International Journal of Science, Strategic Management and Technology
Volume **02** Issue **04** April-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Deep Transfer Learning-Based Ensemble Framework for Alzheimer's Disease Classification using MRI Scans

Mani Deepak Choudhry
Assistant Professor
Department of Computing Technologies
SRM Institute of Science and Technology
manideec@srmist.edu.in

Aaron Fernando
Department of Computing Technologies
SRM Institute of Science and Technology
af8888@srmist.edu.in

M Ganesh
Department of Computing Technologies
SRM Institute of Science and Technology
gm7881@srmist.edu.in

https://doi.org/10.55041/ijsmt.v2i4.423

Cite this Article: Fernando, A. & Ganesh, M. (2026). Deep Transfer Learning-Based Ensemble Framework for Alzheimer's Disease Classification using MRI Scans. International Journal of Science, Strategic Management and Technology, *02*(04). https://doi.org/10.55041/ijsmt.v2i4.423

Abstract**—**Alzheimer is a progressive neural disease which leads to cognitive decline and memory loss as well. The important thing turns out to be making early diagnosis for a more effective medical care and which will improve patient management as well. MRI which is widely used to analyse the structural changes in brain with the early stage abnormalities like in Mild cognitive impairment (MCI), which are often very subtle and quite difficult to identify through general manual interpretation by researchers. Other existing deep learning based systems which rely on single models those suffer from overfitting, poor generalization due to the constrained medical imaging datasets. Now, to address these issues, our project proposes an easily understandable deep transfer learning-based ensemble model for an automated classi- fication of the Alzheimer's disease stages, including Cognitively Normal, Mild Cognitive Impairment, and Alzheimer's Disease. This system follows a proper end to end pipeline, like it starts with preprocessing the MRI scans so that the noise would be reduced and the content quality would be maintained. Then multiple pre-trained convolutional neural network architectures are then utilized to extract the features through transfer learning, which helps in effectively learning the brain structural patterns even with limited data availability.

The above paragraph is the focus of our project for perfor- mance enhancement, the frameworks have been integrated with AI techniques using Grad-CAM visualization to highlight brain regions which will influence the predictions of our model.

Index Terms—Alzheimer's Disease, Magnetic Resonance Imag- ing (MRI), Deep Learning, Transfer Learning, Ensemble Learn- ing, Mild Cognitive Impairment, Explainable Artificial Intelli- gence

I. Introduction

Alzheimer's is a progressive neurodegenerative disorder which affects cognitive abilities primarily leading to memory loss, impaired reasoning and decline in daily functioning which happens gradually [1]. With the growing aged popu- lation and an increasing cases/chances of Alzheimer's across the globe, thus the need for early diagnosis and treatment has become the major concern for the health care and welfare globally [1]. Early identification plays a crucial role in order to maintain patients and which will enable clinical intervention which will help in regulating the disease which will help in improving the patients life [2].

Magnetic Resonance Imaging (MRI) is widely used world- wide as a non invasive way of imaging to detect the structural brain changes associated with Alzheimer's [3]. MRI-based analysis allows the experts to observe the brain neurodegener- ation (particularly in regions which are linked to memory and cognitive functions).

Still recognizing and diagnosing Alzheimer's in its early stages is very challenging. MCI (Mild Cognitive Impairment it is a stage between regular aging and Alzheimer's it shows very subtle abnormalities which makes it difficult to distinguish through manual inspection alone [4]).

Traditional diagnostic approaches rely on radiologists mak- ing the process time consuming and the reliability is low (accuracy can vary) which is a huge variability.

In recent years, deep learning techniques like CNNs have demonstrated very promising and optimal results in medical image classification including Alzheimer's diagnosis using MRI scans [5]. Even being very effective with the results many approaches rely on single model architectures which lead to overfitting and limited generalization due to the constrained size of the medical imaging datasets [6], [7]. Such limitations reduce the reliability and efficiency of these systems when applied to new clinical data.

