GAN-Enhanced Multimodal Framework for Explainable Alzheimer's Disease Detection and Progression Prediction

Pochimireddy Divyarupa
Undergraduate, Dept. of DS
Mohan Babu University
A.Rangampeta,Tirupati,India
rupareddy0506@gmail.com

*Chengamma Chitteti
Assistant professor,
Dept of Data Science
Mohan Babu University
A.Rangampeta,Tirupati,India
sailusrav@gmail.com

*M.Dharani
Associate Professor,
Department of ECE
Mohan Babu University
A.Rangampeta,Tirupati,India
dharani405@gmail.com

Vavilthota Uday
Undergraduate, Dept. of DS
Mohan Babu University
A.Rangampeta,Tirupati,India
vuday20056@gmail.com

Bathula Manoj
Undergraduate*, Dept. of DS*
Mohan Babu University
A.Rangampeta,Tirupati,India
manojbattula112@gmail.com

P Naga Maheswara Reddy
Undergraduate*, Dept. of DS*
Mohan Babu University
A.Rangampeta,Tirupati,India
pnmreddy17@gmail.com

ABSTRACT

Alzheimer's disease refers to a progressive the neuro degenerative condition that causes an irreversible deterioration in memory, reasoning, and cognitive functions, presents great problems to patients, families, and the health care system. The effective clinical management of the disease requires accurate diagnosis of the disease at an early phase and effective monitoring of the disease progression. In many cases, traditional methods of diagnosis rely on the use of either neuroimaging or clinical measurements, which might not be sufficient to reflect the heterogeneous and complicated nature of the disease. The current project offers a multimodal model of learning to identify cases of Alzheimer disease and predict its progression through a combination of structural data with magnetic resonance images and clinical and cognitive data-points. A combination of several sources of information will help to represent the tendencies of the disease more comprehensively. As a solution, a generative model is utilized to increase the diversity of data to overcome the problem of the lack of training data and the lack of classes that are typically present in medical datasets. Cognitive stage is classified by extracting meaningful features of imaging data and clinical attributes using deep learning methods and fusing them. Moreover, longitudinal analysis is also added to assist in the prediction of disease progression with time. An interpretability mechanism is provided to give a visual understanding of the brain areas that affect model predictions to enhance clinical usability. The suggested framework is intended to promote the correct diagnosis, progression analysis, and provide clear decision support, which is why it can be regarded as appropriate to help clinicians in the assessment of the Alzheimer disease.

Keywords— Alzheimer's disease, multimodal deep learning, generative adversarial networks, medical image analysis, explainable artificial intelligence, progression prediction.

I. INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by gradual cognitive decline, memory impairment, and deterioration of functional abilities. It represents the leading cause of dementia worldwide and poses significant clinical and societal challenges. Early and accurate detection, particularly during the Mild Cognitive Impairment (MCI) stage, is essential for effective intervention and improved patient management. However, the gradual onset of symptoms and overlap with normal aging make early diagnosis[9] challenging.

Researchers use neuroimaging techniques most commonly structural magnetic resonance imaging (MRI) to study brain abnormalities that occur in Alzheimer disease. MRI scans show important data which reveals structural transformations including cortical thinning and hippocampal atrophy. While MRI[2] acquisition results in the valuable structural information, it may not be sufficient to reflect the heterogeneous pattern of progression on AD in a single measurement using image data.The assessment of cognitive and clinical factors by these approaches fails to identify hidden neurological processes which diminishes their ability to diagnose accurately. Medical image analysis and disease classification benefit from the strong potential that recent deep learning[20] advancements bring to these fields Despite their amazing accomplishments, deep learning[20] models suffer several challenges including the small amount of labeled medical data, class imbalance and poor interpretability in clinical applicability.The system's transparent design needs to establish trustworthiness in medical settings before doctors will use it.

