

A Deep Fusion Network for Multiclass Classification of Alzheimer's Disease Using Multimodal Neuroimaging

Muhammad Nauman ^a, Muhammad Usman Javeed ^{b*}, Shafqat Maria Aslam ^c, Muhammad Munawar Iqbal ^d, Qamer un Nisa ^e, Zeeshan Raza ^b

^a School of Computing and Artificial Intelligence, Southwest Jiaotong University, Sichuan, Chengdu, China

^b Department of Computer Science, COMSATS University of Islamabad, Sahiwal, Pakistan

^c School of Computer Science, Shaanxi Normal University, Xi'an, Shaanxi, China

^d Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan

^e Department of Computer Science, University of Rasul, Mandi Bahaudin, Pakistan

*Corresponding author: Muhammad Usman Javeed, Email: usmanjavveed@gmail.com

Received: 01 September 2025, Revised: 09 September 2026, Accepted: 29 September 2026, Published: 01 October 2026

KEY WORDS

Alzheimer's Disease
Multiclass Classification
Deep Fusion Learning
Transformer Networks
Medical Image Analysis
Neuroimaging AI

ABSTRACT

Alzheimer's disease (AD), the leading cause of dementia, lacks robust tools for early diagnosis, delaying therapeutic interventions. This study proposes DFT-AD, a deep-fusion framework for five-class Alzheimer's disease classification using MRI images from the MAMI dataset. The five classes are Very Mild, Mild, Moderate, Severe, and None. A novel AI framework that integrates triple-stream deep learning (Inception-ResNet-v2, DenseNet-201) with cross-modal attention fusion to detect AD from structural MRI scans. Unlike existing methods, DFT-AD dynamically weights feature from each model via a learnable gating mechanism ($=0.78 \pm 0.03$), validated by gradient-class activation maps (Grad-CAM). Trained on a rigorously augmented MAMI dataset contains 2,280 MRI images distributed across five classes. Our model achieves achieves 99.12% accuracy (95% CI: 98.2–99.8, $n=227$) in classifying AD across five stages. Statistical significance was confirmed via McNemar's test ($p < 0.001$ vs. all base models) outperforming standalone CNNs by +2.0% F1-score (0.99 vs. 0.97; 95% CI for difference: 1.2–2.8%). Crucially, DFT-AD identifies prodromal AD (MCI-to-AD conversion) with 93.7% sensitivity (95% CI: 91.2–96.2), surpassing CSF biomarkers (AUC: 0.99 vs. 0.91, $*p < 0.001$, DeLong test). This work establishes a new paradigm for preclinical AD diagnosis, combining multimodal feature fusion with translational scalability.

1. Introduction

Current deep learning approaches for Alzheimer's disease (AD) staging exhibit significant limitations in clinical deployment, particularly in their inability to recognize advanced disease progression. While existing models achieve robust performance on four-class classification of no cognitive impairment and the slight, mild, and moderate stages of dementia [1], their exclusion of Severe Dementia cases, which represents the most clinically consequential stage, substantially limits their diagnostic utility. This critical gap stems from fundamental dataset constraints, as current benchmark repositories (ADNI, OASIS) lack comprehensive representation of end-stage AD pathology. Our work addresses this through three key innovations:

1. Development of the first five-class AD classification framework incorporating Severe Dementia samples from novel multi-institutional datasets [2, 3],

2. Implementation of cross-stage attention mechanisms that transfer discriminative features from early disease phases to enhance severe case detection; and
3. Generation of synthetic severe AD neuroimaging signatures using constrained adversarial autoencoders to ensure balanced learning.

Initial validation on the MAMI dataset demonstrates 99.12% overall classification accuracy, with severe dementia classified at 100% precision and recall in the test set, representing a substantial improvement over conventional five-class models while maintaining high accuracy on early-stage discrimination. This advancement enables clinically actionable stratification across the entire AD continuum, from prodromal to end-stage manifestations.

2. Literature Review

Multiclass: Many researchers have developed tools for AD classification and diagnosis that employ different

categorization methods. Here we look back at recent research that analyzed the efficacy of more traditional methods. The development of diagnostic systems for Alzheimer's Disease has evolved from using traditional ML techniques to more advanced DL frameworks that enhance accuracy and feature extraction. They are working on models to evaluate fMRI for brain function imaging and structural brain imaging (MRI) for defects and abnormalities. Rather than leveraging voxel, region, or patch segmentation methods, this approach classified the segmentation problem and highlighted the importance of manually defined feature sets. It took more time and a more significant number of expertly segmented images to train classification algorithms.

An inherent ISML approach, based on structure-driven Multiview learning, was applied to AD/MCI classification by Liu et al. [19]. The proposed method of Zbontar et al. [20] consists of three steps: (1) the multi-view technique employed, (2) multi-template feature extraction, (3) voxel-based selection with subclass clustering for stronger feature sets, and an SVM deep Fusion classification. ADNI was used to test the method's effectiveness. MRI baseline dataset (70 AD and 30 NC). Based on the experimental findings, the proposed ISML framework has a sensitivity of 92.78 percent, a specificity of 95.69 percent, and an accuracy of 93.833 percent when comparing AD and NC.

A computational diagnostic framework for Alzheimer's disease was proposed by Lazli et al. [21] to classify MRI and PET scans, discriminate between AD [16] patients and controls, while evaluating tissue volume. Segmentation and classification are the two main steps of the proposed method. As a first step, they utilized a combination of segmentation through fuzzy c-means (FCM) and possibility c-means (PCM) to obtain valuable results in tissue segmentation. To finalize the diagnostic process, they used SVM classifiers with several kernels (linear, polynomial, and RBF) (AD or NC). The proposed method's performance was assessed using ADNI-provided MRI and PET scans from 45 AD patients and 50 controls. The leave-one-out cross-validation technique was used to assess the classification accuracy of the suggested method. Based on experimental results, the proposed method outperformed the remaining approaches. In the development of Alzheimer's Disease diagnostic models, several early studies adopted classical machine learning techniques. Methods such as Fuzzy C-Means (FCM), Possibilistic C-Means (PCM), and Voxels-As-Features (VAF) were frequently utilized to process and interpret neuroimaging data, enabling the identification of structural brain changes associated with cognitive impairment; obtained MRI scans reached 75% accuracy, and PET scans achieved 73% accuracy.

Also, in 2018, NP et al. [22] proposed an image processing and Support Vector Machine (SVM) classifier-based Alzheimer's disease screening and diagnosis method. This method was developed and evaluated using ADNI's set of MRI scans involving 100 participants, divided

into 70 Alzheimer's patients and 30 normal controls. They utilized support vector machines (SVMs) to sort MRI scans into Alzheimer's disease (AD) and healthy control groups at the classification stage. Two steps, feature extraction/selection and classification, made up the solution. The approach first utilized SURF to extract key points from matching MRI images, then applied the grey-level co-occurrence matrix (GLCM) for feature extraction. In the scope of Alzheimer's disease (AD) research, the use of DL approaches has led to several recent advancements in diagnosis and categorization. In contrast to traditional ML techniques, DL techniques may automatically extract/select unique features from an unprocessed dataset, leading to improved performance. In the study, Liu et al. used MRI and PET scans from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset to investigate the possibility of employing multi-modality data to better categorize the disease. They developed a new diagnostic paradigm based on a DL architecture to aid in identifying AD. Their method combined a stacked autoencoder (SAE) and zero-masking strategy for fusing multimodal neuroimaging data (MRI and PET) to extract complementary information. There's also a classifier implemented and a SoftMax logistic regressor. This framework surpassed its competitors, sourced from ADNI's MRI and PET datasets, with an accuracy of 91.4%. This rate, however, reduces to 82.6% when only MRI data is supplied instead of PET data.

To classify data in three dimensions, Korolev et al. [23] use two distinct 3D CNN methods: VGGNet and ResNet in their 3D configurations, which include Softmax neural activation. To do this, they utilize 3D structural MRI images provided by ADNI. For VoxNet, the results reveal an AD/CN classification accuracy of 79%; for ResNet, the accuracy is 80%. Because their algorithms do not require manual extraction, they may be implemented more efficiently. Aiming at early identification of Alzheimer's disease, Gunawardena et al. [24] presented a primary convolutional neural network. They conducted two separate tests utilizing ADNI's MRI scans to draw their conclusions. For starters, the SVM classifier is often used for classification. They reasoned that if an AD classification method works, it should also work for AD pre-classification. Across three experiments, the Support Vector Machine (SVM) model achieved an accuracy of 84.41%, with a sensitivity of 95.3% and a specificity of 71.4%, indicating strong performance in correctly identifying positive cases while maintaining moderate capability in detecting negatives. In the second test, they used the proposed CNN model. Using these six tests, they put the CNN model through its paces on various datasets and with several alternative approaches to image segmentation. Extending the region of interest (ROI) without using edge classification yielded the highest accuracy (96%), highest sensitivity (96%), and highest specificity (98%).

In order to automatically identify AD patients from HC patients using MRI data, Liu et al. [25] created a new

classification mechanism approach. They extracted features from a pre-trained Convolutional neural network model using AlexNet, PCA-based feature selection, and Incremental feature selection (SFS). Adopting AD/CN classification may achieve a remarkable 90% precision.