To address these challenges transfer learning and ensemble learning are being explored in order to increase the efficiency in turn to face these challenges. Transfer learning enables the adaptation of pre trained CNNs to domain specific tasks which improves feature extraction efficiency in data constrained re- gions [8]. Ensemble learning further increases the predictivity by combining multiple outputs from different models which improves generalization capability [9], [10]. Explainability has become a very crucial requirement in healthcare AI systems as the experts require transparency and it should be easy to interpret the information to trust automated decisions [12].

Several studies show that the ensemble based deep learning frameworks for Alzheimer's disease classification demonstrating improved robustness compared to other individual models [13]. Visualization techniques such as Grad-CAM provide insights into the brain region which in turn influences the

model prediction which helps by supporting interpretability in clinical settings [11].

Based on these observation and the results of these reference our project proposes a deep transfer learning based ensemble framework for the classification of Alzheimer's stages includ- ing MCI, CN and Alzheimer's using MRI scans. The system integrates a very structured preprocessing, feature extraction using multiple pre trained CNN architectures and a weighted ensemble strategy to enhance reliability and robustness.

II. Related Works

Many research studies have showed and explored deep learning based approaches for Alzheimer's classification us- ing MRI scans. CNN architectures such as VGG, ResNet, DenseNet and Inception which are being widely used for automated feature extraction from the given structural brain images input [1]. Most of the works which were proposed earlier focused mainly on binary classification like Alzheimer's vs Normal control instead of multi stage classification including MCI which is more difficult clinically and which is more important for early diagnosis.

Transfer learning has become a very common approach in medical image analysis because training deep networks from starting requires very large datasets which are not available in most of the cases in healthcare domains [2]. Pre trained CNN models which are trained on a very large scale data sets helps in capturing the general image features which can be later fine tuned for medical tasks. Thus improving convergence speed which in turn reduces the training time and minimizing the overfitting issue. But still many papers rely on single model architecture which may perform well on training the data but will fail to generalize in a more effective manner on unseen data especially when class imbalance is present [3].

To improve the robustness some recent research and studies has introduced ensemble learning techniques where outputs from multiple models are combined together to obtain a final prediction [4]. Equal weight averaging is used commonly but this is not consider the validation performance of individual models. Some advanced research works were going through fusion strategies but limited research focuses on validation based weight assignment importantly for the three stage Alzheimer's classification. In addition explainability techniques such as Grad-CAM are integrated in some systems to visualize the brain regions reliably for prediction which will in turn increase transparency and trust in AI based medical systems [5].

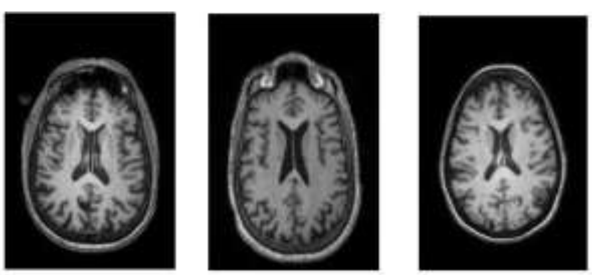
Despite these above specified works there is still a need for a custom integrated framework that combines transfer learning, weighted ensemble fusion, class imbalance handling and a simple and efficient AI which can explain properly and a reliable three stage classification of CN, MCI and Alzheimer's diseases.

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **04** April-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

database. ADNI is a large scale research project this was launched in the year 2003 with the main objective being the study of the progression of Alzheimer's disease using clinical assessments and biological markers. This provides a high qual- ity brain structural images which are widely used for research of the early diagnosis and classification of Alzheimer's disease. Structural Magnetic resonance imaging (MRI) scans were used to analyze anatomical changes associated with neurodegenera- tion in the brain. MRI is a non invasive technique which makes it better than other techniques that can be used and it allows a detailed visualization of the brain structure and helps to detect abnormalities for example like hippocampal atrophy and cortical thinning which are a few which are strongly associated with AD progression.
The dataset contains MRI scans and those scans are divided into three categories:

- **Cognitive normal (CN):** No signs of cognitive impair- ment.
- **Mild cognitive impairment (MCI):** Individuals showing early signs of memory loss and cognitive decline but not as severe for it to be diagnosed as AD.
- **Alzheimer's disease (AD):** Patients who are diagnosed with AD with exhibiting significant cognitive impairment and brain degeneration.