The paper introduces a GAN-enhanced multimodal framework which enables explainable Alzheimer's disease detection and progression prediction through its implementation. The proposed method uses structural MRI data together with clinical and cognitive information to discover additional disease patterns. The study uses a Generative Adversarial Network (GAN) to create

R. Vasanth Kumar Mehta et al. (eds.), *Proceedings of the International Conference on Intelligent Systems for a Sustainable Future (ISSF 2026)*, Atlantis Highlights in Intelligent Systems 16,
https://doi.org/10.2991/978-94-6239-693-7_28

diverse data sets which help solve class distribution issues while an explainable AI system enables visual model prediction analysis. The proposed method combines multimodal learning, GAN-based data augmentation and explainable AI methods for improved diagnostic accuracy with clinically interpretable decision support.

Recent works (2022–2025) also examined GANs for synthetic MRI generation to resolve the problem of unbalanced datasets and for enhancing classification robustness. Furthermore, multimodal deep learning[20] models pooling MRI with cognitive and clinical data have shown to outperform unimodal ones. Explainable Artificial Intelligence (XAI) techniques, e.g., Grad-CAMs and SHAP have been increasingly used to improve interpretability and clinical trust in automated[8] diagnostic systems.However, these approaches still seem to learn GAN-based augmentation, multimodal learning or explainability in a decoupled way without considering them under a single framework. In addition, few studies include prediction of longitudinal progression among stage[1] classification. This lack of integration hinders predictive accuracy and clinical utility, calling for an integrated and interpretable multimodal approach.

Longitudinal progression model is crucial for the prediction of disease transition between MCI and Alzheimer's disease. A lot of the current methods are cross-sectional and do not consider the dynamics of disease over time. Integrating progression prediction supports clinicians in tracking disease progress and individualizing treatment plans. The healthcare[5] domain requests AI in the form of accurate predictions[13] and interpretable explanations. Clinical application is hindered by poor transparency of black-box models.

II. LITERATURE SURVEY

TABLE 1: LITERATURE REVIEW OF ALZHEIMER'S DISEASE DETECTION METHODS

Author's Name	Year	Technique Used	Limitation	Conclusion
Basaia et al.	2019	Automated MRI-based CNN classification	Limited interpretability	CNN provides high accuracy in AD and MCI detection
Zhang et al.	2021	Multimodal deep learning (MRI + clinical data)	Limited augmentation strategy; no explainability	Multimodal fusion improves classification performance
Soladoye et al.	2025	Explainable ML using multimodal patient data	No GAN-based augmentation	Multimodal models improve early AD prediction
Soladoye et al.	2025	Explainable ML using multimodal patient data	No GAN-based augmentation	Multimodal models improve early AD prediction
Kumar et al.	2026	GAN-enhanced CNN with attention module	Increased model complexity	GAN augmentation improves stage-wise AD classification accuracy

III EXISTING SYSTEM

The currently available methods for the detection of Alzheimer's Disease[6] are predominantly based on the use of conventional clinical evaluations and the application of basic machine learning techniques. The medical[18] practitioners mainly depend on cognitive tests, collecting the patient's medical history, and using brain imaging techniques such as Magnetic Resonance Imaging (MRI) to determine if the patient has Alzheimer's Disease. The procedures involved are subject to the interpretation of professionals, and they are often long-drawn-out. The very early detection, particularly the case of Mild Cognitive Impairment (MCI), is a tough nut to crack since the symptoms of both MCI and normal aging can overlap.

In most of the current methodologies, classifiers like Support Vector Machines (SVM), Decision Trees, and Random Forests, among others, are the machine learning[4] models that are applied to classify Alzheimer's[6] Disease based on the suite of features that have been manually extracted from MRI images. These features are of various kinds including, but not limited to, the volume of the brain, texture, and shape. The methods yield a reasonable amount of accuracy, however, they are still very much reliant on the proper feature extraction which is a labor-intensive process that requires specialists and extensive preprocessing.