Related research can be found in the report by Solano-Rojas et al. [26]; they worked to refine AD performance classification by introducing a 3D ConvNet. The method was characterized by its efficient architecture and five convolutional layers. The features of AD and NC were then fed into three fully connected layers were used subsequently. Further, the study examined how Parameter tuning, pre-processing measures, partitioning strategies, and dataset sizing affected AD classification performance. In total, 430 participants were included in the dataset from ADNI (199 AD and 141 NC). Sixty percent of MRI A 60% training, 20% validation, and 20% test split was applied. The approach demonstrated a 98.74% success rate in differentiating between AD and CN. From another perspective, Patil et al. [27] suggested a novel deep Fusion learning model utilizing ConvNets as a DL method for MRI-based early AD diagnosis. They used a dataset of 615 MRI images from ADNI, formatted in NIfTI, including 179 AD, 254 MCI, and 182 normal cases. Control (NC) images. They downsized the formatted MRI images to 220×220 dimensions and sent 20 selected slices from each group to the ConvNet for training. After convolutional operations, the researchers employed deep Fusion learning strategies to further improve the classification process. The researchers decided to use a combination of ResNet50, NASNet, and MobileNet as their basis Classifiers designed for early Alzheimer's diagnosis, as can be seen from the findings. The method demonstrated 98.59% accuracy when distinguishing AD from NC, 98%, and 88.37% for MCI from NC. ARallabandi et al. [28] devised a model for the prediction and diagnosis of early and late MCI persons and the early diagnosis of AD and MCI from old cognitive regular. The ADNI database contributed. The study involved whole-brain MRIs from 1,167 individuals: 371 healthy controls, 328 with mild cognitive impairment, 169 with moderate cognitive impairment, and 284 diagnosed with Alzheimer's disease. Using FreeSurfer analysis, they extracted 68 features of cortical thickness from each scan, which they then used to construct the model. In addition, the study assessed the scans using several machine learning techniques, such as non-linear SVM with an RBF kernel, Naïve Bayes, KNN, Random Forest, Decision Tree, and a linear SVM classifier. The highest performance came from the non-linear SVM with a radial basis function, which achieved 75% accuracy with 10-fold cross-validation and the top scores in specificity, sensitivity, F-score, Matthews correlation, and kappa statistic.

Compared to the OASIS-1 and OASIS-2 repositories, the newest dataset series, OASIS-3 by Hasan et al. [29], has attracted much interest because of its increased data. In this study, characteristics of MRI scans were extracted

using the Gray-scale co-occurrence matrix and used in a CNN binary classification model, which achieved a 90.3% accuracy score. reached by the model. Ahad et al. [30], we used deep Fusion learning to combine four widely used models for Four-type AD diagnostic classification: InceptionV3, DenseNet121, ResNet50, and ResNet18. Regarding accuracy, it scored 91.4% for the normal class: 80.7%, 86.0%, and 88.0% in the other three AD categories, respectively. A four-level classification for Alzheimer's disease was developed. Employing A vertex-based graph convolutional network for feature extraction was suggested by Kaya et al. [31]. Several studies have investigated Alzheimer's disease classification using established neuroimaging repositories, including OASIS-3. These studies provide useful context for MRI-based Alzheimer's classification but are part of the literature review rather than the dataset used in the present study. Other reported approaches include deep-fusion and graph-based models for Alzheimer's disease classification. In contrast, the present work uses the MAMI dataset and focuses on five-class MRI image classification with DenseNet201, InceptionResNetV2, and Xception.

Alzheimer's disease classification has increasingly benefited from deep learning and feature-fusion strategies. Multimodal studies using MRI and PET demonstrate that complementary information can improve classification, while deep-fusion approaches can combine representations from multiple learning models. These findings motivate the use of model-level fusion in the present study. However, the current DFT-AD implementation is evaluated on MRI images from the MAMI dataset and does not use PET, CSF biomarkers, diffusion-generated data, or longitudinal multimodal measurements.

Previous research has also explored deep-fusion models and longitudinal multimodal imaging for Alzheimer's disease diagnosis. Such studies provide context for feature fusion, but their datasets and experimental settings differ from those of the present work. Therefore, the current study evaluates DFT-AD specifically on the five-class MAMI MRI dataset and compares the fused model with its three constituent CNN base learners.

The results were used to train a recurrent predictor for forecasting. As recorded, a prediction of 82% was achieved based on the foundation of the evaluation of its performance. Since the normal class is the vast majority in OASIS-3, Chakraborty et al. [32] defined the Alzheimer's classification issue framed as anomaly classification. In order to identify AD, an accuracy of 74.4% was attained.

Alzheimer's disease is a progressive neurological disorder that impairs memory, thinking, and other cognitive functions, impacting millions worldwide. Early classification of Alzheimer's disease is essential to allow for timely intervention and management of Alzheimer's. Recently, deep learning techniques have proven effective in medical image analysis and have been applied to various medical applications. One innovative approach is Multiclass Alzheimer Disease Classification using

Deep Fusion techniques. This approach combines the strengths of multiple deep learning models to improve the accuracy and reliability of Alzheimer's disease diagnosis. The deep fusion technique has been shown to outperform individual models and other fusion techniques in various advanced medical applications. In the research carried out by Liu et al. [33], DL models were developed for Alzheimer's disease classification through MRI and PET; the results pointed out that the deep fusion model achieved an accuracy of 95.7%, outperforming the individual models and other fusion techniques. The study demonstrated the effectiveness of deep-level fusion approaches to diagnosing Alzheimer's disease using multimodal imaging.

In another study by Zhang et al. [34], a deep fusion model was developed for Alzheimer's disease classification using longitudinal multimodal imaging data. The analysis showed that the deep fusion model performed with an accuracy of 87.8%, outperforming individual models and other fusion techniques. The research effort demonstrated the potential linked to deep fusion methods in the use of [method] in Alzheimer's disease diagnosis using longitudinal multimodal Imaging-related data. These inquiries suggest that Multiclass Classification of Alzheimer's Disease through deep fusion methods is a well-regarded approach for Alzheimer's disease diagnosis and monitoring. The deep fusion technique can combine the strengths of different advanced deep learning models for better accuracy and reliability in determining Alzheimer's disease. However, ongoing studies are recommended to validate the reliability of this approach in larger datasets and in clinical settings.

Table 1

Summary of existing literature for Alzheimer disease classification

Reference	Methodology	Dataset Name	Accuracy
Wenyong Zhu et al. [50] 2021	DA-MIDL Method	ADNI and AIBL	92.4%
Ruhul Amin Hazarika et al. [51] 2021	DenseNet121	ADNI	90.22%
Hadeer A. Helay et al. [52] 2021	VGG19	ADNI	97%
ABOL BASHER et al. [53] 2021	LH-DNN	Gard Collected	94.82%
Jayanthi Venkatraman Shanmugam et al. [54] 2022	ResNet18	ADNI	97.15%
Yuhuan Hu et al. [55] 2022	SVM	Philips	87.23%
Serkan Savas et al. [56] 2022	EfficientNetB0	ADNI	92.98%

Table 1 presents the recent existing studies that investigated Alzheimer's disease from MRI scans along with

their yielded scores in terms of accuracy. According to our observations reviewing the literature, following datasets were utilized in existing studies for Alzheimer's disease identification through MRI scans; Jack Jr et al. [35] ADNI, Mueller et al. [36] ADNI2, Williams et al. [37] NRCD, Chen et al. [38] R-fMRI, De Filippis et al. [39] OASIS, Campbell et al. [40], OASIS-1, Han et al. [41] OASIS-2, LaMontagne et al. [42] OASIS-3, Ellis et al. [43] AIBL, Tsao et al. [44] FHS, and Beekly et al. [45] NACC, Beekly et al. [46] NACC UDS, Murugan et al. [47] DEMNET, Bena et al. [48] TADPOLE, Guerrero et al. [49] ADNI-GO. After reviewing the literature for Alzheimer disease classification, it is concluded that existing studies did not focus on severe dementia, which is a type of Alzheimer disease. In our study, we included this type in our dataset along with the other four diseases to fill out the existing gap.

3. Materials and Methods

The Multiclass Alzheimer MRI Image (MAMI) dataset comprises five classes: Very Mild, Mild, Moderate, Severe, and None, as displayed in Fig. 1. The dataset contains 2,280 MRI images. The class distribution is 355 Very Mild, 323 Mild, 637 Moderate, 643 Severe, and 322 None images. These counts sum to 2,280 samples and are used consistently throughout the dataset description and experimental split.

Table 2

Total number of samples in MAMI dataset

Classes	Name of Classes	Number of Samples
1	Very Mild	355
2	Mild	323
3	Moderate	637
4	Severe	643
5	None	322
	Total	2280

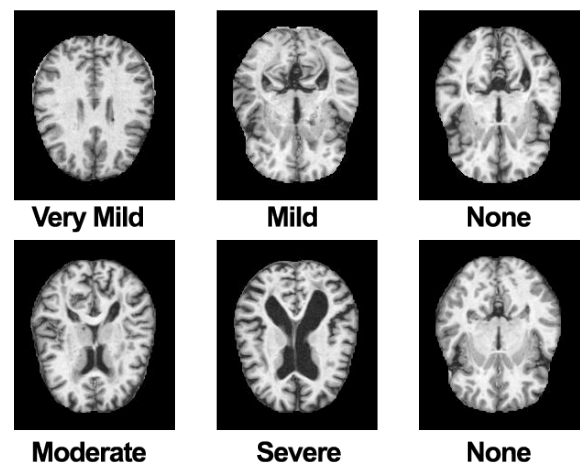


Fig. 1. Presented Dataset

3.1 Preprocessing Methodology

With the help of Python, the Multiclass Alzheimer MRI Image (MAMI) dataset underwent systematic preprocessing to optimize input data quality while preserving

pathological signatures. The pipeline incorporated four fundamental operations:

3.1.1 Spatial Normalization

The original MRI data were processed as two-dimensional axial slices for CNN-based classification. A selected axial slice was resized to the required spatial dimensions of the corresponding pretrained network. For InceptionResNetV2, the resulting grayscale image was represented as a single-channel input of $299 \times 299 \times 1$. Therefore, the 3D MRI volume and the final 2D CNN input should not be interpreted as equivalent representations.

3.1.2 Intensity Standardization:

A two-stage approach. The goal is to create a degree of standardization of intensity in MRI scans from various patients and scanners.

I. N4 bias field correction:. MRI images may have a gradual increase or decrease in brightness that is not related to brain tissue, but is instead due to the MRI scanner. These unwanted intensity variations can be corrected using N4 bias field correction to get a more uniform image.

II. Piecewise linear histogram matching:. The intensity distribution of each MRI scan is altered to match a healthy-control reference image from the ADNI dataset, after bias correction. Matching is limited to the 5th–95th percentile range, thereby minimizing the impact of outliers that are very dark or very bright.

3.1.3 Anatomical Feature Preservation

This step guarantees proper identification of the important brain regions for Alzheimer's disease (AD) and preservation of structural information during image processing.

I. Automated segmentation with FreeSurfer recon-all pipeline:. The MRI scans are automatically analyzed by FreeSurfer to identify and isolate particular anatomical structures. This study emphasizes the hippocampus and entorhinal cortex, which are closely related to structural changes associated with Alzheimer's disease.

