KSÜ Mühendislik Bilimleri Dergisi, 28(1), 2025
Araştırma Makalesi

KSU J Eng Sci, 28(1), 2025
Research Article

Kahramanmaras Sutcu Imam University
Journal of Engineering Sciences

Geliş Tarihi : 20.08.2024
Kabul Tarihi :20.12.2024

Received Date : 20.08.2024
Accepted Date : 20.12.2024

ADNet: A CNN MODEL FOR ALZHEIMER'S DISEASE DIAGNOSIS ON OASIS-1 DATASET

ADNet: OASIS-1 VERİ KÜMESİ ÜZERİNDE ALZHEİMER HASTALIĞI TEŞHİSİ İÇİN BİR CNN MODELİ

Ahmet Samed SARAÇOĞLU[1] (ORCID: 0009-0003-7835-4191)
Ayşe Merve ACILAR[1,2*] (ORCID: 0000-0002-0133-2694)
Özlem ERDAŞ ÇİÇEK[1] (ORCID: 0000-0003-4019-7744)

[1] Necmettin Erbakan University, Engineering Faculty, Computer Engineering Department, Konya, Türkiye
[2]Necmettin Erbakan University, Trauma Intervention and Research Center, Konya, Türkiye

*Sorumlu Yazar / Corresponding Author: Ayşe Merve ACILAR, m.acilar@erbakan.edu.tr

ABSTRACT

Alzheimer's disease (AD) is a chronic neurodegenerative disorder affecting memory, thinking, and behavior. Deep learning models, particularly CNNs, have shown promise in detecting AD at initial stages using the brain's magnetic resonance images (MRI). In this study, a CNN model called ADNet, trained using the OASIS-1 dataset, was proposed. The experimental approaches for evaluating the performance of ADNet are as follows: First, three different datasets were prepared using slices taken from the first quarter, middle, and third quarter of the sagittal plane from each MRI, to determine the most informative slice among the 128 slices. Each dataset was split into 80% training and 20% testing. It was found that the first quarter slice showed the best performance. The potential use of the obtained model as a transfer learning model was also examined. For this, a low-performance model was retrained using ADNet as a transfer learning model, and significant improvements in the results were observed. At last, the model's robustness was evaluated in a more detailed evaluation, using 5-fold cross-validation repeated three times, resulting in a mean accuracy of 97.05%. As a result, ADNet can be used for Alzheimer's screening in clinical settings and could enable patients to receive earlier treatment.

Keywords: Alzheimer's disease diagnosis, deep learning, convolutional neural network, magnetic resonance imaging data, transfer learning, open access series of imaging studies brain database.

ÖZET

Alzheimer hastalığı (AD), hafıza, düşünme ve davranış üzerinde ciddi etkileri olan kronik bir nörodejeneratif hastalıktır. Evrişimli Sinir Ağları (CNN) gibi derin öğrenme modelleri, beyin manyetik rezonans görüntüleri (MRI) kullanılarak AD'nin erken aşamalarda tespit edilmesinde umut verici sonuçlar göstermektedir. Bu çalışmada, Alzheimer teşhisi için OASIS-1 veri seti kullanılarak eğitilen ADNet adlı bir CNN modeli önerilmiştir. ADNet'in performansını değerlendirmek için, ilk olarak, her bireyin MR görüntüsünden alınan sagittal düzlemdeki 128 dilimin ilk çeyreğinden, ortasından ve üçüncü çeyreğinden alınan dilimler kullanılarak üç farklı veri seti hazırlanmış ve en bilgilendirici dilim hangisi araştırılmıştır. Her veri seti %80 eğitim ve %20 test olarak ayrılmış ve ilk çeyrek dilimin en iyi performansı gösterdiği saptanmıştır. Ek olarak, elde edilen modelin transfer öğrenme modeli olarak kullanılıp kullanılamayacağı incelenmiştir. Bunun için düşük performanslı bir model, ADNet transfer öğrenme modeli kullanılarak yeniden eğitilmiş ve sonuçların oldukça iyileştiği gözlemlenmiştir. Son olarak, modelin dayanıklılığı 5 katlı çapraz doğrulama ile üç kez tekrarlanarak daha ayrıntılı bir değerlendirmeye tabi tutulmuş ve %95,36 ortalama doğruluk elde edilmiştir. Sonuç olarak, ADNet'in klinik ortamlarda Alzheimer taramasında kullanılabileceği ve hastaların daha erken tedavi almasını sağlayabileceği düşünülmektedir.

Anahtar Kelimeler: Alzheimer hastalığı teşhisi, derin öğrenme, evrişimli sinir ağı, manyetik rezonans görüntüleme verisi, transfer öğrenme, açık erişimli görüntüleme çalışmaları serisi beyin veri tabanı.

ToCite: SARAÇOĞLU A.S.; ACILAR A. M. & ERDAŞ ÇİÇEK. Ö., (2025). ADNet: A CNN MODEL FOR ALZHEIMER'S DISEASE DIAGNOSIS ON OASIS-1 DATASET. *Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi*, 28(1), 487-504.

INTRODUCTION

A neurodegenerative disease characterized by the progressive death and loss of the function of nerve cells in the brain is referred to as Alzheimer's Disease (AD). Memory, language, reasoning, and problem-solving may be affected by this disease (Scheltens et al., 2022). In the beginning, patients usually have minor memory problems, however as time goes on, they might have difficulty in performing their daily activities, show behavior changes, or fail to communicate (Afzal et al., 2021) (Porsteinsson, Isaacson, Knox, Sabbagh, & Rubino, 2021).

AD is a growing global health problem that causes impairment in cognitive abilities. The World Health Organization's 2023 Report predicts that 55 million people were affected by the disease in 2021 (Alzheimer's Association, 2023). This number is expected to rise significantly in the coming years. Alzheimer's disease not only affects the quality of life of patients but also places a huge strain on healthcare systems. Researchers are constantly working to improve the comprehension and management of this complex disease. In recent years, considerable progress has been made in understanding the causes, diagnostic methods, and approaches to potential treatment.

Although its exact cause remains unknown, the development of AD is attributed to the accumulation of amyloid plaques and tau deposits in the brain (Breijyeh & Karaman, 2020). These protein deposits impair neurons' function and ultimately result in their death. While aging is the most important risk factor for developing AD, genetic susceptibility, diabetes, heart disease, and low education level are associated with an increased risk (Bendlin et al., 2011).

AD currently has no definitive cure, but some life changes may help in preventing or delaying AD. Exercising regularly, eating healthy food, engaging oneself mentally, and interacting socially can all help maintain cognitive health and reduce the risk of AD (Bendlin et al., 2011). Symptoms can be managed by diagnosing early and slowing down the progression of the disease. Diagnosis normally involves thorough clinical evaluation which includes neuropsychological testing and neuroimaging (Porsteinsson et al., 2021). There is extensive research on novel therapeutic interventions for combating ADs. The aim is to find new drugs that inhibit or eliminate amyloid plaques and tau proteins (Ovsepian, Leary, Zaborszky, & Ntziachristos, 2019), decrease inflammation, and save nerve cells (Breijyeh & Karaman, 2020). Timely diagnosis is crucial for efficient management/treatment of the disease and the design of new drugs. Besides, progress in artificial intelligence and imaging technology presents new prospects for early detection of the disease and monitoring treatment response (Alzheimer's Association, 2023). The application of classic machine learning algorithms and deep learning techniques to medical image analysis has garnered significant attention in recent years. Various studies have explored different approaches to improve the accuracy and reliability of diagnostic tools, each contributing unique methods and findings to the field. Baglat et al. (2020) applied a Random Forest classifier to T1-weighted Magnetic Resonance Imaging (MRI) data with a dataset of 150 samples into two categories and used a 5-fold cross-validation method Test accuracy of 86% was obtained (Baglat, Salehi, Gupta, & Gupta, 2020). In another study, Logistic Regression on a dataset of 373 samples, classifying them into three categories was employed by Alroobaea and colleagues. They used 5-fold cross-validation method and achieved a test accuracy of 84.33% (Alroobaea & Bragazzi, 2021). Extra Tree Classifier on longitudinal MRI data with a dataset of 373 samples was performed and achieved a test accuracy of 85% by Jadhao et al. (Jadhao et al., 2023). Another machine learning algorithm Random Forrest used by Shrivastava et al. to detect dementia on a dataset of 136 samples divided into two classes: non-demented and demented. The dataset was split into training and testing sets with a 75:25 ratio, achieving the best test accuracy of 84%(Shrivastava, Singh, & Kaur, 2023). Rajayyan and Mustafa (2023) utilized the Gaussian Naïve Bayes method on a dataset of 373 samples, classified into three categories: non-demented, moderate dementia, and demented. The study used an 80:20 hold-out cross-validation technique, achieving a test accuracy of 95%(Rajayyan & Mustafa, 2023).

