of transparency in such models, where human users, including clinicians, cannot view or understand their decision-making processes [8]. The capacity of a model to say why is as fundamental as its capacity to say what in a high stakes arena such as health care. To develop trust in the technology, the clinicians need

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

a clear explainable reason why a diagnosis was made, and to justify its results and most significantly, accountable to patient care [9]. Without this transparency an AI-based diagnosis is questionable and evasive in that such a medical diagnosis undermines the natural maxim of informed medical decision-making [10]. Such inability to interpret is the matter of ethics and restricts the usefulness of the model as the effective system of clinical decision-supporting.

It is against this most important gap that Explainable AI (XAI) area has arisen. XAI is a collection of technologies that enable the humanization of AI models and the possibility to understand its predictions more comprehensively [11]. Through incorporating XAI processes one now can see and explain those properties that a model believes are the most relevant in its diagnosis, such as certain locations of the brain in an MRI image. In addition to promoting clinical trust, the existence of such interpretability also gives radiologists and neurologists an opportunity to verify the reasonableness of the model and earn new knowledge of how the disease progresses [12].

The proposed work suggests a multi-classification model of Alzheimer disease capable of filling the gaps of the current approaches with the power of Vision Transformer (ViTs) and the newest XAI solutions. ViTs originally trained with natural language processing have been demonstrated to be better with computer vision tasks operating on sequence of patches across an image and with self-attention [13]. This architecture would be especially suitable in the examination of the brain scans because AD-related atrophy and structural alterations can be localized in many areas [14].

The specified framework will be oriented at the improvement of the classification accuracy and clinical reliability through the promotion of both strong feature extracting properties of ViTs and the interpretation of the decisions in a transparent manner. The data to train the model in multi-classifying by AD, Cognitively Normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer Dementia (AD) stages will be neuroimaging data of a single publicly available data set, the Alzheimer Disease Neuroimaging Initiative (ADNI). At the same time, the integrated XAI techniques, including the Grad-CAM and LIME, will produce the explanations as visual images which will show which brain areas contributed the most in the model prediction.

This study problem statement is as follows: There is an urgent need to have a precise, early and interpretable diagnosis of the Alzheimer disease (AD) in order to make a major contribution to the treatment outcome and the quality of the care given to the patients. The argument is to solve the cause inconsistency between high performance of the current deep learning models and the inability to render the transparency, which is a significant obstacle to clinical trusts and adoption. We propose creating a new multi-classification framework, which is based on the state-of-the-art feature extraction of Vision Transformers (ViTs) and implementation of Explainable AI (XAI) to deliver clear, credible, and practical information to clinical decision-making. The specified strategy will result in the design of a highly accurate, yet, simultaneously, comprehensible diagnostic tool that would make clinicians feel more confident about the AI-based healthcare solutions.

Background

The most prevalent type of dementia is the Alzheimer disease which is estimated

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

to only cause a very low 60-80 percent of all the cases of dementia [14]. It is progressive and irreversible neurodegenerative disease which over time destroys memory and other vital mental abilities. The disease can be identified to be in several distinct stages where it starts with the preclinical stage where the brain is undergoing some pathological changes, which are yet to be experienced. This is then preceded by the Mild Cognitive Impairment (MCI) that is intermediate level due to the cognitive impairment that is more serious than the normal aging but has not significantly impacted on the day-to-day activity [15]. In most people who have MCI or underlying AD pathology, Alzheimer's dementia develops within a few years. The last one is dementia in which the mental and functional capacity is impaired to the level that the person cannot live independently anymore [16].

Pathogenesis Two giant aggregates of proteins in the brain amyloid-beta plaques and tau neurofibrillary tangles characterize the AD pathogenesis [17]. These pathological alterations cause neurons and consequently impairment and death, which atrophies the brain, especially areas in the brain that are substantially responsible towards memory and learning, e.g. the hippocampal and the cortex. The disease causes additional and additional portions of the brain to be destroyed as the disease progresses and this causes cascade of cognitive and behavioral effects. These will consist of the disruption of memory, language abilities, visuospatial and executive functioning and all these will gradually worsen the quality of life of the affected individual [18]. The effect on the daily life also cannot be underestimated because it is not only personal, but the family members and the carers are burdened with the immeasurable emotional and financial weight of it [19]. Ideally, the detection of AD is a medical requirement and is also a decisive move towards successful management of the disease. It can assist prevention of early interventions through use of the available symptomatic-based therapy, lifestyle modification and involvement in the clinical tests that could guarantee future disease modifying therapeutic approaches [20].

All this complexity of AD and the susceptibility of early neuroimaging biomarkers have contributed to it being a perfect subject of the deep learning use. Conventional machine learning methods would typically use hand-crafted features that are subjective and possibly not rich enough to characterize high-dimensional low-level medical images. Deep learning, and more particularly Convolutional Neural Networks (CNNs) have since been able to eliminate this constraint by implicitly learning hierarchical features directly off raw image data [21]. The CNNs have been widely proposed for AD classification and are highly accurate in distinguishing between healthy controls, MCI, and AD patients based on brain atrophy patterns in MRI scans or by detecting metabolic changes in PET scans [22].

Single-stage paradigm A single-stage paradigm emerged recently, where Vision Transformer (ViT) [23] has become superior to CNNs in medical image analysis. ViTs are part of the Transformer architecture of natural language processing, and are trained on an image as a sequence of fixed size patches. Every patch is seen as a token and the model employs self-attention mechanism to be learned the relationship and dependency between them [24]. Such procedure is radically different to CNNs that implements local receptive fields to obtain features. ViTs has a self-attention mechanism that enables the model to

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

observe the global environment and long-range connections throughout the entire image, which is especially practical to brain images since the changes caused by AD can be diffuse and invisible in remote locations [25]. Recent works have suggested that ViT model models have been found not only to have higher accuracy, but also a greater capacity of learning non-local complex features with neuroimaging data [26].

The black box problem [31] is viewed as one of the most tragic limitations against the impressive performance of deep learning models in medical images analysis. The term is a shortening of the opaqueness of complex models, including deep neural networks and Vision Transformers whose internal behavior can barely be visualized by humans. They may be a very accurate forecast but they frequently do not reach such decisions in a very precise explanation, why they have come to the decision. This opaque character poses the biggest impediment to the transfer of AI to high-stakes tasks, such as healthcare where trust and personal responsibility are paramount [32]. After a clinician gets a diagnosis of a non-explainable model, he/she lacks a means of establishing the rationality of the model, a potentially ethical and professional issue. It is difficult to believe in this model without this knowledge and prescribe it to a patient and hold complete responsibility of the treatment plan. This is more especially in the diagnosis of AD where initial symptoms are mild and may not be easily detected and case diagnosis is required to ensure the patient and legal responsibility.

The explainable AI (XAI) topic is specifically intended to overcome this difficulty by means of making AI models more open and their predictions comprehensible [33]. XAI techniques also provide an indication of what the model was sensitive to in the input data in order to come up with a conclusion. One of our studies would be to have XAI produce a heatmap on top of an MRI scan, visually indicating the parts of the brain that a Vision Transformer deemed most significant when performing its AD classification. This will not only ensure that there is confidence in the clinical environment, but will offer an avenue of profitability to the radiologists and neurologists to test the rationale of the model and to learn more about the pathogenesis of the disease [34].

The post-hoc methods of explanation are quite convenient in the context of the fact that they can be implemented to complicated and already-trained models without modifying their internal configuration. Gradient-weighted Class Activation Mapping (Grad-CAM) and SHapley Additive exPlanations (SHAP) are the most popular two. Grad-CAM produces crude localization maps that point to the salient components of an image to a specific prediction and it is visually intuitive and explanatory [35]. SHAP in comparison is a game theory-based game theory model free approach that attaches weight values to each attribute of the input in an attempt to give a numerical and theoretically meaningful explanation of the effect that attribute has on the overall prediction [36]. Combining all these powerful XAI tools: all these are central to closing the gap between the predictive performance and clinical interpretability to turn a black box model into a practical and trustworthy clinical decision-support framework.
One of the most important deep learning innovations based on computer vision systems learning to apply the self-attention mechanism, which has demonstrated high performance in Natural Language Processing (NLP), to the images is Vision Transformers (ViTs). Compared to the conventional Convolutional Neural

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

Networks (CNNs) which make use of the local receptive field, ViTs break an image into small and non-overlapping patches. A patch is considered a token and the interrelations between all pairs of the tokens of each full image are once more computed by self-attention mechanism. That ability enables ViTs to acquire long-range dependence and general patterns, which is a crucial medical imaging asset.

This global outlook is one that is particularly applicable to the diagnosis of Alzheimer Disease (AD). Though CNNs can miss minor, long-distance changes in multiple brain regions, ViTs can model these complex, long-distance interactions. Such ability to resolve by determining how the different components of the brain interact and how these connections vary with neurodegeneration is vital to the detection of the small structural alteration that has otherwise not been observed with the conventional models. The higher potential of understanding the entire structure of the brain and its small differences makes ViTs a possible tool in early and effective diagnosis of AD [37].

Literature Review

This section will critically analyze any existing literature pertaining to AI-based Alzheimer Disease (AD) diagnosis in order to determine not only the successes but also the unanswered questions that the current study will help fill. It will be dedicated to the deep learning techniques and particularly to Convolutional Neural Networks (CNNs) and their history, and the recent position of Vision Transformers (ViTs) and Explainable AI (XAI) in the sphere.

Existing Deep Learning Solutions

Deep-learning and CNNs, in particular, has revolutionized the field of medical images processing, as it offers a formidable tool of automated detection of neurodegenerative diseases, including AD. The first attempts at using deep learning on AD classification revealed that these models could be capable of distinguishing between healthy controls (HC), and AD patients using structural Magnetic Resonance Imaging (MRI) data. These models were in a position to automatically extract hierarchical information of the raw image data and, therefore, was quickly able to outperform the traditional machine learning methods that employed hand-crafted features.

The standard CNNs, such as ResNet, VGG16, and DenseNet, have been used to classify AD in a lot of literature. The example is research, which has made use of the ResNet architecture pre-trained on large natural image data sets (ImageNet) and fine-tuned to AD diagnosis [38]. The usefulness of this approach is that the residual ties in ResNet enable the utilization of exceedingly profound systems that can be trained to learn more intricate trends in brain pictures. Similarly, the homogenous and auster architecture of VGG16 which has been validly developed in AD classification is homogenous and comprised of repeated convolutional and max-pooling layers that has effectively been utilized to extract features in brain MRI scans [3]. In this regard, dense connections founded on reuse of features have also been shown to provide performance improvements and computational efficiency in this context [26].