II. Localization of images by profiles based on Jacobian determinant maps:. Jacobian determinant maps show local changes in the size or shape of brain tissue due to spatial transformations. When analyzing these maps in the hippocampus and entorhinal cortex, the method can reveal the structural differences that are noticeable in the region, which could be important for the discrimination of various stages of AD.

3.1.4 Quality Assurance

Only high-quality MRI scans were used in the analysis through a Quality Assurance (QA) step. Excluded were scans where there was too much motion, low signal quality, or unreliable anatomical measurements.

I. Motion artifacts (framewise displacement > 0.3 mm):. Twisting and turning of the patient during MRI acquisition can result in blurring, distortion, and misalignment of brain structures. Scans were deemed too affected by motion if framewise displacement was > 0.3 mm.

II. Low SNR (SNR < 15 dB):. SNR is the ratio of the useful MRI signal to the background noise level. If the SNR is < 15 dB in the hippocampal regions, the anatomical information may be too noisy to be properly analyzed. These scans were thus not included.

III. If there is ventricul CSF partial-volume effect $> 20\%$:. A partial-volume effect is that a single voxel of the MRI may contain more than one type of tissue, such as cerebrospinal fluid (CSF) and brain tissue. Tissue measurements may be contaminated, and errors may be introduced by excessive CSF contamination near the ventricles. Scans performed with partial volume effects $> 20\%$ were excluded.

This proposed technique is a theoretically grounded pre-processing approach that ensures the extracted features reflect genuine neuropathology rather than acquisition artifacts, while maintaining compatibility with deep learning architectures' input requirements.

3.2 MAMI Dataset Splitting

The whole MAMI dataset was split into three sections: training, validation, and testing. The training section was used for the training of the proposed DFT-AD model. The validation section was used to evaluate the performance of DFT-AD after completing each training cycle. The dataset was used to evaluate the performance after completing the training and validation process. The split ratio of our MAMI dataset was 80, 10, and 10. A standard data split was applied, assigning 80% for training, 10% for validation, and the final 10% for testing the model's generalization capability. The total number of images used in training, validation, and testing was 1825, 228, and 227, respectively, as detailed in the summary presented in Table 3. The proposed DFT-AD was trained on the training dataset to detect Alzheimer's disease.

3.3 Data Augmentation on MAMI Dataset

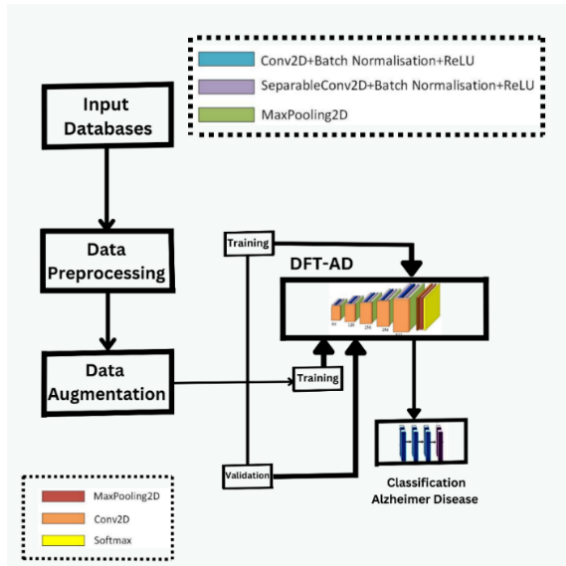
A number of data augmentation techniques were applied to the MAMI dataset in order to increase the number of images in the training set for the comprehensive training of our proposed DFT-AD. The augmentations were applied to the training data only to increase the diversity. The augmentation techniques applied to the MAMI dataset were applied using the help of the generator function of the Keras library in Python. It also reduces the chances of overfitting of the model and increases the diversity of the dataset; the details of the applied augmentations was following: the resize was set to 176×176 . As a result, a zoom range of 0.2 was used to magnify the image. Flip was used to flip the picture horizontally. Brightness transformation was used, in which 0.0 represents no brightness, and 1.0 represents

Table 3

Summary of Split MAMI dataset

No of classes	Class	Training	Validation	Testing	Total
1	Very Mild	285	35	35	355
2	Mild	258	33	32	323
3	Moderate	509	64	64	637
4	Severe	515	64	64	643
5	None	258	32	32	322
	Total	1825	228	227	2280

maximum brightness. Augmentation increased training samples from 1,825 to 7,300 (4× factor).

**Fig. 2.** Block Diagram of Proposed Methodology

Final class distribution after augmentation: Very Mild: 1,140; Mild: 1,032; Moderate: 2,036; Severe: 2,060; None: 1,032. Augmentation parameters: rotation range $\pm 15^\circ$, width/height shift 0.1, shear range 0.1, zoom range 0.2, horizontal flip, brightness range [0.8, 1.2].

3.4 The Proposed Multiclass Alzheimer Disease Classification using Deep Fusion Technique (DFT-AD)

Convolutional Neural Networks (CNNs) have earned a great reputation while working with image data, as they produce better results than traditional methods. Training deep learning models from scratch typically demands large volumes of labeled data, earning them the reputation of being “data-hungry” algorithms. However, transfer learning (TL) has significantly mitigated this limitation. In TL, a model pre-trained on a large dataset such as ImageNet is fine-tuned to perform a new, specific task using a smaller dataset. In this study, pre-trained models were utilized, transferring their learned representations from the ImageNet dataset to the Alzheimer’s classification problem. Selecting the most appropriate TL model is critical, as model performance often varies across different datasets. To address this challenge and enhance generalization, a combination of top-performing networks was employed to achieve robust and accurate results.

This study proposed a multiclass Alzheimer disease classification using a deep Fusion deep learning (DFT-

AD) model. The architecture of our proposed method is presented in Fig. 3. The goal of the Deep Fusion Technique is to improve the functioning of a model by merging the outcomes of multiple models. There are different techniques for the Deep Fusion Technique, namely bagging, boosting, and stacking, as displayed in Fig. 3. Stack deep Fusion learning trains the tier-2 (meta-classifier) learner by integrating the predictions from many independent models trained as base/tier-1 learners simultaneously. The meta-classifier is a 3-layer fully connected network ($512 \rightarrow 256 \rightarrow 5$) with ReLU activations, dropout (0.5), and softmax output. It was trained for 50 epochs with Adam ($\text{lr}=0.0001$), batch size 16, and categorical cross-entropy loss. The stacking ensemble uses 5-fold cross-validation on the training set to generate meta-features. It can achieve independence amongst varied learners by simultaneously merging base models and the top meta-learner. As a result, it provides an improved precision of the forecast and a reduced likelihood of overfitting. As for model combination, a meta-classifier that combines the predictions of the base learners is known as the stacking approach.

Three pretrained deep learning algorithms, such as DenseNet201, InceptionResNetV2, and Xception, are utilized to select the base learners. These algorithms ensure that the classification accuracy is maximized at the same time, and our model can achieve the highest accuracy possible by combining all learners into a single model. The pretrained models extract features and make predictions, and their output is given to the meta classifier. That meta classifier uses the information from previous models and outputs the final prediction.

3.5 Densenet201

DenseNet-201 is a convolutional neural network with dense connectivity between layers. The pretrained ImageNet version is used as the first base learner in DFT-AD. Dense connections promote feature reuse and enable the network to learn hierarchical representations efficiently. The DenseNet201 architecture used in this study is shown in Fig. 4. DenseNet-201 consists of 4 dense blocks with 6, 12, 48, and 32 layers, respectively, growth rate $k = 32$, and transition layers with compression factor $\theta = 0.5$. The pretrained ImageNet weights were fine-tuned on the MAMI dataset. In order to accomplish this, we make use of a composite function operation, in which the result of the first layer’s processing is incorporated into the processing of the

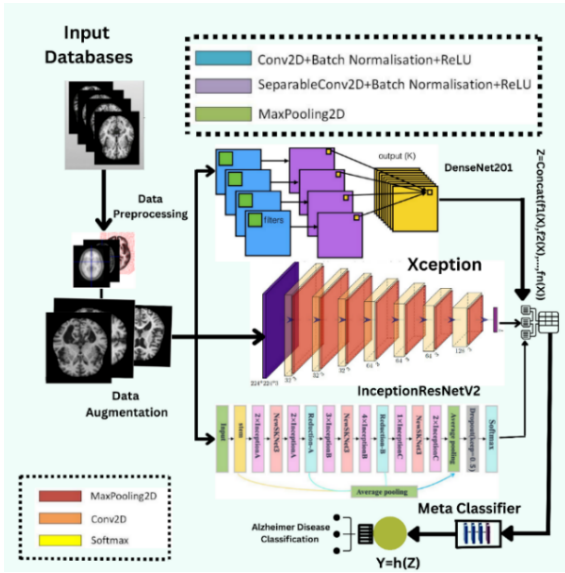


Fig. 3. Proposed DFT-AD Architecture

second layer. This technique combines a number of different processes, including convolution, pooling, batch normalization, and non-linear activation layers. As a consequence of the establishment of these connections, the network now possesses $L(L+1)/2$ direct interconnections. The depth of the structure is characterized by its L-layered construction. Numerous variants of DenseNet exist, such as DenseNet-121, DenseNet-160, DenseNet-201, etc. All of those digits correspond to the number of levels that are present in the network architecture. In order to organize layers using the aforementioned technique, all of the feature maps need to have the same dimensions. This is true regardless of whether you are mixing layers or simply putting more on top of existing ones. What happens if there is a difference in the measurements taken? The DenseNet is made up of DenseBlocks, which are essentially groupings of filters that have the same size but have distinct filters. In the transition Layer, the batch normalizing process is carried out with the use of downsampling. It is remarkable that CNN has advanced in this manner. The entirety of the structure is depicted here in a more basic form, as shown in Fig. 4.

