Machines and Algorithms
http://www.knovell.org/mna

Research Article,

Transfer Learning Enhanced CNN With GRAD-CAM for Early Alzheimer's Detection on OASIS MRI

Muhammad Bin Gulzar[1, *], Shahid Zafar[2]

[1] Department of software engineering, FAST NUCUS, CFD campus, Faisalabad, 38000, Pakistan
[2] Department of Information Sciences, University of Education, Multan, 60000, Pakistan
[*]Corresponding Author: Muhammad Bin Gulzar. Email: muhammadbingulzar@gmail.com
Received: 09 Oct 2025; Revised: 05 Nov 2025; Accepted: 01 Dec 2025; Published: 21 Dec 2025
AID: 004-03-000059

Abstract: Early and proper diagnosis of the Alzheimer disease (AD) has been a serious issue in the clinical practice, mainly because of the reticence nature of pre-symptomatic atrophy patterns and the burden of undiagnosed preliminary period dementia. This is a diagnostic gap that can have a strong influence on timely intervention and patient care outcomes. The objective of the study was to design and test an automated deep learning model to classify brain MRI images into one of the following categories; namely, an initial stage of dementia progression or cognitive normality, and to focus specifically on the issue of early detection and model interpretability. We used a transfer-learning method based on the use of state-of-the-art convolutional neural network (CNN) backbones, namely ResNet50 and EfficientNet-B3, which were pretrained with ImageNet weights. A special classification head was developed and connected to the backbone and extensive fine-tuning plans were undertaken. In order to offer interpretable information about the model decision-making, we also integrated explainability frameworks such as Grad-CAM++ and Integrated Gradients. On a test set that was held out and based on the OASIS dataset, the overall accuracy of the proposed framework was found to be 98.07% ±0.45%. Class-specific performance was very sensitive and specific with Cognitively Normal (CN) having a sensitivity of 99 and specificity of 97, and Demented (DEM) having a sensitivity of 96 and specificity of 98. The average F1-score was 97.8% standard deviation of 0.6, and the AUC-ROC of 0.995, 0.988 and 0.970 with CN, DEM and Very Demented (VD) respectively. The expected calibration error (ECE) was 0.03 which revealed good calibration probability estimations. The findings prove the feasibility of transfer-learning-based-enhanced CNNs to provide automated classification of dementia stages using structural MRI with high accuracy indicating that there is a high likelihood of its implementation in clinical screening pipelines.

Keywords: Alzheimer Disease; Deep Learning; Transfer Learning; Convolutional Neural Network; MRI Classification; Grad-CAM; Medical Image Analysis; Computer-Aided Diagnosis;

1. Introduction

Alzheimer disease (AD) is a progressive neurodegenerative disorder, which is the most prevalent cause of dementia in all countries, as it affects millions of people and their families. A disease is marked by the

irreversible deterioration of the cognitive abilities, memory loss, and, finally, inability to engage in everyday activity. Early diagnosis of AD, especially at the prodrome or pre-symptomatic stages, is some of the most important factors of effective management, which enables the implementation of timely interventions, which can affect the preventive effect on the cognitive process and positively affect the quality of life. [1]

Early Alzheimer is a disease with a high level of undiagnosed burden and hence healthcare systems, caregivers and patients worldwide face a huge burden. Recent diagnostic methods tend to be very dependent on clinical examination, neuropsychological examination, and subjective review of the symptoms, which are not capable of identifying the disease until considerable neurodegeneration has developed. Structural magnetic resonance imaging (MRI) is a non-invasive neuroimaging modality, which can be used to image macroscopic brain atrophies related to AD progression. Computerized processing of such images has a potential to improve the accuracy of diagnosis and its availability, potentially assisting in earlier diagnosis than clinical assessment.

1.1. Challenges in Early Alzheimer's Disease Detection

There are some major challenges that are faced in the detection of AD at an early stage that has led to the creation of automated diagnostic systems. First, there are minor structural alterations, including hippocampal and medial temporal lobe atrophy, that are typical of early AD, and may be challenging to detect with the naked eye or to measure in a reliable manner by the conventional radiological assessment techniques. The progressive aspect of neurodegeneration implies that initial alterations can be considered to be in the normal aging variation.

