Open Access
Full Length Article

ORIGINAL ARTICLE

Journal of Information Communication Technologies and Robotic Applications:A Novel Architecture Optimization Using Deep Learning for Alzheimer's Multiclass Disease Detection

Zoobia Fatima[1], Iram Noreen[2], Rimsha Butt[3], M. Faizan Chaudary[4]

[1]Department of Computer Science, Bahria University, Lahore, Pakistan (e-mail: zoobiafatima.bulc@bahria.edu.pk)
[2]Department of Computer Science, Bahria University, Lahore, Pakistan (e-mail: iram.bulc@bahria.edu.pk)
[3]Department of Computer Science, Bahria University, Lahore, Pakistan (e-mail: rimsha.nahid@gmail.com)
[4]Department of Computer Science, Bahria University, Lahore, Pakistan (e-mail: faizanch1703@gmail.com)

Neurodegenerative Disorders (NDD) directly affect the nervous system, causing the gradual loss of neurons and cognitive decline. Alzheimer's disease (AD) and its precursor, Mild Cognitive Impairment (MCI), affect a large proportion of the world's population. Deep learning models, such as Convolutional Neural Networks (CNNs), are primarily employed for AD detection using MRI-based biomarkers. However, their performance is constrained by challenges such as overfitting, inter-class similarity, and limited labelled data. The number of convolutional layers, the number of filters per layer, and combinations of hyperparameters also affect the model's performance. Trying every possible combination, including critical hyperparameters, to obtain the best model requires substantial training time and effort. This study explores the performance of different CNN architectures on the Alzheimer's Kaggle dataset. It is a multiclass dataset with different stages of Alzheimer's disease, including Mild-Demented, Moderate-Demented, Non-Demented, and Very-Mild Demented. A thorough performance-based analysis was carried out to determine the optimal CNN architecture. The findings offer important insights into refining and optimizing deep learning algorithms for accurate Alzheimer's disease detection. Various combinations of parameters, including learning rate, batch size, dropout rate, and optimizer, have been tuned to mitigate overfitting and improve generalization. The impact of all optimization techniques is analysed to select the best model among eleven CNN versions. The proposed approach achieves 90.6% accuracy, 96% specificity, and an F1-score of 90%, demonstrating a balanced and therapeutically appropriate performance quality.

Keywords: Alzheimer's, Convolutional Neural Network (CNN), Deep learning, Hyperparameters, Mild Cognitive Impairment (MCI), Optimizationand adaptable lo lhe constantly changing demands of modern WSNs applications in terms of sensors' data acquisition.

Author's Contribution
1,'2Methodology,Writing:Original Draft Preparation, Review,and Editing, Validation, Formal Analysis, Result Reporting,3Visualization, Data Analysis, Result Reporting, Writing:Original Draft Preparationparticipation in Email: Zoobia Fatima1, Iram Noreen2, Rimsha Butt3,M. Faizan Chaudary4

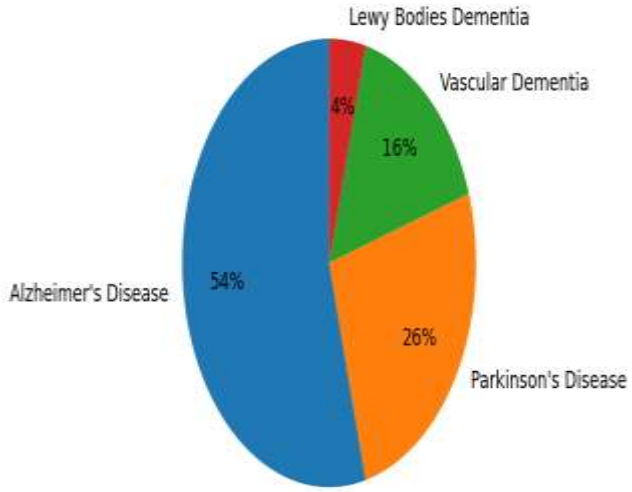
Address of Correspondence Author
Iram Noreen[2]
iram.bulc@bahria.edu.pk

Article info.
Received: May15,2026
Accepted: June 10,2026
Published: June 30,2026

Funding Source. Nil
Conflict of Interest: Nil

1. Introduction

Neurodegenerative disorders impair cognitive and motor abilities gradually and quietly. There are more than 600 cognitive disorders in the world today, including degenerative diseases like Alzheimer's disease, Parkinson's disease, blood vessel diseases, brain injuries, seizures, brain strokes, brain tumours, and disorders like schizophrenia, as well as genetic diseases like Huntington's disease, autism spectrum disorder, etc. [1]. The majority of the population is affected by Alzheimer's disease, accounting for 60-80% of dementia cases, as depicted in Figure 1. According to Alzheimer's Disease International (ADI), every 3 seconds, someone in the world is affected by dementia. According to the WHO, the number of dementia sufferers worldwide is expected to rise from 55 million in 2019 to 139 million by 2050, and this number may double in the coming years [2],[3]. The specific etiology of the disorder remains unclear, and there is no effective cure. Experts estimate that 10-15% of people with Mild Cognitive Impairment (MCI) may eventually acquire Alzheimer's disease. Early detection of MCI can prevent or postpone the progression of Alzheimer's disease [4].



Magnetic Resonance Imaging (MRI) scans generate comprehensive 3D images of millions of voxels, enabling precise imaging of brain structure. The majority of Alzheimer 's-related lesions may be detected on MRI scans, and radiologists' competence typically determines their severity. Digital image processing technologies improve the evaluation process by allowing more precise measurements of brain structures, soft tissues, and lesion features. Therefore, AI and computer-aided techniques assist physicians in the early diagnosis of Alzheimer's [5]. CNNs, a deep learning technique, are commonly utilized for image classification and segmentation in medical image biomarkers [6],[7],[8]. Image analysis in medicine has various problems, particularly a lack of labelled datasets, noise, uneven class distributions, and class similarities [9]. Mostly, transfer learning is widely used whenever a dataset is limited [10], [11]. However, modern CNN architectures often suffer from overfitting due to the large number of parameters [12]. Given the extensive training process, manual tuning of hyperparameters is a challenging and time-consuming process [13]. A study of existing research indicates a gap in the automatic optimization of essential architectural components, such as the number of convolutional layers and the number of filters per layer [14]. These components are critical for improving CNN performance in accurately classifying Alzheimer's disease [15].

The proposed study focuses on how deep learning models, specifically Convolutional Neural Networks (CNNs), may be used to classify MRI scans into four phases of Alzheimer's disease: mild dementia, moderate dementia, non-demented, and very mild dementia. The proposed research aim was to assess the effect of various hyperparameters, such as optimizer approaches (Adam, SGD, and AdamW) and training strategies (early halting and dropout), on the model's classification accuracy and other metrics. The study includes training CNNs with various designs, including thick layers and dropout methods to reduce overfitting and improve generalization. The dataset consisted of MRI scans [16] that were modified to create a robust training set, enabling the model to

accurately capture complex details in brain imaging. The proposed architecture accurately classifies Alzheimer's disease, reducing the burden on healthcare specialists and accelerating decision-making processes.