As the field progressed, researchers began to leverage deep learning methods to capture the intricate features of neuroimaging data. Inception-v3 models, as Salami et al. (2022) proposed, represented a significant leap, achieving an accuracy of 87.75%. More recent studies have focused on advanced CNN architectures and transfer learning. Saratxaga and colleagues (2021) applied deep learning methods to a dataset of 1,114 samples, focusing on 10 central MRI slices. They classified the data into three categories: non-demented, moderate dementia, and demented, with a 70:10:20 split between training, validation, and test sets. They obtained the highest test accuracy of 89% using the ResNet18 model (Saratxaga et al., 2021). Khagi and colleagues (2019) employed a Scratch Trained CNN method on a smaller dataset of fifty-six samples, using 30 MRI slices per sample. The study classified the data into two

categories: moderate dementia and demented, achieving a high test accuracy of 98.51% with a 60:20:20 data split for training, validation, and testing (Khagi & Kwon, 2019). Chui et al. (2022) used a GAN-CNN-TL model on a large dataset of 2,168 samples, classifying them into four categories: non-demented, very mildly demented, mildly demented, and moderately demented. The study employed a 5-fold cross-validation method, achieving a test accuracy of 97.5%(Chui, Gupta, Alhalabi, & Alzahrani, 2022). Balasundaram et al.(2023) applied a combination of CNN, multilayer models, and ResNet50 to a dataset of 373 samples, focusing on a single MRI slice. The data was classified into four categories, and the study reported a test accuracy of 94% (Balasundaram, Srinivasan, Prasad, Malik, & Kumar, 2023). In another study, the MobileNet model was applied to a dataset of 436 coronal plane MRI images, classifying them into two categories: non-demented and demented. The study employed an 80:10:10 split for training, validation, and testing, achieving a test accuracy of 95.24% (Ghosh et al., 2023). Hajamohideen et al. (2023) used a Siamese CNN model with a triplet-loss function on a large dataset of 6,400 axial plane MRI images. The study classified the data into four categories and reported a test accuracy of 93.85% (Hajamohideen et al., 2023). Another study that uses 2 deep learning approaches in the diagnosis of AD, examines Simplistic CNN Architectures and Transfer Learning. Simplistic CNN Architectures uses simple CNN architectures with 2D and 3D convolutions to process 2D and 3D structured brain scanning. This approach achieves 93.61% and 95.17% accuracy for 2D and 3D multi-class AD classifications, respectively (Lu et al., 2022). The second method applies the principles of transfer learning to take advantage of pre-trained models. This approach uses pre-trained models for classification of the medical images such as the VGG19 model (Helaly, Badawy, & Haikal, 2022). The pretrained VGG19 model achieves 97% accuracy for multiclass AD classification. Continued endeavors to combat this disease give hope that it may be possible to prevent or treat AD in the future.

This paper aims to use deep learning to develop a new method for diagnosing AD. For this purpose, answers to the following research questions were sought:

- How can high accuracy be obtained in AD diagnosis with deep learning models using RAW MRI data in the Oasis-1 dataset from the Oasis Brain Database?
- How does the use of different quartiles and median slices impact CNN model performance?
- Could the obtained model be used as a transfer learning method to improve the performance of low-performing models?

In this study, a new CNN model called ADNet was proposed for Alzheimer's disease diagnosis on the OASIS-1 dataset. This paper is organized as follows: The Materials and Methods section details and prepares the dataset, description of performance metrics, and the model architecture of ADNET proposed in this study. Next, the Experimental Results section is given. In the Discussion section, these results were interpreted in the context of existing literature. Finally, the Conclusion section summarizes the key findings and highlights the contributions of the study.

MATERIALS AND METHODS

MRI (Magnetic Resonance Imaging) images are obtained by placing the patient within a strong magnetic field, which aligns the hydrogen protons in the body and allows for precise sectional views of specific body regions by capturing images in slices. Slicing refers to the acquisition of cross-sectional images of the body or brain in thin layers. They are typically generated in different types of planes, including axial, sagittal, and coronal, to provide a comprehensive view of internal structures from multiple perspectives. Examples of the MRI slices of the axial, sagittal, and coronal planes are given in Figure 1.

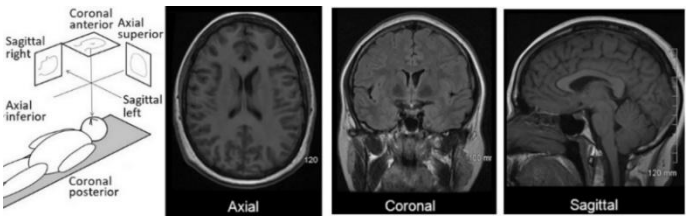


Figure 1. Example of the MRI Slices of the Axial, Sagittal, And Coronal Planes(Avots, Jafari, Ozcinar, & Anbarjafari, 2024)

Axial, coronal, and sagittal are the three main MRI imaging planes. As shown in Figure 1, the axial plane captures cross-sectional images from top to bottom, allowing visualization of structures as if viewed from above. The coronal

plane provides front-to-back sectional views, giving a perspective like looking directly at the face. The sagittal plane divides the body into left and right sections, offering a side view of the brain and other structures. Each plane allows for a distinct anatomical perspective, aiding in a comprehensive analysis of structural details within the nervous system. Slices are taken from these planes. Slice thickness, usually ranging from 1 to 5 mm, is selected depending on the required resolution, with thinner slices providing more detail but requiring longer scan times.

Dataset Descriptions

OASIS data sets are open-source brain MRI datasets namely OASIS-1, OASIS-2, OASIS-3, OASIS-3 TAU, and OASIS-4. The OASIS-1 dataset, which consisted of a cross-sectional collection of 416 respondents aged 18 to 96 years, was used in the present study. There are three to four individual T1-weighted MRI scans for each subject. The subjects were both men and women and all were right-handed. All data have been anonymized to accommodate public distribution. The data are available at http://www.oasis-brains.org (Marcus et al., 2007). A total of 1688 data was made available in this dataset. After the removal of the missing data, 905 images remained for training and testing. The data has three classes: images of non-AD, subjects at the initial phase of AD, and subjects with AD. The technical specifications for all MRIs used in this study are as follows: The MRI data had dimensions of **256 x 256 x 128**, indicative of the resolution of the scan. This indicates that the image comprises 256 pixels in width, 256 pixels in height, and 128 slices, thus facilitating a comprehensive representation of the brain's structure. The **voxel size** was **1x1x1.25 mm**, where each voxel, the three-dimensional equivalent of a pixel, has a width of 1 mm, a height of 1 mm, and a depth of 1.25 mm. The smaller the voxel size, the higher the spatial resolution, enabling more precise visualization of anatomical structures. The scan was obtained in the **sagittal plane** (Sag), meaning the images were taken along a vertical plane that divides the brain into left and right hemispheres. This orientation provides a clear view of midline brain structures, which is critical for assessing various neurological conditions. Slice examples of MRI according to classes from the Sagittal Plane are depicted in Figure 2.

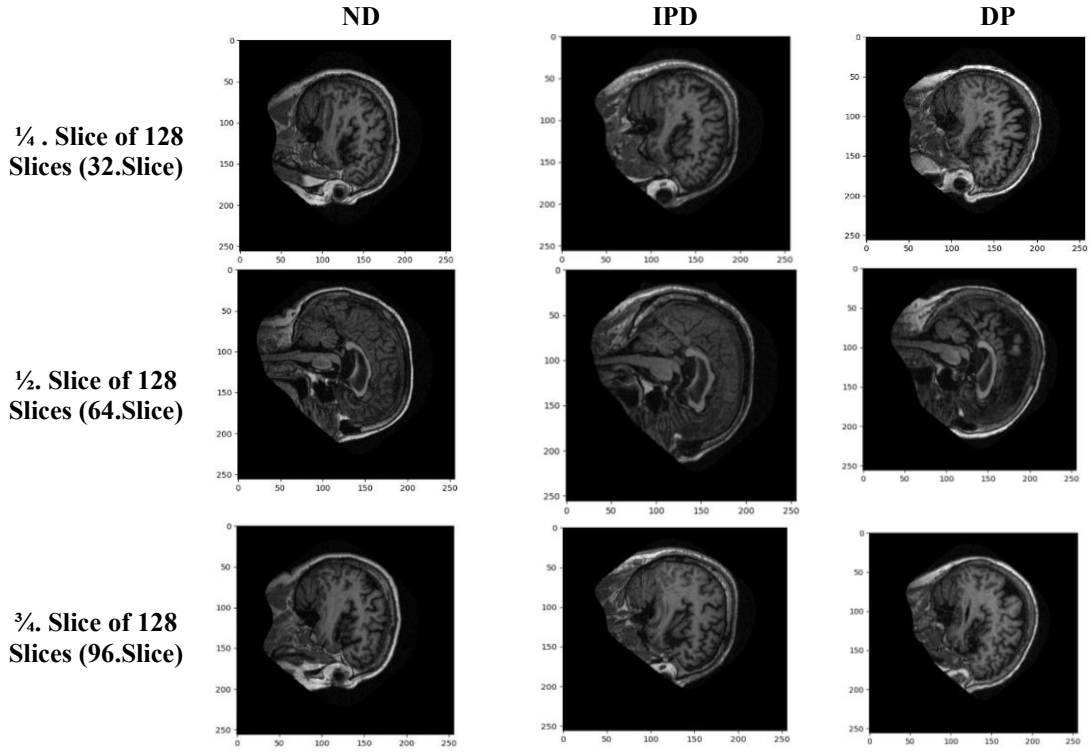


Figure 2. Slice Examples of MRI According to Classes from the Sagittal Plane

MRI slice examples are depicted in Figure 2 from the sagittal plane for three different classes: ND, IPD, and DP. The slices are taken from three distinct parts of the brain, representing the ¼., ½., and ¾. slice of the total 128 slices in the scan for each subject. Each column corresponds to a different condition, while each row illustrates a different slice location in the sagittal plane. These images help visualize and compare the brain structures across these three diagnostic categories.