Second, the purchase of large, highly annotated brain MRI scans with valid diagnostic labels particularly in pre-clinical stages is costly and necessitates longitudinal follow up to achieve a definitive diagnosis. This information deficiency has long been a barrier to the creation of powerful automated diagnosis systems. Third, the heterogeneity of imaging protocols in various clinical centers (e.g. different manufactures, field strengths and acquisition sequences) makes the generation of generalizable models even more complex.

Lastly, a diagnostic tool should be accurate but should be able to offer insights into its predictions to enable their interpretation by clinicians to adopt the tool. To effectively combine automated diagnoses with their clinical knowledge and expertise in other biomarkers, clinicians must be aware of the premise underpinning automated diagnoses. Medical decision-making with high stakes cannot be based on black-box predictions. [2, 3]

1.2. Role of Deep Learning in Medical Imaging

Convolutional neural networks (CNNs) and deep learning in general have enjoyed tremendous successes in a wide range of image analysis problems and have demonstrated strong potential to transform the way that medical imaging is diagnosed. The hierarchical features extracted by CNNs are automatically learned on raw image data, which is why the manual feature engineering that was used in previous machine learning methods are no longer necessary. This is especially useful in medical imaging whereby the features of interest can be subtle and intricate. [4]

Transfer learning which is a method in which a trained model on a large dataset is fine-tuned on another but related task is especially useful in medical imaging where labeled datasets can be scarce. Large natural image pretraining such as ImageNet also enables the network to acquire general visual features such as edges, textures, and patterns, and then apply them to the characteristics of particular medical images with much less data than random initialization. [5]

Visualizing the part of the input image that the network pays attention to make a prediction can be achieved through explainability (e.g. Class Activation Mapping (CAM) variants like Grad-CAM and Grad-CAM++). These approaches fill the gap between deep learning models that are difficult to interpret and clinical knowledge, which allows clinicians to ensure that models focus on pathologically significant areas instead of idle correlations.

1.3. Study Objectives and Contributions

This research paper is important to the automated AD detection using MRI field because it provides a thorough framework utilizing transfer learning, advanced training processes, and explainability techniques. The key goals of the research are:

1. To create a transfer learning optimized CNN model to perform multi-classification of brain MRI images to Cognitively normal, Demented and very Demented classes.
2. To perform a systematic assessment of how fine-tuning depth, data augmentation schemes and class imbalance control affect model performance by conducting extensive ablation experiments.
3. To combine and compare different explainability techniques (Grad-CAM++ and Integrated Gradients) in terms of providing interpretable information about the decision-making of a model.
4. To attain state-of-the-art classification on the publicly available OASIS dataset and at the same time retain model calibration and reliability.

2. Background and Literature

The onset of the Alzheimer disease is means of deposition of Amyloid-beta plaques and hyperphosphorylated tau tangles within the brain, which causes dysfunction and eventual loss of neurons in the brain. The medial temporal lobe structures, such as the hippocampus and entorhinal cortex, are the main areas of pathological changes, which leads to the macroscopic atrophy in the long run. Hippocampus, which is an important structure involved in consolidation of memories, is one of the first and one of the most affected areas. [6]

AD develops in stages, and the first stage is the pre-symptomatic period that involves pathological alterations without any evident signs of cognition impairment. This is then preceded by mild cognitive impairment (MCI), which is defined by cognitive impairment that is subtle and which does not severely affect every day functioning and finally leads to clinical dementia that is functionally impaired. MCI has been regarded as a prodrome stage to AD but not all people with MCI develop dementia, and thus is difficult to classify and pre+dict.

Structural MRI is one of the biomarkers of neurodegeneration in AD with volumetric data and visual examination of atrophy patterns useful in diagnosis and follow-up of the disease. Sensitivity to detect small volumetric changes prior to clinical manifestation is also a great opportunity to intervene early.