1.1 Research Contributions

The following are some novelties of the proposed research work.

i. Eleven distinct CNN architectures are tested for multiclass AD classification.
ii. The impact of important parameter combinations, i.e., Learning Rate (LR), dropout, Batch Size (BS), Optimizers (e.g., ADAM), filter sizes, etc., is thoroughly investigated.
iii. Several regularization and optimization techniques were applied to increase the model's robustness and reduce overfitting.
iv. Performance-based analysis is done to select the best and most efficient model.
v. The findings provide valuable insights for developing and fine-tuning CNN models for Alzheimer's disease classification.
vi. The suggested model has a high specificity (96%) and F1-score (90%), indicating a fair trade-off between accuracy, recall, and true-negative rate, which is crucial for reliable medical diagnosis.

The rest of the paper is arranged as follows: Section 2 provides a thorough overview of related literature. Section 3 describes the materials and methods utilized in this work, such as data acquisition, preprocessing, the proposed deep learning algorithm, and the experimental phase. Section 4 covers the results, providing a detailed analysis of the results. Finally, Section 5 provides a quick review of major findings and recommended paths for further research.

1.2 Abbreviations and Acronyms

This section lists the abbreviations and acronyms used throughout the article for ease of reference (see Table 1).

Table 1: List of Abbreviations

Abbreviation	Definition
AD	Alzheimer’s Disease
ADNI	Alzheimer's Disease Neuroimaging Initiative
AI	Artificial Intelligence
CN	Cognitive Normal
CNN	Convolutional Neural Network
DL	Deep Learning
HC	Healthy Control
MCI	Mild Cognitive Impairment
MRI	Magnetic Resonance Imaging
NDD	Neurodegenerative Disorders
OASIS	Open Access Series of Imaging Studies

2. Related Work

Researchers and practitioners have made various advances in the field of Alzheimer’s classification and detection, which is illustrated in this section.

Fatahi et al. [17] used ensemble learning to aggregate the performance of six CNN-based classifiers with different hyperparameters on the ADNI dataset and a local dataset. The authors achieved 88.46% accuracy for three-way classification. E. Hussain et al. [18] suggested a CNN-based approach to classify Alzheimer's disease as "Alzheimer's" or "healthy" based on brain MRI data. The proposed 12-layer CNN model was assessed on the OASIS dataset and compared to pre-trained architectures such as InceptionV3, Xception, MobilenetV2, and VGG. The suggested 12-layer CNN model achieved accuracy rates of 97.75% and 97.50% F1-score.

Erdogmus et al. [19] suggested a CNN with 12 layers, and researchers have optimized 12 hyperparameters. The DARWIN dataset was utilized to assess the model's performance. This dataset contains 1D data. The data was translated from 1D to 2D for transfer learning models. The model achieves a binary classification accuracy of 90.4% (patient vs healthy).

In [20], A new deep learning ensemble method was used that included Bi-GRU, MLP, and QDNN classifiers with

sophisticated feature selection and optimization. It enables researchers to find early signs of Alzheimer's with up to 98.7% accuracy.

Huanhuan et al. [21] suggested a method for diagnosing the early stages of dementia utilizing Convolutional Neural Networks (ConvNets) and a subset of the ADNI dataset comprising 615 MRI images. Statistical Parameter Mapping (SPM) was used in preprocessing to reduce head movement artifacts. The detection method made use of powerful classifiers, such as a Residual Network (ResNet) with 50 layers and the Neural Architecture Search Network. To solve overfitting in the fully linked layer, a dropout layer was added. The accuracy rates for identifying AD/MCI were 97.65% and 88.37% for MCI/HC.

Lahmiri et al. [22] created a CNN model integrated with k-nearest neighbour (KNN) for binary classification on the OASIS dataset. The KNN parameters were adjusted using the Bayesian optimization (BO) technique. The investigation yielded results of 94.96 ± 0.0486% accuracy, 92.05 ± 0.0746% sensitivity, and 96.62 ± 0.0350% specificity.

Wang et al. [23] noted that current CNN networks usually demand massive amounts of data and have quite complex topologies. The article proposes an accurate and portable densely connected 3D convolutional neural network (DenseCNN) for Alzheimer's classification based on hippocampus segmentation. DenseCNN uses 746 training and 187 testing data points from the ADNI dataset. The accuracy of the suggested Dense CNN model was 89.8%.

Razavi et al. [24] suggested an unsupervised feature learning method that includes two crucial stages. First, features were extracted using sparse filtering and unsupervised neural layer networks. Boltzmann machines were used to organize the data. The study used MRI images from the ADNI dataset. The maximum classification accuracy that SoftMax regression provided was 98.3%.

Yildirim et al. [25] modified the ResNet model to offer a hybrid method. They created and extracted the last five levels of ResNet and added ten more layers. The accuracy rates per class were 90% for non-demented, 70% for moderately demented, 90% for very mildly demented, and 96.6% for mildly demented. Similarly, Venugopalan et al. [26] presented a unique method to denoise MRI data with an emphasis on the hippocampal area. The proposed approach integrated the ADNI dataset, analyzing the evolution of moderate cognitive impairment using biological indicators, including MRI, PET, and neuropsychological tests. The dataset comprised 220 patients' 2004 HER records, 503 MRI images, and 808 SNPs. The MRI data covered 8209 voxels at 18 different locations. A three-tier autoencoder was used for noise filtration and the extraction of 1680 common characteristics. The accuracy of this approach was 78%.

Katabathula et al. [27] examined the categorization of Alzheimer's disease using the Hippocampal3D tool, which divided the hippocampus into left and right sections. The Dense-CNN model was used to extract deep visual features, whereas the LB spectrum was used to acquire global shape information. Their model resulted in 92.52% classification accuracy, 88.20% sensitivity, 94.95% specificity, and a 97.89% AUC score. An ensemble learning method utilizing five pre-trained and fine-tuned models (VGG19, Inception-ResNetv2, ResNet152v2, EfficientNetB5, EfficientNetB6, and a scratch model) was presented by Sadat et al. [28]. Researchers used the OASIS dataset with an 80-20 ratio as training, validation, and testing subsets, respectively. A weighted average approach was then used to complete the entire procedure. They achieved 96% accuracy using the ensemble approach. An optimized VGG-16 model was presented by Deepa and Chokkalingam [15] for the multiclass categorization of Alzheimer's disease. The batch size and dropout rate were adjusted using the Arithmetic Optimization Algorithm (AOA). Three datasets, ADNI, ASIS, and SIMON, were used after preprocessing and segmentation. Their optimized VGG-16 model achieved an accuracy and F1-score of 97% and 95.78%, respectively. Baghdadi et

al. [29] suggested the Gorilla Troops Optimizer (GTO) method for parameter optimization with transfer learning. Eight cutting-edge transfer learning models were used with the ADNI dataset and the Alzheimer's Dataset. MobileNet achieved the highest accuracy (96.65%) on the Alzheimer's Dataset, while the Xception model had the best accuracy (96.25%) on the ADNI dataset. Furthermore, the available research shows significant progress in applying deep learning approaches to Alzheimer's disease detection. However, several studies rely on binary classification, ignoring the complexities of multiclass problems among neurodegenerative diseases. Furthermore, issues such as model generalizability, interpretability, and optimization have yet to be satisfactorily addressed. These restrictions highlight the need for a superior, scalable solution. Motivated by these shortcomings, the next sections present an optimized CNN model to improve multiclass detection for Alzheimer's disorder.

3. Methodology

This section describes the proposed methodology for neurodegenerative disease classification as demonstrated in Figure 2. The procedure starts with careful data preparation, which includes MRI image preprocessing, normalization, and augmentation. The dataset is divided into subsets for testing, training, and validation. Iteratively adjusting hyperparameters reduces errors and improves model performance. The refined version is then used to identify AD phases accurately. In-depth explanations of demented (i.e., cognitive normal), and very mild dementia. The dataset includes T1-weighted brain MRI images. All images are uniformly resized and stored in JPEG format. Details of the dataset are mentioned in Table 2, and sample images are displayed in Figure 3.