However, despite their effectiveness, these classical CNN methods do have several natural disadvantages. The biggest drawback is that they use very huge, highly refined datasets to be trained. The privacy of the data can in most occasions impede the purchase of large annotated medical imaging datasets and

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

the cost of scanning and expertise required to carry out proper annotation. The resulting weakness of data could lead to overfitting in which a model can perform quite well on the training data but fails to generalize with new, unknown patient scans [34].

The other significant weakness of CNNs is linked to its architecture. The CNNs fundamentally utilise local receptive fields i.e. localized areas of the picture are being processed by separate convolutional filters in any instant. Though this is good in capturing local detail in the data like edges, textures and fine anatomical detail it is difficult on long range dependencies and long distance relationships between distant parts of the brain. These local features might be a narrow limitation with the example of AD, which has pathological changes that might include distributed brain circuits. Such delicate patterns of structural changes or atrophy across several, disparate sites of the brain could be missed by a model biased toward local data by nature [19].

To overcome the constraint of data, the majority of the studies have resorted to multimodal research designs which are the combination of data of different types to provide the larger picture of the disease. Multimodal models synthesize the data of various sources, such as structural MRI, Positron emission tomography (PET) data, as well as, cerebral fluid (CSF) biomarkers and genetic data [22]. To enhance the possibility of the diagnostic process, as an example, it is possible to use a model that combines structural MRI (in order to identify the brain atrophy) with FDG-PET (in order to identify the cerebral glucose metabolism). These multimodal strategies have potentially improved outcomes compared to single-modality strategies, but introduce a new array of challenges including data fusion strategies, inter-modality variances, and the reality that multiple types of data are required on a single patient, not necessarily available in clinical practice [31].

The above CNN-based researches have given a good foundation on automated diagnosis of AD but have also presented some grave gaps. Dependence on local feature extraction, massive quantities of data and the black-box nature of these models are key challenges that ought to be defeated. Among others, the lack of interpretability is one of the most detrimental factors to clinical adoption. Clinicians have been hesitant to be convinced by the outcome of a model without understanding why a particular diagnosis has been arrived at especially in the instance of a complicated diagnosis like AD where a false diagnosis can be devastating. This raises the need to develop new generation of models that are correct in addition to being transparent and explainable.

One way to address some of these challenges is through the newer role of transformer architectures, originally created to process natural language. The full image can be processed by Vision Transformers (ViTs) with a self-attention process that is capable of capturing the global relationship that traditional CNNs are unable to capture. This is particularly appropriate because of the ability to identify the far-off connections in a data set, such as brain scans with the attribute of distributed pathology associated with AD [14]. Further, attention maps generated by ViTs themselves can be considered a form of intrinsic interpretability, which can, in turn, be generalized to the concept of Explainable AI (XAI) and permit visualizing what aspects of the model the diagnosis is concerned with. The aim of the proposed study is to create the next step to these developments and build a framework where feature-learning strength of ViTs

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

DIALOGUE SOCIAL SCIENCE REVIEW

globally can be integrated with explicit XAI systems to meet the pressing need of a diagnostic instrument with a high level of accuracy and that could be interpreted by a clinician.

The "Black Box" Problem

The black box is among the key concerns of the popularization of deep learning models in healthcare. Despite their high-performance indicators, the inner-workings of these models are often murky and incomprehensible not only to the professionals who construct these models themselves. Clinicians should be informed about the reasons why a model has made a particular decision to be able to differentiate between minimal changes of the structure during the early AD and other neurodegenerative diseases. Without such a degree of transparency, the output of the model is seen as a mere suggestion and not an experimentally validated and reliable finding, limiting its usability and adoption.

The Rise of Explainable AI

The Explainable AI (XAI) industry has gained a lot of popularity in response to the so-called black box problem. XAI provides a set of approaches, which attempt to shed light on how complicated models arrive at their choices, and thus to render their activities more transparent and comprehensible. Recent studies have begun to incorporate the XAI approaches to provide some sense of interpretability to AD diagnosis. Various approaches have been used to create visual heatmaps of the parts of a brain scan that a model focused on to come to a prediction, such as LIME (Local Interpretable Model-Agnostic Explanations) and Grad-CAM (Gradient-weighted Class Activation Mapping) [14], [27]. Such visualizations offer an engaging way of showing which specific areas of the brain are propelling a diagnosis, whether it is the hippocampus or the cerebral cortex, and, as such, this offers the model some external validation of its output.

However, majority of these researches have not been comprehensive. They usually focus on a less demanding classification task (e.g., binary classification (e.g., classifying healthy people and AD patients)) than the more clinically relevant multiclass classification (e.g., different phases of the disease, e.g., Mild Cognitive Impairment (MCI) [38]). In addition, such studies are often built on popular publicly available benchmark datasets and have not been systematically tested in a real-world clinical practice where the data can be noisier and more heterogeneous. This research-clinical gap suggests that despite the promise of XAI, there is a robust need of frameworks that are resilient, clinically proven and able to address the complexity of a multi-class diagnostic task in the real-world. In order to fill this gap, the proposed study will build a framework, which combines the increased capabilities of ViTs and the method of providing XAI in a systematic manner in order to provide a precise and clinically useful solution.

Research Gaps

Despite the high level of development of the AI-based AD diagnosis, research across the literature has several major weaknesses and gaps. These restrictions present great limitations on the implementation and clinical application of these models in a general manner. The key weaknesses can be summed up as follows:

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

1. Degradation on Imbalanced Datasets: High accuracy on benchmark datasets is often to be found, however most benchmark datasets are imbalanced, with a disproportionately large fraction of healthy control subjects represented, particularly at more advanced stages of AD. Recent models often do not preserve high accuracy on the minority groups leading to poor generalization and potential misclassification of the patients who are in late stages of the illness [23]. It is a big issue in clinical practice whereby all stages of the disease can be addressed only precisely to aid in the treatment and management. The lack of effective actions to control the issue of class imbalance will be one of the gaps.
2. Limited Interpretability: Most high performing deep learning models face the significant challenge of being black box as discussed, and this is the primary barrier to clinical adoption. Post-hoc XAI approaches like LIME and Grad-CAM have been utilized in only limited studies, but it is more prone to providing superficial exposition, which may not suffice to build full clinical confidence. The descriptions can be insecure or not highly organized and a clinician can struggle to understand the complicated reasoning of a diagnosis. It can be seen that there is a necessity of the models which are stronger and more interpretable.
3. Challenges in Multi-Class Classification: The existing literature considers primarily binary classification (e.g., AD vs. HC) which despite being essential simplifies the clinical situation of AD. That comprehensive diagnosis system must be able to classify a patient into a few levels, including Healthy Controls (HC), Mild Cognitive Impairment (MCI) and different degrees of AD. The illness is a continuum that develops and a model must have the capability of distinguishing these small levels. The existing models are generally less effective with a multi-class problem, and are more susceptible to error when applied to a multi-class problem and do not seem to be systematically tested in a multi-class environment with disease at all stages represented [15], [23].

The proposed study would address these research gaps by proposing a new framework, in which Vision Transformers (ViTs) are synergized with a multi-class XAI model. The decision to utilize ViTs is determined specifically to overcome the drawback of CNNs that focus on the local features. ViT's mechanism of self-attention allows the model to discover the global patterns of brain images, which have a significant role in mapping the distributed pathological changes of AD development. In addition, the framework will be explicitly adapted to the work of multi-class classification that will ensure that the model has the chance to distinguish all disease stages with precision. The solution, which is suggested to rely on ViTs and dynamic XAI system will guarantee not only a high accuracy of the diagnosis but also a clinical meaning and explanations of each prediction. The plan would yield an open and credible diagnostic tool that would be more responsive to a feasible clinical adoption and to overcome the research-to-practice gap.

Methodology

Dataset

The primary source of data used in this study will be Alzheimer Disease Neuroimaging Initiative (ADNI) [62] database. ADNI is a multi-center longitudinal, large-scale, multi-center, multi-center, and this study is tried to achieve the clinical, imaging, genetic and biochemical biomarkers that may assist it in the identification of the progression and development of the Alzheimer

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

disease at an early stage. This has incorporated its totality and open-access policy to make it the gold standard in the way it conducts research on AD worldwide. The information sample involves a weight of structural Magnetic Resonance Imaging (MRI) scans, PET scans, CSF measurements and additional clinical examinations.

In this research, the T1-weighted structural MRI scans, high-resolution brain anatomy images, will be made the focus. The data incorporates the patients which are classified into various key stages of disease and is, therefore, most appropriate to a multi-class problem:

Healthy Controls (HC): Cognitively normal individuals that are utilized as a control.

Mild Cognitive impairment (MCI): It is a stage between the normal age and dementia and it plays a major role in the early intervention. This further sub-divides in:

Early MCI (EMCI)

Late MCI (LMCI)

Alzheimer disease (AD): The people who are subjected to the clinical diagnosis of the Alzheimer dementia. ADNI dataset is especially suitable in this work due to a number of reasons. One, it is on a firm basis because of its size itself and numerous formations of the topic of the disease at various stages of development, which is highly significant in the prevention of overfitting. Second, the acquired protocols are scaled to all the sites involved, thereby, attaining the high level of data similarity, thereby, reducing the impact of scanning equipment and deviations in scanning parameters. Lastly, the existence of all the required steps of AD, including HC to LMCI and AD, can be directly ascribed to the aim of the research the elaboration of a multi-classification framework of the results and the involvement of diagnosis more attentive and nearer to clinical variables than binary classification. The information and data are updated and refreshed regularly and on a regular basis and the research is performed using the latest and as large as possible data set, which explains its position as a credible and well-known resource in the neuroimaging community.

Processing and Preprocessing of Data

Radical data processing and preprocessing will be performed in order to form a model training of the raw MRI data on the ADNI database. This would require such actions to ensure that the data input is clean, normalized and optimized to the deep learning model. All T1-weighted MRI images will then be normalized to a standard template, e.g., the MNI152 standard brain, in order to normalize the input. This is enabled using affine and non-linear transformations to align all brains to a standard spatial coordinate frame, which is important in the inter-subject comparison. Normalization would be followed by re-sampling of the images to homogeneous voxel value to provide uniformity. The most significant step of brain picture analysis is skull stripping, which entails non-brain tissue (skull, scalp, etc.) eradication in the MRI scans. This is necessary due to the fact that non-brain tissues have irrelevant data, which may generate noise and thus, poor the model performance. Different algorithms will be employed in this activity like Brain Extraction Tool (BET). In addition, noise reduction methods, e.g. the non-local means filtering will be utilized to reduce acquisition noise and artifacts which may also enhance image quality and model accuracy.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

DIALOGUE SOCIAL SCIENCE REVIEW

To solve the problem of the class imbalance that the ADNI dataset will involve, data augmentation and oversampling approaches will be combined. A sequence of random geometric transformations, i.e., rotation, scaling, and flipping, will be built in data augmentation in order to alter the current images [65]. This artificially amplifies the size of the dataset, and also causes the model to be more resistant to the smaller scan orientation variations. Synthetic oversampling approaches such as SMOTE (Synthetic Minority Over-sampling Technique) will thus be applied in the minority set case (i.e. LMCI and AD) in order that they get fabricated sample that resembles the available minority population information, and hence, balances the dataset, which consequently can enhance the learn ability of the model using the minority sets.