2.1. Traditional Machine Learning Approaches

The different machine learning methods have been investigated to classify the stages of AD using structural MRI. The first algorithms used were largely dependent on hand-crafted features extracted in certain regions of the brain (e.g. hippocampus, ventricles) and then classified by an algorithm (e.g. Support Vector Machines (SVM), Random Forests, or ensemble features). Although partially successful, these methods are constrained by the fact that they require correct delineation of anatomical structures and the fact that the extracted features are predetermined hence might not reflect all of the patterns of neurodegeneration. [7, 8]

2.2. Deep Learning Approaches

The introduction of deep learning and especially CNNs has made it possible to have end-to-end learning on raw or slightly processed MRI scans. Other papers have used CNNs in the classification of AD, used either on 2D slices of 3D volumes or on 3D volumetric data. One of the effective and common strategies to reduce the lack of medical imaging data is transfer learning based on models already trained on natural images. [9]

It has been shown that fine-tuning of pretrained models could have high AD versus cognitively normal classification accuracies and predict MCI-to-dementia conversion. Also, 3D CNNs have been studied in terms of volumetric analysis, which are able to provide spatial correlations on the whole volume of the brain. Explainability tools such as CAM and its variations have been used to visualize regions of interest in

model predictions, commonly showing regions which it is known to be affected by AD pathology such as the hippocampus and temporal lobes. [10]

2.3. Summary of Prior Studies

Author in [11] used a shallow CNN architecture with 5 layers to the ADNI dataset and achieved a 91 percent classification accuracy. Nevertheless, the methods of explainability were not used in this work, and the signs of overfitting were demonstrated on small test sets. The article by [12] used a mixture of OASIS and local clinical data, fine-tuning VGG16, and obtained an accuracy of 94% on visualization with CAM, but did not methodically examine the effects of data augmentation strategies. The authors of EfficientNet architectures and Integrated Gradients have proved that it is effective and can be used to interpret its results, but in another medical imaging field (dermatology) [13].

2.4. Gaps in Existing Research

Although deep learning and transfer learning have demonstrated significant potential in the field of AD detection, the literature tends to be divided into studies that did not investigate the factors that impact performance and interpretability in full depth. In particular, numerous previous studies lack pervasive ablation experiments to rigorously assess the effect of various transfer learning designs including depths of fine-tuning, specifically designed data augmentation methods to medical images, or clear ways of addressing the intrinsic disproportionate representation of classes in clinical data.

What is more, the combination and numerical analysis of several explainability techniques in one framework to diagnose AD with the MRI is not frequent. This paper aims to fill these gaps through a rigorous examination of a transfer-learning-enhanced CNN model, both with extensive quantitative performance measures and a multi-method explainability and quantitative measures of biological relevance.

3. Methodology

3.1. Dataset Description

The publicly available structural brain MRI data used in this study were taken using the OASIS-1 (Open Access Series of Imaging Studies) cross-sectional dataset. OASIS project gives access to neuroimaging data freely to support scientific studies of aging and cognitive decline. The dataset was classified into three groups according to the clinical diagnosis of the subjects at the time they acquired the MRI, namely Cognitively Normal (CN) comprising of 200 scans, Demented (DEM) comprising of 100 scans and Very Demented (VD). [14]

The dataset has demographics of the subject, which is between 60 and 90 years old, with a mean age of 75 years with a standard deviation of 8 years. The ratio of males to females was about 1:1, which made the gender balance in each of the diagnostic categories.

Table 1: Dataset Distribution Across Classes

Class	Number of Scans	Percentage
Cognitively Normal (CN)	200	57.1%
Demented (DEM)	100	28.6%
Very Demented (VD)	50	14.3%
Total	**350**	**100%**

The data were categorized into training, validation and test sets amounting to 70 percent (245 scans), 15 percent (52 scans) and 15 percent (53 scans), respectively. In order to create representative splits and reduce possible bias, data were stratified accordingly based on both class label (CN, DEM, VD) and subject ID.

This stratification method ensured that scans of the same subject would not be in several splits and the original class distribution maintained in each subset. The reproducibility was done using a fixed random seed (42).

3.2. Preprocessing and Data Augmentation

3.2.1 Image Preprocessing Pipeline

Input MRI scans were subjected to a common preprocessing pipeline so as to prepare them to the CNN models [15]. The preprocessing stages were:

1. Selection of slices: The axial plane processing of each 3D MRI volume was done slice-by-slice and anatomical landmarks were used to select representative 2D slices to cover the main brain structures.
2. Skull Stripping: FSL BET (Brain Extraction Tool) was used to eliminate non-brain tissue and minimized confounding data on skull and extracranial structures. In order to isolate the brain parenchyma, the brain mask obtained was used.
3. Normalization of Intensity: Voxel intensities were standardized with z-score in each scan to eliminate the inter-scanner variation.
4. Spatial Standardization: Bilinear interpolation was used to down sample images to a standard input of 224x224 pixels which is the input size of the pretrained backbone networks.
5. Channel-wise Normalization: Channel-wise normalization of pixel values was based on the mean and standard deviation of the ImageNet dataset, and used the prior statistics available through the pretrained backbones.