Table 2: Metadata summary of the Original Alzheimer's MRI Dataset

Class Label	**No. of Images**	**Image Format**	**Image Size (px)**	**Augmentation**
Non-Demented	3200	JPEG (.jpg)	176 × 208	Yes (flip, rotate, zoom, etc.)
Very Mild Demented	2240	JPEG (.jpg)	176 × 208	Yes (flip, rotate, zoom, etc.)
Mild Demented	896	JPEG (.jpg)	176 × 208	Yes (flip, rotate, zoom, etc.)
Moderate Demented	64	JPEG (.jpg)	176 × 208	Yes (flip, rotate, zoom, etc.)
Total	6400	JPEG (.jpg)	176 × 208	Yes (flip, rotate, zoom, etc.)

each step's contributions are provided in the next subsections.

3.1 Dataset Acquisition

The proposed study uses a publicly available dataset known as the Augmented Alzheimer MRI dataset [16], which was published on Kaggle and is made up of 180 × 180, 96dpi, and 40,384 pre-processed MRI images. The dataset comprises two types: i) original and ii) augmented. It is divided into four categories: mild dementia, moderate dementia, non-

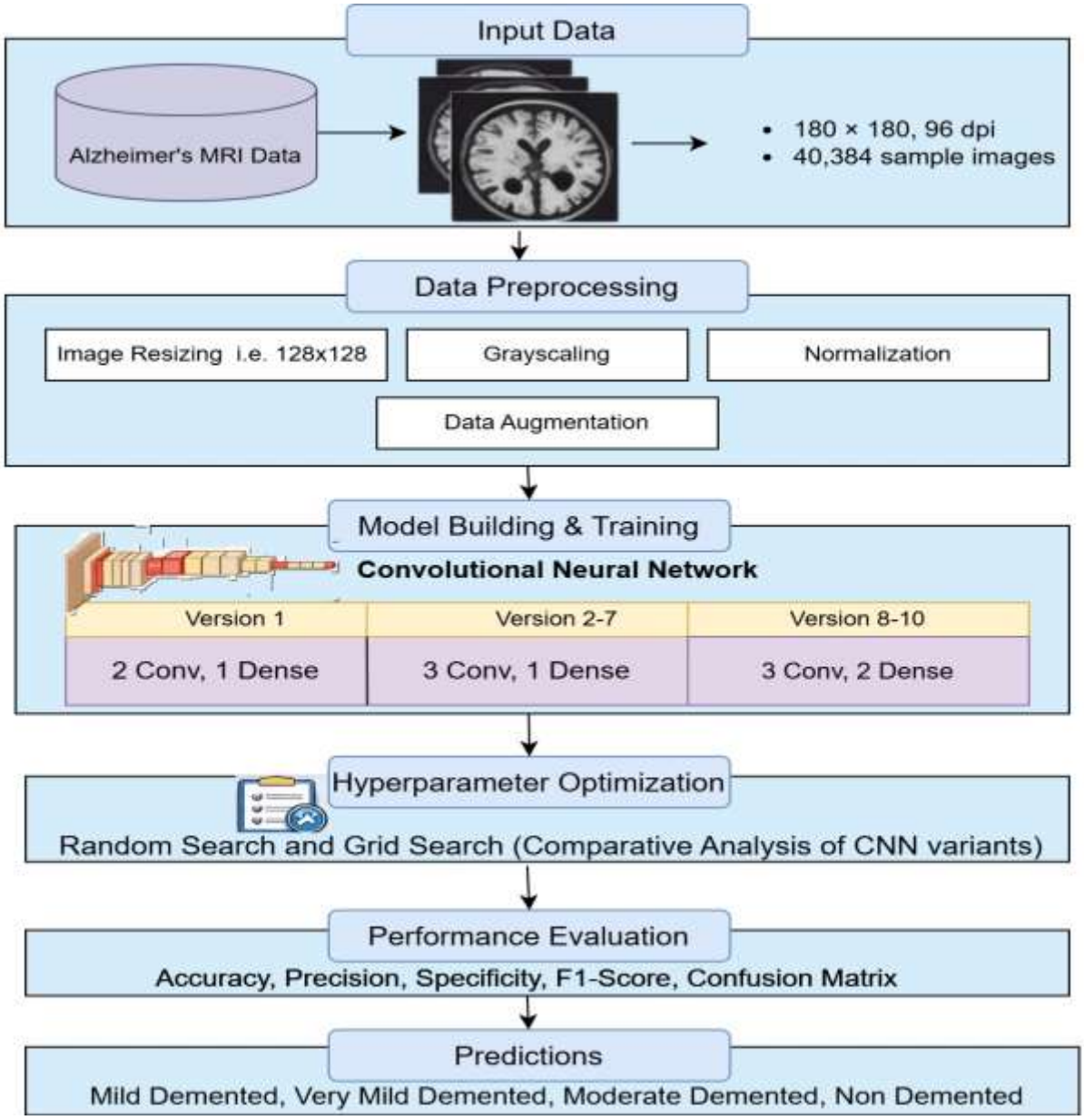


Figure 2: Proposed Framework for Alzheimer's Disease Classification Using Convolutional Neural Network (CNN)

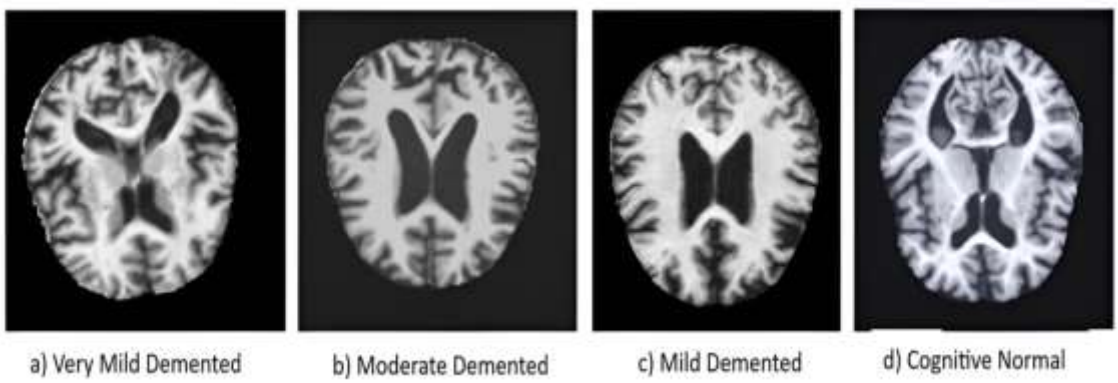


Figure 3: Classes included in the Alzheimer's Dataset: (a) Very Mild Demented (b) Moderate Demented (c) Mild Demented (d) Cognitive Normal.