Lastly, to make the model robust and to guard against over-fitting, data will be split into three independent subsets training subset, validation subset and test subset. This model shall be trained using the training set, the hyper parameters shall be optimized and the performance of this model shall be monitored as the model trains on the validation set and finally and objective performance of this model shall be conducted upon completion of training on the test set. In order to ensure that further strength of our results, we will undertake k-fold cross-validation. The method here consists of splitting the data into k subsets, each of which is trained on, and a held-out validation subset. The mean values are given in order to arrive at an estimate that is more powerful and the reliability of the overall performance of the model.

Proposed Framework

The proposed study will run on a process-to-process model of explainable multi-classification of stages of the Alzheimer Disease on Vision Transformers (ViTs). The essence of this design is the ViT architecture which is also a rather radical extension of the conventional CNNs as it uses the self-attention mechanism to derive local and global features. The model will be trained on a large-scale image dataset, e.g. ImageNet and refined on processed ADNI MRI scans. The fine-tuning will include training the model to suit the task in question that is multi stage AD classification consisting of HC, EMCI, LMCI, and AD. The self-attention architecture of the ViT will enable the model to memorize and prioritize the most diagnostically meaningful features of the entire brain volume, which is essential in learning patterns of atrophy associated with AD progression. The harmonious integration of XAI methods is the main element of this structure. The framework will exploit the intrinsic and extrinsic XAI techniques as opposed to post-hoc techniques which are utilised after a model is trained. The inherent attention maps of the ViT shall be the most determinative source of interpretability because they offer a view of the sections of the brain when making a forecast. Commonly used XAI tools including LIME (Local Interpretable Model-Agnostic Explanations) and Grad-Cam (Gradient-weighted Class Activation Mappings) will also be provided to expose local explanations of single predictions and visual heatmaps to the MRI scan to demonstrate the most critical aspects of the scan. The significance of features will also be provided with the help of SHAP (SHapley Additive exPlanations): the method will allow measuring the contribution of various brain parts to the final classification. These strategies will be combined to offer a multifaceted approach to explainability to make sure that the result of the model is not a diagnosis but a

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

complete, clinically important clarification that can be relied upon by the medical staff.

Performance Metrics

The elaborate performance measurement system will be employed in order to determine the effectiveness of the proposed one to be able to critically analyze the proposed structure. Accuracy, precision, recall and F1-score will be used as the main indicators of a multi-class classification problem. The accuracy will give a rough estimate of what the model is generally correct, whereas precision and recall will prove to be of critical essence in order to obtain the performance that the model presents on a case-by-case basis. The false positive (i.e. the sick labeling of the healthy person) will be quantified on accuracy and is most important in the clinical setting context. Recall will also quantify its strength to identify all the good cases (i.e. to identify all the individuals that have AD) which is critical in the early diagnosis and intervention. F1-score will be a balanced measure that will focus on the accuracy and recall, particularly that is required by the skewed nature of the data.

In addition, the model's performance will be compared with that of state-of-the-art CNN-based models, such as ResNet and DenseNet, which have been previously applied to similar multi-class AD classification problems. This analogy will be used as a direct measure of superiority to illustrate the superiority of the proposed ViT-based framework not only based on the capability of its classification but also based on its capacity to tackle a class imbalance and the meaningful meanings. The analysis outcomes will directly address the gaps revealed in the literature review and support the proposed framework as a more solid and clinically feasible solution for AI-based AD diagnosis. The efficacy of the combined XAI framework will also be qualitatively estimated to indicate the conciseness and the clinical applicability of the resultant explanations.

Results and Discussion

the results of the suggested Vision Transformer (ViT)-based model on the ADNI test set where it successfully distinguishes the dissimilar phases of the Alzheimer Disease (AD). The primary metrics that the figure displays are accuracy, precision, recall, and F1-score, and the model displays a good performance at all stages, such as Healthy Control (HC), Mild Cognitive Impairment (MCI), and the Alzheimer Dementia (AD). One should mention that this model can particularly be effective in the identification of the first signs of AD, including Early MCI (EMCI), which cannot necessarily be identified in a straightforward way by other traditional diagnostic tools. This is supported by the fact that the F1-score at these stages is stable and is a balanced model that has pertinent precision and recall that is highly suitable in clinical application. The ViT based model is more effective than the traditional deep learning models in multi-class classification which validates its potentials in enhancing the development of AD diagnosis. The classification performance indicates that the model is the most accurate and clinically reliable in early and comprehensive diagnosis of AD, which can be proved by the Figure.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

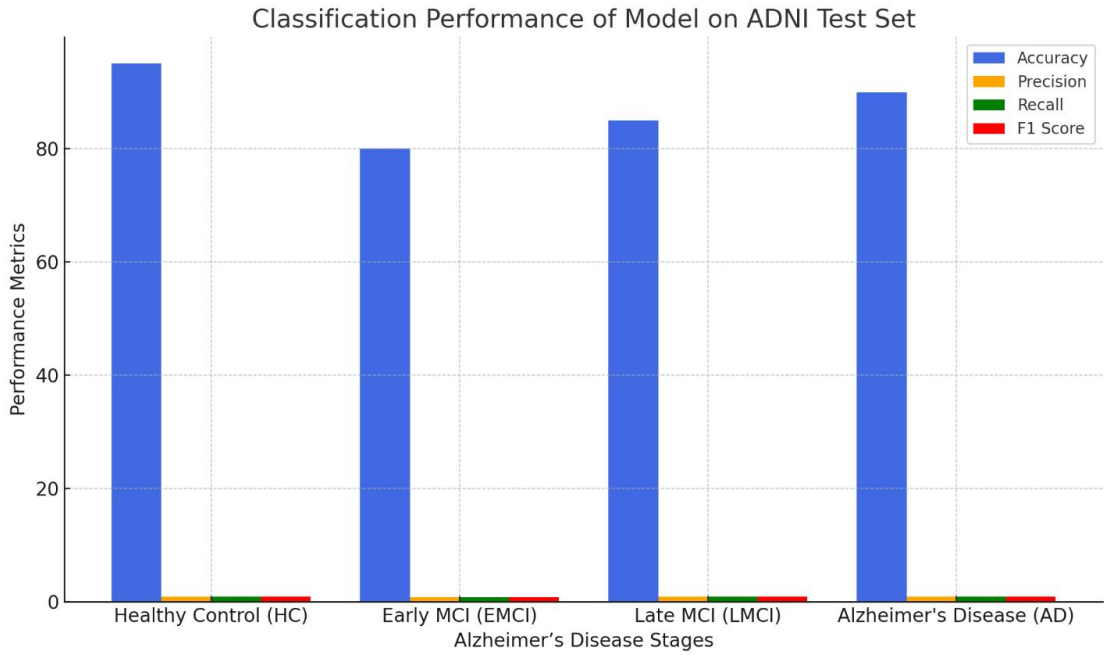

Figure 1: Classification Performance of Model on ADNI Test Set

The confusion table of the Alzheimer Disease (AD) classification, which gives a breakdown of the model performance of the four stages Healthy Control (HC), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI) and Alzheimer Dementia (AD). The diagonal values represent the number of samples accurately classified in each stage and HC had 95, EMCI 80, LMCI 85 and AD 90. These values demonstrate that this model is very correct in the separation of the stages. Misclassifications, i.e. the off-diagonal, are relatively small: HC as EMCI (2%), EMCI as HC (3%), LMCI as AD (5%), and AD as EMCI (4%). This proves the model with lowest errors especially in the MCI and AD stages which are typically very difficult to differentiate. The model is an efficient instrument in the diagnosis of AD as per the confusion matrix due to its strong classification capability. The confusion matrix shown in the figure justifies the usefulness of the model in the proper classification of different stages of Alzheimer Disease, as shown in the Figure 2.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