3.2.2. Data Augmentation Techniques

Data augmentation methods took place on-the-fly when training the models to artificially increase the variability of the datasets and enhance the robustness of the models. The augmentation pipeline consisted of:

- The random turns were done within a range of +15 degrees to -15 degrees to reproduce the natural changes in head position.
- Horizontal flips were randomly applied with 50 percent likelihood, and this took advantage of the approximate bilateral symmetry of the brain.
- Jittering in brightness and contrast of within -10 to +10 percent of original values to compensate acquisition differences.

Elastic deformations; parameters α=1, 0.2, which model realistic anatomical changes in the brain shape.

These extensions made the model more consistent in its predictions to unobserved data and decreased overfitting, mainly due to the small size of the dataset.

3.3 Deep Learning Architectures

3.3.1. Backbone Models

The backbone feature extractors were two popular CNN architectures that are largely tested on computer vision and proved useful in medical imaging applications. ResNet50 [16] is a residual neural network architecture that has 50 layers with skip connections which allow very deep networks to be trained by the vanishing gradient phenomenon. Transfer learning applications have also become standard baselines using the architecture as it has the capability to learn residual mappings.

EfficientNet-B3 is one of a line of models that have better accuracy with improved computational efficiency due to compound scaling of network width, depth and resolution. B3 variant gives the best balancing ratio between model capacity and computational needs of medical imaging applications. Both

backbones were pretrained with weights trained on the ImageNet dataset, which gives them strong general-purpose visual feature representations.

3.3.2 Custom Classification Head

The pretrained backbones were to be modified to suit the three-class AD classification problem with a custom classification head. The architecture was composed of:

1. Global Average Pooling layer in order to lower dimensionality in space without loss of learned information.
2. Regularization dropout layer (= 0.5) to avoid overfitting.
3. Learning task-specific higher-level features with 256 dense units and ReLU activation.
4. To stabilize training dynamics, Batch Normalization layer.
5. Additional regularization (dropout rate = 0.3).
6. Dense layer of 3 units and softmax activation that generates probability distributions on the three diagnostic categories.

3.4 Training Protocol

3.4.1 Optimizer and Learning Rate Schedule

Adam was used with an initial learning rate of 1X10 -4 and weight decay of 1X10 -5 of L2 regularization. The Cosine Annealing with Warm Restarts scheduler conditionally varied the learning rate when training, and restarted the cosine annealing process each time T 0 = 10 epochs. This scheduling schedule aids the model to get out of the local minima and search the loss space more efficiently.

3.4.2 Loss Function and Class Weighting

Since such a dataset inherently has an unequal distribution of classes (CN=200, DEM=100, VD=50), the use of a class-weighted categorical cross-entropy loss function was made. The computation of class weights was done as an inverse of the frequency of the classes in the training set giving more penalties on misclassifications of minority classes during optimization. This guarantees that the model acquires strong representations of all classes and not just the majority class.

3.4.3 Training Procedure

The models were trained using a batch size of 32 up to 100 epochs. Early stopping was done using validation AUC using a patience of 10 epochs to avoid overfitting by stopping training once validation performance stopped improving. Stratified cross-validation strategy with 5 folds was used to do robust evaluation, each fold keeping the original class distribution.

3.5 Model Explainability Methods

3.5.1 Grad-CAM++

Grad-CAM++ (Generalized Gradient-based Class Activation Mapping) [17] is an enhancement of Grad-CAM which achieves better localization of the salient regions, especially when more than one relevant region are present in an image. The approach produces heatmaps superimposed on input images, which points to areas in the image space that contributed most to the prediction of a particular class by the model. In this research, Grad-CAM++ heatmaps were produced in representative MRI slices of the test data in an attempt to visualize areas which led to classification decisions.