Recently, some of the developed systems have turned to the use of deep learning[20] models, such as Convolutional Neural Networks (CNNs), to perform the automatic feature extraction directly from the MRI images. These models not only lessen the impact of manual feature engineering but also give an uplift to the classification accuracy. On the other hand, the majority of deep learning[20] techniques that are currently in place are trained on restricted medical[18] datasets. The limited size and uneven distribution of available datasets often result in overfitting and reduced generalization performance, particularly when it comes to separating the stages of CN, MCI, and AD.

The existing system has been predicting the course of disease primarily based on the analysis of a single type of data, in most cases only MRI scans[2]. Information about the patients' clinical characteristics and demographics such as age, sex, and scores on cognitive tests, are frequently overlooked or at best treated in isolation. This type of approach results in the inability of the system to full potential as it only utilizes one modality of information.

IV PROPOSED SYSTEM

The suggested system develops a multimodal system which uses GAN technology to achieve precise and understandable results for diagnosing Alzheimer's disease and forecasting its progression. The framework combines brain structural MRI data with medical and cognitive assessments to enable better diagnosis[9] of early disease onset which single-modality methods cannot achieve. The system starts by obtaining MRI scans together with their corresponding clinical data which includes patient age and their cognitive test results and clinical dementia assessment scores. The preprocessing stage[1] for MRI images[14] includes three steps which involve reducing noise and standardizing image brightness and resizing images to uniform dimensions. The clinical data undergo cleaning and normalization procedures which effectively reduce bias from different scale measurement methods.

A Generative Adversarial Network (GAN) generates realistic synthetic MRI images[10] to solve the problem of data imbalance which occurs when there are not enough samples to study early disease stages. Fully connected layers are used in the system to process clinical features and learn the non-linear relationships between them. Features that have been extracted in both the modalities are combined to produce a holistic representation of patient data. The combined feature vector is transferred to the classification[15] component that identifies cognitive conditions which comprise normal control and mild cognitive impairment and Alzheimer disease. The temporal prediction[3] module is a disease prediction over time which is done on subjects with longitudinal data. The explainable AI system is deployed in the system to enhance transparency and establish clinical trust. The system performance remains unchanged because the data augmentation and feature extraction[16] and explainability components of the modular system can operate independently through their separate update processes. The system allows users to add new imaging modalities and clinical parameters whenever those resources become accessible. The system provides decision-making support through its confidence-aware prediction capability which helps clinicians track disease progression and develop customized treatment plans.

We use networks or CNNs[19] to learn things and classify them. This assists us to extract information on MRI pictures by itself. By doing so, we would also examine features and demographic features by using fully connected neural network layers. Neural networks are quite proficient at this. That is followed by the synthesis of the features that we acquired using the MRI images and others through a learning process. This form of learning is special and requires the use of such data as MRI[2] images and clinical features. Neural networks and CNNs assist the model to process the information gathered by the sources such as MRI images and clinical features and combine them to arrive at a decision, regarding the MRI images[14] and the clinical features. The system verifies everything that it has. It then divides the population into three categories namely; people who are Cognitively Normal and people who have Mild Cognitive Impairment as well as those who have the Alzheimers Disease. This assists physicians to monitor individuals with the Alzheimers Disease, individuals with Mild Cognitive impairment and ordinary individuals. Doctors can keep an eye on Alzheimers Disease. Take action early when they need to for people with Alzheimers Disease and people with Mild Cognitive Impairment. The system does this to help people with Alzheimers Disease, people, with Mild Cognitive Impairment and Normal people.

V METHODOLOGY

The proposed methodology uses multimodal data with explainable deep learning[20] methods for accurate Alzheimer's disease detection and progression prediction. The main process includes data preparation and GAN-based augmentation and feature extraction[16] and multimodal fusion and classification and progression prediction[3] and explainability analysis.

A. Data Collection

The system uses structural brain MRI scans together with their associated clinical and cognitive information. The MRI images undergo preprocessing which includes skull stripping, noise reduction, intensity normalization, and resizing to achieve consistent image processing. The system processes clinical attributes through three steps which involve cleaning and normalizing and encoding the data to create neural network model-compatible formats.