3.2 Data Pre-processing

The dataset has a problem of unbalanced data classes. To resolve this issue, the samples from the original data are augmented in the Alzheimer's Augmented dataset. Using these techniques, a more equal distribution of images across classes was produced to rectify the class imbalance in the original dataset. The original dataset was organized as follows: Non-Demented (3200 images), Very Mild Demented (2200 images), Mild Demented (896 images), and Moderate Demented (64 images). However, the model training may produce biased predictions due to the notable imbalance, where the non-demented class has the most images. A more even distribution of the dataset—Mild Demented (8960 pictures), Moderate Demented (6464 photos), non-demented (9600 images), and Very Mild Demented (8960 images) —was achieved by applying augmentation techniques to balance the dataset. Each class is better represented with balanced distribution, which improves model performance and generalization and makes model training more efficient. Figure 4 shows the class distribution of the original and augmented data. The possibility of bias against any one class is decreased after augmentation. No additional preprocessing procedures were required because the Kaggle Augmented Alzheimer MRI Dataset is already formatted, scaled, and normalized to make it directly compatible with the CNN architecture. Additionally, augmentation was applied, only increasing the variety of training examples, to improve the performance, fairness, and generalization across various scenarios of learning distinctive characteristics for all classes. The approach was streamlined, and implementation was made easier, which ensured the dataset was ready for successful model training.

3.3 Proposed Model

The experiments investigated several variations in CNN architectures to determine the best parameters to be used for Alzheimer's prediction. Initially, we experimented with

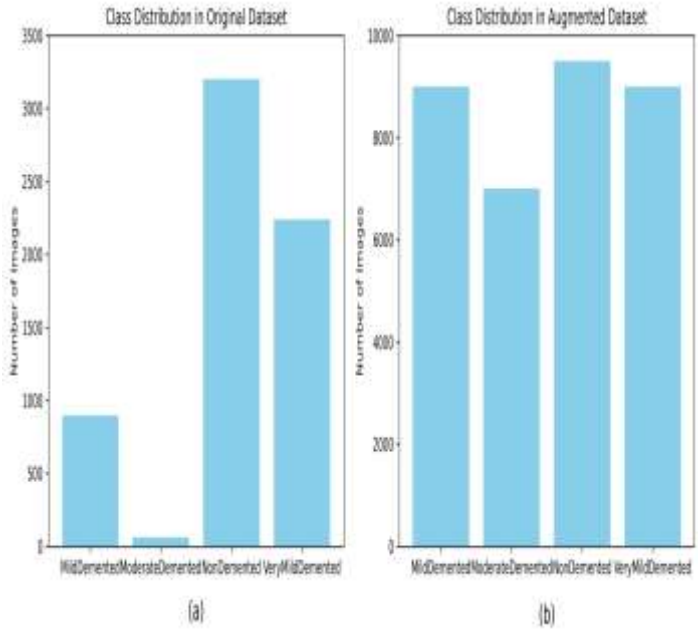


Figure 4: Class distribution—(a) Original Data (b) Augmented Data

version-1, which is a simple two-layer CNN design. This design included input layers of 128x128 grayscale brain scan images, followed by conv layers with max pooling layers. Another convolutional layer and a pooling layer were added before flattening the features, which were then processed via a fully connected dense layer to produce the output. Figure 5. shows the 2-layered CNN architecture used to predict multiclass Alzheimer's.

In Version-2, the network's depth was raised by adding another third convolutional layer, with the goal of improving hierarchical feature extraction. However, while this increased the model's capacity to capture more complicated spatial features, the performance improvement was minimal (see Figure 6).

Finally, throughout the testing phase for Versions 8–10, we tried a 3-layer CNN with an extra dense layer (see Figure 7). This phase contained input layers containing 128x128 grayscale brain scan pictures, three convolutional layers each, after the pooling layer, a flattening layer, and two dense classification layers.

This performance improvement is due to the function of substantial amounts for learning nonlinear variants of high-level characteristics retrieved by convolutional layers. Because the conv layers have already extracted enough spatial and structural details from brain MRI images, increasing convolutional depth resulted in declining returns and probable overfitting. In contrast, increasing thick layers improved the model's ability to transfer extracted data to class labels, resulting in greater classification across Alzheimer's stages.

Models with two dense layers exhibited higher classification accuracy. This was especially noticeable in the 3-layer CNN architectures examined in the last phase of our tests. As a result, these architectural choices reflect the optimal designs for diagnosing dementia stages using brain scan AD images.

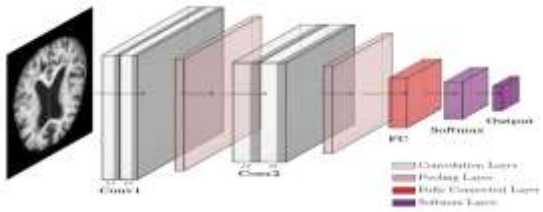


Figure 5: CNN architecture (Version 1) - Two-layered architecture

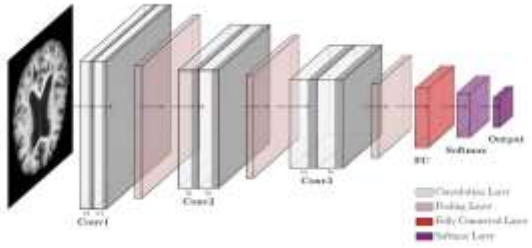


Figure 6: CNN architecture (Versions 2-7)- Three-layered architecture

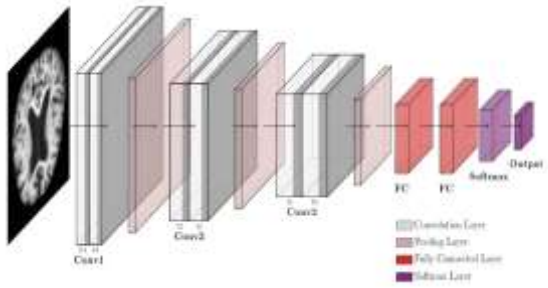


Figure 7: CNN architecture (Version 8, 9, 10) -Three-layered architecture with 2 FC.

3.4 Experimental Setup

This section of the study details the experiments and evaluation techniques used to evaluate the effectiveness of the suggested model. A well-thought-out proposed framework was used for experiments to guarantee robust results. The model's performance was thoroughly assessed in various scenarios using validation techniques and evaluation metrics. CNN models were constructed and trained using TensorFlow and Keras deep learning libraries, in a Jupyter Notebook on a conventional PC system, which included RAM: 16 GB, GPU: 11 GB, VRAM, and Processor: Intel i5. The dataset was divided into an 80-20 split ratio, where 80% is for training data and validation, and 20% for testing. Because the data set utilized in this study is already augmented, the issue of overfitting with fewer data samples during training was already solved. Table 3 shows the experimental parameters. The complete training, model-selection, and evaluation procedure is summarized in Algorithm 1 (See Table 3).