The OASIS-1 dataset has individual folders for each MRI scan. Each MRI scan has two different versions (preprocessed and raw) placed in these folders. In this study, the raw data in the RAW folder was preferred. In this folder, the MRI scans are in IMG file format and the *measures* of the relevant scans are in CSV format. These parameters are: "*ID* (unique identification number of each subject), *M/F* (gender of the subject), *Hand* (dominant hand of the subject), *Age* (age of the subject), *Education* (education level of the subject), *SES* (socioeconomic status of the subject), *MMSE* (Cognitive function test scores), *CDR* (Clinical Dementia Ratings), *eTIV* (Brain volume estimation), *nWBV* (Normalized brain size), *ASF* (Brain scan size correction factors), *Delay* (Time delay between brain scan). Only the CDR value was considered in this study for labeling the MRI images. The dataset was divided into three classes using the Clinical Dementia Rating (CDR) parameter as Nondemented – ND (CDR=0), Initial Phase of Dementia -IPD (CDR=0.5), and Dementia Patient -DP (CDR=1). The plot given in Figure 3 shows the class distribution based on the CDR parameters in the OASIS dataset across three classes: Nondemented (ND), Initial Phase Dementia (IPD), and Dementia Patient (DP).

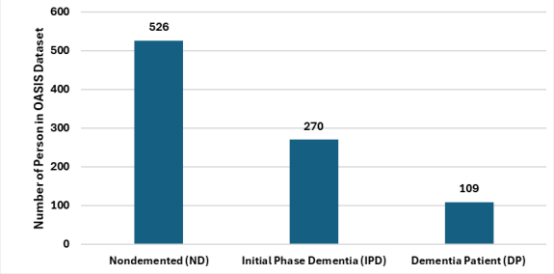


Figure 3. Distribution of Dataset According to Classes

As shown in Figure 3, The Nondemented, Initial Phase Dementia, and Dementia Patient classes have 526, 270, and 109 individuals, respectively.

Data Preparing

In this work, raw MR data from the OASIS-1 dataset were utilized to explore the potential of deep learning models in AD detection. The data preparation stage is important so that this raw data can be given as input to deep learning models. The steps below were performed in this phase:

Conversion to NII File Format: The images in the OASIS-1 dataset are saved in IMG format which cannot be directly processed by the deep learning model. In the first stage, IMG files were transformed into NII file format to render them as a Numpy array in Python. This was performed by the Nibabel library which is a commonly used tool for the analysis of brain imaging arrays (Salhi et al., 2023).

Selection of MRI Slices: After the conversion of the NII files as a Numpy array, it was determined that the MRI scans consist of 128 slice images, with each slice representing a distinct axis of the brain image. The human brain is a complicated organ, and imaging from different angles provides a better insight into brain function and pathology. Quarter slices (1/4, 3/4) and center slices (1/2) were selected for the experiments, in this study. This selection will enable brain images to be analyzed from different angles and model performance to be assessed. The slice choice was modeled after this study (Gramfort et al., 2013) which examined the effect of the distinct slice selections on model performance for the analysis of MEG and EEG data.

Conversion from NII to PNG: NII files were transformed into PNG files for each slice selected using the Nibabel library. This transformation served two major purposes: size reduction and deep learning model adaptability. The total size of the NII file decreased from 15,183,699,040 bytes to 86,262,583 bytes for PNG files. This substantial size reduction has significantly decreased data storage and maintenance costs. PNG format is one of the accepted formats for training the CNN model. This conversion makes the data processable by the model directly.

Methodology and Classification

In this work, a Convolutional Neural Network (CNN) deep learning architecture is designed and trained. CNN is a deep-learning algorithm generally used in image processing. This is a highly effective mechanism for image recognition and classification. The proposed CNN architecture aims to accurately detect AD from MRI. The model

is designed in Python within the Keras framework. It consists of two conventional layers, two maximum pooling layers, a flattening layer, a fully connected layer, and an output layer. The schema of the model architecture of ADNet is depicted in Figure 4.

Table 1. Hyper Parameter Tuning of ADNet

Hyperparameter Values	Metric	ND	IPD	DP	Avg.
Kernel Size for Conv. Layers = 3x3 *Number of Units for 1. Conv. Layer = 32* *Number of Units for 2. Conv. Layer = 64* *Activation Function = Adam*	Prec	0.867	0.533	1.000	0.800
	Rec	0.684	0.800	0.833	0.773
	F1	0.765	0.640	0.909	***0.771***
	Acc				***0.743***
Kernel Size for Conv. Layers = 3x3 Number of Units for 1. Conv. Layer = 32 Number of Units for 2. Conv. Layer = 64 Activation Function = Rmsprop	Prec	0.667	0.429	0.857	0.651
	Rec	0.737	0.300	1.000	0.679
	F1	0.700	0.353	0.923	0.659
	Acc				0.657
Kernel Size for Conv. Layers = 5x5 Number of Units for 1. Conv. Layer = 32 Number of Units for 2. Conv. Layer = 64 Activation Function =Adam	Prec	0.731	1.000	1.000	0.910
	Rec	1.000	0.300	1.000	0.767
	F1	0.844	0.462	1.000	0.769
	Acc				0.800
Kernel Size for Conv. Layers = 5x5 Number of Units for 1. Conv. Layer = 32 Number of Units for 2. Conv. Layer = 64 Activation Function = Rmsprop	Prec	0.750	0.500	0.857	0.702
	Rec	0.632	0.600	1.000	0.744
	F1	0.686	0.545	0.923	0.718
	Acc				0.686
Kernel Size for Conv. Layers = 3x3 Number of Units for 1. Conv. Layer = 64 Number of Units for 2. Conv. Layer = 128 Activation Function = Adam	Prec	0.708	0.600	1.000	0.769
	Rec	0.895	0.300	1.000	0.732
	F1	0.791	0.400	1.000	0.730
	Acc				**0.743**
Kernel Size for Conv. Layers = 3x3 Number of Units for 1. Conv. Layer = 64 Number of Units for 2. Conv. Layer = 128 Activation Function = Rmsprop	Prec	0.688	0.417	0.833	0.646
	Rec	0.579	0.556	0.833	0.656
	F1	0.629	0.476	0.833	0.646
	Acc				0.618
Kernel Size for Conv. Layers = 5x5 Number of Units for 1. Conv. Layer = 64 Number of Units for 2. Conv. Layer = 128 Activation Function = Adam	Prec	0.542	0.375	0.333	0.417
	Rec	0.684	0.300	0.167	0.384
	F1	0.605	0.333	0.222	0.387
	Acc				0.486
Kernel Size for Conv. Layers = 5x5 Number of Units for 1. Conv. Layer = 64 Number of Units for 2. Conv. Layer = 128 Activation Function = Rmsprop	Prec	0.667	0.400	0.857	0.641
	Rec	0.632	0.400	1.000	0.677
	F1	0.649	0.400	0.923	0.657
	Acc				0.629

Since using the complete dataset for hyperparameter tuning would be a weighty and time-consuming process, 30% of the dataset was selected as a sample using the Stratified Sampling Method, and the experiments were conducted using this sample set. Since only 30% of the data set was studied, the results were relatively low. The primary hyperparameters adjusted in these experiments were the kernel size of the convolutional layers, the number of units in the first and second convolutional layers, and the choice of activation function). The experimental design involved using both 3x3, and 5x5 for kernel sizes, testing different numbers of units for the convolutional layers (either 32 and 64 units or 64 and 128 units), and comparing the effect of Adam and Rmsprop optimization functions. This systematic

approach permitted an assessment of the impact of each combination of hyperparameters on the model's performance across different dementia stages. For each configuration, precision, recall, F1 score, and accuracy performance metrics were recorded across three classes. The mean value for each metric was calculated to evaluate the overall performance of each configuration. These results are given in Table 1.

Based on the hyperparameter tuning results presented in Table 1, the optimal configuration appears to be the model with a 3x3 kernel size, 32 units in the first convolutional layer, 64 units in the second convolutional layer, and the Adam optimizer. This configuration achieves the highest mean accuracy (0.743), indicating superior overall performance across all classes. Moreover, the highest F1 score of 0.771 suggests a well-balanced model in terms of precision and recall, in datasets such as this. The use of the Adam optimizer enhances the model's performance compared to Rmsprop. The default hyperparameter values for the Adam optimization algorithm in Keras (learning_rate = 0.001, beta_1 = 0.9, beta_2 = 0.999, epsilon = 1e-7) were used. The Sparse_Categorical_Crossentropy that computes the difference between the model's outputs and the observed labels as a loss function, ReLU, and Softmax as activation functions were also used in this study. A detailed description of the ADNet layers depicted in Figure 4 is given below.