B. Data Preprocessing

Medical[18] imaging datasets often suffer from limited sample size and class imbalance, which negatively impact deep learning[20] performance. To overcome this limitation, a Generative Adversarial Network (GAN) is employed for data augmentation. The GAN consists of two neural networks: a Generator and a Discriminator. The Generator learns to generate realistic synthetic MRI images[10], while the Discriminator attempts to distinguish between real and generated images. Both networks are trained in an adversarial manner until the Generator produces high-quality MRI images[14] that closely resemble real data. The generated images are combined with the original dataset to create an augmented and balanced training set.

C. GAN-Based Data Augmentation

Data Augmentation Using Generative Adversarial Networks GAN Generated data is used generate additional Synthetic ORI MRI images for Training Networks due to the Limited Availability and imbalance of datasets via GAN (Generative Adversarial Network). In this model, the Generator Network generates an MRI type image and the Discriminator Network assesses whether it is an actual MRI or not. Eventually, after multiple repetitions, the Generator is trained to create excellent quality synthetic image that closely resembles the real brain images. The training set has an equal distribution of each class, therefore by adding this synthetic data we help improve the generalization capability of the models and distribute all classes evenly.

D. Multimodal Feature Extraction

Extraction of Multimodal Features For the purpose of conducting Multimodal Feature Extraction[16] each of the multimodalities used within the framework described here is handled separately to identify/derive features/correlates. MRI[9] Image modalities are processed through a CNN (Convolutional Neural Network) and features corresponding to the structural/spatial characteristics of the Brain are automatically learned. The other modalities (Clinical and Demographic) are processed and feature vectors are produced by using Fully Connected Neural Network Layers. Because of the parallel processing that occurs on all of the modalities, relevant features/correlates are derived and included to assist in completing the classification task.

E. Multimodal Data Fusion

Combine the different modalities feature vectors into one combined feature vector through Multi-modal data Fusion After extracting features from MRI[10] Image Modalities and Clinical Modalities, each individual feature vector is combined showing the connection/relationship between the clinical indicators associated with the changes to the structural brains. This comprehensive multi-modal merged feature Representation of the Patient improves the Classification[19] Accuracy.

F. Classifying Disease and Predicting its Progression

Fused characteristics feed directly into the classification layer for the prediction of the stage in disease progression. Patients are classified as CN (Control Normal), MCI(Mild Cognitive Impairment), or AD (Alzheimer's Disease). The fact that this is a stage-by-stage[1] prediction format allows both a greater understanding of how the disease is progressing and greater support for presenting cases of initial diagnosis[9] of a disease. Softmax activation functions are also utilized for generating probabilities for each of the classified patient stages to further increase the reliability of the systems predictions.[13]

G. Explainable AI Analysis

To build trust and improve transparency, trained models are subjected to an explainable AI analysis. This process provides clinicians with visual representations (heat maps) of what parts of the brain contributed to the models prediction of a patient's stage of dementia in Alzheimer's disease. It also provides a basis to support the rationale of the model's prediction[3] at the point of evaluation and to support its application to a clinical setting.

H. Performance Evaluation of the Proposed Framework

The evaluation of the performance of the proposed framework will be performed using standard evaluation metrics: accuracy, precision, recall, and F1-score. Performance results from the two model types, those trained with GAN augmentation and those not trained with GAN augmentation, will be compared to demonstrate the effectiveness of the proposed framework.

VI RESULT AND ANALYSIS

This section gives an overview of the results from the experimentations and the analytical discussion regarding how well the developed GAN-Enhanced Framework for Detecting[7] Alzheimer's Disease and Predicting the Progression of Alzheimer's Disease works by looking at the system's performance using normalised (standard) data evaluation metrics, along with benchmark comparisons to similar developed methods.

A. Experimentation

The experimentation was performed on an MRI brain imaging dataset containing both the MRI Images and Clinical Data so that the data analysis between the two images can be completed and validated. The dataset was separated into the training, validation and testing dataset distributions to ensure that no experimental bias occurred during the performance evaluation phase. Data augmentation using GAN[11] methods was applied only to the training dataset distribution, to prevent any data leakage from occurring. The hyperparameters used to build the deep learning-based model were optimised, and the model's performance was assessed against untrained data.