In general medical imaging datasets often suffer from class imbalance that is not the case for AD because the the number of samples in the AD disease category is smaller compared to the other classes. To address this issue during the modeling phase the class weights were applied to make sure minority class samples get appropriate importance during the learning process. Before the training process begins the datasets are divided into three subsets for experimenting. 70% of the images are used for training, 15% of the images are used for validation and the other 15% for testing. This split makes sure that the models are trained in an effective manner. The ADNI dataset provides very reliable MRI scans collected under some standardized protocols which makes it very suitable for developing deep learning based systems for automated AD classification and early diagnosis.



(a) Alzheimer's Disease (AD) (b) Mild Cognitive Impairment (MCI) (c) Normal Control (NC)

III. Dataset Description

The dataset which we used to train our model is obtained from the Alzheimer's disease neuroimaging initiative (ADNI)

Fig. 1. 3 types of image categories present in the dataset: (a) Alzheimer's Disease (AD), (b) Mild Cognitive Impairment (MCI), (c) Normal Control (NC).

IV. System Architecture

A. Proposed Methodology

Our weighted ensemble framework for classifying Alzheimer's disease is presented in this section. We outline the ensemble fusion approach, explainability mechanism, individual model configurations, and the overall system architecture.

B. System Architecture Overview

Through weighted probability fusion, our suggested system combines two different CNN architectures, VGG16 and Xcep- tion, using a heterogeneous ensemble approach. The entire process from the initial MRI scan to the final diagnosis, com- plete with explainability visualization, is shown in Figure 2.
The workflow is divided into four primary steps:

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **04** April-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

1) Preprocessing brain MRI scans to standardize input format and minimize noise;
2) Parallel inference using VGG16 and Xception models that were independently trained;
3) Weighted ensemble fusion using coefficients that were determined empirically;
4) Grad-CAM visualization generation for decision support that can be understood.

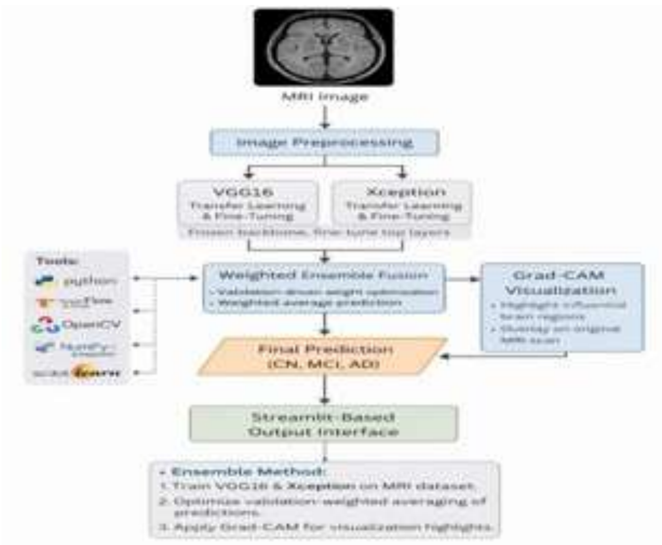


Fig. 2. System Architecture Diagram

C. Image Preprocessing Pipeline

format to ensure uniform intensity, removal of unnecessary color shades and to reduce redundancy. The pretrained CNNs which will be used later require a fixed input size to be dealt with. So we will resize all the images to 224 × 224 to ensure compatibility with the system. So then we have to normalize the intensity from 0–255 to 0–1 to stabilize training and is crucial for transfer learning.