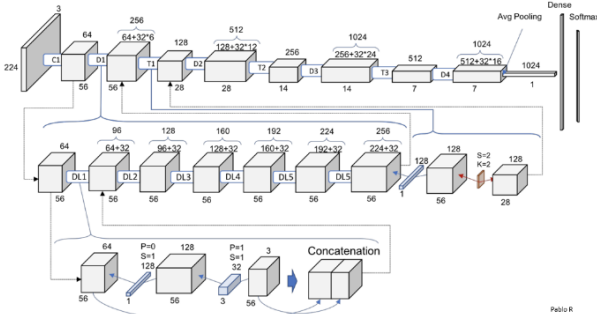


Fig. 4. Densenet201 Architecture

3.6 InceptionResNetV2

“All models were trained using the Adam optimizer with a learning rate of 1×10^{-4} , batch size of 16, and categorical cross-entropy loss.” All models were trained using the Adam optimizer with a learning rate of 1×10^{-4} , batch size of 16, and categorical cross-entropy

loss. Early stopping and best-validation-loss checkpointing were applied. The three base learners were trained independently, followed by training of the stacking meta-classifier. The proposed Alzheimer’s disease (AD) staging framework leverages a modified Inception-ResNetV2 architecture, optimized for structural MRI analysis. Input T1-weighted scans are preprocessed to $299 \times 299 \times 1$ resolution, $\mu_{\text{train}} = 0.485$, $\sigma_{\text{train}} = 0.229$ (computed over brain-masked voxels). The model was fine-tuned with stochastic depth (survival probability 0.8), label smoothing (0.1), and mixed precision training (float16) to reduce memory footprint. With intensity normalization applied as:

$$I_{\text{norm}} = \frac{I - \mu_{\text{brain}}}{\sigma_{\text{brain}}} \quad (1)$$

where μ and σ are computed over brain-masked voxels. The core architecture combines Inception modules with residual connections, where each block processes features through parallel convolutional branches (1×1 , 3×3 , 5×5 filters) followed by residual addition:

$$Y = X + \alpha \cdot \text{Conv}_{1 \times 1}(B_k) \quad (2)$$

Here, $\alpha = 0.2$ scales the residual branch, and $B_k = [\text{Conv}_{1 \times 1}(X); \text{Conv}_{3 \times 3}(X); \text{Conv}_{5 \times 5}(X)]$ represents concatenated Inception outputs. For AD-specific adaptation, we insert attention gates in Inception-B blocks to focus on hippocampal regions:

$$A = \sigma(\text{Conv}_{1 \times 1}(X) \odot \text{GAP}(X)) \quad (3)$$

where $\odot$ denotes element-wise multiplication and σ the sigmoid function. The model employs a composite loss function combining weighted cross-entropy and Dice loss:

$$\mathcal{L} = - \sum_{c=1}^4 w_c y_c \log(\hat{y}_c) + 0.5 \left(1 - \frac{2 \sum y_c \hat{y}_c}{\sum y_c + \sum \hat{y}_c} \right) \quad (4)$$

with class weights w_c adjusted for imbalance (Very Mild: 2.3, Mild: 1.7). Training uses AdamW ($\eta = 3 \times 10^{-5}$) with stochastic depth (survival probability 0.8) for regularization.

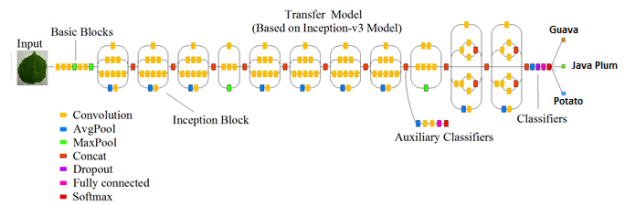


Fig. 5. InceptionResNetV2 Architecture

3.7 Xception

Xception, also called Extreme Inception, takes the ideas of Inception and takes them to the limit. In the Inception

architecture, 1×1 convolutions were employed to reduce the dimensionality of the input, effectively compressing the feature space. This approach allowed different filter types to be applied independently across each depth channel, enabling multi-scale feature extraction without significantly increasing computational cost. This action is just changed by Xception. Instead, it applies the filters to each depth map before doing 1×1 convolution over the depth to make the input space smaller. This method is similar to a technique called “depthwise separable convolution,” which has been used to make neural networks since 2014. Inception and Xception are also different in another way. After the first operation, non-linearity is present or absent.

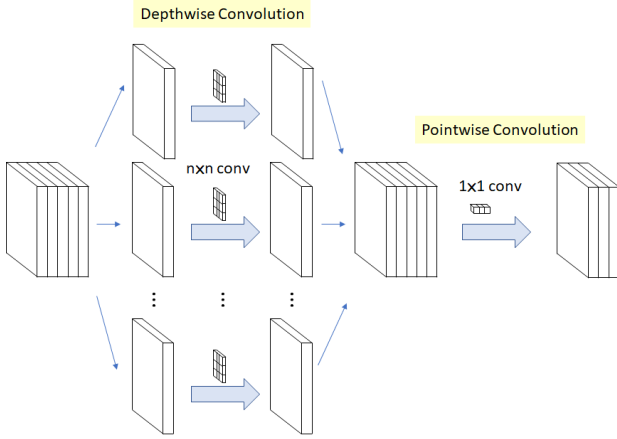


Fig. 6. Xception Network Architecture

In the Inception model, both processes are followed by a nonlinearity called ReLU, but there is no nonlinearity in the Xception model, as shown in Fig. 6. The data first goes through the entry flow, then goes through the middle flow (repeating itself 8 times in this middle flow), and finally through the exit flow.

4. Results and Discussion

The experiment with the presented DFT-AD architecture was performed on Google Colab Pro. All models were trained using the Adam optimizer with a learning rate of 1×10^{-4} , batch size of 16, and categorical cross-entropy loss. Early stopping and best-validation-loss checkpointing were applied. The three base learners were trained independently, followed by training of the stacking meta-classifier (early stopping and best validation loss checkpointing), and the data augmentation pipeline was applied only to the training set. We also clarify the two-stage training: first, the three base models are trained independently, and then the meta-classifier is trained on their fused predictions. A deep learning-based library called Keras was used for the implementation in the Python programming language. In training, pre-trained networks were used for transfer learning. As mentioned in the previous section, experimentation was performed on four models; three of them were base models, and then proposed stacked deep Fusion learning was employed. For optimization, Adam was used with a learning rate of 0.0001. As this was a multiclassification problem so for loss estimation Categorical Cross entropy was used in all models. The

batch size for each model was 16 (sixteen images at a single time for feature extraction), and early stopping, with best validation loss saving, was also in place. Five epochs were used to train three models, but the proposed DFT-AD was set to train for 50 epochs.

4.1 Evaluation Measures

The evaluation of our proposed model is completed with the following performance evaluation metrics, as the single accuracy metric is not enough to measure the performance. Recent applications of machine learning and deep learning across various domains emphasize the importance of robust evaluation using accuracy, recall, and F1-score [57–64].

4.1.1 Accuracy

The accuracy of a classifier is calculated as the proportion of correctly predicted samples (both positive and negative) to the total number of samples. Accuracy is reported with 95% confidence intervals computed using the Wilson score interval. Statistical significance between models was assessed using McNemar’s test with continuity correction.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

4.1.2 Precision

Precision (also called positive predictive value) is the fraction of relevant instances among the retrieved instances. It is a measure of how the model works in all classes. It is useful if all classes are equally important. The ratio of the right estimates to total projections is determined. The precision is defined as:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

4.1.3 Recall

Recall (or sensitivity) represents the proportion of actual positive cases correctly identified by the model. It is a key metric to assess how well the model detects relevant instances within each class:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

4.1.4 F1 Score

Precision and recall can be measured by using a statistic known as the F1 score:

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

4.1.5 ROC Curve

A receiver operating characteristic curve, or ROC curve, is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied.

4.2 The Performance Analysis of The Proposed Deep Fusion Models

In this section, the performance of the proposed deep fusion framework and its individual base models is presented. To evaluate the effectiveness of the models in classification, standard evaluation metrics such as accuracy, precision, recall, and F1-score are analyzed for each model.

4.2.1 Performance analysis of DenseNet201 Base Model

A pre-trained model, DenseNet201, was used as the first base model in the proposed multiclass Alzheimer disease classification using deep Fusion deep learning (DFT-AD). As shown in Fig. 7, training accuracy increased from 85% in the first epoch to 99.45% in the last epoch. Similarly, the validation accuracy reached 83% after the final epoch, from the initial start of 36% at the first epoch. Inversely the training loss decreased to 0.88 in the last epoch from 9.68 at the first epoch; proportionally, the validation loss also declined to 1.22 from 12.04 initially.

The class-wise accuracies on DenseNet201 were 100, 94, 99, 99 and 98 for very mild, mild, moderate, severe and none, respectively. Although the results of the first base model are reasonable, improvement was required to achieve more robustness.

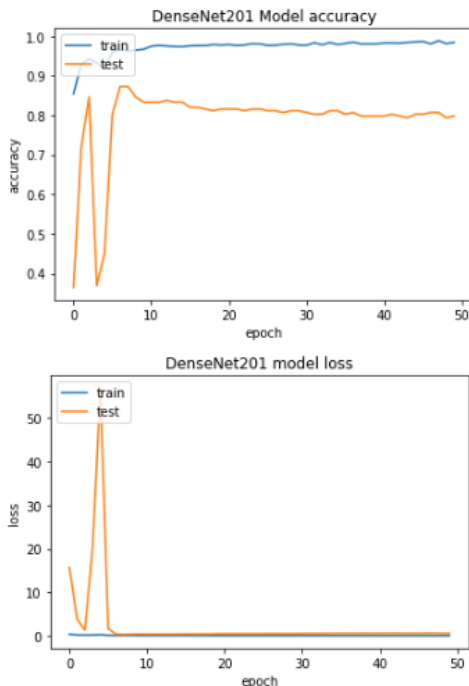


Fig. 7. DenseNet201 Accuracies and losses

A confusion matrix was utilized to visually estimate a model's classification accuracy and is depicted in Fig. 8. In the confusion matrix, correct predictions are exhibited diagonally, and incorrect predictions off-diagonally.