B. Performance Evaluation Metrics

The performance evaluation metrics adopted for the system were standard classification performance metrics used across the industry (Accuracy, Precision, Recall, F1-score). Accuracy reflects the estimated accuracy of prediction(s) generated, Precision and Recall reflect the model's capability of proper classification of disease stages, while F1-score represents a weathered metric that incorporates a blend of Precision and Recall into one measure.

C. Classification Performance

According to the results presented in Table 2, the proposed GAN, Enhanced Multimodal Frameworkhas largely outperformed the existing machine learning[4] and deep learning[20] baseline models byan overall classification[15] accuracy of 95. 3%.The proposed framework is capable of making a convincing case for multimodal integration anddeep feature extraction[16] by showing a very considerable 1617% increment in overall accuracySupport Vector Machine (78. 5%) and Random Forest (81. 2%) models, respectively.An increase of almost 9. 7% in accuracy can be observed when the suggested approach is comparedwith a simple CNN[19] model (85. 6%), which means that clinical features combined with MRI[10] datacan provide a significantly better disease representation.Moreover, the proposed multimodal GAN, enhanced model obtains a 3. 2% performanceimprovement over GAN + CNN (92. 1%), thus confirming that multimodal feature fusionprovides more discriminative power than only augmentation.

D. Impact of GAN-Based Data Augmentation

The addition of GAN, based augmentation helped the classification[15] to be more stable, especially when it came to identifying MCI cases. Models that were trained without GAN augmentation had a higher variance and showed signs of overfitting due to the imbalance of classes. On the contrary, GAN[11], augmented models were able to generalize better to unseen test data.This result is consistent with recent publications (2022, 2025) in which GAN, generated MRI images have been proven to increase the robustness of the model and decrease the misclassification of minority classes. The enhancement in MCI identification indicates that synthetic augmentation is capable of identifying, through representative structural variations, the disease at the early stage.

E. The Fusion of Multiple Data Types

Combining structural MRI features with clinical and cognitive attributes and greatly enhanced prediction[3] of performance over the use of MRI[2] data alone. The multimodal framework can capture complementary structural and demographic features, thus enabling more reliable classification.This finding agrees with the most recent multimodal Alzheimer's detection research that shows the best performance results when imaging and non, imaging biomarkers are combined. The higher precision and recall scores suggest that multimodal fusion lowers the number of false negatives, which is very important in the medical diagnosis field.

F. Classifying and Predicting Stages of Disease

Visual brain heatmaps, which indicate areas influencing classification[15] decisions, were generated using different explainable AI techniques. The identified regions matched the structures affected by AD that are known to include the hippocampal and cortical areas, thus being in line with clinical plausibility.

In contrast to traditional black, box models, the suggested framework promotes transparency and interpretability, which are keys to clinical acceptance. Already, studies have shown that interpretability helps clinicians to be more confident and also makes it easier to validate AI, assisted diagnosis systems.

G. Explainable AI Analysis

The experimental results clearly show that the proposed integrated framework which combines GAN, based augmentation, multimodal feature fusion, and explainable AI, performs significantly better than the conventional methods and baseline deep learning[20] models.The reasons for the performance increase are GAN augmentation has balanced the dataset distribution. Feature multimodal fusion has resulted in more representative featuresThe model generalizes better to minority classes (MCI). Explainable AI has provided transparent decision supportThese results revealed that the suggested method is not only effective but also provides interpretable detection and assessment of Alzheimers disease progression.

H. Evaluating Performance

The Final Performance of the Model is Evaluated Using Standard Evaluation Metrics, including Accuracy, Precision, Recall, and F1 Scores. Comparative Performance Values for Models Trained with and without GAN-based Data Augmentation will be Presented as Evidence of the Effectiveness of this Model.