Since MRI is grayscale now, we will now have to duplicate the single channel to three dimensions, as the pretrained CNNs require 3-channel input. This makes the shape become 224 × 224 × 3, without disturbing the actual structure. To match the Neural Networks expected input, we need to add an extra dimension to ensure smooth operation making the shape 1 × 224 × 224 × 3. We have distribute the data into training (70%), test (15%) and validation (15%). Now we will augment the training data only to reduce overfitting and to improve more anatomical variation. After analyzing our data, we found out that the AD samples were quite a few. So we applied class weights and applied them during training, which improved minority influence and improve AD recall.

D. VGG16 Architecture Configuration

Classical machine learning algorithms assume that the train- ing and testing data are generated by the same probability distribution. This assumption might not hold in scenarios where the size of training data available is very small. Thus, we may be interested in reusing a model already trained on some other (related) data. So we used a VGG16 which is a deep convolutional neural network (CNN) architecture developed by the Visual Geometry Group (VGG) at the University of Oxford.

It is favorable for our architecture as it has uniform 3 × 3 convolutions and a smooth gradient flow, it works for small datasets as medical datasets are usually available in small size due to confidentiality. We used a five Convolution Blocks setup, as this structure is a major step in medical imaging. In our classification later, block5 is the part where the disease pattern emerge to distinguish between the three stages. We also introduced Global Average Pooling (GAP) to replace the traditional flattening in order to prevent overfitting, forcing spatial summarization and improve Grad-CAM

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **04** April-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

which will be introduced later. We select Dense(128) Layer to ensure all the connected layers get accustomed to our medical imaging data, also to ensure feature recombination. We had to ensure Dropout has been set up to 0.5 to reduce memorization and help them generalize minority class to ensure stability.

The class-weighted cross-entropy loss is defined as:

$$\mathcal{L} = -\sum_{c} (w_c \cdot y_c \cdot \log(\hat{y}_c)) \qquad (1)$$

The experiments were performed on the data acquired from the Alzheimer's Disease Neuroimaging Initiative (ADNI) 2003 with the aim of analysing the efficacy of biological markers clinical neuropsychological tests and neuroimaging techniques like MRI and PET for diagnosis of AD in early stages. For further details on ADNI, visit www.adni-info.org.

By checking the data acquired, we got to know that all of them are not in grayscale format, some of them were stored in RGB format. So we convert them into the required grayscale One of the crucial parts of our System is that we used Two-Phase Progressive Fine-Tuning. In Phase 1, we froze the backbone, kept the learning rate (1×10^{-4}) and train only the head. This is done to ensure we preserve the Pretrained ImageNet features and ensure stable feature transfer. In Phase 2, we do not unfreeze the entire network we only adapt to block5 conv layers, as this is where the domain adaptation happens. During Fine-Tuning, we reduce the learning rate (1×10^{-5}) to ensure classification learning, than re-learning again and again. We brought in Class-Weighted Cross Entropy to punish the misclassification of AD and encourage minority sensitivity.

E. Xception Architecture Configuration

Transfer learning typically deals with the scenarios wherein we have to transfer the knowledge learned from a source task in a source domain to a target task in a target domain. Xception is a deep convolutional neural network (CNN) architecture developed by Francois Chollet at Google, introduced as a successor to the Inception model. The normal CNNs detect pattern and mix color information. But the model here uses Depthwise Separable Convolution, where we first look into various spatial patterns and then we mix the information what should be preferred and by how much by comparing the evaluation metrics of the individual models.