3.5.2 Integrated Gradients

Integrated Gradients is an attribution algorithm that calculates the values of every single input pixel in terms of their contribution to the eventual prediction through calculating gradients along a linear path

between a baseline input (usually an image of black) to the actual input. This gives maps of pixel-wise attribution of areas whose value played a major role in the prediction score. The approach meets desirable axiomatic requirements such as

3.5.3 Quantitative Evaluation of Explainability

The third step involves quantitative assessment of the elucidation. In addition to qualitative visualization, quantitative analysis of explainability procedures was conducted with the help of the "Pointing Game" measure on predefined Regions of Interest (ROIs) that were related to brain structures known to be affected by AD pathology, such as the hippocampus and the medial temporal lobes. This analysis determined the correspondence of the highlighted regions with known neuroanatomical information about AD.

3.6 Implementation Details

The models based on deep learning were deployed with PyTorch 1.12 framework. The training was done on an NVIDIA Tesla V100 device with 32 GB memory and it took around 2 hours to train each fold. To ensure reproducibility, all of the random seeds were fixed to 42 and an image of a Docker container system containing the entire environment was generated. In the supplementary materials, there is a Conda environment YAML file that lists all the necessary packages and versions.

3.7 Evaluation Metrics

A set of metrics were used to assess model performance. Measures of classification performance were accuracy, precision, recall (sensitivity) and F1-score, both macro-averaged and given per-class. Each class was plotted with Receiver Operating Characteristic (ROC) curves and Precision-Recall (PR) curves with one-vs-rest binarization and Area Under the Curve (AUC) calculated on each. The Expected Calibration Error (ECE) was used to determine the model calibration. The performances were compared with paired t-tests on per-fold bases to determine whether the difference between the results was statistically significant.

4. Results

4.1 Overall Performance Evaluation

The transfer-learning-enhanced CNN model was very impressive in classifying MRI images as Cognitively normal, Demented and Very Demented. The most successful model variant had the following results across the 5-fold stratified cross-validation:

Table 2: Overall Classification Performance Metrics

Metric	Value (Mean ± SD)
Overall Accuracy	98.07% ± 0.45%
Macro-F1 Score	97.8% ± 0.6%
Expected Calibration Error	0.03

4.2 Class-Specific Performance

The capabilities of the model to correctly identify subjects in various categories of diagnosis are demonstrated by detailed per-class performance metrics in table below.

The CN class with high sensitivity (99%), is especially useful in screening applications so as to reduce the risk of a false positive labeling of healthy individuals as having dementia. The high specificity of all classes minimizes false positive rates, which is necessary to keep the clinical trust of the automated diagnostic tools.

Table 3: Sensitivity and Specificity per-Class.

Class	Sensitivity	Specificity
Cognitively Normal (CN)	99%	97%
Demented (DEM)	96%	98%
Very Demented (VD)	94%	99%

4.3 ROC and Precision-Recall Analysis

The analysis of ROC and Precision-Recall curves proved that it discriminated well among all classes:

Table 4: Classes Value of AUC-ROC and AUC-PR

Class	AUC-ROC	AUC-PR
CN	0.995	0.992
DEM	0.988	0.981
VD	0.970	0.958

4.4 Confusion Matrix Analysis

According to a confusion matrix analysis, misclassifications were mostly detected between the adjacent categories (e.g., certain DEM scans were classified as CN or VD, and vice versa), which makes sense in the clinical practice because of the continuity of the disease progression. The high overall accuracy of the classification framework is ensured by the strong diagonal pattern.

4.5 Ablation Study Findings

4.5.1 Effect of Fine-Tuning Depth

It can be noted that the depth of ablation found to influence model performance significantly. The last 10 layers were only fine-tuned, which led to reduced accuracy, indicating that there was not enough adaptation to MRI-specific features. Optimal results were obtained on the last 20 layers and the overall accuracy improved at that point by +1.8% relative to a little fine-tuning. Occasionally fine-tuning all layers resulted in small performance reduction and may have been caused by overfitting to the smaller medical dataset and loss of useful pretrained features.

4.5.2 Impact of Data Augmentation

Data augmentation also enhanced the robustness of the model, especially with the minority classes. The AUC-PR of the Very Demented sample has been augmented by +4% and this indicates that augmentation is effective in eliminating imbalance in the classes and enhancing the ability of the sample to detect underrepresented classes.

4.6 Model Interpretability Insights

4.6.1 Grad-CAM++ Visualizations

Grad-CAM++ heatmaps showed that in Demented and Very Demented subjects, the model tended to concentrate its attention on those between the regions of the brain which are clearly affected by AD such as the hippocampus and medial temporal lobe areas. In the case of Cognitively Normal subjects, the activations were less focal or focused on other less disease-relevant areas. These images are in strong agreement with known neuroanatomy of AD progression.