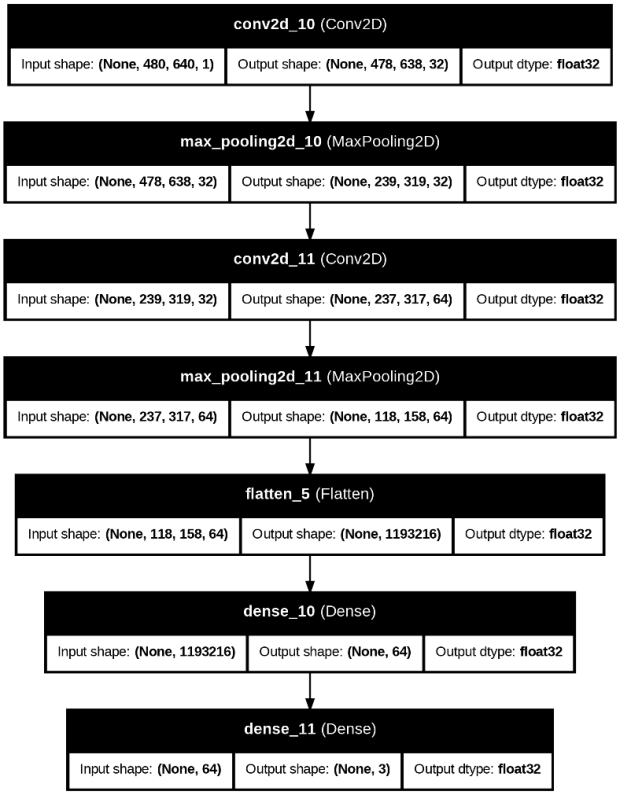


Figure 4. The Schema of the Model Architecture Called ADNet

Input Layer: 480 x 640 x 1 gray scale MRI is acquired. Every image is represented by a single channel (gray scale) containing 480 rows and 640 columns.

First Convolutional Layer: It has 32 kernels of size 3x3. Each kernel produces a new image feature map by shifting a 3x3 window on the image. The ReLU activation function discards values less than or greater than zero while retaining values greater than zero. This layer helps to extract edges, textures, and other significant details in the images.

First Maximum Pooling Layer: Reduces the image using a 2x2 window. It minimizes the size of the image by selecting the largest value in each window. This layer helps the model avoid over-learning and be more stable.

Second Convolutional Layer: It has 64 kernels of size 3x3. It works similarly to the first convolutional layer but aims to extract more complex features. The ReLU Activation Function is utilized.

Second Maximum Pooling Layer: It works similarly to the first Maximum Pooling Layer but minimizes the output of the 2nd Convolutional Layer.

Flatten Layer: It transforms three-dimensional vectors into a one-dimensional vector. This conversion is required for processing by the fully connected layer.

First Fully Connected Layer: It contains 64 neurons. Each Neuron is connected to all the inputs from the preceding layer and uses a weights matrix to convert them into an output. ReLU is selected as the activation function.

Output Layer: It contains 3 neurons. Each neuron predicts the probability of each of the three classes in an image. Softmax activation function makes sure that the sum of the output of each neuron is equal to 1 and each of them generates a likelihood value from 0 to 1.

Performance Evaluation Metrics

The accuracy (Acc), precision (Prec), recall or sensitivity (Rec), F1 score, specificity(spec.), Matthew's Correlation Coefficient (MCC), and Jaccard similarity Index (JSI) metrics obtained from the confusion matrix were used for evaluating the performance of the proposed model. The diagonal members of the matrix show the number of instances correctly classified by the model. The other cells show the number of instances that the model misclassified. For all metrics, a higher value indicates a more successful classification. While evaluating the results, success performance metrics obtained from the confusion matrix were used for multiple classifications which is illustrated in Figure 5.

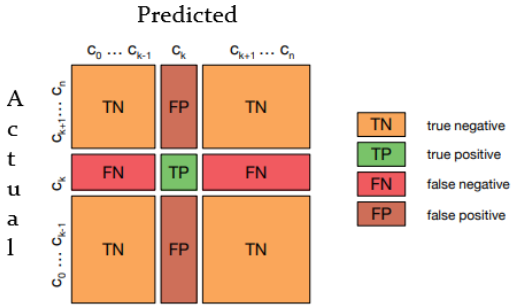


Figure 5. Confusion Matrix for Multi-Class Classification with N Classes, $0 \leq k \leq n$ (Krüger, 2016)

Figure 5 provides a visual representation of the confusion matrix for multi-class classification with n classes, demonstrating the distribution of true and predicted classes across categories. The formulas utilized in this study and calculated from the confusion matrix are also provided in equations (1), (2), (3),(4),(5),(6) and (7).

$$Accuracy\ (Acc) = (TP + \sum_i TN_i) \Big/ (TP + \sum_i TN_i + \sum_i FP_i + \sum_i FN_i) \quad (1)$$

$$Precision\ (Prec) = TP \Big/ (TP + \sum_i FP_i) \quad (2)$$

$$Recall\ (Rec\ or\ Sensitivity) = TP \Big/ (TP + \sum_i FN_i) \quad (3)$$

$$Spec = \sum_i TN_i \Big/ \sum_i TN_i + \sum_i FP_i) \quad (4)$$

$$F1 = (2 * Prec * Rec)/(Prec + Rec) \quad (5)$$

$$MCC = (TP.\sum_i TN_i) - (\sum_i FP_i . \sum_i FN_i) \Big/ \sqrt{(TP + \sum_i FP_i)(TP + \sum_i FN_i)(\sum_i TN_i + \sum_i FP_i)(\sum_i TN_i + \sum_i FN_i)} \quad (6)$$

$$JSI = TP \Big/ (TP + \sum_i FN_i + \sum_i FP_i) \quad (7)$$

where,

True Positives (TP): The number of instances correctly predicted as positive,
True Negatives (TN): The number of instances correctly predicted as negative,
False Positives (FP): The number of instances incorrectly predicted as positive,

False Negatives (FN): The number of instances incorrectly predicted as negative,
i: index of the class.

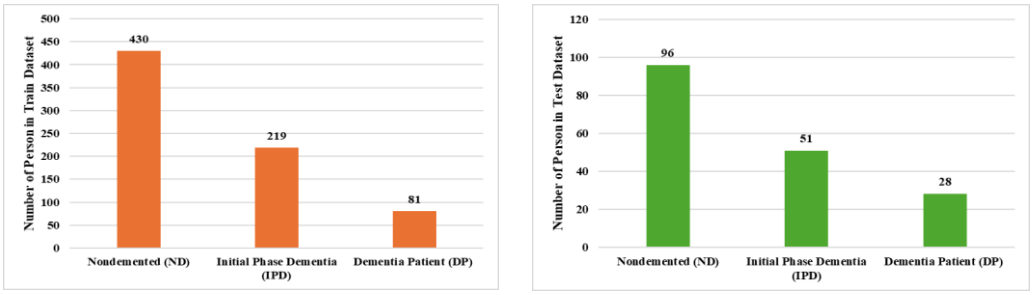
As seen from the equations, accuracy indicates overall effectiveness by measuring the proportion of correctly classified instances out of all instances. Precision assesses the model's exactness by calculating the proportion of true positive predictions among all positive predictions. Complementing this, Recall (or Sensitivity) measures the model's ability to capture relevant instances by identifying the proportion of actual positives correctly classified. The F1 Score, as the harmonic means of Precision and Recall, offers a balanced measure of accuracy, particularly valuable in the context of imbalanced datasets. MCC (Matthews Correlation Coefficient) provides a comprehensive measure by considering all elements of the confusion matrix, which makes it suitable for datasets with class imbalance. Specificity focuses on the model's ability to correctly identify true negatives, providing insight into its effectiveness at identifying negative instances. Lastly, the JSI (Jaccard Similarity Index) quantifies the overlap between true positives and predicted positives, measuring the similarity between predicted and actual positives, which is crucial for evaluating the model's ability to capture relevant instances accurately. High values are acceptable for all metrics.

EXPERIMENTAL RESULTS

Three different experimental studies have been conducted in this study to evaluate the performance of the proposed ADNet model. First, an experimental study was conducted to decide which slice of the 128 slices taken from the sagittal panel would be used. For this, three different datasets were composed using slices taken from the first quarter, middle, and third quarter of the sagittal plane for each individual MRI, respectively. Slice examples of MRI according to classes are given in Figure 2. Each dataset was split into 80% training and 20% testing using a hold-out method. The model was trained with the training set and was evaluated with the test set, and the results were reported in Table 2. As a result of this experimental study, it was decided to use images on the first quarter slice (1/4) of each MRI. In experiment 2, it was investigated whether the obtained model could be used as a transfer learning model. For this purpose, a model with low performance was retrained using ADNet as a transfer learning model, and the results were examined. The last experimental study involved evaluating the proposed model in more detail, the data set was trained using 5-fold cross validation and this process was repeated 3 times. The results were also reported in Table 4 and interpreted.