So we gave more weight to the model that performed better on validation data. So first based on validation accuracy, we can check that the VGG16 (69.97%) which is more sensitive to AD and a bit aggressive, while Xception (76.63%) is more balanced and has better overall discrimination with better MCI detection. Clearly, Xception is strong overall, so equal weighting would ignore their performance differences, that's why it is statistically sensible. This method is quite transparent and also avoids overfitting risks. Here no additional training required, fully reproducible and no hyperparameterization. After using the calculation formula here, we got to know that Xception (52.3%) is clearly got more influence than VGG16 (47.7%) on the predictions and thus maintain the overall balance

$$W_i = \frac{A_i}{\sum_j A_j} \qquad (2)$$

In this we use the Entry–Middle–Exit Flow Design, as this increases recall, precision and the overall stability of the System. The Entry Flow ensures it reduces image size step-by- step, then learns basic patterns and detects simple brain struc- tures. The Middle Flow repeats the same block multiple times, this in turn helps refine patterns and strengthens mid-level features, which focuses more on fine details. The Exit Flow combines all the learned features, produces high level features from what it learned so far and then prepares information for classification. In addition, we use Global Average Pooling (GAP) to reduce overfitting and compress spatial maps. We also implement Batch Normalization at Head, this stabilizes gradient updates and reduces internal covariate shift. This model uses shortcut connections, this in turn helps keep the training stable, prevents information loss and it allows deeper networks to learn properly.

So we implement here also the Two-Phase Fine-Tuning to ensure MRI specific adaptations. The model itself produces a rich feature vector (2048 numbers) to convert the MRI into a detailed numerical representation of brain health. In Phase 1, we freeze the backbone and only adhere to the head part. In Phase 2, we unfreeze the last six layers and reduce the overall learning rate (1×10^{-5}) to ensure we contain more semantic features. During Fine-Tuning, we reduce the learning rate (1×10^{-5}) to ensure classification learning, than re-learning again and again. We brought in Class-Weighted Cross Entropy to punish the misclassification of AD and encourage minority sensitivity.

F. Weighted Ensemble Fusion Strategy

The reasoning behind using an ensemble is to by pass the weakness of individual models. Each model has its own hypothesis about the given input. By stacking different models with different assumptions about a class label, we can find a better classification that may not be possible with individual models. So instead of giving equal importance, we decided So we also happen to use Probability Fusion, this is to create an ensemble prediction using the weights we calculated above. So instead of choosing one model's output, we multiply each probability vector [P (CN), P (MCI), P (AD)] by the model's individual weight we calculated earlier, then add them together. Each model outputs a probability distribution:

$$P(x) = [P(\mathrm{CN}),\ P(\mathrm{MCI}),\ P(\mathrm{AD})] \qquad (3)$$

Instead of hard voting, probabilities are fused:

$$P\mathrm{ensemble} = W1 \cdot P1 + W2 \cdot P2 \qquad (4)$$

Final decision:

$$\hat{y} = \arg\max (P_{\mathrm{ensemble}}) \qquad (5)$$

At last, this ensemble maintains 100% AD recall, improves AD precision to 50%, reduces false positives (by almost 50%) compared to VGG16. So this in turn improves diagnostic reliability and better risk stratification.

G. Explainability Through Grad-CAM

In medical field, you have to be always not only be right about your prediction, but also explain them how you came to that conclusion, and what factors influenced it. In medical AI, doctors show how their predictions were made using Grad-CAM. It shows which parts of the MRI actually influenced their decision.

So what it actually does is that, it takes the last convolutional feature maps. We used the final conv layer, as it contains highly influential features. They represent meaningful features, as the earlier layer detects edges and textures. ReLU is also applied here so that we remove irrelevant noise or signals and focus on the region that actually helps the prediction.

This also checks how strongly each part of the MRI actually affects the chosen class, assigns importance scores and then creates heatmaps that showcases or highlights the regions that influenced the decision. The indications include the Red areas

show strong influence and Blue areas show weak influence. So the Heatmaps are built and resized to 224 × 224, which is laid on the original MRI. Here Jet Colormap is used to ensure proper transparency.