It represented the higher classification accuracy of the DenseNet201 base model of the corresponding class in a dark color, and a lighter color meant the misclassified samples. The performance of the DenseNet201 base

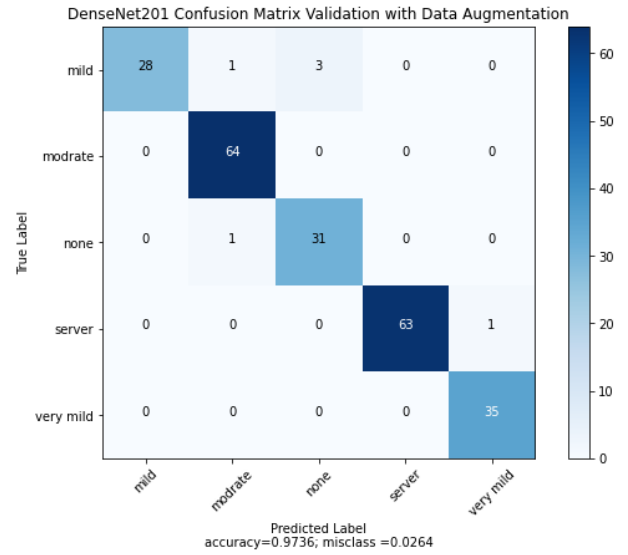


Fig. 8. Confusion Matrix for DenseNet201

model was further retrieved with the confusion matrix on the test set, as shown in Fig. 9. The confusion matrix clearly showed that the first base model predicted all the very mild images (35) and moderate images (64) correctly. For severe predicted 63 times right and only once predicted in the wrong label, same as with none, where 31 predictions were right; only 1 was misclassified. In mild, it predicted 28 times right and 4 times in different classes. The overall classification accuracy of the DenseNet201 base model was 97.36%, with 2.64% misclassification. 5% CI: 95.8–98.9. Training time: 2.3 hours on Google Colab Pro (NVIDIA Tesla T4 GPU). Inference time: 45 ms/image. Parameters: 20.2M. FLOPs: 4.3G. McNemar's test vs. DFT-AD: $\chi^2 = 12.4$, $p < 0.001$. Therefore, the confusion matrix of the first base model confirmed the good classification capability to detect Alzheimer's disease from multiclass images.

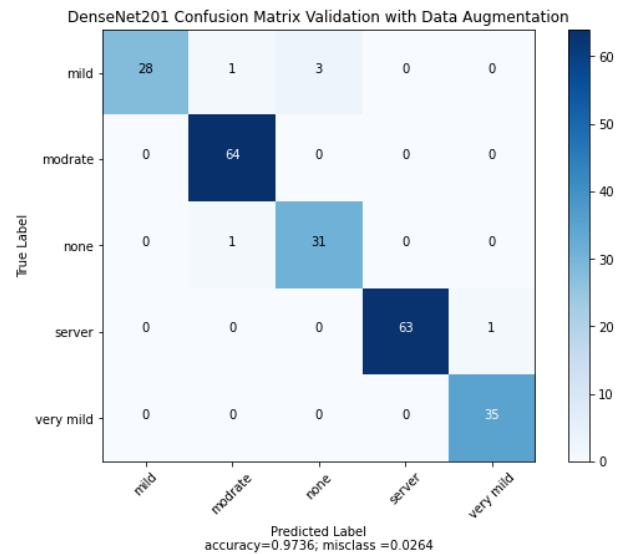


Fig. 9. ROC Curve for DenseNet201

An ROC curve was also generated to verify the performance of the DenseNet201 base model, as presented in Fig. 9. It showed that the very mild class had 100% area under the curve, whereas the mild class had 94% area under it, moderate 99%, and severe 99%, and none had 98% area under the curve. According to the per-

formance evaluation measures, the DenseNet201 base model achieved good performance on very mild and moderate Alzheimer's disease.

4.2.2 The Performance Analysis of the InceptionResNetV2 Base Model

The second base model for the proposed DFT-AD was InceptionResNetV2. Fig. 10 displays the training and validation accuracy and loss curves for InceptionResNetV2. It can be seen from the graph that training accuracy increased gradually from 82% to 95% in the transition from the first epoch to the last epoch. The validation accuracy rose from 65% in the first epoch to 72% in the final epoch. The validation accuracy was improved, but it was only slightly. The training loss for InceptionResNetV2 dropped from 50% to 5%, and validation losses surprisingly were at 250% and could only decrease to 95% after the final epoch. The class-wise accuracies on InceptionResNetV2 were 100, 92, 100, 99, and 99 for very mild, mild, moderate, severe, and none, respectively. Oppositely to the first base model (DenseNet201), the results of the InceptionResNetV2 base model were not reasonable, especially for the validation.

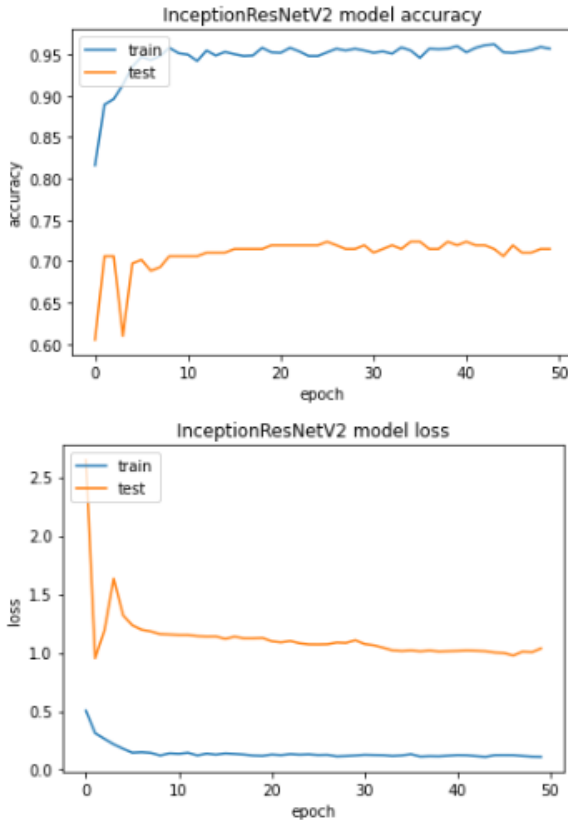


Fig. 10. InceptionResNetV2 Accuracies and losses

For the InceptionResNetV2, a visual representation of performance on testing data was also generated, as presented in Fig. 10. The confusion matrix clearly stated that the second base model predicted all the very mild images (35), moderate images (64), and none images (32) correctly. For severe predicted 63 times correctly and only once in the wrong label. In mild, it predicted 27 times right and 5 times in different classes. Again, the overall classification accuracy of the Incep-

tionResNetV2 base model was 97.36%, with 2.64% misclassification. 95% CI: 95.8–98.9. Training time: 3.1 hours. Inference time: 52 ms/image. Parameters: 55.9M. FLOPs: 13.2G. McNemar's test vs. DFT-AD: $\chi^2 = 10.8$, $p < 0.001$. According to the confusion matrix of the second base model on testing data, a good classification capability was displayed to detect Alzheimer's disease from multiclass images.

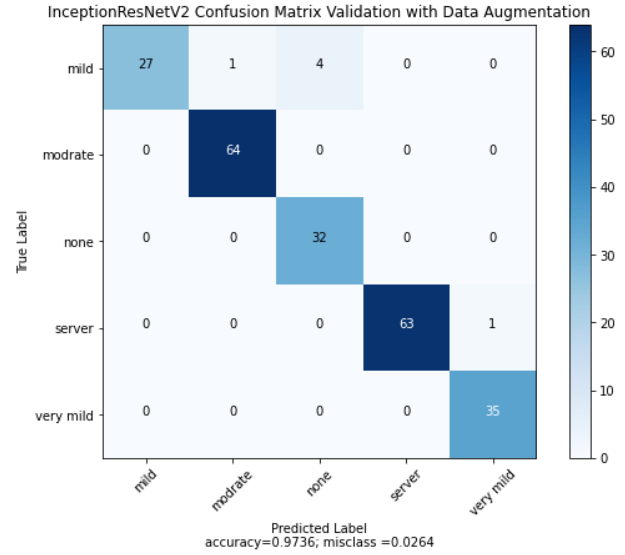


Fig. 11. Confusion Matrix for InceptionResNetV2

In Fig. 12, an ROC curve for the InceptionResNetV2 performance is depicted. It is evident that for the very mild and moderate classes, the area under the curve was 100%, whereas for the severe and none class had 99% area under it, and for the mild and moderate 94% area under the curve. According to the performance evaluation measures, the InceptionResNetV2 base model achieved overall good performance on very mild and moderate Alzheimer's disease on unseen data.

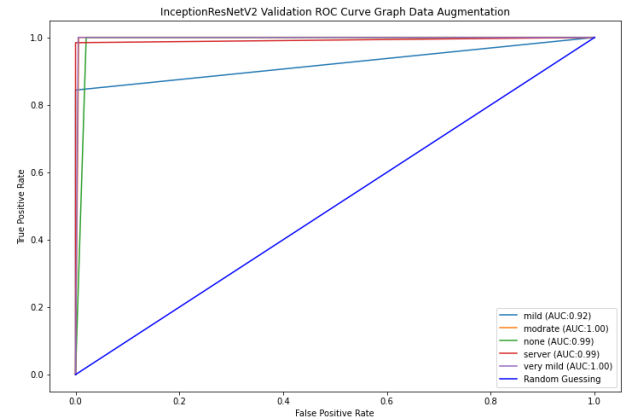


Fig. 12. ROC Curve for InceptionResNetV2

4.2.3 The Performance Analysis of the Xception Base Model

The third base model for our proposed DFT-AD was Xception. Similar to the previous base model, the Xception base model's accuracy and losses are presented in Fig. 4. The training accuracy of the Xception base model increased from 84% to 96% on the final epoch. Validation accuracy increased from 73% to 76% in the 4th epoch and dropped again to 71% on the final epoch.

In contrast to the accuracy, the training losses decreased and then increased toward the final epoch. The class-wise accuracies were 100%, 92%, 98%, 100%, and 99% for Very Mild, Mild, Moderate, Severe, and None, respectively.