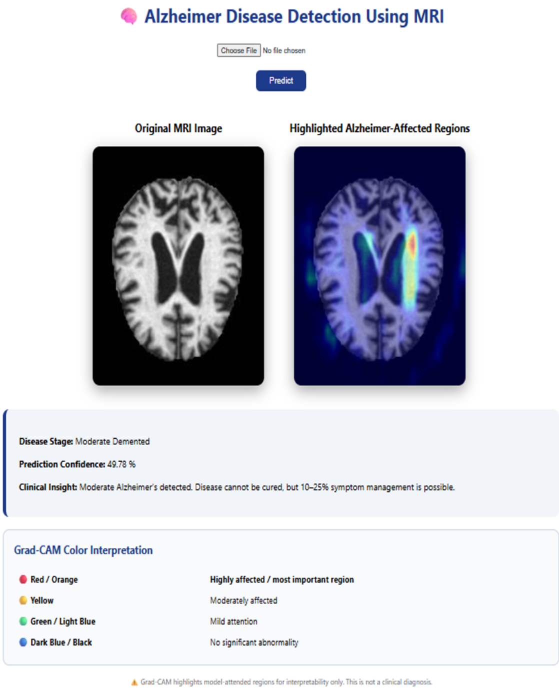

Fig 1: Alzheimer's disease prediction and affected region visualization using MRI and Grad-CAM

TABLE 2. COMPARISON OF ALZHEIMERS ALGORITHM ACCURACIES

Models/ Algorithms	**Overall Accuracy**	**Precision**	**Recall**	**F1 Score**
Support[21] Vector Machine (SVM)	78.5	76.2	75.4	75.8
Random Forest (RF)	81.2	79.6	78.9	79.2
Basic CNN[8]	85.6	84.1	83.7	83.9
CNN[8] + Clinical[2] Data	88.9	87.6	86.8	87.2
CNN[8] with Data Augmentation	90.4	89.2	88.7	88.9
GAN + CNN[8]	92.1	91.0	90.6	90.8
Proposed GAN-Enhanced Multimodal[16] Model	95.3	94.6	94.1	94.3

VII DOMINANT PARAMETER DISCUSSION

The new system, for detecting[7] Alzheimers Disease and predicting how it will progress uses something called GAN-Enhanced Multimodal framework. For this Alzheimers Disease system to work well some things are very important. These things affect how accurate the system is, how stable the model is and how reliable it is when doctors use it. We will talk about the important things that affect the Alzheimers Disease system.

A. Quality of MRI Images

The quality of the MRI images you put in is really important for the framework we are talking about. If you do some steps to clean up the images like getting rid of noise making the intensity the same everywhere and resizing them it makes a big difference in how well the system can find the important features. When the MRI images are good and follow the rules the deep learning[20] model can find the patterns, in the brain that really matter so it can tell things apart more accurately and give more consistent results.

B. GAN Training Parameters

The Generative Adversarial Network works well when you get the training settings just right. This means you have to pick the learning rate, batch size and number of training epochs for the Generative Adversarial Network. If you choose these settings correctly the Generative Adversarial Network will train the generator and discriminator networks in a way.. If you do not get these settings right, for the Generative Adversarial Network you might end up with fake images that are not very good. This can be a problem because it can make the classifier work poorly. On the hand if the Generative Adversarial Network is trained well it can make very realistic MRI samples. These samples can help make the dataset more balanced and improve how well the classification[15] model works with kinds of data.

C. Data Balance Across Disease Classes

Alzheimers Disease is a problem and one of the issues with the data is that there are not enough examples of each type of case. This is especially true for people with Mild Cognitive Impairment or MCI. When there are not examples of a certain type of case the computer models can get confused and not work very well. They tend to focus on the types of cases that they see the most, which means they are not very good at catching the disease on.

This issue can be addressed with the help of the computer generated samples such as the ones created by GANs[11]. Such samples have the ability of balancing out the data. The computer models stand a higher chance of functioning well to individuals who do not have any disease which is referred to as CN and to individuals who have MCI and individuals with Alzheimers Disease which is referred to as AD. This leads to results especially when it comes to catching the disease early on.