Table 3: Algorithm 1: Model Training and Evaluation Pipeline

Algorithm 1: CNN Training, Model Selection, and Final Testing Input: Dataset D Output: Selected model M* and its test metrics
1: D ← Preprocess(D) 2: D_aug ← Augment(D) 3: (Train, Val, Test) ← Split(D_aug, 0.80, 0.10, 0.10) 4: Initialize candidate models {M1, …, M10} 5: best_val ← −∞ ; M* ← null 6: for each model Mi in {M1, …, M10} do 7: ConfigureArchitecture(Mi) 8: for each optimizer O in {Adam, AdamW, SGD} do 9: Train(Mi, Train, optimizer=O)

```
        val_score ← Validate(Mi, Val)
        if val_score > best_val then
            best_val ← val_score
            M* ← CopyBestCheckpoint(Mi)
        end if
    end for
end for
metrics_test ← Evaluate(M*, Test)
return (M*, metrics_test)
```

Table 4: Network Parameters

Parameters	**Values**
Classes	i. Very Mild Demented ii. Mild Demented iii. Moderate Demented iv. Non-Demented
Total Number of training images	32307
Total Number of testing images	8076

3.5 Hyperparameter Tuning

The CNN structures usually consist of multiple layers. Several hyperparameters affect these designs, such as the number of convolutional layers, the number of filters per layer, the learning rate, the batch sizes, the epochs, the optimizer, the number of fully connected layers, the neurons in those layers, and the dropout rates. The ideal values for the hyperparameters must be identified to maximize the model's performance. Numerous previous studies have done manual optimizations. Exploring every potential combination of hyperparameters manually is tedious and laborious. Hence, automatic optimization algorithms such as Random Search and Grid Search algorithms are used to find the optimal configurations. They are non-adaptive hyperparameter optimization methods, and they do not retain past knowledge in memory. However, they are less biased and more systematic than manual optimization methods. In addition, they are more likely to find good combinations very quickly. The proposed model was trained on multiple epochs, with each epoch refining the weights and learning the distinguishing features that are linked to the various stages of Alzheimer's disease. It was optimized by minimizing prediction errors while improving the accuracy of classification during training. The model in this phase was trained using multiple optimizers and configurations, including early stopping to prevent overfitting. Rapid experiments were conducted with multiple ranges and various combinations of hyperparameters to explore the best combination and to find the best performance. This research focuses on improving the layout of model parameters to create a lightweight CNN. Unlike previous studies, which largely focus on manual optimization techniques and learning rates, we emphasize hitherto unexplored hyperparameters, i.e., the number of convolutional layers, filters, optimizers, dense layers, dropout, etc. Table 4 and 5 shows the network parameters and value ranges for these hyperparameters, which were established during a preliminary manual optimization procedure. CNN models include one or two dense layers, with the second layer deleted if the neurons are given zero. A larger range is provided for the first dense layer to allow for flexibility, which frequently results in a greater neuron count than the second layer. This technique promotes adaptation while balancing efficiency and performance.

Table 5: Hyperparameters Used in Manual Optimization

Parameter	**Range**
Convolutional Layers	2–3
Filter	16, 32, 64
Optimizers	Adam, AdamW, SGD
Learning Rates	0.0001
Dense Layers	0, 1, 2
Dropout Layers	0.2, 0.5
Batch Size	32

3.6 Evaluation Metrics

The proposed Alzheimer’s disorder detection framework was evaluated using several key evaluation metrics, i.e., precision, accuracy, recall, specificity, and F1-score. These metrics provide a comprehensive evaluation of the model's classification ability, especially in medical imaging tasks.

Despite accuracy being a widely used metric in scientific research and medical image

analysis, it doesn't always represent a model's real performance, particularly in the case of class imbalance. For example, a model may obtain high accuracy by precisely predicting the major class but fail to detect minor classes, i.e., early-stage Alzheimer's. Therefore, more assessment criteria is required to understand the model's predictive behaviour better.

Precision (Positive Predictive Value) is the fraction of accurately predicted positive cases among all anticipated positives. It reflects the dependability of favourable forecasts.

Precision = $\frac{TP}{(TP+FP)}$ ..... *Equation (1)*

A high positive predicted value in Equation (1) indicates a minor False-Positive-Rate (FPR), which is important for avoiding misdiagnosing cognitively normal patients as AD patients.

Recall (True Positive Rate or Sensitivity) is the fraction of genuine positive instances accurately detected by the model:

Recall= $\frac{TP}{(TP+FN)}$ ..... *Equation (2)*

High recall is important in medical diagnosis since it assures that most Alzheimer's cases are accurately identified, reducing missed diagnoses.

Specificity assesses the model's ability to accurately detect negative situations.

Specificity= $\frac{TN}{(TN+FP)}$ ..... *Equation (3)*

A high specificity suggests that the model successfully distinguishes between healthy persons and those with Alzheimer's, decreasing false alarms.

F1-score is the harmonic mean of recall and accuracy, giving a balanced metric when there is an unequal class distribution.

F1-Score $=2 * \frac{Precision*Recall}{Precision+Recall}$ ..... Equation (4)

A high F1-score implies that the model achieves a strong mix of precision and recall, making it ideal for medical classification tasks that need both false negatives and positives.

Accuracy assesses the model's overall performance.

Accuracy= $\frac{(TP+TN)}{(TP+TN+FP+FN)}$ .....Equation (5)

However, as previously mentioned, accuracy alone can be deceiving in the context of class imbalance.

4. Results

This section describes the experimental results of models designed to categorize Alzheimer's disease stages using MRI data. Several designs and optimization strategies were tested, and their performance was measured using accuracy, precision, recall, F1-score, and specificity. The findings show each model's usefulness and the effect of training procedures on classification performance. Table 6 shows the detailed performance measures of all models.

Table 6: Performance measure

Model	**Accuracy %**	**Precision %**	**Sensitivity%**	**F1-score %**	**Specificity %**
Model 1	86	86	86	86	99
Model 2	84	84	84	83	99
Model 3	79	80	81	80	96
Model 4	89	90	89	89	98
Model 5	89	89	88	89	99
Model 6	86	86	87	86	98
Mod	85	85	86	85	98

el 7					
Model 8	89	90	89	89	99
Model 9	88	88	88	88	98
Model 10	88	88	88	88	99
Model 11 (i.e., Test)	90	90	90	90	96

The performance indicators for the models employed to categorize Alzheimer's disorder stages are persuasive, as shown in Table 6. Each statistic gives a thorough assessment of the models' efficiency and dependability. For example, the models' F1-score is an excellent 90%, demonstrating an equal balance between accuracy and recall that keeps erroneous positives and false negatives to a minimum. This suggests that the models are extremely good at properly detecting situations that do not belong in the target class. Overall, the models obtained 89% accuracy, demonstrating their dependability and efficacy in discriminating between different phases of Alzheimer's disease.

However, all 11 evaluated versions showed better results in terms of specificity, particularly those using the ADAM optimizer and early stopping. Versions 1, 4, 8, and 10 achieved a specificity of 99%. This suggests that the models are extremely good at properly detecting situations that do not belong in the target class. These models accurately classified non-demented (healthy) patients, with fewer false positives. The SGD optimizer performed well in both versions 2 and 3 with 40 epochs, obtaining 99% specificity. However, performance decreased marginally with longer epochs (60 and 80). Dropout lowered specificity to 96% in version 11, indicating that although it effectively prevents overfitting, it may also hamper the model's ability to identify healthy controls. ADAM optimizer models, with early halting and dropout, provide the optimum balance of accuracy and generalization, particularly for non-demented classification with high specificity. These findings demonstrate that the architectural choices and optimization methodologies used in the trials produced robust models capable of exact categorization, which is critical in medical diagnostics.

4.1 Accuracy/Loss Representations

Furthermore, we examined the accuracy and loss graphs for each model to acquire a better understanding of training processes. These graphs show the model's performance across several epochs, giving a visual depiction of how effectively it learns from data over time. The accuracy graphs show the model's capacity to accurately classify brain scan pictures, whilst the loss graphs show the error decrease during the training phase. Examining these graphs allows us to identify tendencies like overfitting or underfitting, as well as measure model stability and convergence. The accuracy and loss curves for each architecture provide useful information for improving model performance and ensuring accurate categorization of Alzheimer's disease stages. After a certain epoch, the validation and training accuracy & loss plots of Models 1 and 2 in Figure 8 continue to overlap, indicating that both models reached their full learning capacity. The graphs then flatten and stabilize, indicating that the models are successfully learning from the data and generalizing to the validation dataset. Models 4, 5, and 6, on the other hand, show quicker learning curves. These models are also doing very well with the validation data, as seen by the validation accuracy curves, which often closely resemble the training accuracy curves. A similar pattern can be seen in the loss graphs, as the validation and training loss steadily decline throughout several epochs. The flattening of the graphs indicates that, despite differences in learning pace, all models typically completed much of their training by 80 epochs. Residual models exhibit steady and reliable learning tendencies, continuing to expand on the earlier designs.