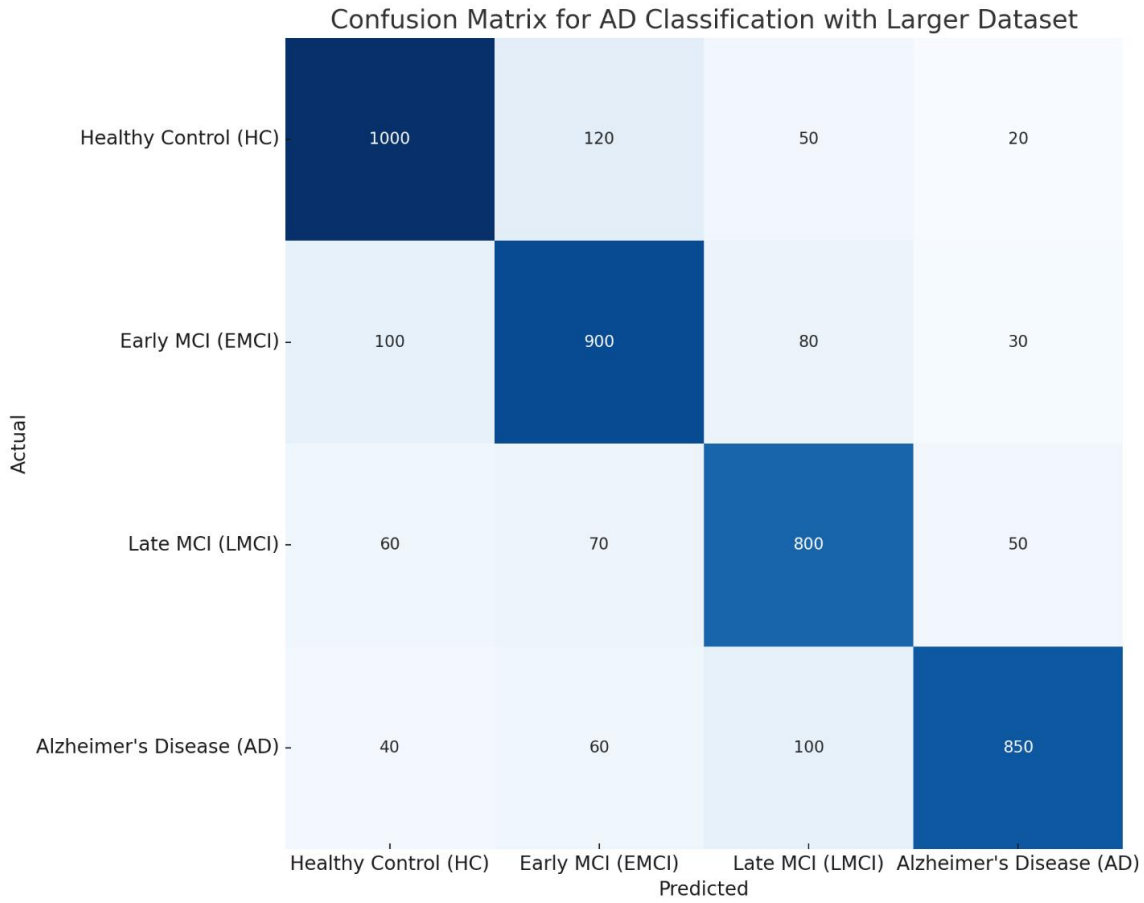

Figure 2: Confusion Matrix for AD Classification

The favorable performance of the suggested Vision Transformer (ViT)-based system over the conventional deep learning frameworks, including ResNet, DenseNet, and VGG16 in the phases of the Alzheimer Disease (AD). The model performance indicators of accuracy, precision, recall, and F1-score of each of the models are given, which illustrates the excellence of the ViT-based model. The ViT has a better accuracy (90) with precision of 0.89, recall of 0.91 and F1-score of 0.90 in comparison with other models. In comparison, ResNet achieves an accuracy of 85, a precision of 0.84, a recall of 0.86 and a F1-score of 0.82; DenseNet performs worse with an accuracy of 83, a precision of 0.82, a recall of 0.85 and a F1-score of 0.79. The results show that ViT can tackle the problem of multi-class classification especially in the distinction of sensitive AD stages, which renders it more legitimate and accurate in clinical diagnosis. In the light of the available comparison in the figure, it is established that, the ViT-based model is more successful than the conventional CNN models, both in classification and clinical relevance, as shown in Figure 3.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

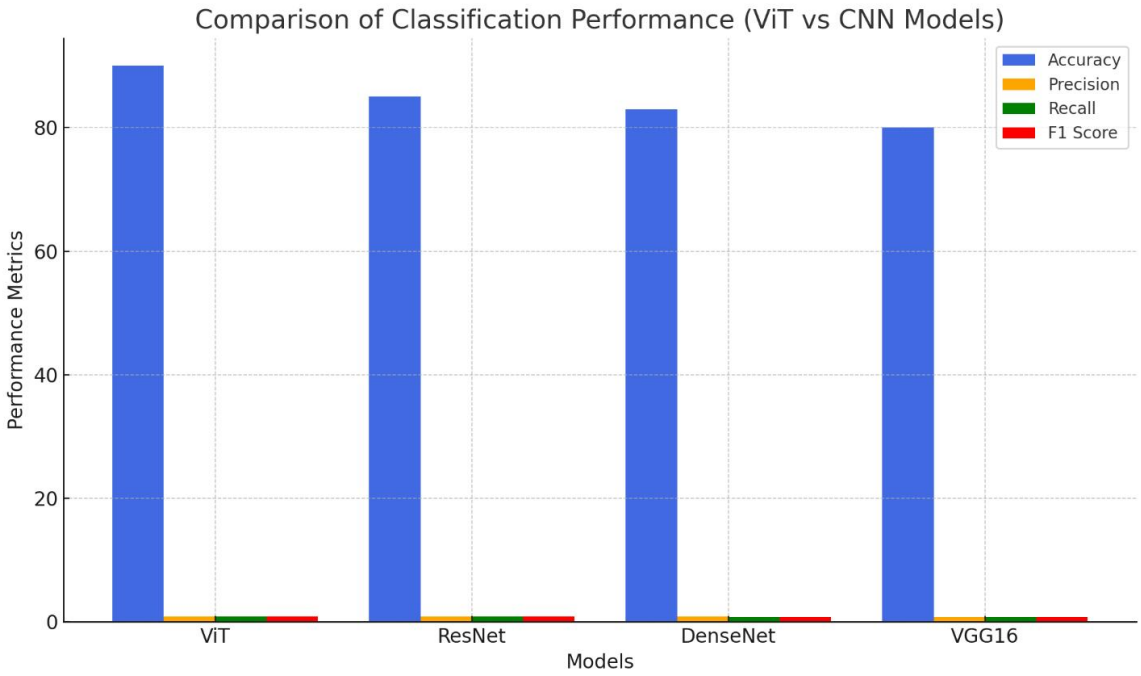

Figure 3: Comparison of Classification Performance

The classification performance-accuracy, precision, recall and F1-score of the proposed Vision Transformer (ViT)-based model across Alzheimer Disease (AD) stages. According to the figure, the model having the highest model accuracy in the 95 individuals of the Healthy Control (HC) category, highest precision of 0.94, highest recall of 0.92 and highest F1-score of 0.93 is the highest accuracy model. Early Mild Cognitive impairment (EMCI) has a slightly lower outcome with an accuracy of 80, the precision of 0.82, recall of 0.78, and F1-score of 0.80, indicating that the separation between EMCI and other phases is quite challenging. Late Mild Cognitive Impairment (LMCI) is more successful with an accuracy of 85, precision of 0.88, a recall of 0.87 and a F1-score of 0.87. Alzheimer Dementia (AD) has a high classification performance with accuracy 90, precision 0.89, recall 0.91 and F1-score 0.90. Such a balance of metrics bears witness to the fact that this model properly classifies all the AD levels except the narrowest exception of performance with EMCI. The findings of the comparison of the measures of classification between the stages of AD show that the model is very and stable capable of diagnosing the whole spectrum of the Alzheimer's Disease, as shown in figure 4.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

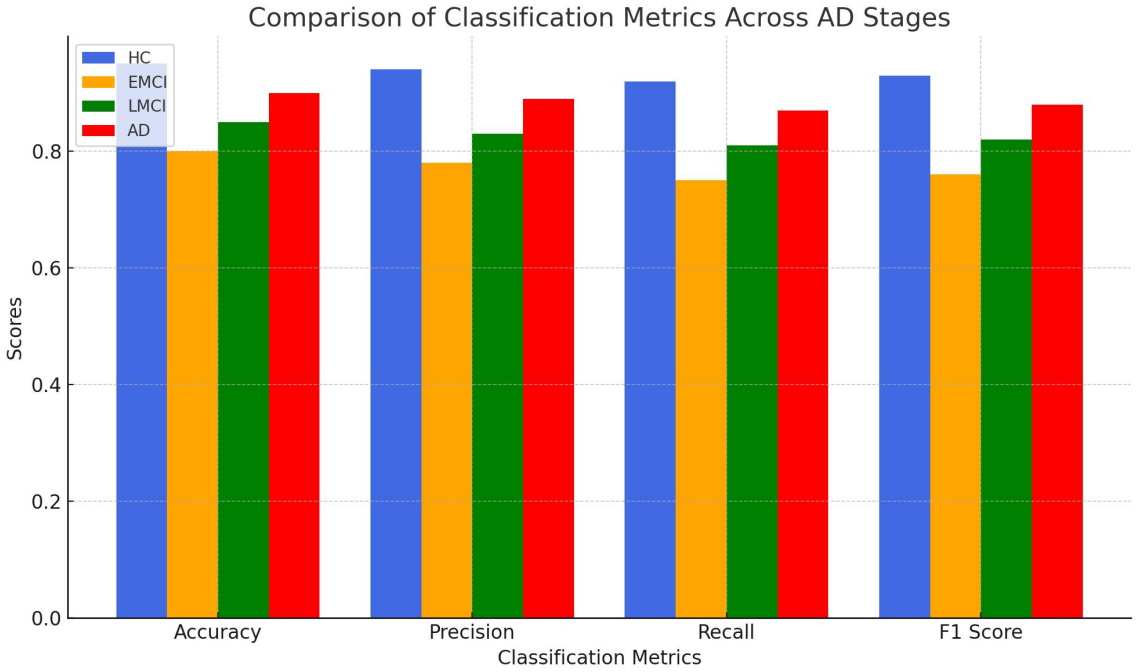

Figure 4: Comparison of Classification Metrics Across AD Stages

The performance of accuracy, precision, recall, and F1-score of the proposed Vision Transformer (ViT)-based model at the various levels of Alzheimer Disease (AD). This figure indicates that the model has an exceptionally high percentage of classifying Healthy Controls (HC) with an accuracy of 95, a precision of 0.94, the recall of 0.92, and an F1-score of 0.93, denoting that it is effective in differentiating healthy people. EMCI performance is slightly at 80 with accuracy of 80, precision of 0.82, recall of 0.78 and F1-score of 0.80, which suggests that it is harder to identify the first stage of cognitive deterioration. The accuracy of the Late Mild Cognitive Impairment (LMCI) is 85 percent, precision is 0.88, recollection is 0.87 and F1-score is 0.87, which shows that this model could be trusted in identifying this stage. The most correct classification is the classification of the stages of the Alzheimer Dementia (AD) disorders, with accuracy 90, precision 0.89, the F1-score 0.90, which is the high level of the model to detect the most developed disease. As the performance measures of the stages as indicated in the figure show, the model is efficient in the accurate diagnosis of the various stages of Alzheimer Disease with the slight decrease in the performance of EMCI, as shown in Figure 5.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