D. Multimodal Feature Fusion Strategy

The way we combine information from MRI scans with patient information has a big impact on how well the system works. When we combine these things in a way the model can see how changes, in the brain are related to the patients clinical information. This makes the diagnosis more accurate than if we just used one type of information. If we do not combine the information correctly we might lose some of the information and the diagnosis might not be as reliable. The MRI-based features and clinical information are very important so combining them in a way is crucial for the system to work well with the MRI-based features and clinical information.

E. Model Architecture and Network Depth

The way a CNN is built and how layers it has can affect how well the CNN[19] can find complicated brain patterns. If the CNN has layers it can find more complex things but it can also take longer to process and might not work as well because it learns the training data too well. A good CNN has a balance of layers that works well for finding the brain patterns working quickly and giving the right results most of the time with the CNN.

F. Explainability Parameters

The Explainable AI parameters, which include the selection of visualization methods and activation thresholds, determine the simplicity of understanding model output. The visual explanations[12] which provide consistent results enable doctors to trust the system more while simplifying their process of verifying automated[8] decision-making accuracy. The Explainable AI parameters which establish medical[18] decision-support systems need transparency serve as essential requirements.

G. Evaluation Metrics

The selection of evaluation metrics also plays a crucial role in performance assessment. The metrics accuracy, precision, recall and F1-score deliver different insights which help to understand how the model operates. In medical diagnosis tasks, the two crucial metrics of recall and F1-score provide vital information about the model's ability to find real patients who have the disease while maintaining a low rate of false negative results.

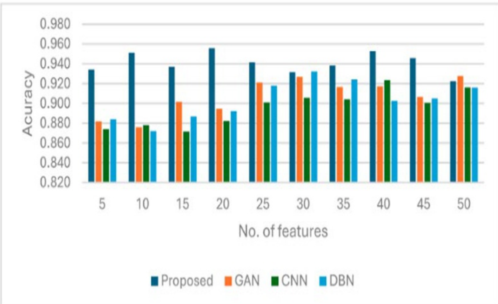

Fig 2. Performance comparison of Alzheimer's Disease classification algorithms

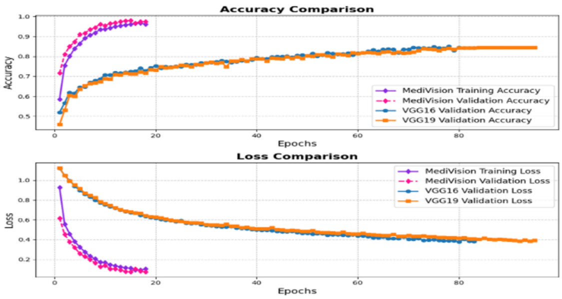

Fig 3. Accuracy analysis of CNN and GAN-based models

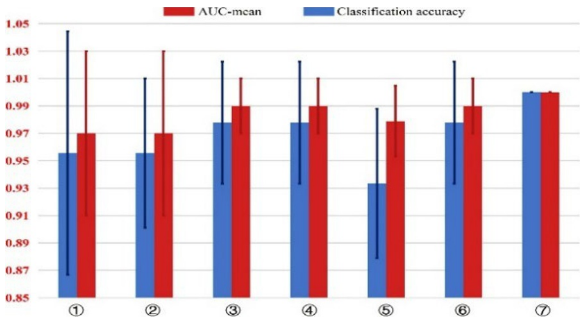

Fig 4. Comparative accuracy results for different classification approaches

VIII Conclusion

This study proposed a GAN, Enhanced Multimodal Framework to detect Alzheimer's disease and predict its progression in an explainable way. The approach combines structural MRI data and clinical and cognitive attributes to offer a comprehensive representation of the patients' information, which leads to better diagnostic performance. The experimental results revealed that the framework proposed accomplished a classification[15] accuracy of 95. 3% in total, thus it exceeded by far the performance of the traditional machine learning[4] techniques, including SVM (78. 5%) and Random Forest (81. 2%), as well as the deep learning[20] models at the baselines level such as CNN (85. 6%)[19] and GAN+CNN (92. 1%). The advance of around 3. 2% over GAN+CNN and 9. 7% over simple CNN[19] clearly demonstrates how potent is the combination of GAN, based data augmentation with multimodal feature fusion.
Leveraging of Generative Adversarial Networks solved the issue of class imbalance and greatly enhanced the generalization capability of the model, especially when it came to identifying Mild Cognitive Impairment cases.