These models have effectively attained their maximum learning capability, as evidenced by the large overlap between their training and validation accuracy charts. The models' ability to learn from the data and lower error rates is further supported by the loss graphs for other models, which also exhibit a

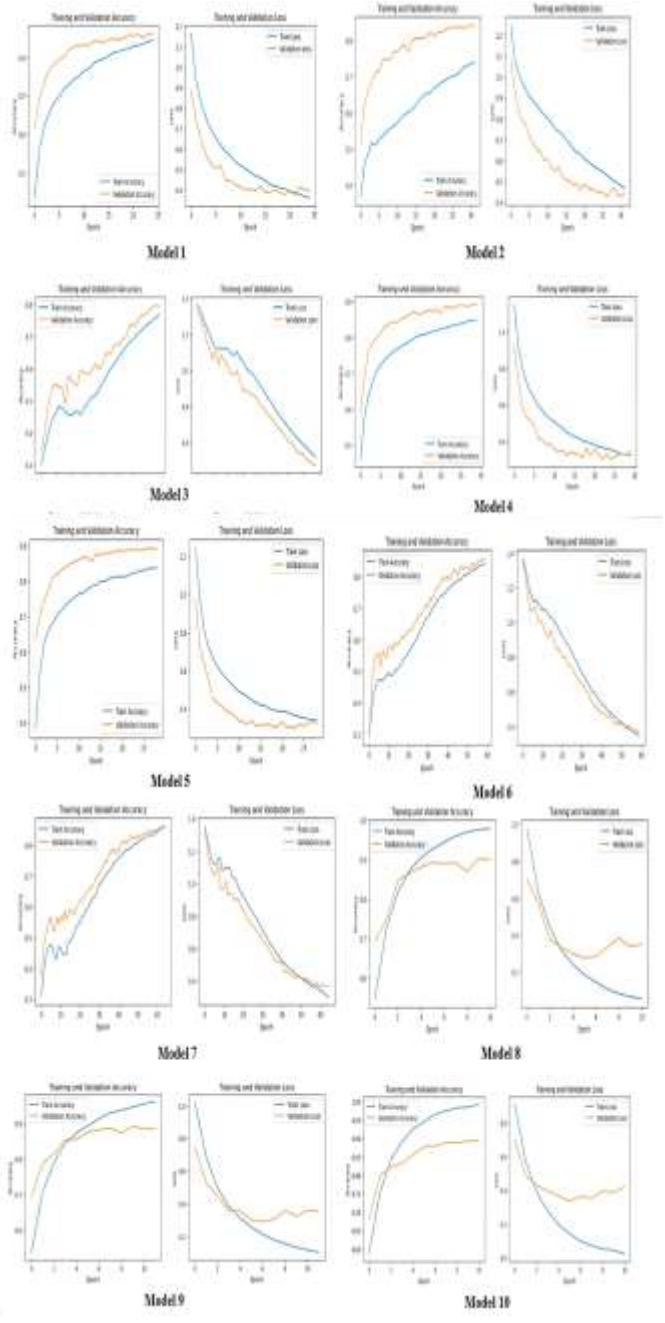


Figure 9: Accuracy/ Loss representation of variations in CNN

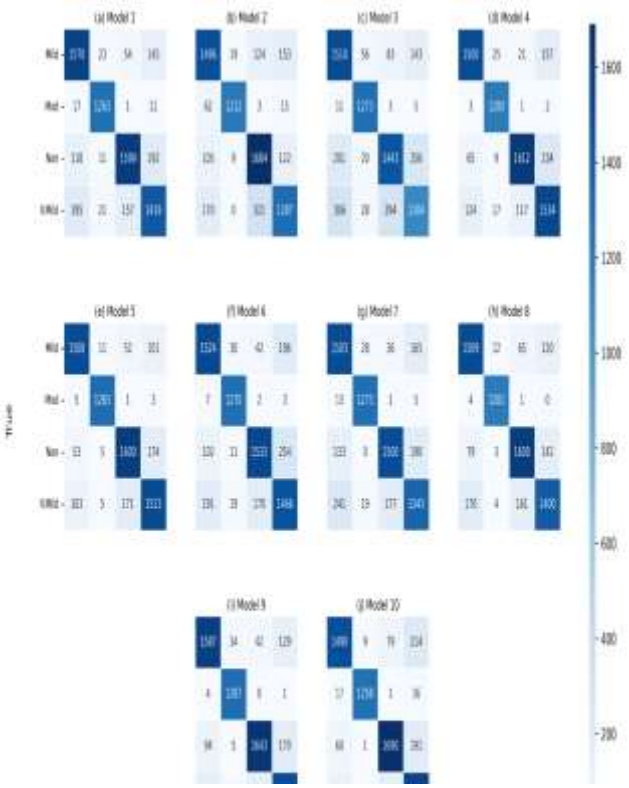


Figure 8: Confusion Matrix of CNN Architectures. Each matrix represents predictions and true classifications for Mild Demented, Moderate Demented, Cognitive Normal (i.e., Non-Demented), and Very Mild Demented categories across different models.

consistent and ongoing decline. The tight alignment of training and validation measures shows that these models have good generalization potential. These findings demonstrate how well the models categorize the phases of Alzheimer's disease, demonstrating their resilience and dependability throughout the learning process.

An essential tool for assessing how well our models perform in categorizing the phases of Alzheimer's disease is the confusion matrix. By offering a thorough analysis of TP (True positive), TN (True negative), FP (False positive), and FN (False negative), it enables us to recognize the models' advantages and shortcomings. Each model's confusion matrix in our study shows how effectively it distinguishes between the different phases of Alzheimer's disease. The number of cases that are successfully or wrongly categorized is represented by each cell in the matrix. The off-diagonal elements show misclassifications, and the diagonal elements show the number of accurate predictions.

4.2 Confusion Matrix

Figure 8 shows the detailed confusion matrix of each model.

4.3 Ablation Study

The ablation study investigated the impact of various optimization tactics, training setups, and regularization approaches on model performance. Model 1, which employed the ADAM optimizer with early halting, served as the baseline, obtaining 86% accuracy and 99% specificity. Model 2 replaced ADAM with SGD, resulting in a little drop in performance (84% accuracy, 83% F1-score) while keeping excellent specificity. Extending SGD training to 80 epochs in Model 3 decreased accuracy to 79% and specificity to 96%, demonstrating lower generalization. Model 4 reintroduced ADAM with more epochs, increasing performance to 89% accuracy and 98% specificity. Model 5 improved the learning rate in the same configuration while keeping good metrics on all measurements.

Model 6 continued training without stopping early, resulting in comparable accuracy but a small reduction in specificity. Model 7 incorporated dropout regularization, resulting in 85% accuracy while maintaining 98% specificity. Model 8 combines ADAM, early halting, and batch optimization to get one of the greatest results (89% accuracy, 99% specificity). Model 9 eliminated early halting, resulting in a minor loss in generalization. Model 10 duplicated the baseline and confirmed its stability. Model 11 raised dropout rates, somewhat improving accuracy (90%) but decreasing specificity to 96%.