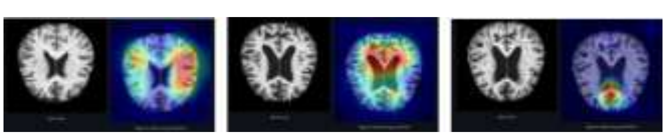


Fig. 3. Grad-CAM Visualizations: Input Scan and Regions Influencing Prediction

In short, Grad-CAM helps the ensemble focus on the Hip- pocampus, temporal lobes and ventricular areas of the MRI. This highlighted areas, proves that the model is trustworthy enough to be used in real-time at clinics, which this might help the medical staff to make the decision sooner or might even replace them. This adds transparency, detects model bias and shows us the understanding of ethical AI.

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **04** April-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

V. EXPERIMENTAL RESULTS

Thus the proposed transfer learning based ensemble frame- work was trained and evaluated using MRI images categorized into CN, MCI and Alzheimer's classes. Image preprocessing methods like resizing, normalization and augmentation were done in order to improve model generalization and reduce the overfitting issue. The individual transfer learning models VGG16 and Xception were trained using a structured two phase strategy including frozen backbone training and then fine tuning of the higher layers.

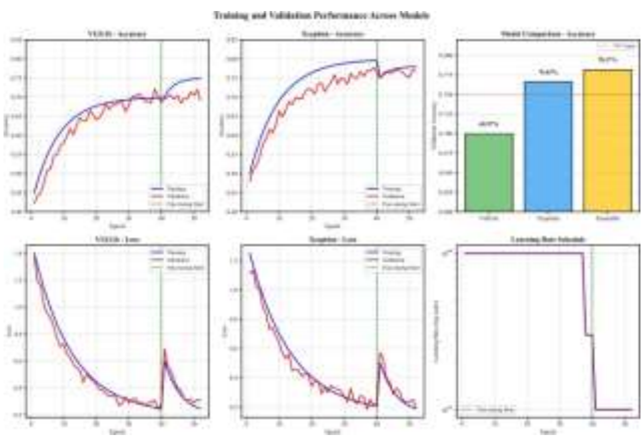


Fig. 4. Training and Validation Performance Across Models.

Performance was calculated using accuracy, precision, re- call, F1 score and confusion matrix. Each individual model achieved a satisfactory accuracy in classification with strong performance in detecting Alzheimer's cases. Some misclassi- fications were observed between CN and MCI classes due to subtle structural differences in the early stages of neurodegen- eration.

Although the ensemble model achieves the highest accuracy of 78.17%, a slight reduction in precision, recall and F1 score is being observed compared to the Xception model and VGG16. This happens primarily due to class imbalance and because of increasing the sensitivity of the ensemble model for detecting AD cases. The ensemble model provides a better trade-off between the overall accuracy and disease detection capability.

After implementing validation-driven weighted ensemble fusion, the overall classification performance improved com- pared to individual models. The ensemble approach reduced false negatives in the Alzheimer's class and improved recall for MCI which is a very crucial stage for early intervention. The weighted fusion strategy allowed better contribution from the model with higher validation performance leading to improved generalization capability.

Confusion matrix showed improved class balance and re- duced prediction bias towards majority classes. Grad-CAM visualization demonstrated that the model focuses on relevant brain regions like hippocampal areas and ventricular enlarge- ment patterns which are associated with Alzheimer's. This makes the system perform well numerically and also aligns with medical understanding of the disease. From the above discussed results, we can confirm that combining transfer learning, weighted ensemble fusion and proper explainable AI provides robustness and reliability compared to standad deep learning models.

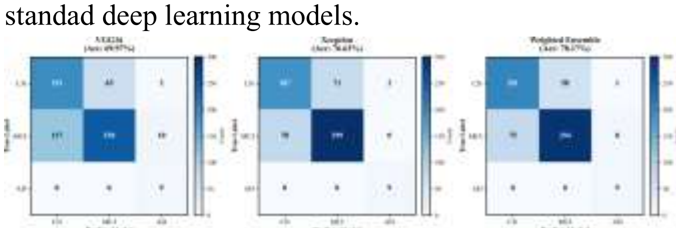


Fig. 5. Confusion Matrix.