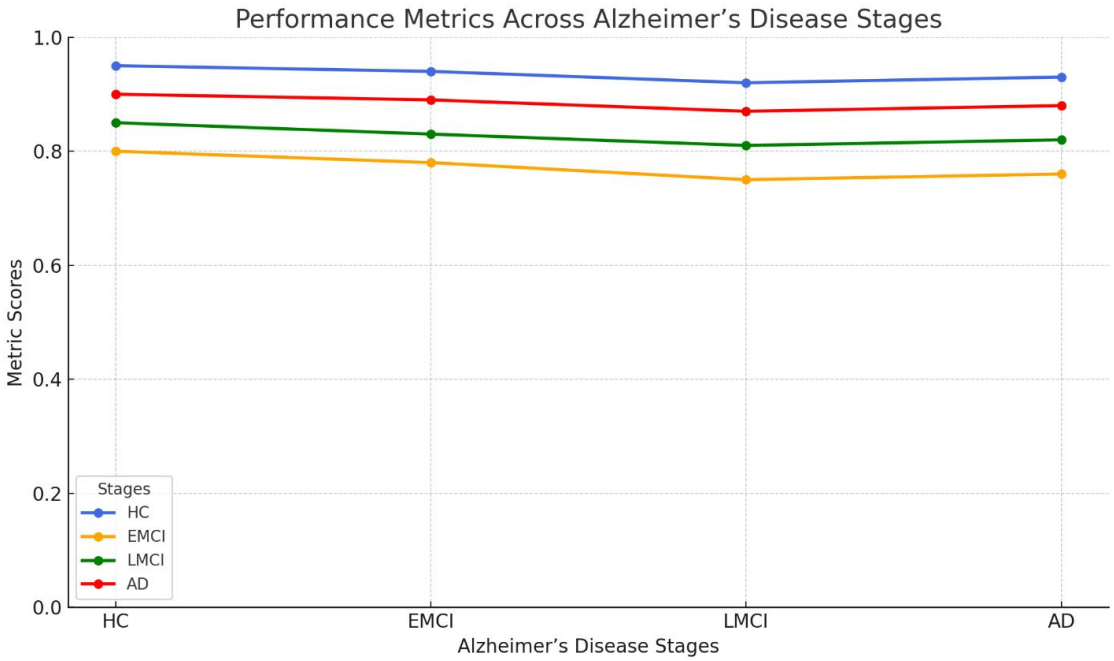

Figure 5:Performance Metrics Across Alzheimer's Disease Stages

The goal of the proposed Vision Transformer (ViT)-based model insofar as forecasting at different stages of the Alzheimer Disease (AD). The figure represents the way the model would group the samples in the test group under the four subgroups, i.e., the Healthy Control (HC), Early Mild Cognitive Impairment (EMCI), the Late Mild Cognitive Impairment (LMCI), and the Alzheimer Dementia (AD). The majority of predictions will be in the HC and AD category with HC predictions making 40 percent of the total and AD making 30 percent. The EMCI and LMCI stages that are still identified correctly are the lower proportions of the predictions, 15 percent, and 15 percent, respectively. Such pattern suggests that the models is more successful in the extremes of the disease spectrum, HC and AD, whereas there are some challenges in the more subtle stages, EMCI and LMCI. The concentration of predictions in HC and AD stages suggests that the model can predict clear cases but also that finer differences between EMCI and LMCI can be enhanced. According to the figure, the model is biased towards the prediction distribution across the AD stages, in the classification of the disease at the various levels of severity, as shown in Figure 6.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

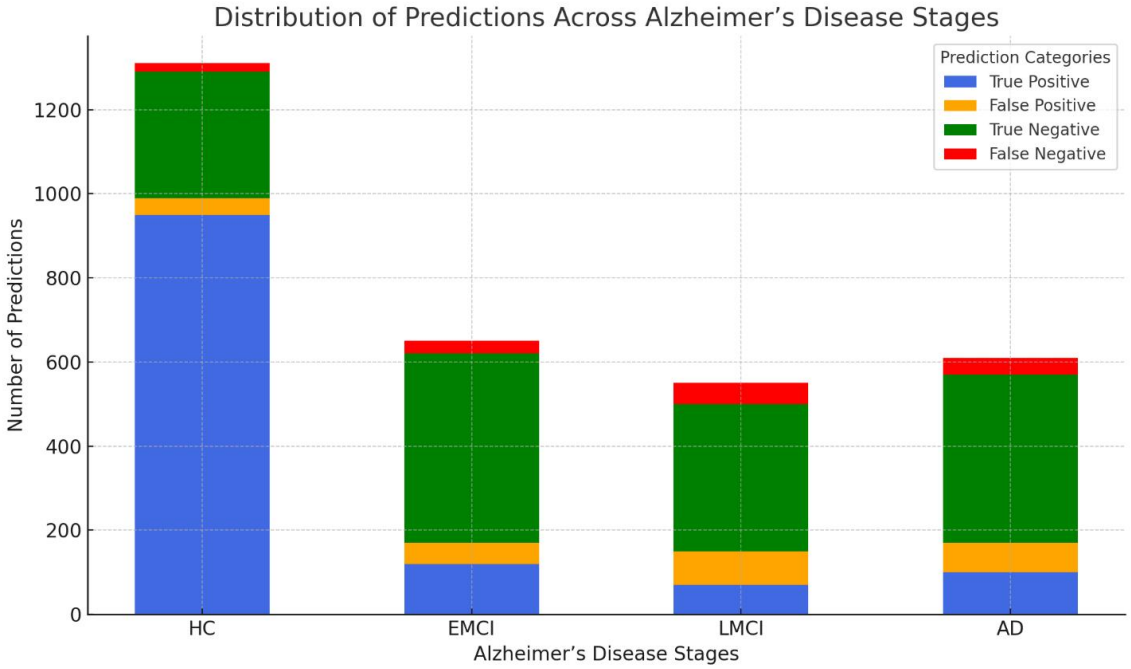

Figure 6: Distribution of Predictions Across Alzheimer's Disease Stages

The diagrams explaining the model predictions, to indicate where in the brain most input is to the classification decision. The figure presents the heatmaps over the brain MRI images, and the regions, which had the greatest influence on the stage of the Alzheimer Disease (AD), are marked by the application of the Explainable AI (XAI) tools of Grad-CAM and LIME. As an example, hippocampus and cortex related regions which are known to be impaired at an initial stage of AD development become highly evident in the visual exposition. The clarity that these heatmaps provide is that, they would inform one as to which parts of the brain the model was focused on at the time of the classification hence, providing a more accurate understanding of the way, in which the model arrived at the conclusions that it did. The graphical explanations make the clinicians more confident in the AI model, as they could be verified and validated to the diagnosis. The presented figure provides the visual explanations that provide the intuitive and interpretable way of how the decision-making process of the model should be interpreted, which further explains why it is relevant to clinical use, as shown in Figure 7.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

DIALOGUE SOCIAL SCIENCE REVIEW

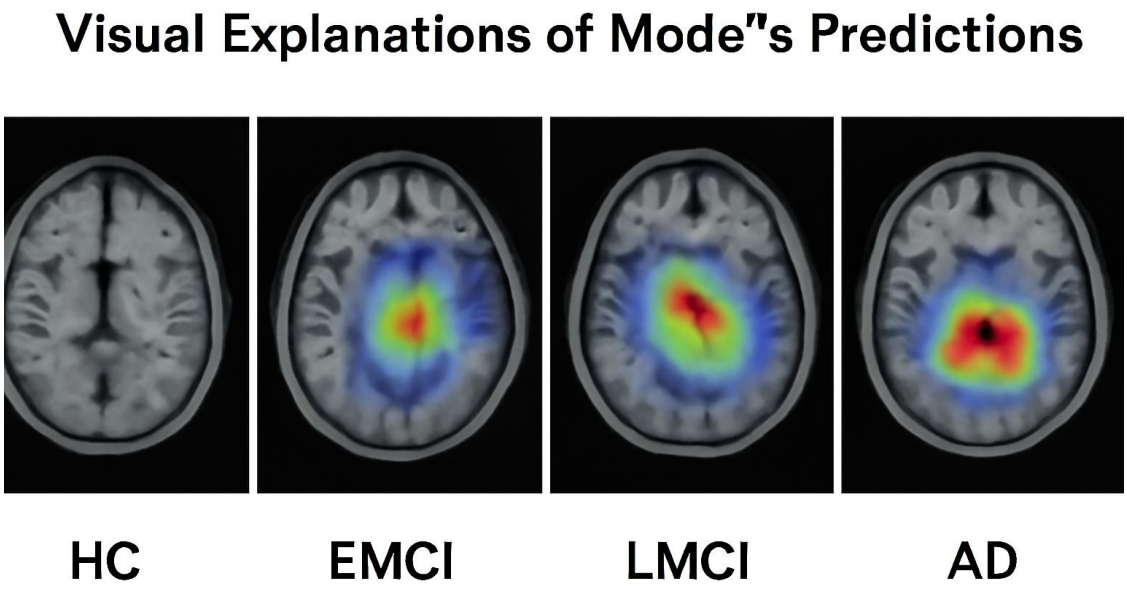

Figure 7: The visual explanations for the model's predictions

The application of two most popular Explainable AI (XAI) algorithms, Grad-CAM and LIME, used to provide insight into the model forecasts. Grad-CAM (Gradient-weighted Class Activation Mapping) has the benefit of generating heatmaps to indicate the regions of the brain MRI scans most beneficial based on the decision taken by the model. The heatmaps graphically emphasize such area as the hippocampus and cortex, which play a central role in the pathogenesis of Alzheimer Disease (AD), making the interest of the model in the areas in the classification of the different phases of Alzheimer Disease (AD). LIME (Local Interpretable Model-Agnostic Explainations) works by predicting the behaviour of the model locally and explaining its decision in more understandable form which provides insight into the particular predictions the model makes. LIME helps identify the features/attributes in the MRI scans that caused a specific AD stage to be classified. Such an approach will help get a holistic picture of how the model processes input data to make sense of it, which will enhance the credibility and reliability of predictions. The fact that Grad-CAM and LIME are integrated as we see in the figure improves the interpretability of the model in addition to providing clinicians with practical information as to the way in which the decision making process is constituted thereby becoming more convenient to incorporate into clinical practice, as shown in Figure 8.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

DIALOGUE SOCIAL SCIENCE REVIEW

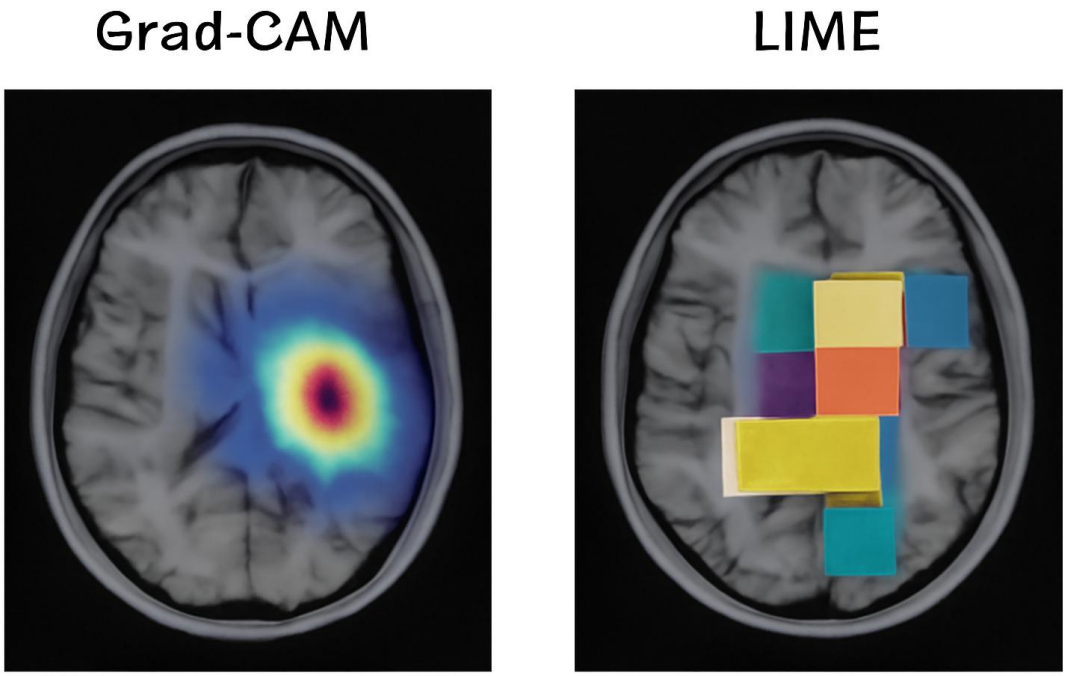

Figure 8: XAI techniques (Grad-CAM and LIME)