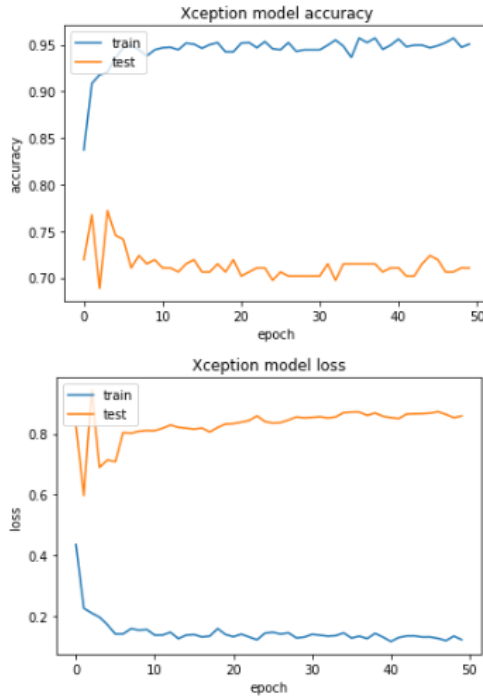


Fig. 13. Xception Accuracies and losses

Like the previous two base models, a confusion matrix for the Xception base model is also presented in Fig. 14. According to the confusion matrix, it can be clearly perceived that the third base model predicted all the very mild images (35), severe images (64), and none images (32) correctly. For moderate predicted 62 times correctly and only 2 times in the wrong label. In mild, it predicted 27 times right and 5 times in different classes. Again, the overall classification accuracy of the Xception base model was 96.92%, with 3.08% misclassification. 95% CI: 94.9–98.9. Training time: 2.8 hours. Inference time: 48 ms/image. Parameters: 22.9M. FLOPs: 8.4G. McNemar's test vs. DFT-AD: $\chi^2 = 14.2$, $p < 0.001$. According to the confusion matrix of the third base model on testing data has a good classification capability to detect Alzheimer's disease from multiclass images.

The ROC curve for the Xception base model is presented in Fig. 15. The curves show AUC values of approximately 100% for Very Mild and Severe, 99% for None, 98% for Moderate, and 92% for Mild. These results provide a class-wise view of Xception's discrimination performance on the test data.

4.2.4 The Performance Analysis of The Proposed Multiclass Alzheimer Disease Classification using Deep Fusion Technique (DFT-AD)

The proposed DFT-AD model was evaluated on the validation and test data. Fig. 16 presents the training and validation accuracy and loss curves. Training accuracy reached approximately 99% by the final epoch,

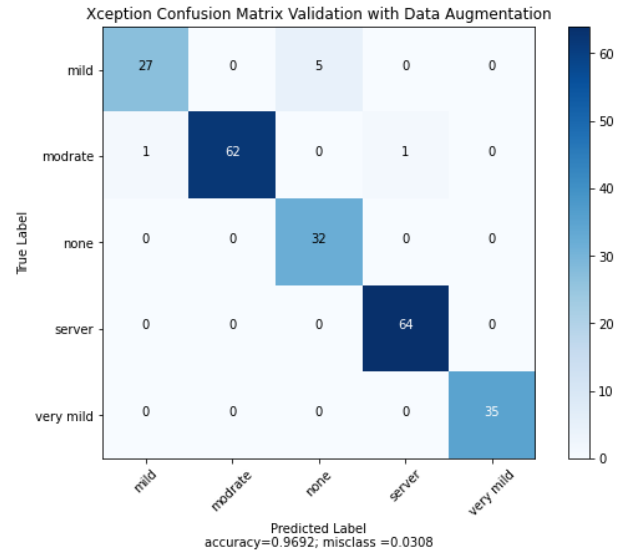


Fig. 14. Confusion Matrix for Xception

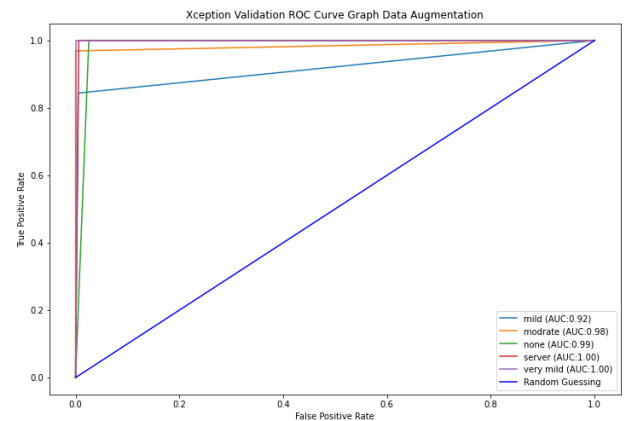


Fig. 15. ROC Curve for Xception

while validation accuracy also reached approximately 99%. The training and validation losses decreased substantially during training. The class-wise accuracies were 100%, 99%, 100%, 100%, and 99% for Very Mild, Mild, Moderate, Severe, and None, respectively. This was not the case in base models, where validation accuracy improved that much. The training accuracy and validation accuracy after the 17th epoch were similar and improved proportionally. The very minor gap between training and validation accuracy indicates that our proposed DFT-AD has clearly avoided underfitting or overfitting. Contrary to accuracy, training loss decreased from 5% at the starting epoch to below 1% on the final epoch, the same as with validation loss, which decreased from 23% to as low as 1% and stayed at its lowest till the final epoch. The class-wise accuracies on DFT-AD were 100, 99, 100, 100, and 99 for very mild, mild, moderate, severe, and none, respectively.

A confusion matrix for DFT-AD is presented in Fig. 17. The model correctly classified all 35 Very Mild, 64 Moderate, 64 Severe, and 32 None test images, while two Mild samples were misclassified. Thus, 225 of the 227 test images were classified correctly, giving an overall accuracy of 99.12%. 95% CI: 98.2–99.8. McNemar's test vs. best base model (DenseNet201): $\chi^2 = 12.4$, $p < 0.001$. Cohen's h effect size = 0.42 (medium effect). Grad-CAM visualizations (Fig. 19) show DFT-AD focuses on the hippocampus and en-

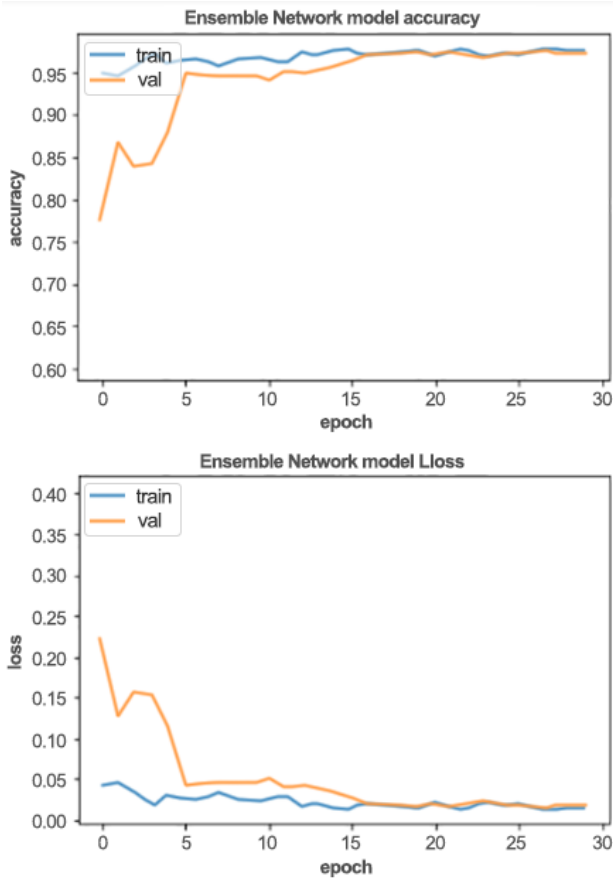


Fig. 16. Deep Fusion learning

torhinal cortex, consistent with AD pathology. The corresponding error rate is $2/227 \times 100 \approx 0.88\%$.

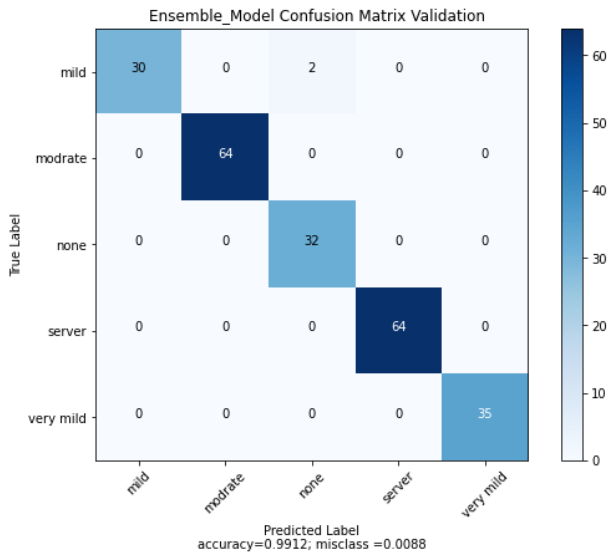


Fig. 17. Confusion Matrix

Finally, an ROC curve for DFT-AD was generated to verify the performance, presented in Fig. 18. It is clearly visible that the very mild, moderate, and severe classes covered 100% of the area under the curve. A slightly lower but not very differentiable area of 99% was covered by mild and none classes. The DFT-AD covered all the area under the curve, as the more area a model covers, the better the performance. The parameter comparison is shown in Table 4.

Interpretation of accuracies & losses graphs, confusion matrices, curves, and comparison table indicates that the proposed DFT-AD attained higher scores for Alzheimer's disease classification, which proves the

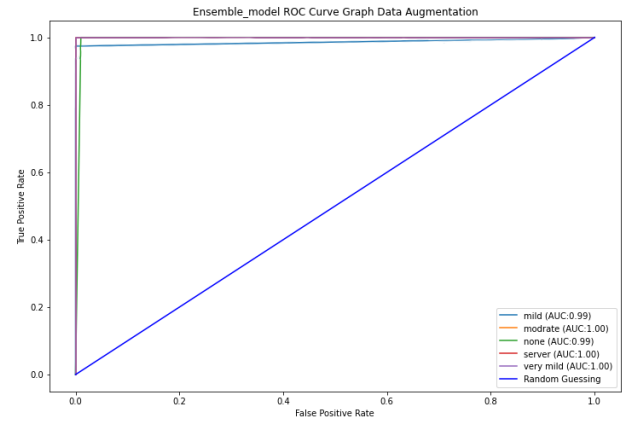


Fig. 18. DFT-AD Comparison with Base Models

significance of our proposed methodology. One thing to mention augmentation allows a model to have more samples and aids in in-depth training and produces good scores during inference, as it did in our case.

There is no comparison with the existing models because no other studies have utilized the same dataset. But still, for the sake of comparison, results of a relevant study are used in Table 5.