The Interpretation of Data from the Middle (1/2) and Quarter Slices (1/4, 3/4) of Sagittal Plane (Experiment 1)

Initially, three different datasets were created from MRI data for this study. As previously mentioned, the MRI data had dimensions of 256 x 256 x 128, meaning that there were 128 sagittal plane slices available for each individual. The three datasets were formed using specific slices from the 128 available MRI slices: the ¼. slice (32nd slice), the ½. slice (64th slice), and the ¾. slice (96th slice). Each dataset was split into 80% training and 20% testing using a hold-out method and the distribution of train and test datasets is given in Figure 6. To determine which of these slices would be used in this study, experimental analyses were conducted. The test results obtained from three slices trained using the same model architecture showed that the slicing techniques had a major effect on the model's performance.



(a) (b)

Figure 6. Distribution of (a) Train and (b) Test Dataset

Figure 6(a) shows that the training dataset is composed of 430 non-demented individuals, 219 with initial phase dementia, and 81 diagnosed as dementia patients. Figure 6(b) presents the testing dataset, which includes 96 non-demented individuals, 51 in the initial phase of dementia, and 28 dementia patients. This distribution highlights the class imbalance, with a higher proportion of non-demented individuals in both datasets. As a result of experimental

studies given in Table 2, it has been seen that ADNet could deal with this problem using ¼ . slice of 128 Slices (32.Slice).

Table 2. Performance Metrics of ADNet Models Trained with Different Slices

		Acc.	Prec.	Rec.(Sens.)	F1
¼ . Slice of 128 Slices (32.Slice)	ND	***0.9829***	1.0000	0.9792	0.9895
	IPD		0.9615	0.9804	0.9709
	DP		0.9655	1.0000	0.9825
	Avg.		***0.9757***	***0.9865***	***0.9809***
½. Slice of 128 Slices (64.Slice)	ND	0,9282	0,9720	0,9541	0,9630
	IPD		0,8824	0,9184	0,9000
	DP		0,8261	0,8261	0,8261
	Avg.		0,8935	0,8995	0,8963
¾. Slice of 128 Slices (96.Slice)	ND	0,5556	0,5556	1	0,7143
	IPD		-	0	-
	DP		-	0	-
	Avg.		-	0,333333	-

* The - sign indicates that the relevant metric cannot be calculated.

Accuracy, precision, recall, and F1 score were calculated for each class and depicted in Table 2. The 32. slice exhibits the highest performance across all metrics, with accuracy, precision, recall (sensitivity), and F1 scores close to or exceeding 0.98, suggesting that this slice provides highly informative features for the model. In contrast, the performance metrics for the 64. slice is somewhat lower, with an average accuracy of 0.8935, and precision, recall, and F1 scores in a similar range, indicating moderate model efficacy for this region. Notably, the 96. slice yields substantially lower results, with accuracy and precision values around 0.55, and recall and F1 metrics not calculable for certain classes. The accuracy rates by epochs during training are shown in Figure 7.

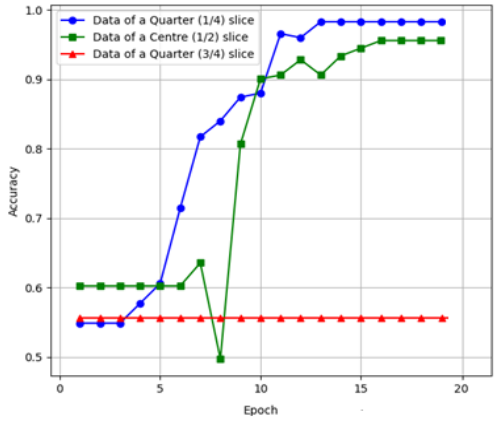


Figure 7. Training Accuracy of ADNet According to Quarter and Center Slices of MRI

A significant disparity in accuracy performance depending on the MRI slice location is seen in Figure 7. The blue line represents the ¼. Slice of 128 slice data, exhibits a sharp increase in accuracy, achieving near-perfect classification by epoch 10. The green line, indicating the accuracy on center slices, also shows an upward trend, though it stabilizes at a slightly lower level compared to the 1/4 slice. In contrast, the red line represents the ¾. Slice of 128, remains consistently low across all epochs, suggesting minimal learning from this slice region. This variation in performance suggests that different regions of MRI data provide varying degrees of discriminative information, potentially due to anatomical and pathological differences across slices. This observation emphasizes the importance of slice selection and its impact on model performance.

Effect of Transfer Learning (Experiment 2)

Another question tackled in this article is whether the obtained most successful model could be used as a transfer learning method to improve the performance of low-performing models. For this, transfer learning was performed using the ADNet model trained with the most successful quartile (1/4), and then the data from the lowest-performing quartile (3/4) was retrained using this obtained model. This technique increased the test accuracy of the lowest successful quartile (3/4) from 55.56% to 96.11%. These results are given in Table 3. Also, the comparisons of the accuracy rates according to the epochs during the training are presented in Figure 8.

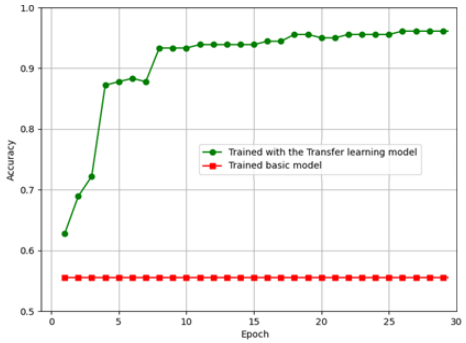


Figure 8. Accuracy Comparison with Transfer Learning in AD Diagnosis

As seen in Figure 8, the transfer learning method enabled the model to adjust to the new dataset and significantly improved its performance. The green line represents the model trained with transfer learning, showing a rapid increase in accuracy within the first 10 epochs, reaching over 90%, and stabilizing close to 1.0. This rapid convergence and high accuracy indicate the effectiveness of transfer learning in accelerating model training and improving performance by leveraging pre-existing knowledge. In contrast, the red line, representing the basic model trained from scratch, maintains a low and stable accuracy of around 0.55 throughout all epochs, suggesting limited learning and poor classification ability. It is said that the ADNet could be used as a transfer learning method to improve the performance of low-performing models. These significant findings have the potential to improve AD recognition to be more precise and reliable.

Table 3. Performance Metrics after using ADNet as transfer learning on the composed of ¾. Slice of 128 Slices

		Acc	Prec	Rec(Sens)	F1
¾. Slice of 128 Slices (96.Slice) using ADNET as Transfer Learning	ND	0,9611	0,9709	1,0000	0,9852
	IPD		1,0000	0,8727	0,9320
	DP		0,8621	1,0000	0,9259
	Avg.		0,9443	0,9576	0,9477

Table 3 presents the performance metrics of the ADNet model utilizing transfer learning on the ¾. slice (96th slice) of the 128-slice. The model achieved high accuracy, precision, recall (sensitivity), and F1 scores across the three classes. For the ND class, recall reached 1.0 with an accuracy of 0.9611, indicating excellent classification capability for this category. In the IPD class, precision was 1.0, but recall dropped slightly to 0.8727, reflecting a minor trade-off between precision and recall. The DP class exhibited strong performance, with a precision of 0.8621 and perfect recall at 1.0, yielding an F1 score of 0.9259. The average metrics for accuracy, precision, and recall (0.9443, 0.9576, and 0.9477, respectively) highlight the robustness of the transfer learning approach in enhancing model performance on this slice.

Cross Validation Results (Experiment 3)

To evaluate the proposed model in more detail, the data set was trained using 5-fold cross validation and this process was repeated 3 times. Figure 5 presents the confusion matrices obtained from a 3x5 cross-validation (CV) scheme applied to evaluate the model's classification performance across different repeats and folds. Each row in the figure represents a distinct repeat of the cross-validation, while each column corresponds to a specific fold within that repeat. The matrices illustrate the counts of actual versus predicted labels for three classes: ND, IPD, and DP. These results highlight the model's consistency and robustness in correctly identifying each class across multiple cross-validation folds, allowing for a comprehensive assessment of its performance stability.