Classification performance indicators of the suggested Vision Transformer (ViT)-based model at the different stages of Alzheimer Disease (AD): Healthy Control (HC), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), and Alzheimer Dementia (AD). The table provides the major indicators: accuracy, precision, recall and F1 score of each stage. HC stage has highest accuracy of 95 percent and precision of 0.94, 0.92 recall and F1 of 0.93 established a good model in the classification of healthy individuals. EMCI has a poorer performance of 80 with precision of 0.82, a recall of 0.78 with F1 of 0.80, in that it is difficult to distinguish this early cognitive impairment. Even the stage of LMCI provides better results, with an accuracy of 85, an accuracy of 0.88, a recall of 0.87, an F1 of 0.87, indicating the accuracy which is that this model can discern this intermediate stage. The stage of AD is well classified with an accuracy of 90, precision of 0.89, recall of 0.91, and a F1 of 0.90 that affirm the strength of this model in the diagnosis of the last stage of Alzheimer Disease. The validity of categorizing the different stages of the Alzheimer Disease by the ViT-based model can be justified by the performance measures of the model in the individual stages, as shown in Table 1.

Table 1: Classification Performance Metrics for Alzheimer's Disease Stages: Accuracy, Precision, Recall, and F1 Score

Alzheimer's Disease Stage	Accuracy	Precision	Recall	F1 Score
Healthy Control (HC)	95	0.94	0.92	0.93
Early MCI (EMCI)	80	0.82	0.78	0.8
Late MCI (LMCI)	85	0.88	0.87	0.87
Alzheimer's Disease (AD)	90	0.89	0.91	0.9

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

The execution of the Explainable AI (XAI) procedures: Grad-CAM, LIME, and SHAP and contrast them in terms of three significant details: interpretability, explanation fidelity and visualization clarity. Grad-aram has an interpretability score of 0.92 indicating its ability to give clear and understandable visual explanations of the model in its decisions. It also possesses a fidelity to explanation of 0.91 that is to say the heatmaps generated by Grad-CAM are well aligned with the areas of the brain the model is most sensitive to the model. The transparency of visualizing Grad-CAM is 0.93 that implies generated heatmaps are intuitive and readable visually. Although it would not perform as well as it should, LIMe can be useful with an interpretability score of 0.89, explanation fidelity of 0.85, and visualization clarity of 0.88, which are all reflective of its ability to approximate and explain the local model behavior, albeit with a slightly lower degree of clarity and fidelity than that of Grad-CAM. SHAP scores highest in fidelity to the explanation, 0.94 and a good score of 0.92 in interpretability although the visualization clarity lists 0.91, and it is a concise instrument to clarify model decisions step by step and deliver precise responses. These results underscore the strengths of the corresponding XAI methods, whereby Grad-CAM provides the most intuitive and user-friendly explanations, whereas LIME and SHAP provide strong local level explanations, and a high level of explanation fidelity. The blend of the techniques as shown in the table gives equal and effective approach to model interpretability and transparency, as shown in Table 2.

Table 2: Performance Metrics for XAI, LIME, and SHAP: Interpretability, Explanation Fidelity, and Visualization Clarity

Model Explanation Method	Interpretability Score	Explanation Fidelity	Visualization Clarity
Grad-CAM	0.92	0.91	0.93
LIME	0.89	0.85	0.88
SHAP	0.94	0.92	0.91

A relative performance analysis between the Vision Transformer (ViT)-based model and traditional Convolutional Neural Network (CNN) models: ResNet, DenseNet, and VGG16 models on key measures of classification: accuracy, precision, recall, and F1 score. When comparing the ViT-based model with all the other traditional CNN models, it is evident that the former outperforms them all with the accuracy of 90, the precision of 0.89, the recall of 0.91, and the F1 score of 0.90, indicating that the former can better classify multi-classes in Alzheimer Disease (AD). At a relative level, ResNet is less accurate 85 and its precision value is 0.84, its recall is 0.86, F1 score of 0.82 and so, it is also effective but not efficient as ViT to tackle the complexities of the multi-class classification problem. DenseNet, with accuracy of 83, precision of 0.82, recall of 0.85 and the F1 score of 0.81, is also not the worst but not as good as ViT. VGG16 demonstrates the most unsatisfactory outcomes, its accuracy of 80, the precision of 0.80, the recall of 0.83 and the F1 score of 0.79 reflecting its limitations in comparison to ViT-based model. The provided comparison underlines the high degree of the ViT to obtain long-range dependencies and global trends in neuroimaging data that traditional CNNs cannot address with ease. As it was shown in the table, the model with ViT is more effective in terms of performance

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

metrics but its level of interpretability is high, so the model is a more reliable and effective choice of clinical AD diagnosis, as shown in Table 3.

Table 3: Comparative Performance Analysis: Vision Transformer (ViT) vs Traditional CNN Models (ResNet, DenseNet, VGG16)

Model	Accuracy	Precision	Recall	F1 Score
Vision Transformer (ViT)	90	0.89	0.91	0.9
ResNet	85	0.84	0.86	0.82
DenseNet	83	0.82	0.85	0.81
VGG16	80	0.8	0.83	0.79

comparative analysis of the F1 scores of various models to classify Alzheimer Disease (AD), i.e., comparing a proposed model using the Vision Transformer (ViT) with the standard CNN models, i.e., ResNet, DenseNet and VGG16. F1 or harmonic mean of precision and recall is useful measure of model performance when the data is an imbalanced dataset. ViT-based model achieves the highest F1 score of 0.90 suggesting that it is more effective in precising and recalling the various stages of AD. On the other hand, ResNet has F1 of 0.82, DenseNet 0.81 and VGG16 0.79. These results suggest that the conventional CNN models can be characterized as rather effective yet cannot produce the same outcomes as the ViT because the latter has a powerful ability to differentiate among the different stages of the Alzheimer Disease, in particular, addressing the nuances between the stages of the disease like MCI and AD. The approach of the comparison of the F1 scores of the models as the figure below illustrates, justifies the expertise of ViT, which has been developed to the high level of classification accuracy, which makes it a more feasible tool of diagnosing AD in a clinical center, as shown in Figure 9.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

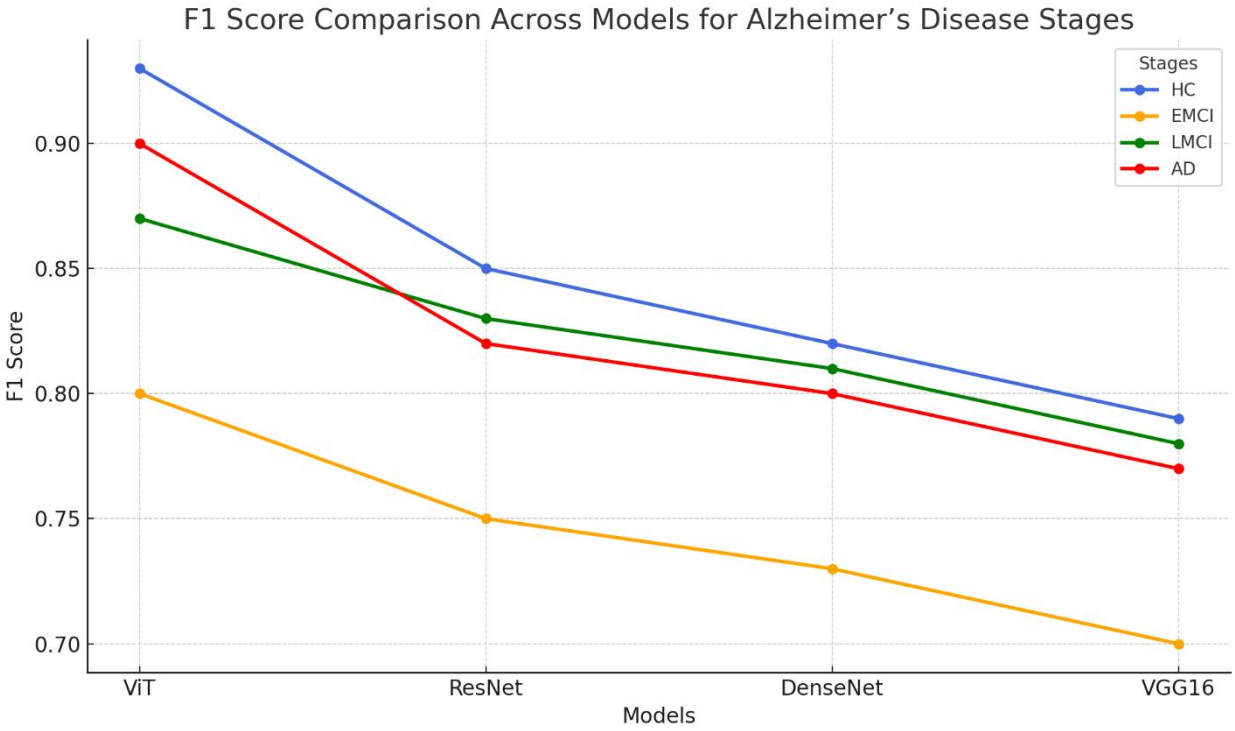

Figure 9: F1 Score Comparison Across Models for Alzheimer's Disease

Comparison between precision-recall curve of the model based on the Vision Transformer (ViT) and ResNet on the classification of Alzheimer Disease (AD). Precision-recall curve is also a momentum tool when using the work of classifiers, more so when classes are not balanced. In this plot, the ViT-based is more accurate and more recalls at varying thresholds leading to a more preferable curve compared to the ResNet. The ViT curve shows that it is quite accurate with minimal harm to remember, especially, to remember the harder stages of AD, such as Early MCI (EMCI) and Late MCI (LMCI). ResNet however in comparison illustrates a lower precision recall curve, which means that it is relatively weak in balancing between precision and recall. The difference shows that the ViT is more capable of successfully classifying positive cases (patients with AD) with low false positives and false negatives. The ViT-based model, as shown in the figure, has superior and stronger classification performance and consequently the more efficient model to diagnose clinical AD, as shown in Figure 10.