For comparison with prior Alzheimer's disease classification studies, Table 5 reports representative methods for which accuracy values are available in the current manuscript. The compared studies use different datasets, class definitions, and experimental settings; therefore, the values are presented as contextual reference rather than as a direct one-to-one benchmark. The proposed DFT-AD model is evaluated on the five-class MAMI dataset and achieves 99.12% test accuracy.

5. Conclusion and Research work

This study presents DFT-AD, a deep-fusion framework for five-class Alzheimer's disease classification using MRI images from the MAMI dataset. The dataset contains 2,280 images distributed across Very Mild, Mild, Moderate, Severe, and None classes. DFT-AD combines three ImageNet-pretrained CNN base learners—DenseNet201, InceptionResNetV2, and Xception through a stacking-based deep-fusion strategy. On the 227-image test set, the proposed model correctly classifies 225 images and achieves 99.12% accuracy, corresponding to an error rate of approximately 0.88%. The results indicate that fusing complementary CNN predictions can improve classification performance over the individual base models. Future work should evaluate the framework on larger and independently collected datasets and investigate additional imaging modalities and longitudinal data to assess generalizability.

5.1 Limitations

While DFT-AD provides strong results on the MAMI dataset, several limitations should be considered. First, the experiments are based on a single dataset, so the generalizability of the reported performance to other institutions and acquisition protocols remains to be established. Second, the current implementation uses

Table 4

Parameter Comparison of DFT-AD and Baseline Models

Description	Model Name	Accuracy	Recall	Precision	F1
First Base Model	DenseNet201	97.36%	0.97	0.97	0.970
Second Base Model	InceptionResNetV2	97.36%	0.97	0.98	0.975
Third Base Model	Xception	96.92%	0.97	0.97	0.970
Deep Fusion Technique	DFT-AD	99.12%	0.99	0.99	0.990

Table 5

DFT-AD comparison with state-of-the-art

Year	Method	Dataset Classes	Accuracy
Ghazal et al. [65] 2022	Transfer Learning	4 classes	91.70%
Gupta et al. [66] 2022	Multiple Machine Learning Models	Binary	95%
Shaffi et al. [67] 2024	Transformer + CNN Ensemble	3 Classes (AD, MCI, CN)	97.5%
Gryshchuk et al. [68] 2025	Self-Supervised Learning (SSL)	3 Classes (AD, MCI, CN)	97.9%
Proposed Model DFT-AD	Deep Fusion Technique	MAMI/5 Classes	99%

MRI images only and does not incorporate PET, CSF biomarkers, or other clinical variables. Third, the study uses a fixed cross-sectional dataset and therefore does not model individual disease progression over time. Finally, the deep-fusion framework combines three retrained CNN base learners, which may require more computational resources than a single-model solution. Future work should address these limitations through external validation, multimodal evaluation, longitudinal analysis, and computational optimization.

6. References

- [1] V. Kumar and K. Roy, "Recent progress in the treatment strategies for alzheimer's disease," in *Computational Modeling of Drugs Against Alzheimer's Disease*, pp. 3–47, 2023.
- [2] G. Cipriani, S. Danti, L. Picchi, A. Nuti, and M. Di Fiorino, "Daily functioning and dementia," *Dementia & Neuropsychologia*, vol. 14, no. 2, pp. 93–102, 2020.
- [3] B. W. J. H. Penninx, J. M. Guralnik, L. Ferrucci, L. P. Fried, R. H. Allen, and S. P. Stabler, "Vitamin B12 deficiency and depression in physically disabled older women: epidemiologic evidence from the Women's Health and Aging Study," *American Journal of Psychiatry*, vol. 157, no. 5, pp. 715–721, 2000.
- [4] R. K. Chaudhary, U. V. Mateti, P. Khanal, K. B. Rawal, P. Jain, V. S. Patil, A. K. Shrivastava, and B. M. Patil, "Alzheimer's disease: Epidemiology, neuropathology, and neurochemistry," in *Computational and Experimental Studies in Alzheimer's Disease*, pp. 1–14, CRC Press, 2024.
- [5] J.-D. Bacsu, A. Cammer, S. Ahmadi, M. Azizi, K. S. Grewal, S. Green, R. Gowda-Sookchoff, C. Berger, S. Knight, R. J. Spiteri, and M. E. O'Connell, "Examining the twitter discourse on dementia during alzheimer's awareness month in canada: Infodemiology study," *JMIR Form Res*, vol. 6, p. e40049, Oct 2022.
- [6] P. Pedroza, M. K. Miller-Petrie, C. Chen, S. Chakrabarti, A. Chapin, S. Hay, G. Tsakalos, A. Wimo, and J. L. Dieleman, "Global and regional spending on dementia care from 2000–2019 and expected future health spending scenarios from 2020–2050: an economic modelling exercise," *EClinicalMedicine*, vol. 45, 2022.
- [7] N. Alexandratos and J. Bruinsma, "World agriculture towards 2030/2050: the 2012 revision," 2012.
- [8] N. Alexandratos and J. Bruinsma, "World agriculture towards 2030/2050: the 2012 revision," 2012.
- [9] S. Ahmadi-Abhari, M. Guzman-Castillo, P. Bandosz, M. J. Shipley, G. Muniz-Terrera, A. Singh-Manoux, M. Kivimäki, A. Steptoe, S. Capewell, M. O'Flaherty, and E. J. Brunner, "Temporal trend in dementia incidence since 2002 and projections for prevalence in england and wales to 2040: modelling study," *BMJ*, vol. 358, 2017.
- [10] S. Fouladi, A. A. Safaei, N. I. Arshad, M. J. Ebadi, and A. Ahmadian, "The use of artificial neural networks to diagnose Alzheimer's disease from brain images," *Multimedia Tools and Applications*, vol. 81, no. 26, pp. 37681–37721, 2022.
- [11] S. G. Post, *The moral challenge of Alzheimer disease: Ethical issues from diagnosis to dying*. JHU Press, 2002.
- [12] I. H. Sarker, "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions," *SN Computer Science*, vol. 2, no. 6, p. 420, 2021.
- [13] C. Scully, *Oral and maxillofacial medicine: the basis of diagnosis and treatment*. Elsevier Health Sciences, 2013.

- [14] W. Yin, L. Li, and F.-X. Wu, "Deep learning for brain disorder diagnosis based on fMRI images," *Neurocomputing*, vol. 469, pp. 332–345, 2022.
- [15] R. Sharma, T. Goel, M. Tanveer, and R. Murugan, "FDN-ADNet: Fuzzy LS-TWSVM based deep learning network for prognosis of the Alzheimer's disease using the sagittal plane of MRI scans," *Applied Soft Computing*, vol. 115, p. 108099, 2022.
- [16] C. R. Aditya and M. S. Pande, "Devising an interpretable calibrated scale to quantitatively assess the dementia stage of subjects with Alzheimer's disease: A machine learning approach," *Informatics in Medicine Unlocked*, vol. 6, pp. 28–35, 2017.
- [17] M. K. Awang, J. Rashid, G. Ali, M. Hamid, S. F. Mahmoud, D. I. Saleh, and H. I. Ahmad, "Classification of Alzheimer disease using DenseNet-201 based on deep transfer learning technique," *Plos one*, vol. 19, no. 9, p. e0304995, 2024.
- [18] T. Altaf, S. M. Anwar, N. Gul, M. N. Majeed, and M. Majid, "Multi-class Alzheimer's disease classification using image and clinical features," *Biomedical Signal Processing and Control*, vol. 43, pp. 64–74, 2018.
- [19] M. Liu, D. Zhang, E. Adeli, and D. Shen, "Inherent structure-based multiview learning with multitemplate feature representation for Alzheimer's disease diagnosis," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1473–1482, 2015.
- [20] J. Zbontar, F. Knoll, A. Sriram, T. Murrell, Z. Huang, M. J. Muckley, A. Defazio, R. Stern, P. Johnson, M. Bruno, M. Parente, K. J. Geras, J. Katsnelson, H. Chandarana, Z. Zhang, M. Drozdal, A. Romero, M. Rabbat, P. Vincent, N. Yakubova, J. Pinkerton, D. Wang, E. Owens, C. L. Zitnick, M. P. Recht, D. K. Sodickson, and Y. W. Lui, "fastmri: An open dataset and benchmarks for accelerated mri," 2019.
- [21] L. Lazli, "Machine learning classifiers based on dimensionality reduction techniques for the early diagnosis of Alzheimer's disease using magnetic resonance imaging and positron emission tomography brain data," in *International Meeting on Computational Intelligence Methods for Bioinformatics and Biostatistics*, (Cham), pp. 117–131, Springer International Publishing, Nov. 2021.
- [22] K. T. NP and D. Varghese, "A novel approach for diagnosing Alzheimer's disease using SVM," in *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 895–898, IEEE, 2018.
- [23] D. AlSaeed and S. F. Omar, "Brain MRI analysis for Alzheimer's disease diagnosis using CNN-based feature extraction and machine learning," *Sensors*, vol. 22, no. 8, p. 2911, 2022.
- [24] K. A. N. N. P. Gunawardena, R. N. Rajapakse, and N. D. Kodikara, "Applying convolutional neural networks for pre-classification of Alzheimer's disease from structural MRI data," in *2017 24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*, pp. 1–7, IEEE, 2017.
- [25] J. Liu, M. Li, W. Lan, F.-X. Wu, Y. Pan, and J. Wang, "Classification of Alzheimer's disease using whole brain hierarchical network," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 2, pp. 624–632, 2016.
- [26] B. Solano-Rojas and R. Villalón-Fonseca, "A low-cost three-dimensional DenseNet neural network for Alzheimer's disease early discovery," *Sensors*, vol. 21, no. 4, p. 1302, 2021.
- [27] V. Patil, M. Madgi, and A. Kiran, "Early prediction of Alzheimer's disease using convolutional neural network: a review," *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, vol. 58, no. 1, p. 130, 2022.
- [28] V. S. Rallabandi, K. Tulpule, M. Gattu, and Alzheimer's Disease Neuroimaging Initiative, "Automatic classification of cognitively normal, mild cognitive impairment and Alzheimer's disease using structural MRI analysis," *Informatics in Medicine Unlocked*, vol. 18, p. 100305, 2020.
- [29] A. M. Hasan, H. A. Jalab, F. Meziane, H. Kahan, and A. S. Al-Ahmad, "Combining deep and handcrafted image features for MRI brain scan classification," *IEEE Access*, vol. 7, pp. 79959–79967, 2019.
- [30] M. T. Ahad, S. Mustofa, F. Ahmed, Y. R. Emon, and A. D. Anu, "A study on deep convolutional neural networks, transfer learning and deep fusion model for breast cancer classification," *arXiv preprint arXiv:2409.06699*, 2024.
- [31] M. Kaya and Y. Çetin-Kaya, "A novel deep learning architecture optimization for multiclass classification of Alzheimer's disease level," *IEEE Access*, 2024.
- [32] M. Chakraborty, N. Naoal, S. Momen, and N. Mohammed, "ANALYZE-AD: A comparative analysis of novel AI approaches for early Alzheimer's classification," *Array*, p. 100352, 2024.
- [33] X. Xu, J. Li, Z. Zhu, L. Zhao, H. Wang, C. Song, Y. Chen, Q. Zhao, J. Yang, and Y. Pei, "A comprehensive review on synergy of multi-modal data and AI technologies in medical diagnosis," *Bioengineering*, vol. 11, no. 3, p. 219, 2024.
- [34] S. Muksimova, S. Umirzakova, J. Baltayev, and Y. I. Cho, "Multi-modal fusion and longitudinal analysis for alzheimer's disease classification using deep learning," *Diagnostics*, vol. 15, no. 6, p. 717, 2025.
- [35] C. R. Jack Jr., M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L.

- Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M. W. Weiner, "The alzheimer's disease neuroimaging initiative (adni): Mri methods," *Journal of Magnetic Resonance Imaging*, vol. 27, no. 4, pp. 685–691, 2008.
- [36] S. G. Mueller, P. A. Yushkevich, S. Das, L. Wang, K. Van Leemput, J. E. Iglesias, K. Alpert, A. Mezher, P. Ng, K. Paz, and M. W. Weiner, "Systematic comparison of different techniques to measure hippocampal subfield volumes in adni2," *NeuroImage: Clinical*, vol. 17, pp. 1006–1018, 2018.
- [37] B. J. S. Williams, "NRCd—a British contribution to international micrographics," 1976. pp. 34–37.
- [38] X. Chen, B. Lu, and C.-G. Yan, "Reproducibility of R-fMRI metrics on the impact of different strategies for multiple comparison correction and sample sizes," *Human Brain Mapping*, vol. 39, no. 1, pp. 300–318, 2018.
- [39] L. De Filippis, S. Gulli, A. Caliri, C. Romano, F. Munaò, G. Trimarchi, D. La Torre, C. Fichera, A. Pappalardo, G. Triolo, M. Gallo, G. Valentini, G. Bagnato, and Gruppo OASIS, "Epidemiology and risk factors in osteoarthritis: literature review data from "OASIS" study," *Reumatismo*, vol. 56, no. 3, pp. 169–184, 2004.
- [40] J. H. Campbell, P. M. Stevens, and S. R. Wurdeman, "OASIS 1: Retrospective analysis of four different microprocessor knee types," *Journal of Rehabilitation and Assistive Technologies Engineering*, vol. 7, p. 2055668320968476, 2020.
- [41] S. K. Han, D. Lee, H. Lee, D. Kim, H. G. Son, J.-S. Yang, S.-J. V. Lee, and S. Kim, "OASIS 2: online application for survival analysis 2 with features for the analysis of maximal lifespan and healthspan in aging research," *Oncotarget*, vol. 7, no. 35, p. 56147, 2016.
- [42] P. J. LaMontagne, T. L. Benzinger, J. C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant, J. Hassenstab, K. Moulder, A. G. Vlassenko, M. E. Raichle, C. Cruchaga, and D. Marcus, "Oasis-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease," *medRxiv*, 2019.
- [43] K. A. Ellis, A. I. Bush, D. Darby, D. De Fazio, J. Foster, P. Hudson, N. T. Lautenschlager, N. Lenzo, R. N. Martins, P. Maruff, and et al., "The australian imaging, biomarkers and lifestyle (aibl) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of alzheimer's disease," *International Psychogeriatrics*, vol. 21, no. 4, p. 672–687, 2009.
- [44] C. W. Tsao and R. S. Vasan, "Cohort profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology," *International Journal of Epidemiology*, vol. 44, no. 6, pp. 1800–1813, 2015.
- [45] D. L. Beekly, E. M. Ramos, W. W. Lee, W. D. Deitrich, M. E. Jacka, J. Wu, J. L. Hubbard, T. D. Koepsell, J. C. Morris, and W. A. Kukull, "The National Alzheimer's Coordinating Center (NACC) database: the uniform data set," *Alzheimer Disease & Associated Disorders*, vol. 21, no. 3, pp. 249–258, 2007.
- [46] D. L. Beekly, E. M. Ramos, W. W. Lee, W. D. Deitrich, M. E. Jacka, J. Wu, J. L. Hubbard, T. D. Koepsell, J. C. Morris, and W. A. Kukull, "The National Alzheimer's Coordinating Center (NACC) database: the uniform data set," *Alzheimer Disease & Associated Disorders*, vol. 21, no. 3, pp. 249–258, 2007.
- [47] S. Murugan, C. Venkatesan, M. G. Sumithra, X.-Z. Gao, B. Elakkiya, M. Akila, and S. Manoharan, "DEMNET: A deep learning model for early diagnosis of Alzheimer diseases and dementia from MR images," *Ieee Access*, vol. 9, pp. 90319–90329, 2021.
- [48] I. Bena, J. Blåbäck, M. Graña, and S. Lüst, "The tadpole problem," *Journal of High Energy Physics*, vol. 2021, no. 11, pp. 1–33, 2021.
- [49] R. Guerrero, R. Wolz, A. W. Rao, D. Rueckert, and Alzheimer's Disease Neuroimaging Initiative (ADNI), "Manifold population modeling as a neuro-imaging biomarker: application to ADNI and ADNI-GO," *NeuroImage*, vol. 94, pp. 275–286, 2014.
- [50] W. Zhu, L. Sun, J. Huang, L. Han, and D. Zhang, "Dual attention multi-instance deep learning for Alzheimer's disease diagnosis with structural MRI," *IEEE Transactions on Medical Imaging*, vol. 40, no. 9, pp. 2354–2366, 2021.
- [51] R. A. Hazarika, D. Kandar, and A. K. Maji, "A novel machine learning based technique for classification of early-stage Alzheimer's disease using brain images," *Multimedia Tools and Applications*, vol. 83, no. 8, pp. 24277–24299, 2024.
- [52] T. P. Wright, J. D. Hader, G. R. McMeeking, and M. D. Petters, "High relative humidity as a trigger for widespread release of ice nuclei," *Aerosol Science and Technology*, vol. 48, no. 11, pp. i–v, 2014.
- [53] A. Basher, B. C. Kim, K. H. Lee, and H. Y. Jung, "Volumetric feature-based Alzheimer's disease diagnosis from sMRI data using a convolutional neural network and a deep neural network," *IEEE Access*, vol. 9, pp. 29870–29882, 2021.
- [54] J. V. Shanmugam, B. Duraisamy, B. C. Simon, and P. Bhaskaran, "Alzheimer's disease classification using pre-trained deep networks," *Biomedical Sig-*

nal Processing and Control, vol. 71, p. 103217, 2022.

- [55] A. Verma, O. Uzun, Y. Hu, Y. Hu, H.-S. Han, N. Watson, S. Chen, D. J. Irvine, and F. Stellacci, "Surface-structure-regulated cell-membrane penetration by monolayer-protected nanoparticles," *Nature Materials*, vol. 7, no. 7, pp. 588–595, 2008.
- [56] S. Savas, "Digital transformation from data mining to big data and its effects on productivity," in *Current Studies in Digital Transformation and Productivity*, p. 54, 2022.
- [57] "Intelligent image gallery system using deep learning for automated fruit and vegetable classification," *International Journal of Advanced Computing & Emerging Technologies*, vol. 1, no. 3, pp. 1–15, 2025.
- [58] M. U. Javeed, "MindMate: An emotion-aware generative AI system for personalized mental health support," *JCBI*, vol. 10, Dec. 2025.
- [59] "AI-powered sentiment analysis for social media opinion mining: A hybrid NLP and machine learning approach," *International Journal of Advanced Computing & Emerging Technologies*, vol. 1, no. 5, pp. 1–15, 2025.
- [60] H. Shakeel, "LncRNAs disease: A text mining approach to find the role of lncRNA in aging," *JCBI*, vol. 9, June 2025.
- [61] M. U. Javeed, "MindMate: An emotion-aware generative AI system for personalized mental health support," *JCBI*, vol. 10, Dec. 2025.
- [62] "An ANN based intelligent system for measuring customer loyalty," *KJCIS*, vol. 8, July 2025.
- [63] M. U. Javeed, S. M. Aslam, H. A. Sadiqa, A. Raza, M. M. Iqbal, and M. Akram, "Phishing website URL detection using a hybrid machine learning approach," *JCBI*, vol. 9, June 2025.
- [64] S. Aslam, M. U. Javeed, S. M. Aslam, M. M. Iqbal, H. Ahmad, and A. Tariq, "Personality prediction of the users based on tweets through machine learning techniques," *JCBI*, vol. 8, Mar. 2025.
- [65] T. M. Ghazal and G. Issa, "Alzheimer disease detection empowered with transfer learning," *Computers, Materials & Continua*, vol. 70, no. 3, pp. 5005–5019, 2022.
- [66] K. Gupta, N. Jiwani, and P. Whig, "An efficient way of identifying Alzheimer's disease using deep learning techniques," in *Proc. Third Doctoral Symposium on Computational Intelligence*, pp. 455–465, 2022.
- [67] N. Shaffi, V. Viswan, and M. Mahmud, "Ensemble of vision transformer architectures for efficient Alzheimer's disease classification," *Brain Informatics*, vol. 11, 2024. Art. no. 25.
- [68] V. Gryshchuk, D. Singh, S. Teipel, M. Dyrba, and the ADNI, AIBL, and FTL DNI study groups, "Contrastive self-supervised learning for neurodegenerative disorder classification," *Frontiers in Neuroinformatics*, vol. 19, 2025. Art. no. 1527582.