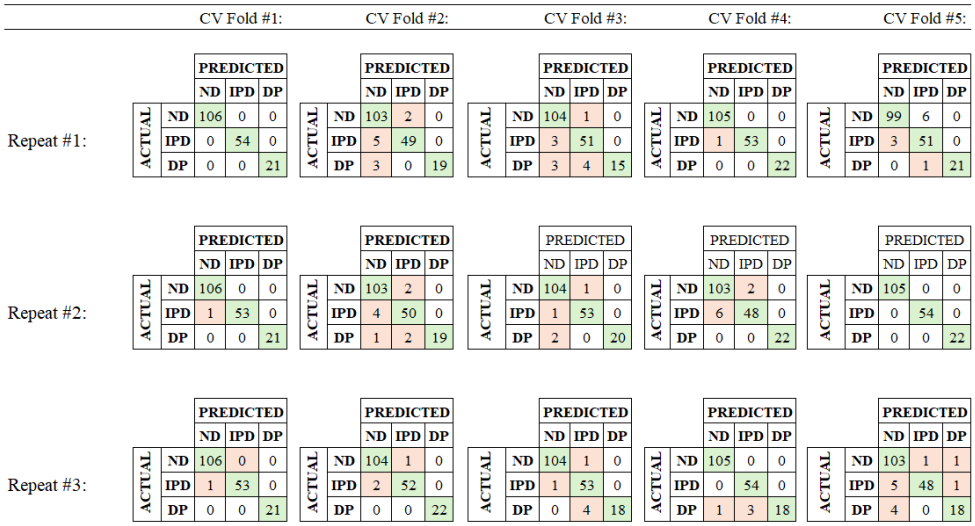


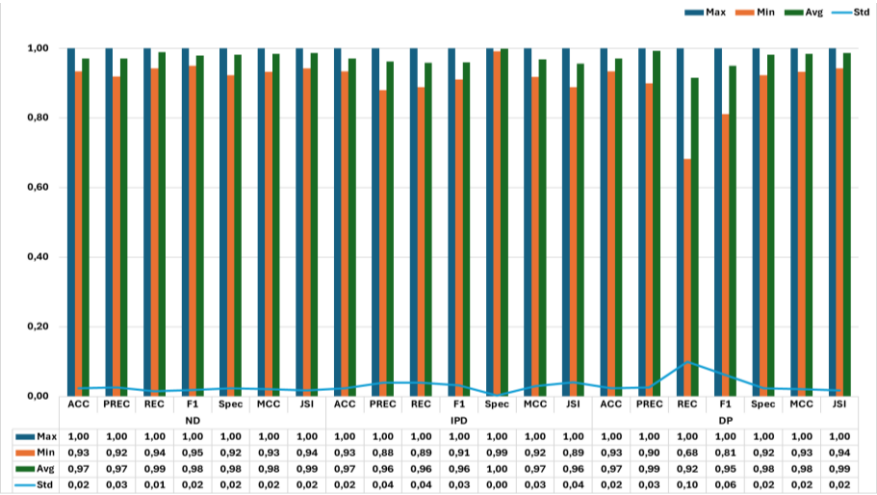
Figure 9. Confusion Matrices of 3x5cv for Dementia Classification Classes for Test Sets

As seen in Figure 9, the results demonstrate high consistency in classification accuracy with minimal misclassifications across folds and repetitions. Overall, the results suggest the model's robustness and reliability across different CV folds and repetitions, highlighting its potential effectiveness in multi-class medical image classification tasks. The performance metrics given in equations (1)-(7) are calculated from these confusion matrices shown in Figure 9 according to each fold of repeats. The average results of folds placed in repeats and the average of metrics obtained from repeats are reported for test sets in Table 4.

Table 4. Performance Metrics of 3x5cv for Dementia Classification Classes for Test Sets

Repeat #1:							
	Acc.	**Prec.**	**Rec.**	**F1**	**Spec.**	**MCC**	**JSI**
ND	0.9646	0.9669	0.9829	0.9746	0.9759	0.9794	0.9829
IPD		0.9502	0.9556	0.9524	1.0000	0.9680	0.9556
DP		1.0000	0.9000	0.9429	0.9936	0.9366	0.9000
Avg.		0.9724	0.9461	0.9566	0.9899	0.9613	0.9461
Repeat #2:							
	Acc.	**Prec.**	**Rec.**	**F1**	**Spec.**	**MCC**	**JSI**
ND	0.9757	0.9723	0.9905	0.9812	0.9861	0.9883	0.9905
IPD		0.9735	0.9556	0.9642	1.0000	0.9683	0.9556
DP		1.0000	0.9545	0.9758	0.9974	0.9730	0.9545
Avg.		0.9819	0.9669	0.9738	0.9945	0.9765	0.9669
Repeat #3:							
	Acc.	**Prec.**	**Rec.**	**F1**	**Spec.**	**MCC**	**JSI**
ND	0.9713	0.9745	0.9924	0.9832	0.9916	0.9905	0.9924
IPD		0.9644	0.9630	0.9628	0.9984	0.9731	0.9630
DP		0.9800	0.8909	0.9314	0.9913	0.9350	0.8909
Avg.		0.9730	0.9488	0.9592	0.9938	0.9662	0.9488
Avg. Results of all the Repeats							
	Acc.	***Prec.***	***Rec.***	***F1***	***Spec.***	***MCC***	***JSI***
ND	*0.9705*	*0.9716*	*0.9769*	*0.9740*	*0.9845*	*0.9860*	*0.9886*
IPD		*0.9715*	*0.9577*	*0.9637*	*0.9995*	*0.9698*	*0.9580*
DP		*0.9873*	*0.9193*	*0.9494*	*0.9941*	*0.9482*	*0.9152*
Avg.		*0.9727*	*0.9474*	*0.9579*	*0.9927*	*0.9680*	*0.9539*

Table 4 provides an in-depth performance analysis of the model across three repeated tests for the ND, IPD, and DP categories, evaluating metrics including Accuracy (Acc.), Precision (Prec.), Recall (Rec.), F1 Score (F1), Specificity (Spec.), Matthews Correlation Coefficient (MCC), and Jaccard Similarity Index (JSI). This multi-metric approach demonstrates the model's stability across varying conditions. In terms of Accuracy, ND and IPD consistently exhibit high values across each repeat, with averages of 0.9705 and 0.9715, respectively. DP achieves the highest overall accuracy at 0.9873, indicating reliable classification performance throughout. Precision metrics are similarly consistent, with ND maintaining an average of 0.9716, IPD at 0.9577, and DP slightly lower at 0.9193. Although DP's precision remains high, this relative drop implies a minor discrepancy in true positive predictions, potentially due to feature overlap or subtle distinctions within the DP class. The Recall results show the model's effectiveness in capturing true positives. ND achieves the highest Recall average (0.9769), followed by IPD (0.9637) and DP (0.9494). These high Recall scores confirm the model's ability to detect relevant samples accurately. The F1 Score, which balances Precision and Recall, shows optimal scores for ND and IPD, averaging 0.9740 and 0.9995, respectively, signifying strong balance. DP, with an F1 Score of 0.9941, maintains robust performance though it indicates a slightly less consistent balance between Precision and Recall. Specificity values highlight the model's proficiency in identifying true negatives and minimizing false positive rates. The consistency in Specificity across repeats underscores the model's reliability. MCC, a key indicator for imbalanced data performance, also reflects this trend. ND achieves the highest MCC at 0.9860, followed by IPD (0.9580) and DP (0.9152). JSI remains high, with averages of 0.9886 for ND, 0.9580 for IPD, and 0.9152 for DP, indicating a strong overlap between predicted and actual classifications. The high JSI values reflect the model's ability to align closely with true distributions, which is crucial for tasks demanding precise class correspondence. Overall, the model demonstrates robust performance across all metrics. Collectively, the results confirm the model's reliability and capability in diverse classification scenarios. Descriptive Statics of the performance metrics of 3x5 fold cross validation results were given in Figure 10. The accuracy and loss graph according to epoch numbers for training the best model among these models is given in Figure 11.

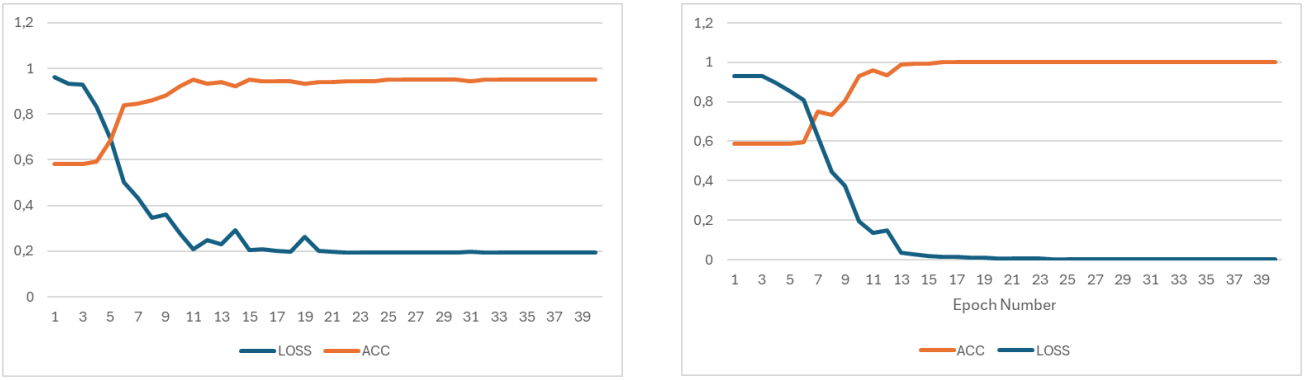


	ACC	PREC	REC	F1	Spec	MCC	JSI	ACC	PREC	REC	F1	Spec	MCC	JSI	ACC	PREC	REC	F1	Spec	MCC	JSI
	ND							IPD							DP						
Max	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Min	0,93	0,92	0,94	0,95	0,92	0,93	0,94	0,93	0,88	0,89	0,91	0,99	0,92	0,89	0,93	0,90	0,68	0,81	0,92	0,93	0,94
Avg	0,97	0,97	0,99	0,98	0,98	0,98	0,99	0,97	0,96	0,96	0,96	1,00	0,97	0,96	0,97	0,99	0,92	0,95	0,98	0,98	0,99
Std	0,02	0,03	0,01	0,02	0,02	0,02	0,02	0,02	0,04	0,04	0,03	0,00	0,03	0,04	0,02	0,03	0,10	0,06	0,02	0,02	0,02

Figure 10. Descriptive Statics of the Performance Metrics of 3x5 Fold Cross Validation Results

Figure 10 presents graphs of the descriptive statistics for the performance metrics of ADNet evaluated through 3x5 fold cross-validation. The minimum, maximum, average, and standard deviation of accuracy, precision, recall, F1 score, specificity, Matthews correlation coefficient, and Jaccard similarity index.