The confusion matrix shows the performance of the en- semble model among the three classes CN, MCI and AD. From the confusion matrix that has been extracted we can say that the model identifies most of the CN and MCI cases. The model shows high sensitivity towards AD which means it will be correctly identifying most AD cases with very minimal misclassification into other classes. This is very useful in medical diagnosis where detecting the disease cases correctly is more important.

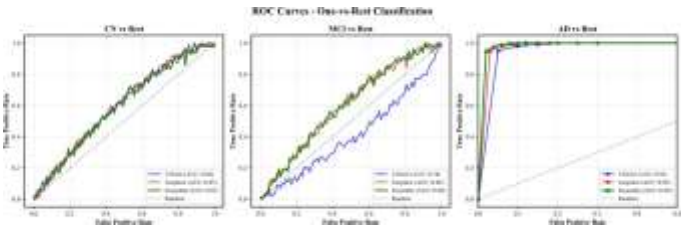


Fig. 6. ROC Curves - One-vs-Rest Classification.

The table 2 shows the difference between the accuracy of the different models. The ensemble model achieves the

TABLE I
COMPARISON OF MODELS FOR ALZHEIMER'S CLASSIFICATION

Model / Method	**Architecture Type**	**Acc (%)**	**AD Rec (%)**	**Key Strength**
CNN-based Model (Baseline)	Basic CNN	65–70	50	Simple implementation
ResNet-based Model	Residual Network	70–75	60	Deep feature learning
DenseNet (DEMNET)	Dense Connections	75–78	65	Better feature reuse
3D-CNN Models	Volumetric CNN	78–82	70	Uses full 3D MRI data
VGG16 (Proposed)	Transfer Learning	69.97	100	High AD sensitivity
Xception (Proposed)	Depthwise CNN	76.63	100	Strong overall performance
Weighted Ensemble (VGG16 + Xception)	Hybrid Ensemble	**78.17**	**100**	Balanced and robust performance

TABLE II
PERFORMANCE METRICS OF PROPOSED MODELS

Model	**Acc (%)**	**Prec (%)**	**Rec (%)**	**F1 (%)**
VGG16	69.97	73	71	72
Xception	76.63	75	76	75
Proposed Ensemble	**78.17**	68	63	64

highest performance with an accuracy of 78.17%, which outperforms both the Xception model and the VGG16 model. The Xception model demonstrates a higher accuracy compared to VGG16. However, by combining both models through weighted ensemble learning, the overall accuracy is improved which showcases the effectiveness of the proposed approach.

VI. CONCLUSION

This paper is a deep learning based framework for the automated classification of AD stages using MRI scans by integrating the transfer learning with ensemble learning the system which we have proposed aims to address the draw backs and limitations of single model approaches particularly in terms of overfitting and generalization on the limited medi- cal datsets which are available. The framework completely fo- cuses on classifying the three stages congitively normal(CN), mild cognitive impairment (MCI) and Alzheimer's disease (AD) which is crucial for early diagnosis and effective clinical intervention.

The proposed methodology combines both VGG16 and Xceptions which are both powerful convolutional neural net-work architectures. Using a validation driven weighted en- semble strategy. The system incorporates a proper structured pipeline, two phase fine tuning and class weighted loss functions to handle the data imbalance and imporve models sensitivity which helps out in differentiating between the three classes. Additionally we have used Grad CAM based visu-alization to enhance interpretability by highlighting the brain regions influencing the model's predictions,thereby supporting trust in clinical applications.

Experimental results demonstrate that the ensemble model achieves better performance compared to other individual models with an overall accuracy of 78.17 and improved detection capability for Alzheimer's cases. The model shows strong sensitivity towards AD while maintaining a balanced performance across other classes. The confusion matrix and ROC analysis further confirm the robustness and generaliza- tion capability of our proposed approach.