Overall, the best-performing combinations used the ADAM optimizer with early halting, modest training epochs, and carefully controlled dropout. These results, shown in Table 4, emphasize the importance of optimization and regularization procedures in improving classification reliability for Alzheimer's disease staging.

Table 7 Ablation study summarizing the impact of various architectural and training modifications across model versions

Model	Optimizer	Epochs	Early Stopping	Dropout	Accuracy %	F1-score %	Specificity %
1	ADAM	40	✓	✗	86	86	99
2	SGD	40	✓	✗	84	83	99
3	SGD	80	✗	✗	79	80	96
4	ADAM	60	✓	✗	89	89	98
5	ADAM	60	✓	✗	89	89	99
6	ADAM	80	✗	✗	86	86	98
7	ADAM	60	✓	✓	85	85	98
8	ADAM	60	✓	✗	89	89	99
9	ADAM	60	✗	✗	88	88	98
10	ADAM	40	✓	✗	88	88	99
11 (i.e., test)	ADAM	60	✓	↑	90	90	96

5. State of the Art Comparison

The proposed approach's performance is compared with several SOTA deep learning models described in the literature review for Alzheimer's disease classification. Most of the previous research uses standard transfer learning models or CNN architectures with minimal hyperparameter optimization, focusing mostly on binary classification. In contrast, the proposed approach assesses different CNN variants and optimizes key parameters to address overfitting in a multiclass context. As demonstrated in the comparison study, the proposed model achieves 90% accuracy and 96% specificity, exceeding previously reported approaches on the KAGGLE MRI dataset. These findings show that the suggested optimization-driven CNN framework is useful for accurately and finely classifying Alzheimer's disease stages. Presented below is a comprehensive, reviewer-friendly state-of-the-art comparison table, accompanied by a succinct, impactful statement for inclusion in the text.

Table 8: Performance Comparison of Proposed Model with State-of-the-Art Methods

Paper (Year)	CNN Model(s)	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	Limitations
Ebrahimi & Luo (2021) [30]	2D CNN (slice-based), 3D CNN (voxel-based, transfer learning)	~90 (2D); 96.88 (best 3D)	100	94.12	NA	Slice-based ≤90; 3D improves performance but risk of overfitting; no augmentation
Priyatama et al., (2023) [32]	CNN, VGG-16, and VGG-19	75	NA	NA	50	Imbalanced data and lower model performance
Ghasemi Dakdareh & Abbasian (2024) [33]	DenseNet (2D)	88	74.63	95.18	74.63	2D slices lose spatial info; no 3D due to compute limits
Saddam Bekhet et al. (2025) [34]	1D CNN + Random Forest (hybrid)	92.2	NA	NA	NA	Hybrid model, CNN-only performance borderline ≤90
Proposed	CNN variants	90.6	90	96	90	Balanced dataset, augmentation applied, improved generalization

To assure a fair comparison, key performance measures such as specificity, sensitivity, and F1-score are provided where possible. As demonstrated in Table 8, although the proposed approach achieved a slightly acceptable accuracy of 90.6% when compared to other previous research, it has a substantially better (96%) specificity and (90%) F1-score, making it an important parameter in Alzheimer's disease diagnosis. Specificity measures the model's ability to properly identify cognitively normal persons (true negatives). In clinical contexts, high specificity is critical to reducing false-positive diagnoses, which can cause unneeded psychological stress, further medical testing, and improper therapy. As a result, the suggested model prioritizes dependability in differentiating healthy participants, albeit at the expense of somewhat lower overall accuracy. Moreover, the suggested model's better F1-score shows an enhanced balance among false-positives and negative results, which is especially relevant in multi-class medical classification problems. This trade-off results in a more clinically reliable procedure for real-world applications.

6. Conclusion

This study shows that deep learning, specifically Convolutional Neural Network (CNNs), can accurately classify the several stages of Alzheimer's disorder using structural MRI data. To determine the most optimal and clinically reliable model, a complete performance-based study was performed, taking into account architectural variants, hyperparameter-based optimization, and regularization strategies like as dropout and early stopping.

Model Version 11 outperformed the other configurations, achieving an accuracy of 90.6%, an F1-score of 90%, and a specificity of 96%. This set of measurements reveals a balanced trade-off between overall precision, recall, and accuracy of the model with strong specificity, lowering false positives. However, overall accuracy is slightly lower than in other state-of-the-art research. The

better balance of metrics guarantees strong and dependable multiclass classification, especially for intermediate stages of dementia.

The findings give helpful insights for adopting and optimizing CNN architectures, underlining the significance of assessing models using many performance criteria other than accuracy alone. The study also emphasizes current limitations, such as difficulty distinguishing between extremely mild and mild illness stages. Future research should concentrate on expanding dataset variety, adding multimodal imaging data, utilizing transfer learning, testing frameworks for practical healthcare contexts, and dealing with interpretability and ethical concerns to enable appropriate deployment in medical practice.

Overall, this study adds to the growing amount of data supporting the use of AI-driven diagnostics in neurodegenerative disease evaluation, providing a dependable framework for the early and accurate identification of Alzheimer's disease.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

References

[1] J. Chaki and M. Woźniak, "Deep learning for neurodegenerative disorder (2016 to 2022): A systematic review," *Biomed. Signal Process. Control*, vol. 80, p. 104223, 2023, doi: https://doi.org/10.1016/j.bspc.2022.104223.

[2] S. Longs, C. Benoist, and W. Weidner, "World Alzheimer Report 2023 Reducing dementia risk: never too early, never too late," 2023.

[3] S. Long, C. Benoist, and W. Weidner, "World Alzheimer Report 2023: Reducing Dementia Risk: Never Too Early, Never Too Late," 2023, *Alzheimer's Disease International, London, England*.

[4] M. Kaya and Y. Cetin-Kaya, "A Novel Deep Learning Architecture Optimization for Multiclass Classification of Alzheimer's Disease Level," *IEEE Access*, vol. 12, pp. 46562–46581, 2024, doi: 10.1109/ACCESS.2024.3382947.

[5] B. S. Rao, M. Aparna, S. S. Kolisetty, H. Janapana, and Y. V. Koteswararao, "Multi-class Classification of Alzheimer's Disease Using Deep Learning and Transfer Learning on 3D MRI Images," *Traitement du Signal*, vol. 41, no. 3, pp. 1397–1404, Jun. 2024, doi: 10.18280/ts.410328.

[6] J. Bin Bae *et al.*, "Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging," *Sci. Rep.*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-79243-9.

[7] A. M. El-Assy, H. M. Amer, H. M. Ibrahim, and M. A. Mohamed, "A novel CNN architecture for accurate early detection and classification of Alzheimer's disease using MRI data," *Sci Rep*, vol. 14, no. 1, doi: 10.1038/s41598-024-53733-6.

[8] P. K. Mandal and R. V. Mahto, "Deep Multi-Branch CNN Architecture for Early Alzheimer's Detection from Brain MRIs," *Sensors*, vol. 23, no. 19, Oct. 2023, doi: 10.3390/s23198192.

[9] M. Kaya, "Feature fusion-based ensemble CNN learning optimization for automated detection of pediatric pneumonia," *Biomed. Signal Process. Control*, vol. 87, Jan. 2024, doi: 10.1016/j.bspc.2023.105472.