The ND, and IPD classes display consistently high performance across all metrics, with average values close to 1.0 and low standard deviations, indicating stable and reliable performance across folds. However, the DP class exhibits more variability, with lower minimum values for recall and F1 score however, the standard deviation was calculated as 0.1 at most. This shows that the cross-validation results demonstrate robust and consistent classification performance for detection the Alzheimer. The accuracy and loss graphs of the best model from 3x5cv according to epoch number are shown in Figure 11.



(a) (b)

Figure 11. Accuracy and Loss Graphs of the (a) Worst and (b) Best Model from 3x5cv According to Epoch Number

The stability observed in both metrics over subsequent epochs implies that the model has reached a point of minimal overfitting and exhibits consistent performance from Figure 11. This figure highlights the model's efficient convergence and reliability across the cross-validation process, reinforcing the robustness of the selected architecture and training strategy.

DISCUSSION

This study examines the efficacy of ADNet which is a Convolutional Neural Network based Deep Learning model for the diagnosis of Alzheimer's Disease on the OASIS-1 dataset. This section presents a discussion of the implications of the experimental findings and a comparison of ADNet's performance with existing studies on Alzheimer's diagnosis using the OASIS-1 dataset. The comparative analysis presented in Table 5 provides a comprehensive overview of the performance of various state-of-the-art methods applied to the OASIS dataset. All studies given in Table 5 were conducted using the OASIS-1 dataset, with accuracy as the common metric for evaluating success. Consequently, the discussion is primarily focused on this metric. Table 5 also presents detailed information for each study, including the number of records in the dataset, data type, sampling method, method used, class labels used, and test accuracy value.

As seen from Table 5, a comparison with earlier studies shows that traditional machine learning methods such as Random Forest and SVM have demonstrated competitive results, with accuracies ranging from 84.00% to 94.80%. The findings indicate that the deep networks employed in recent methodologies are more adept at capturing the intricate features essential for the tasks pertaining to the OASIS dataset. Deep learning models have demonstrated a notable enhancement, as evidenced by models such as Inception-v3 model (87.75%), Supervised Autoencoder-SSAs (90%), and MobileNet (95.24%) which use holdout as a sampling method. The proposed method, ADNet, exhibits superior accuracy in comparison to alternative approaches, attaining 98.29% test accuracy for the dataset divided into 80% training and 20% testing with the holdout method in experiment 1. It is noteworthy that while the ADNet, when used as a transfer learning approach, achieved a slightly lower accuracy of 96.11%, it still outperforms many of the existing models in the literature. This illustrates the robustness and adaptability of the ADNet architectural design, thereby establishing its potential as a valuable tool for applications beyond the OASIS dataset.

Furthermore, the results obtained from experiment 2 demonstrated that the selection of MRI sections had a considerable influence on the performance of the models. The use of a quarter (1/4) of the data provided resulted in the highest accuracy compared to other slicing strategies, thereby underscoring the significance of data preprocessing techniques. The application of transfer learning resulted in a notable enhancement in the performance of the data with relatively lower accuracy, from 55.56% to 96.11%. This effectively exploited the information from a model that had demonstrated efficacy on a subset of disparate data sets to enhance classification in a novel data set. In addition to the notable performance of the developed convolutional neural network (CNN) model, it is essential to emphasize the originality of our transfer learning approach. The self-training version has been derived from a subset of the OASIS dataset. This data-driven adaptation further optimized the model's performance by ensuring that the information conveyed was consistent with the characteristics and distribution of the target dataset. The introduction

of the self-training model removed the necessity for reliance on pre-trained external models, which may no longer be readily available or optimized for the specific task at hand. This approach encourages self-sufficiency and enables researchers to adapt transfer learning to suit specific datasets and research objectives.

Table 5. Comparison with the State-of-the-Art on OASIS-1 Dataset

Reference	Name of Best Method Used	Data Count	Type of Data Used	Class Labels (number of classes)	Sampling Method (Train(%): Test(%) for Holdout)	Test Accuracy (%)
(Shrivastava et al., 2023)	Random Forest	136	Measures Included in the OASIS dataset	Nondemented, demented (2)	75:25 (Holdout)	84
(Jadhao et al., 2023)	Extra Tree Classifier	373	Measures Included in the OASIS dataset	Nondemented, moderate dementia, demented (3)	85:15 (Holdout)	85,71
(Baglat et al., 2020)	Random forest classifier	150	Measures Included in the OASIS dataset	Nondemented, demented (2)	5-fold cross validation	86.84
(Salami, Bozorgi-Amiri, Hassan, Tavakkoli-Moghaddam, & Datta, 2022)	Inception-v3 model	1094	MRI Data	CN (Cognitively Normal), AD (Alzheimer's Disease)(2)	80:20 (Holdout)	87.75
(Mendoza-Leon, Puentes, Felipe, & Hern, 2020)	Supervised Autoencoder : SSAs	174	MRI Data	Healthy, AD-demented (2)	80:20 (Holdout)	90
(Neffati, Ben Abdellafou, Jaffel, Taouali, & Bouzrara, 2019)	DKPCA + MKSVM	198	MRI Data	No dementia, Very mild AD, Mild AD, Moderate AD(4)	5-fold cross validation	92.50
(Saratxaga et al., 2021)	Deep Learning	1114	MRI Data	Nondemented, moderate dementia, demented (3)	70:30 (Holdout)	93.18
(Mohammed et al., 2021)	AlexNet+SVM hybrid model	6400	MRI Data	Mild dementia, Moderate dementia, Non-dementia, Very mild dementia (4)	80:20 (Holdout)	94.80
(Rajayyan & Mustafa, 2023)	Gaussian Naïve Bayes	373	Measures Included in the OASIS	Nondemented, moderate dementia, demented (3)	80:20 (Holdout)	95
(Ghosh et al., 2023)	MobileNet	436	MRI Data	Nondemented, demented (2)	80:20 (Holdout)	95.24
(Chui et al., 2022)	GAN-CNN-TL	2168	MRI Data	Non-demented, Very Mildly Demented, Mildly Demented, Moderately Demented (4)	5-fold cross validation	96.8
Proposed Method (Exp.2)	*ADNet using as a Transfer Learning*	*905*	*MRI Data*	*Non Demented, Initial Phase Dementia, Dementia Patient(3)*	*80:20 (Holdout)*	*96.11*
Proposed Method (Exp. 3)	*ADNet*	*905*	*MRI Data*	*Non Demented, Initial Phase Dementia, Dementia Patient(3)*	*5-fold cross validation*	***97.05***
Proposed Method (Exp. 1)	*ADNet*	*905*	*MRI Data*	*Non Demented, Initial Phase Dementia, Dementia Patient(3)*	*80:20 (Holdout)*	***98.29***

This considerable improvement demonstrates ADNet's capability to enhance model performance in cases with limited training data or suboptimal initial performance. The robustness of ADNet was also evaluated through a comprehensive testing process, employing 5-fold cross-validation repeated three times in experiment 3. This evaluation produced a mean test set accuracy of 97.05%, indicating that ADNet is a reliable and stable model across varying subsets of data. Furthermore, the results of the ADNet demonstrate higher performance than those of the previously leading method with 5-fold cross validation, GAN-CNN-TL, which achieved an accuracy of 96.8%.

Despite the encouraging outcomes, it is acknowledged that the methodology employed in this study is subject to certain constraints. Firstly, the relatively limited size of the dataset may restrict the model's generalizability. Further

research is required with a larger and more diverse dataset to confirm the model's effectiveness in real-world clinical settings. Secondly, the study focused on three diagnostic categories. Subsequent studies may wish to incorporate supplementary phases of Alzheimer's disease (AD) to achieve a more comprehensive analysis. Furthermore, an investigation of alternative convolutional neural network (CNN) architectures and hyperparameter settings has the potential to enhance the model's performance.

CONCLUSIONS

Alzheimer's disease (AD) is a major global health problem characterized by progressive cognitive decline and functional impairment. Early and accurate diagnosis is of paramount importance for the timely implementation of appropriate interventions that may improve patient outcomes. This study investigates the effectiveness of a convolutional neural network based deep learning model for diagnosing Alzheimer's disease using the OASIS-1 dataset, a widely used neuroimaging dataset.