Despite the promising results certain limitations still remain which includes class imabalance and reduced performance in distinguishing between CN and MCI due to very subtle structural differences. Future work can focus on incorporating larger and more diverse datasets, exploring advanced ensemble strategies, and integrating multimodal data such as PET scans and clinical information. Enchancing the model interpretability and deploying the system into the real world clinical envirom- nments can improve its practical applicability and impact in early Alzheimer's diagnosis.

REFERENCES

[1] R. Brookmeyer, S. Gray, and C. Kawas, "Forecasting the global burden of Alzheimer's disease," *Alzheimer's & Dementia*, vol. 3, no. 3, pp. 186– 191, Jul. 2007. https://doi.org/10.1016/j.jalz.2007.07.002

[2] R. C. Petersen, "Mild cognitive impairment as a diagnostic entity," *Journal of Internal Medicine*, vol. 256, no. 3, pp. 183–194, Sep. 2004. https://doi.org/10.1111/j.1365-2796.2004.01388.x

[3] E. Pellegrini, F. Ballerini, E. Hernandez, M. del C. Valde´s Herna´ndez,
J. M. Gonza´lez-Castro, and D. Rueckert, "Machine learning of neu- roimaging for assisted diagnosis of dementia: A systematic review," *Alzheimer's & Dementia*, vol. 14, no. 9, pp. 1230–1240, Sep. 2018. https://doi.org/10.1016/j.jalz.2018.06.3069

[4] M. Tanveer, B. Richhariya, R. U. Khan, A. H. Rashid, P. Khanna, M. Prasad, and C.-T. Lin, "Machine learning techniques for the diagnosis of Alzheimer's disease: A review," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 16, no. 1s, pp. 1– 35, Apr. 2020. https://doi.org/10.1145/3344998

[5] R. Jain, N. Jain, and P. Aggarwal, "Convolutional neural network based Alzheimer's disease classification from MRI," *Cognitive Systems Research*, vol. 57, pp. 147–159, Oct. 2019. https://doi.org/10.1016/j. cogsys.2019.04.001

[6] H. Wang, Y. Shen, S. Wang, and Y. Liu, "Ensemble learning for early diagnosis of Alzheimer's disease using structural MRI," *Neurocomput- ing*, vol. 365, pp. 12–22, Nov. 2019. https://doi.org/10.1016/j.neucom.
2019.06.047

[7] M. Hon and N. M. Khan, "Towards Alzheimer's disease classification through transfer learning," in *Proc. IEEE Int. Conf. on Bioinformatics and Biomedicine (BIBM)*, Kansas City, MO, USA, 2017, pp. 1166–1169. https://ieeexplore.ieee.org/document/8217829

[8] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao,
D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures,

dataset characteristics, and transfer learning,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016. https://ieeexplore.ieee.org/document/
7426826

[9] G. Brown, J. Wyatt, R. Harris, and X. Yao, “Diversity creation methods for ensemble learning,” *Information Fusion*, vol. 6, no. 1, pp. 5–20, Mar. 2005. https://doi.org/10.1016/j.inffus.2004.04.001

[10] M. A. Ganaie, M. Hu, M. Tanveer, and P. N. Suganthan, “Ensemble deep learning: A review,” *IEEE Access*, vol. 9, pp. 125327–125353, 2021. https://ieeexplore.ieee.org/document/9534064

[11] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and
D. Batra, “Grad-CAM: Visual explanations from deep networks via gradient-based localization,” in *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 618–626. https://arxiv.org/abs/ 1610.02391

[12] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Mu¨ller, “Caus- ability and explainability of artificial intelligence in medicine,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 9, no. 4, 2019. https://doi.org/10.1002/widm.1312

[13] M. Tanveer, B. Richhariya, R. U. Khan, A. H. Rashid, P. Khanna,
M. Prasad, and C.-T. Lin, “Classification of Alzheimer’s disease using ensemble of deep neural networks trained through transfer learning,” *IEEE Access*, vol. 8, pp. 74191–74203, 2020. https://ieeexplore.ieee.org/ document/9076290