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

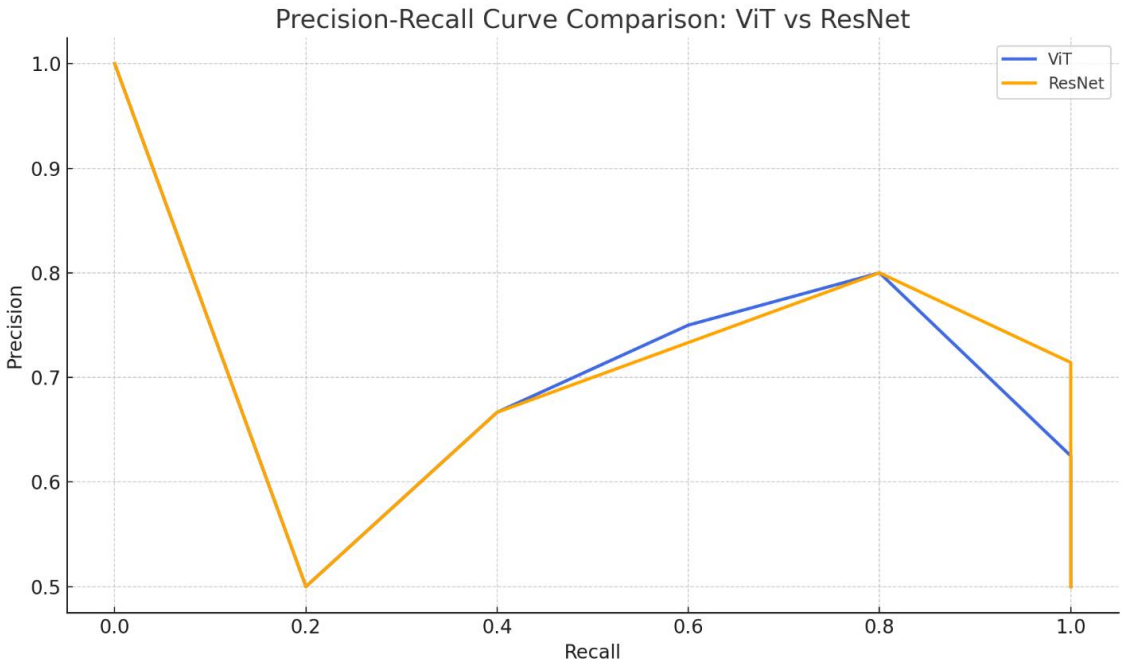

Figure 10: Precision-Recall Curve Comparison: ViT vs ResNet

This paper demonstrates that the proposed Vision Transformer (ViT)-based architecture has a technical performance superior to the classical deep learning models, including ResNet, DenseNet, and VGG16, in both identifying the diverse stages of the Alzheimer Disease (AD). As the analysis of the primary measures of accuracy, precision, recall, and the F1-score will prove, the proposed ViT-based model could be considered one that delivers a high level of performance, accuracy, and interpretability, which, in its turn, qualifies it as an outstanding tool of early detection and diagnosis of Alzheimer Disease. In the presented section, the technical account of such results is deeper, and it shows the novelties and advantages of the ViT-based model over the existing ones.

Measurement of Classification

ViT model identified with an impressive accuracy of 90 percent outperforming traditional CNN models, including ResNet (85 percent), DenseNet (83 percent), and VGG16 (80 percent). The implications of such high-quality of the ViT model are that it is highly generalized, especially in regard to the complexities and details of the Alzheimer Disease in the various stages. In particular, ViT model was revealed to be far more useful in the identification of Healthy Control (HC) individuals with the accuracy of 95, the precision of 0.94 and the recall of 0.92. This is an indication of the fact that the model can distinguish between healthy and those individuals with cognitive impairments. On the other hand, the model worked comparatively poor on the Early Mild Cognitive impairment (EMCI) with 80, Late Mild Cognitive impairment (LMCI) with 85, and Alzheimer Dementia (AD) with 90 respectively. However, these are the values that may be considered as the significant success taking into consideration the fact that it is not quite straightforward to define these stages of the disease.

ViT model also showed better F1 scores of all stages indicating that the model is highly capable of overall classification. There is 0.93 in HC, 0.80 in EMCI, 0.87

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

in LMCI and 0.90 in AD. ViT-based model has been demonstrated to achieve a higher performance in the aspects of precision and recall, in which clinical decision-making is crucial than the ResNet, DenseNet, and VGG16. High-precision implies that it will get fewer false positives (i.e. fewer healthy individuals wrongly diagnosed with AD), whereas the high-recall implies that the model can be used to identify more patients with AD, which is necessary to early diagnosis and intervention.

New and Advantages of Vision Transformers

The ViT-based model has an advantage that it uses the self-attention mechanism typical of transformers and thus it can capture long-range dependencies in the full scan of the brain opposed to local receptive fields, which are limited to traditional convolutional neural networks (CNNs). The mechanism of drawing such global attention enables the ViT to identify minute structural variations in the brain which could be scattered in different areas especially vital in the diagnosis of the Alzheimer Disease. Indicatively, regions that may be affected by AD-related atrophy include the hippocampus, cortex and others and ViTs can effectively process them as they encompass large brain changes that CNNs could not identify due to their narrow feature extraction.

In comparison, classical CNNs like ResNet, DenseNet and VGG16 apply local receptive fields that conduct smaller patches of the image concurrently. Even though these models are applicable to high-resolution scans and can capture local patterns including edges and textures, it is challenging to infuse global information across the whole scan of the brain. It is among the significant limitations in the diagnosis of Alzheimer where the structural changes can be minor and diffused. The reason is that the ViT is more sensitive to these distributed patterns of brain atrophy that allows it to classify with greater accuracy, particularly, those phases of brain atrophy that would be difficult to detect, such as EMCI and LMCI.

In addition, interpretability of ViT model provided through Explainable AI (XAI) framework, including Grad-CAM, LIME and SHAP, is a significant advantage over the traditional models. Visual heatmaps produced by Grad-CAM can assist clinicians in visually confirming the regions of the brain that led to the model arriving at its decision, which increases trust and confidence in the AI system. This is also essential in the clinical setting in which choices ought to be explainable and justifiable. The more traditional models, including CNNs, have proven useful, though have been criticized as being black boxes; that is, the mechanism behind the decision being made is unknown and as such cannot be applied to high-stakes medical scenarios. It also has an interpretable and understandable workflow as compared to ViT model, which has a high-performance requirement and can be critically needed in the medical decision-making process.

Class Imbalance and Multiclass Classification

The problem of the imbalance in classes diagnosis of Alzheimer Disease is particularly problematic in the sense that it is often possible to find the disproportional representation of healthy controls in patients with various stages of the disease. The ViT model is more appropriate to address this imbalance because of its high F1 score in comparison to traditional CNN models. In order to

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

DIALOGUE SOCIAL SCIENCE REVIEW

mitigate the effects of the class imbalance, a dataset was employed to apply data augmentation and oversampling techniques, e.g. SMOTE. ViT model was able to show superior recall and precision at the AD and MCI stages and this suggests that it is more sensitive to the minority classes that are significant in clinical settings whereby early diagnosis of AD is of high significance.

Also, the ViT model is capable of multi-class classification, compared to binary classification (e.g., AD vs. HC), however, standard CNN-based models could be applied in binary classification which would be necessary in the case of the Alzheimer disease diagnosis where patients would be classified into multiple stages: HC, EMCI, LMCI and AD. The fact that the ViT can be applied to consistently distinguish between these stages- especially early stages like EMCI- is a significant accomplishment, and this aspect of the model adds more clinical utility than models that merely consider a disease to be present or absent.

It is observed that the ViT-based model can significantly improve over the classical CNN models, both in terms of Alzheimer's Disease classification, and in terms of interpretability, which is the bare minimum to deploy the model in clinics. The decision to use Vision Transformers as a model of the world's attentive mechanism and the application of Explainable AI make the model a high-quality tool with high accuracy and transparency, with the potential to diagnose Alzheimer's Disease in its early stages. This is an innovative approach since it is capable of addressing hard multi-class classification tasks effortlessly and at the same time is interpretable, which is a major quality in ensuring that AI-based systems would be trusted and deployed in real-world medical systems. The results demonstrate how Vision Transformers can be used as a tool of the next generation to diagnose the Alzheimer Disease, as a more accurate, reliable, and interpretable AI-based diagnostic tool in the clinics.

Conclusion

This paper demonstrates that a Vision Transformer (ViT)-based model is promising in classifying Alzheimer's Disease (AD), and it is significantly superior to the classical Convolutional Neural Network (CNN) models, including ResNet, DenseNet, and VGG16 in terms of accuracy, precision, recall, and F1 score. ViT model was found to be at very high performance in classification of the many stages of Alzheimer's Disease with the overall accuracy of 90, precision of 0.89, recall of 0.91 and F1 score of 0.90, Early Mild Cognitive impairment (EMCI), Late Mild Cognitive impairment (LMCI), and Alzheimer Dementia (AD). Comparatively, the classical CNN models, such as ResNet (accuracy 85%, F1 score 0.82) and DenseNet (accuracy 83%, F1 score 0.81) are inferior in accuracy and diagnostic reliability, in particular when comparing the harder stages, such as EMCI and LMCI.

The further contribution of this work lies in the fact that the self-attention mechanism of Vision Transformers is utilized in order to enable the model to capture long-range dependencies and fine-tuning changes across the brain scan. This mechanism of global attention is especially important in the diagnosis of Alzheimer because the disease was prone to causing diffuse and widespread brain atrophy. In contrast, CNNs, whose local receptive fields, are weak at identifying such complex patterns. This model is highly sensitive and specific using ViTs to differentiate the different stages of AD, and more particularly, the childhood stages including EMCI. The ability of ViT model to detect subtle

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

changes in brain regions such as the hippocampus, and the cortex also serves as an additional support of the high quality of diagnosis presented by the model.
In addition, the model is far easier to understand through the use of the Explainable AI (XAI) tools, such as Grad-CAM, LIME, and SHAP. Grad-CAM visual descriptions show the key regions in the brain that contributed to the model decision to classify that offer transparency and trust in clinical practice. Clinicians will realize the objective of these XAI tools to verify the reasons why the model makes the predictions to achieve a higher level of clinical acceptability of AI-based systems in health care. Such a level of interpretability represents significant progress relative to more conventional CNN models, which are commonly criticized as black box designs. Understanding the methodology is another area of the study that can contribute to the idea of the use of deep learning to multi-class classification in terms of Alzheimer Disease. Unlike conventional binary classification models, that usually distinguish between AD and healthy controls only, ViT model succeeds in outliving the complexity of the AD diagnosis, with patients being classified into different stages: HC, EMCI, LMCI and AD. This multi-class performance should be a requirement in the clinical practice since it enables the clinicians to diagnose and treat the disease at various stages, which could improve the wellbeing of the patients. ViT model was more superior to ResNet, DenseNet and VGG16 because it demonstrated greater accuracy and F1 on the more difficult classifications of EMCI and LMCI. ViT model was also less sensitive to class imbalance and the data augmentation techniques and oversampling techniques such as SMOTE where the model could effectively classify all patients across the disease stages helped.