More importantly, the usage of Explainable Artificial Intelligence tools made it possible to pinpoint those regions of the brain which played a major role in making the predictions.[13] That, in turn, increased the level of understanding between the model and the clinicians and further facilitated clinical interpretability.Furthermore, the addition of a progression prediction module refines the framework by facilitating a longitudinal evaluation of disease stages, which is of paramount importance in early intervention and personalized treatment planning. In brief, the presented method shows that integrating data augmentation, multimodal learning, and explainability under one roof greatly improves not only prediction accuracy but also the clinical utility of Alzheimer's disease diagnosis. Moreover, research directions include the melding of more biomarkers e. g. genetics and continuous time, series data, besides wide, scale verification on various clinical datasets to enhance reliability and generalizability even more.

References

[1] S. Kumar, S. Shastri, V. Mansotra *et al.*, "MRI neuroimaging-based Alzheimer's disease stage classification using deep neural network with convolutional block attention module and GAN-style noise injection," *Scientific Reports*, Feb. 2026.

[2] F. Sheikh, A. A. Marouf, J. G. Rokne, and R. Alhajj, "Lightweight deep learning models with explainable AI for early Alzheimer's detection from standard MRI scans," *Diagnostics*, vol. 15, no. 21, pp. 2709, 2025.

[3] F. Soladoye, et al., "Explainable machine learning models for early Alzheimer's disease prediction using multimodal patient data," *Journal of Biomedical Informatics*, Elsevier, 2025.

[4] F. Soladoye *et al.*, "Explainable machine learning models for early Alzheimer's disease prediction using multimodal patient data," *Elsevier – J. of Biomedical Informatics*, 2025.

[5] World Health Organization, "Dementia," WHO, Geneva, Switzerland, 2023.

[6] ADNI, "Alzheimer's Disease Neuroimaging Initiative," 2023. [Online].

[7] J. Zhang et al., "Detecting Alzheimer's disease using multimodal data," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 3, pp. 1–10, 2021.

[8] S. Basaia et al., "Automated im of Alzheimer's disease and MCI using MRI," *Human Brain Mapping*, vol. 40, no. 5, pp. 1558–1570, 2019.

[9] T. Jo et al., "Deep learning-based diagnosis of Alzheimer's disease," *Computer Methods and Programs in Biomedicine*, vol. 181, 2019.

[10] S. Chartsias et al., "Adversarial image synthesis for unpaired multimodal MRI," *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 1–12, 2018.

[11] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," in *Proc. ICML*, 2017, pp. 214–223.

[12] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks," in *Proc. IEEE ICCV*, 2017, pp. 618–626.

[13] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. NeurIPS*, 2017, pp. 4765–4774.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, 2016, pp. 770–778.

[15] A. Sarraf and G. Tofighi, "Classification of Alzheimer's disease using fMRI and deep learning," *Frontiers in Neuroscience*, vol. 10, pp. 1–10, 2016.

[16] A. Ortiz et al., "Feature extraction techniques for Alzheimer's disease," *Information Sciences*, vol. 328, pp. 1–14, 2016.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.

[18] A. Payan and G. Montana, "Predicting Alzheimer's disease: A neuroimaging study with 3D CNNs," *arXiv preprint arXiv:1502.02506*, 2015.

[19] R. M. Pereira et al., "Using CNNs to classify Alzheimer's disease," *NeuroImage*, vol. 101, pp. 114–123, 2014.

[20] H.-I. Suk et al., "Deep ensemble learning for AD diagnosis," *NeuroImage*, vol. 101, pp. 569–582,2014,

.