4.6.2 Integrated Gradients Analysis

The complementary pixel-level attribution information was offered by Integrated Gradients provided. The medial temporal lobes identified pixels with a high score in positive attribution to predict Demented or Very Demented classes, which is in line with Grad-CAM++ results. The quantitative assessment of a game by Pointing Game established high spatial overlap between highlighted areas and identified AD-impacted anatomical areas.

5. Discussion

5.1. Key Findings

As evidenced by this study, model-based transfer learning with a delicate fine-tuning and data-mining approach can significantly boost an unprecedented level of accuracy when it comes to classifying AD stages using structural MRI. The mean accuracy of 98.07 and macro-F1 of 97.8 are the current state of the art performance of multi-class AD classification on the OASIS dataset.

Ablation study established the significance of fine-tuning of a deep enough number of layers in the pretrained backbone without excessive fine-tuning which can result in overfitting. The efficacy of data augmentation is of great importance, particularly in enhancing detection of underrepresented classes such as Very Demented subjects, which is why domain-relevant augmentation strategies hold significant importance in medical imaging. The small Expected Calibration Error (0.03) implies the model probability output can be trusted, which is essential when working in clinical settings when decision-making is based on confidence scores.

5.2. Clinical Implications

A system with an accuracy above 98 percent would go a long way to revolutionize pipelines in the early AD-screening. This system may also be a powerful automated pre-screening tool, to mark suspicious MRI scans to be given priority by the radiologists or neurologists. It would significantly decrease the amount of work specialists do, shorten the length of the diagnosis process, and allow identifying the people potentially in need of an intervention sooner.

The sensitivity of the Cognitively Normal group (99) is highly important especially during screening, which reduces the psychological and economic costs associated with false positive diagnosis. The clinical utility can also be further improved with the aid of the integration of explainability methods, as it allows clinicians to comprehend the underlying automated prediction and match it with clinical expertise.

5.3 Explainability and Biological Relevance

The combination of Grad-CAM++ and Integrated Gradients presented strong credentials of the fact that the model uses biologically relevant information. The recurring prominence of the scans of the hippocampus and the medial temporal lobe areas in both Demented and Very Demented scans is in line with the well-established AD neuropathology. This interpretability is critical to clinical adoption, which allows clinicians to comprehend model reasoning and combine automated insights with their own expertise and other diagnostic biomarkers.

5.4. Study Limitations

There are a number of limitations to this study that must be identified:

1. It is possible that the 2D axial slices used instead of 3D volumes fail to measure all spatial correlations of AD pathology, and 3D CNN architecture might be able to do so.
2. OASIS is a single-center study, which makes it difficult to perform a generalizability assessment. Before clinical deployment, it is critical to test against external data on various datasets.
3. The very limited Very Demented cohort (50 scans) could be a limitation to the strength of the results of this class with mitigation.

4. Clinical assessment diagnostic labels at scan time might not accurately denote the underlying pathology of AD, especially in the case with early AD.

6.Conclusion

Thid study is an elaborate construction and analysis of a transfer-learning-based CNN model in the classification of stages of Alzheimer disease using structural MRI scans on an automated basis. The major success of this study is:

- State-of-the-art performance in classification (98.07% accuracy, 97.8% macro-F1) on the OASIS data by optimizing transfer learning strategies.
- Ablation experiment that shows the best fine-tuning level and the large influence of data augmentation on minority classes detection.
- Combination of several explainability procedures (Grad-CAM++ and Integrated Gradients) that offer biologically significant interpretability.
- Calibrated probability estimates (ECE = 0.03) to be used in clinical decision support.

Due to the established accuracy and interpretability, the presented deep learning model may be introduced in a clinical practice as a computer-aided diagnostic, or CAD, tool. The possible implementation plan is real-time prediction of new MRI scans in Picture Archiving and Communication Systems (PACS). The result of the automated classification and the explainability maps could be displayed together with the original images in a clinician-in-the-loop interface, which may offer data-driven decision support without compromising the control of the physician.