[10] K. Barrera-Llanga, J. Burriel-Valencia, Á. Sapena-Bañó, and J. Martínez-Román, "A Comparative Analysis of Deep Learning Convolutional Neural Network Architectures for Fault Diagnosis of Broken Rotor Bars in Induction Motors," *Sensors*, vol. 23, no. 19, Oct. 2023, doi: 10.3390/s23198196.

[11] M. Zaabi, N. Smaoui, H. Derbel, and W. Hariri, "Alzheimer's disease detection using convolutional neural networks and transfer learning based methods," in *Proceedings of the 17th International Multi-Conference on Systems, Signals and Devices, SSD 2020*, Institute of Electrical and Electronics Engineers Inc., Jul. 2020, pp. 939–943. doi: 10.1109/SSD49366.2020.9364155.

[12] H. C. Shin, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Trans Med Imaging*, vol. 35, no. 5, pp. 1285–1298,, doi: 10.1109/TMI.2016.2528162.

[13] M. Kaya, S. Ulutürk, Y. Ç. Kaya, O. Altıntaş, and B. Turan, "Optimization of Several Deep CNN Models for Waste Classification," *Sakarya University Journal of Computer and Information Sciences*, vol. 6, no. 2, p. 2023, doi: 10.35377/saucis…1257100.

[14] M. Feurer and F. Hutter, "Hyperparameter Optimization," *Automated Machine Learning*,

pp. 3–33, 2019, doi: 10.1007/978-3-030-05318-5_1.
[15] N. Deepa and S. P. Chokkalingam, "Optimization of VGG16 utilizing the Arithmetic Optimization Algorithm for early detection of Alzheimer's disease," *Biomed. Signal Process. Control*, vol. 74, p. 103455, Apr. 2022, doi: 10.1016/J.BSPC.2021.103455.
[16] "Augmented Alzheimer MRI Dataset," Kaggle. Accessed: Jan. 18, 2025. [Online]. Available: https://www.kaggle.com/datasets/uraninjo/augmented-alzheimer-mri-dataset
[17] S. Fathi, A. Ahmadi, A. Dehnad, M. Almasi-Dooghaee, and M. Sadegh, "A Deep Learning-Based Ensemble Method for Early Diagnosis of Alzheimer's Disease using MRI Images," *Neuroinformatics*, vol. 22, no. 1, pp. 89–105, Jan. 2024, doi: 10.1007/s12021-023-09646-2.
[18] E. Hussain, M. Hasan, S. Z. Hassan, T. Hassan Azmi, M. A. Rahman, and M. Zavid Parvez, "Deep Learning Based Binary Classification for Alzheimer's Disease Detection using Brain MRI Images," *Proceedings of the 15th IEEE Conference on Industrial Electronics and Applications, ICIEA 2020*, pp. 1115–1120, Nov. 2020, doi: 10.1109/ICIEA48937.2020.9248213.
[19] P. Erdogmus and A. T. Kabakus, "The promise of convolutional neural networks for the early diagnosis of the Alzheimer's disease," *Eng. Appl. Artif. Intell.*, vol. 123, p. 106254, Aug. 2023, doi: 10.1016/J.ENGAPPAI.2023.106254.
[20] R. Rajasree and S. Brintha Rajakumari, "Ensemble-of-classifiers-based approach for early Alzheimer's Disease detection," *Multimed. Tools Appl.*, vol. 83, no. 6, pp. 16067–16095, Feb. 2024, doi: 10.1007/S11042-023-16023-3/FIGURES/7.
[21] H. Ji, Z. Liu, W. Q. Yan, and R. Klette, "Early Diagnosis of Alzheimer's Disease Using Deep Learning," 2019, doi: 10.1145/3341016.3341024.
[22] S. Lahmiri, "Integrating convolutional neural networks, kNN, and Bayesian optimization for efficient diagnosis of Alzheimer's disease in magnetic resonance images," *Biomed. Signal Process. Control*, vol. 80, p. 104375, Feb. 2023, doi: 10.1016/J.BSPC.2022.104375.
[23] Q. Wang, Y. Li, C. Zheng, and R. Xu, "DenseCNN: A Densely Connected CNN Model for Alzheimer's Disease Classification Based on Hippocampus MRI Data."
[24] F. Razavi, M. J. Tarokh, and M. Alborzi, "An intelligent Alzheimer's disease diagnosis method using unsupervised feature learning," *J. Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0190-7.
[25] M. Yildirim and A. Cinar, "Classification of Alzheimer's disease MRI images with CNN based hybrid method," *Ingenierie des Systemes d'Information*, vol. 25, no. 4, pp. 413–418, Aug. 2020, doi: 10.18280/isi.250402.
[26] J. Venugopalan, L. Tong, H. R. Hassanzadeh, and M. D. Wang, "Multimodal deep learning models for early detection of Alzheimer's disease stage," *Sci. Rep.*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-020-74399-w.
[27] S. Katabathula, Q. Wang, and R. Xu, "Predict Alzheimer's disease using hippocampus MRI data: a lightweight 3D deep convolutional network model with visual and global shape representations," *Alzheimers Res. Ther.*, vol. 13, no. 1, Dec. 2021, doi: 10.1186/s13195-021-00837-0.
[28] S. U. Sadat, H. H. Shomee, A. Awwal, S. N. Amin, M. T. Reza, and M. Z. Parvez, "Alzheimer's Disease Detection and Classification using Transfer Learning Technique and Ensemble on Convolutional Neural Networks," *Conf. Proc. IEEE Int. Conf. Syst. Man Cybern.*, pp. 1478–1481, 2021, doi: 10.1109/SMC52423.2021.9659179.
[29] N. A. Baghdadi, A. Malki, H. M. Balaha, M. Badawy, and M. Elhosseini, "A3C-TL-GTO: Alzheimer Automatic Accurate Classification Using Transfer Learning and Artificial Gorilla Troops Optimizer," *Sensors 2022, Vol. 22, Page 4250*, vol. 22, no. 11, p. 4250, Jun. 2022, doi: 10.3390/S22114250.
[30] A. Ebrahimi and S. Luo, "Convolutional neural networks for Alzheimer's disease detection on MRI images.," *J. Med. Imaging (Bellingham)*, vol. 8, no. 2, p. 24503, Mar. 2021, doi: 10.1117/1.JMI.8.2.024503.
[31] Q. Li and M. Q. Yang, "Comparison of machine learning approaches for enhancing Alzheimer's disease classification," *PeerJ*, vol. 9, p. e10549, Feb. 2021, doi: 10.7717/peerj.10549.
[32] A. Priyatama, Z. Sari, and Y. Azhar, "Deep Learning Implementation using Convolutional Neural Network for Alzheimer's Classification," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi*, vol. 10, no. 2, pp. 310–317, 2023, doi: 10.29207/resti.v7iX2.4707.
[33] S. G. Dakdareh and K. Abbasian, "Diagnosis of Alzheimer's Disease and Mild Cognitive Impairment Using Convolutional Neural Networks," *J. Alzheimers Dis. Rep.*, vol. 8, no. 1, p. 317, Feb. 2024, doi: 10.3233/ADR-230118.
[34] S. Bekhet, N. Saad, and S. Farag, "Alzheimer disease predicting from clinical and MRI data using DeepALZNET dual pathway framework," *Scientific Reports 2025 15:1*, vol.

5, no. 1, pp. 43250-, Dec. 2025, doi: 10.1038/s41598-025-28221-0.