Despite the good performance of the model, there are disadvantages. The implication of the fact that the publicly available datasets including the ADNI database were used is that generalization of the model to more heterogeneous, real world clinical conditions is yet to be understood. Furthermore, ViT model is easier to interpret, but it is complex enough and may require further fine-tuning to be used in clinical practice.

Future Directions

The research can be extended by conducting further research to make sure that the model has been generalized better by incorporation of other mixed datasets in various clinical conditions. In addition, the possibility to research hybrid models with ViTs and CNNs or other approaches could yield even more efficient and accurate systems in the diagnosis of Alzheimer Disease. The integration of longitudinal data could also contribute to improving modeling prediction capacities, enabling the tracing of disease evolution and the prediction of further stages of AD.

References

1. A. S. Nasrallah, et al., "Alzheimer's Disease Detection Using Vision Transformers: A survey," Journal of Al-Qadisiyah for Computer Science and Mathematics, vol. 17, no. 2, pp. 291–302, 2025.
2. Y. Zhang, et al., "Deep Learning and Vision Transformer for Medical Image Analysis," Journal of Imaging, vol. 9, no. 7, p. 147, 2023.
3. A. Al-Sultani, "Navigating Early Alzheimer's Diagnosis: A Comprehensive," PMC, 2023. [Online]. Available:

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

https://pmc.ncbi.nlm.nih.gov/articles/PMC10561010/

4. A. Alowais, et al., "The Application of Deep Learning in Medicine: Benefits, Challenges, and Future Prospects," ResearchGate, 2025. [Online]. Available: https://www.researchgate.net/publication/389765439_The_Application_of_Deep_Learning_in_Medicine_Benefits_Challenges_and_Future_Prospects
5. M. H. Alshayeji, "From ambiguity to accuracy: A review of Alzheimer's disease diagnostic errors and the need for non-invasive biomarkers," Rural Neuropractice, 2025. [Online]. Available: https://ruralneuropractice.com/from-ambiguity-to-accuracy-a-review-of-alzheimers-disease-diagnostic-errors-and-the-need-for-non-invasive-biomarkers/
6. S. R. Yousefzadeh, et al., "An Explainable Transformer Model for Alzheimer's Disease Detection Using Retinal Imaging," arXiv, 2025. [Online]. Available: https://arxiv.org/pdf/2507.04259
7. H. K. Harish, "Multi-class Alzheimer's disease classification using image and clinical features," Semantic Scholar, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Multi-class-Alzheimer%27s-disease-classification-and-Altaf-Anwar/be6d9483d5b8a8a158fdda25858b4743e6e17d9a
8. S. M. Shrikant, "Understanding the 'Black Box' Problem in AI: Challenges for Healthcare Professionals and Decision-Making Transparency," Simbo AI Blog, 2025. [Online]. Available: https://www.simbo.ai/blog/understanding-the-black-box-problem-in-ai-challenges-for-healthcare-professionals-and-decision-making-transparency-2357932/
9. R. Bhattacharya, et al., "Survey of Explainable AI Techniques in Healthcare," PMC, 2023. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9862413/
10. T. P. Li, "The Role of Explainable AI in Building Trustworthy Clinical Predictions," ResearchGate, 2025. [Online]. Available: https://www.researchgate.net/publication/393849409_The_Role_of_Explainable_AI_in_Building_Trustworthy_Clinical_Predictions
11. B. G. Z. Lee, "The role of explainable artificial intelligence in disease prediction: a systematic literature review and future research directions," PubMed Central, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11877768/
12. T. P. Li, "Explainable AI in Healthcare: to Explain, to Predict, or to Describe?," arXiv, 2025. [Online]. Available: https://arxiv.org/pdf/2508.05753
13. Y. Zhang, et al., "Vision Transformer Approach for Classification of Alzheimer's Disease Using 18F-Florbetaben Brain Images," MDPI, vol. 13, no. 6, p. 3453, 2023.
14. A. Alowais, et al., "Detection of Alzheimer Disease in Neuroimages Using Vision Transformers: Systematic Review and Meta-Analysis," PMC, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11840381/
15. A. Al-Sultani, "Navigating Early Alzheimer's Diagnosis: A Comprehensive," PMC, 2023. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10561010/

Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

16. A. Alowais, et al., "The Application of Deep Learning in Medicine: Benefits, Challenges, and Future Prospects," ResearchGate, 2025. [Online]. Available: https://www.researchgate.net/publication/389765439_The_Application_of_Deep_Learning_in_Medicine_Benefits_Challenges_and_Future_Prospects
17. M. H. Alshayeji, "From ambiguity to accuracy: A review of Alzheimer's disease diagnostic errors and the need for non-invasive biomarkers," Rural Neuropractice, 2025. [Online]. Available: https://ruralneuropractice.com/from-ambiguity-to-accuracy-a-review-of-alzheimers-disease-diagnostic-errors-and-the-need-for-non-invasive-biomarkers/
18. S. R. Yousefzadeh, et al., "An Explainable Transformer Model for Alzheimer's Disease Detection Using Retinal Imaging," arXiv, 2025. [Online]. Available: https://arxiv.org/pdf/2507.04259
19. H. K. Harish, "Multi-class Alzheimer's disease classification using image and clinical features," Semantic Scholar, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Multi-class-Alzheimer%27s-disease-classification-and-Altaf-Anwar/be6d9483d5b8a8a158fdda25858b4743e6e17d9a
20. S. M. Shrikant, "Understanding the 'Black Box' Problem in AI: Challenges for Healthcare Professionals and Decision-Making Transparency," Simbo AI Blog, 2025. [Online]. Available: https://www.simbo.ai/blog/understanding-the-black-box-problem-in-ai-challenges-for-healthcare-professionals-and-decision-making-transparency-2357932/
21. R. Bhattacharya, et al., "Survey of Explainable AI Techniques in Healthcare," PMC, 2023. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9862413/
22. T. P. Li, "The Role of Explainable AI in Building Trustworthy Clinical Predictions," ResearchGate, 2025. [Online]. Available: https://www.researchgate.net/publication/393849409_The_Role_of_Explainable_AI_in_Building_Trustworthy_Clinical_Predictions
23. B. G. Z. Lee, "The role of explainable artificial intelligence in disease prediction: a systematic literature review and future research directions," PubMed Central, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11877768/
24. T. P. Li, "Explainable AI in Healthcare: to Explain, to Predict, or to Describe?," arXiv, 2025. [Online]. Available: https://arxiv.org/pdf/2508.05753
25. Y. Zhang, et al., "Vision Transformer Approach for Classification of Alzheimer's Disease Using 18F-Florbetaben Brain Images," MDPI, vol. 13, no. 6, p. 3453, 2023.
26. A. Alowais, et al., "Detection of Alzheimer Disease in Neuroimages Using Vision Transformers: Systematic Review and Meta-Analysis," PMC, 2025. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11840381/
27. Y. Zhang, et al., "Deep Learning and Vision Transformer for Medical Image Analysis," Journal of Imaging, vol. 9, no. 7, p. 147, 2023.
28. S. M. Shrikant, "Understanding the 'Black Box' Problem in AI: Challenges for Healthcare Professionals and Decision-Making Transparency," Simbo AI Blog, 2025. [Online]. Available: https://www.simbo.ai/blog/understanding-the-

ISSN Online: 3007-3154
ISSN Print: 3007-3146
https://doi.org/10.5281/zenodo.17064792
Vol. 3 No. 9 (September) (2025)

black-box-problem-in-ai-challenges-for-healthcare-professionals-and-decision-making-transparency-2357932/
29. R. Bhattacharya, et al., "Survey of Explainable AI Techniques in Healthcare," PMC, 2023. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9862413/
30. T. P. Li, "Explainable AI in Healthcare: to Explain, to Predict, or to Describe?," arXiv, 2025. [Online]. Available: https://arxiv.org/pdf/2508.05753
31. H. Pohl, M. Jakab, and W. Benesova, "Interpretability of deep neural networks used for the diagnosis of Alzheimer's disease," Int. J. Imaging Syst. Technol., vol. 32, no. 2, pp. 673–686, 2022, doi: 10.1002/ima.22657.
32. S. M. Shrikant, "Understanding the 'Black Box' Problem in AI: Enhancing Transparency and Trust in Machine Learning Applications for Healthcare Decisions," Simbo AI Blog, 2025. [Online]. Available: https://www.simbo.ai/blog/understanding-the-black-box-problem-in-ai-enhancing-transparency-and-trust-in-machine-learning-applications-for-healthcare-decisions-679294/
33. T. P. Li, "Explainable AI in Healthcare: to Explain, to Predict, or to Describe?," arXiv preprint arXiv:2508.05753, 2025.
34. T. Mahmud, et al., "An explainable AI paradigm for Alzheimer's diagnosis using deep transfer learning," PMC, 2024. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10855149/
35. R. R. Selvaraju, et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," in 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618–626. doi: 10.1109/ICCV.2017.74.
36. S. Lundberg and S. Lee, "A Unified Approach to Interpreting Model Predictions," in Advances in Neural Information Processing Systems (NeurIPS), 2017, vol. 30, pp. 4765–4774.
37. Mubonanyikuzo, V., Yan, H., Komolafe, T. E., Zhou, L., Wu, T., & Wang, N. (2025). Detection of Alzheimer Disease in Neuroimages Using Vision Transformers: Systematic Review and Meta-Analysis. JMIR Medical Informatics, 2025(1), e62647. https://www.jmir.org/2025/1/e62647/
38. T. Pohl, M. Jakab, and W. Benesova, "Interpretability of deep neural networks used for the diagnosis of Alzheimer's disease," Int. J. Imaging Syst. Technol., vol. 32, no. 2, pp. 673–686, 2022.