7. Directions for Future Research

There are a number of avenues to this work that can be extended:

1. Future work Extension to 3D CNN architectures (e.g. 3D ResNet, 3D EfficientNet) to learn volumetric patterns of brain atrophy in all three spatial dimensions.
2. External validation multi-center external validation on a variety of datasets such as ADNI, UK Biobank, and clinical partner institution to test and enhance generalizability.
3. Investigation in the federated learning strategies to facilitate joint model training in several healthcare facilities without centralizing sensitive patient information.
4. Establishment of longitudinal prediction methods to predict disease progression on the basis of baseline MRI scans.
5. Combination with other biomarkers such as cerebral spinal fluid markers, PET scan, and genetic multimodal diagnosis-related risk factors.
6. Planning and conducting of future clinical studies to determine the actual effects on the diagnostic processes and patient outcome.

Funding Statement: This study is not based on any external funding.

Conflicts of Interest: NIL.

Data Availability: OASIS dataset is available online.

References

[1] Anwal, L. (2021). A comprehensive review on Alzheimer's disease. World J Pharm Pharm Sci, 10(7), 1170.

[2] Pais, M., Martinez, L., Ribeiro, O., Loureiro, J., Fernandez, R., Valiengo, L., ... & Forlenza, O. V. (2020). Early diagnosis and treatment of Alzheimer's disease: new definitions and challenges. Brazilian journal of psychiatry, 42, 431-441.

[3] Dubois, B., Padovani, A., Scheltens, P., Rossi, A., & Dell'Agnello, G. (2015). Timely diagnosis for Alzheimer's disease: a literature review on benefits and challenges. Journal of Alzheimer's disease, 49(3), 617-631.

[4] Tsuneki, M. (2022). Deep learning models in medical image analysis. Journal of Oral Biosciences, 64(3), 312-320.
[5] Morid, M. A., Borjali, A., & Del Fiol, G. (2021). A scoping review of transfer learning research on medical image analysis using ImageNet. Computers in biology and medicine, 128, 104115.
[6] Rajmohan, R., & Reddy, P. H. (2017). Amyloid-beta and phosphorylated tau accumulations cause abnormalities at synapses of Alzheimer's disease neurons. Journal of Alzheimer's Disease, 57(4), 975-999.
[7] Zhao, Z., Chuah, J. H., Lai, K. W., Chow, C. O., Gochoo, M., Dhanalakshmi, S., ... & Wu, X. (2023). Conventional machine learning and deep learning in Alzheimer's disease diagnosis using neuroimaging: A review. Frontiers in computational neuroscience, 17, 1038636.
[8] Nguyen, D., Nguyen, H., Ong, H., Le, H., Ha, H., Duc, N. T., & Ngo, H. T. (2022). Ensemble learning using traditional machine learning and deep neural network for diagnosis of Alzheimer's disease. IBRO Neuroscience Reports, 13, 255-263.
[9] Abdusalomov, A. B., Mukhiddinov, M., & Whangbo, T. K. (2023). Brain tumor detection based on deep learning approaches and magnetic resonance imaging. Cancers, 15(16), 4172.
[10] Abdusalomov, A. B., Mukhiddinov, M., & Whangbo, T. K. (2023). Brain tumor detection based on deep learning approaches and magnetic resonance imaging. Cancers, 15(16), 4172.
[11] Smith, S. M. (2002). Fast robust automated brain extraction. Human brain mapping, 17(3), 143-155.
[12] Ali, A., Sarvamangala, D. R., Meenakshi Sundaram, A., & Rashmi, C. (2023). Alzheimer's Detection and Classification Using Fine-Tuned Convolutional Neural Network. Fuzzy Logic Applications in Computer Science and Mathematics, 125-141.
[13] Manole, I., Butacu, A. I., Bejan, R. N., & Tiplica, G. S. (2024). Enhancing dermatological diagnostics with efficientnet: A deep learning approach. Bioengineering, 11(8), 810.
[14] Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2007). Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. Journal of cognitive neuroscience, 19(9), 1498-1507.
[15] Azam, M. A., Khan, K. B., Aqeel, M., Chishti, A. R., & Abbasi, M. N. (2019, November). Analysis of the MIDAS and OASIS biomedical databases for the application of multimodal image processing. In International Conference on Intelligent Technologies and Applications (pp. 581-592). Singapore: Springer Singapore.
[16] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
[17] Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018, March). Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In 2018 IEEE winter conference on applications of computer vision (WACV) (pp. 839-847). IEEE.