In this study, three distinct datasets were constructed to identify the most informative slice among the 128 sagittal slices present in each MRI. Slices were selected from the first, middle, and third quarters of the sagittal plane to identify which region held the most valuable information for the diagnosis of Alzheimer's disease. The datasets were divided into two distinct sets: 80% for training and 20% for testing. This allowed for a comprehensive evaluation of ADNet's performance. The results demonstrated that the slice from the first quarter exhibited the highest accuracy, achieving a score of 98.26% on the test set. This finding indicates that slices from the initial sagittal region contain significant information that ADNet effectively utilizes for accurate diagnosis. Moreover, the study investigated the potential of ADNet as a transfer learning model. A model with initially low performance, achieving only 55.56% accuracy, was retrained using ADNet as a transfer learning foundation, resulting in a significantly improved accuracy of 96.11%. This considerable improvement demonstrates ADNet's capability to enhance model performance in cases with limited training data or suboptimal initial performance. The robustness of ADNet was also evaluated through a comprehensive testing process, employing 5-fold cross-validation repeated three times. This evaluation produced a mean accuracy of 97.05%, indicating that ADNet is a reliable and stable model across varying subsets of data. Collectively, these findings suggest that ADNet has promising applications for clinical Alzheimer's screening and serves as a reliable transfer learning model, enabling more accurate and earlier diagnoses.

The ADNet model with transfer learning exhibits superior test accuracy in comparison to the majority of existing studies, particularly those that employ more straightforward machine learning algorithms. The efficacy of this methodology is particularly noteworthy considering the limited size of the dataset and the complexity of the three-class classification task. Notwithstanding the promising results, the study is aware of the limitations intrinsic to the research process. Further research is required with a larger and more diverse dataset to verify the generalizability of the model. Furthermore, incorporating additional AD stages and investigating sophisticated deep learning architectures may potentially result in additional improvements.

Consequently, this study has successfully optimized a convolutional neural network (CNN) model for the diagnosis of Alzheimer's disease (AD) using the OASIS dataset. The model's state-of-the-art test accuracy, the effectiveness of transfer learning from the self-supervised learning model, and its robust performance demonstrate the potential of deep learning for early and accurate Alzheimer's disease (AD) detection. Further research is required to validate the findings with larger datasets, incorporate more AD stages, and explore advanced deep-learning architectures for more robust and reliable AD diagnostics.

ACKNOWLEDGEMENT

This study has been financially supported by the Coordinatorship of Scientific Research Projects of Necmettin Erbakan University [Project no: 23GAP19015].

REFERENCES

Afzal, S., Maqsood, M., Khan, U., Mehmood, I., Nawaz, H., Aadil, F., & Nam, Y. (2021). *Alzheimer Disease Detection Techniques and Methods : A Review*. *6*, 26–38. https://doi.org/10.9781/ijimai.2021.04.005

Alroobaea, R., & Bragazzi, N. L. (2021). *Alzheimer ' s Disease Early Detection Using Machine Learning Techniques*. 1–16.

Alzeimer's Association. (2023). 2023 Alzheimer's disease facts and figures. *Alzheimer's Dement., 19(4)*(February), 1598–1695. https://doi.org/10.1002/alz.13016

Avots, E., Jafari, A., Ozcinar, C., & Anbarjafari, G. (2024). Comparative efficacy of histogram-based local descriptors and CNNs in the MRI-based multidimensional feature space for the differential diagnosis of Alzheimer's disease: a computational neuroimaging approach. *Signal, Image and Video Processing*, *18*(1), 1–13. https://doi.org/10.1007/s11760-023-02942-z

Baglat, P., Salehi, A. W., Gupta, A., & Gupta, G. (2020). Multiple Machine Learning Models for Detection of Alzheimer's Disease Using OASIS Dataset. In S. K. Sharma, Y. K. Dwivedi, B. Metri, & N. P. Rana (Eds.), *Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation. TDIT 2020. IFIP Advances in Information and Communication Technology* (Vol. 617, pp. 614–622). https://doi.org/10.1007/978-3-030-64849-7_54

Balasundaram, A., Srinivasan, S., Prasad, A., Malik, J., & Kumar, A. (2023). Hippocampus Segmentation-Based Alzheimer's Disease Diagnosis and Classification of MRI Images. *Arabian Journal for Science and Engineering*. https://doi.org/10.1007/s13369-022-07538-2

Bendlin, B. B., Carlsson, C. M., Gleason, C. E., Johnson, S. C., Sodhi, A., Puglielli, L., … Wharton, W. (2011). Midlife predictors of Alzheimer's disease. *Maturitas*, *65*(2), 131–137. https://doi.org/10.1016/j.maturitas.2009.12.014.Midlife

Breijyeh, Z., & Karaman, R. (2020). Comprehensive Review on Alzheimer's Disease : Causes and Treatment. *Molecules*, *25*(5789), 1–28.

Chui, K. T., Gupta, B. B., Alhalabi, W., & Alzahrani, F. S. (2022). An MRI Scans-Based Alzheimer's Disease Detection via Convolutional Neural Network and Transfer Learning. *Diagnostics*, *12*(1531), 1–14.

Ghosh, T., Palash, M. I. A., Yousuf, M. A., Hamid, M. A., Monowar, M. M., & Alassafi, M. O. (2023). A Robust Distributed Deep Learning Approach to Detect Alzheimer's Disease from MRI Images. *Mathematics*, *11*(12), 2633. https://doi.org/10.3390/math11122633

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Garcia, S., … I, B. L. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*(December), 1–13. https://doi.org/10.3389/fnins.2013.00267

Hajamohideen, F., Shaffi, N., Mahmud, M., Subramanian, K., Al Sariri, A., Vimbi, V., & Abdesselam, A. (2023). Four-way classification of Alzheimer's disease using deep Siamese convolutional neural network with triplet-loss function. *Brain Informatics*, *10*(1). https://doi.org/10.1186/s40708-023-00184-w

Helaly, H. A., Badawy, M., & Haikal, A. Y. (2022). Deep Learning Approach for Early Detection of Alzheimer ' s Disease. *Cognitive Computation*, (September 2021), 1711–1727. https://doi.org/10.1007/s12559-021-09946-2

Jadhao, P., Palsodkar, P., Raut, R., Chaube, K., Rathod, D., & Palsodkar, P. (2023). Prediction of Early Stage Alzheimer ' s using Machine Learning Algorithm. *2023 4th International Conference for Emerging Technology (INCET)*, 1–5. https://doi.org/10.1109/INCET57972.2023.10170583

Khagi, B., & Kwon, G. R. (2019). CNN model performance analysis on MRI images of an OASIS dataset for distinction between healthy and Alzheimer's patients. *IEIE Transactions on Smart Processing and Computing*, *8*(4), 272–278. https://doi.org/10.5573/IEIESPC.2019.8.4.272

Krüger, F. (2016). Activity, Context, and Plan Recognition with Computational Causal Behaviour Models. *Faculty of Computer Science and Electrical Engineering, University of Rostock, Phd Thesis*, p:71-72.

Lu, B., Li, H., Chang, Z., Li, L., Chen, N., Zhu, Z., & Zhou, H. (2022). *A practical Alzheimer ' s disease classifier via brain imaging-based deep learning on 85 , 721 samples*.

Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., Data, C. M. R. I., … Buckner, R. L. (2007). Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young , Middle Aged , Nondemented , and Demented Older Adults Citation Open Access Series of Imaging Studies (OASIS): Nondemented , and Demented Older Adults. *Journal of Cognitive Neuroscience*, *19*(9), 1498–1507. https://doi.org/10.1162/jocn.2007.19.9.1498

Mendoza-Leon, R., Puentes, J., Felipe, L., & Hern, M. (2020). Single-slice Alzheimer ' s disease classification and disease regional analysis with Supervised Switching Autoencoders. *Computers in Biology and Medicine*, *116*(October 2019), 1–14. https://doi.org/10.1016/j.compbiomed.2019.103527

Mohammed, B. A., Senan, E. M., Rassem, T. H., Makbol, N. M., Alanazi, A. A., Al-Mekhlafi, Z. G., … Ghaleb, F. A. (2021). Multi-method analysis of medical records and mri images for early diagnosis of dementia and alzheimer's disease based on deep learning and hybrid methods. *Electronics (Switzerland)*, *10*(22). https://doi.org/10.3390/electronics10222860

Neffati, S., Ben Abdellafou, K., Jaffel, I., Taouali, O., & Bouzrara, K. (2019). An improved machine learning technique based on downsized KPCA for Alzheimer's disease classification. *International Journal of Imaging Systems and Technology*, *29*(2), 121–131. https://doi.org/10.1002/ima.22304

Ovsepian, S. V, Leary, V. B. O., Zaborszky, L., & Ntziachristos, V. (2019). *HHS Public Access*. *25*(4), 288–297. https://doi.org/10.1177/1073858418791128.Amyloid

Porsteinsson, A. P., Isaacson, R. S., Knox, S., Sabbagh, M. N., & Rubino, I. (2021). *Diagnosis of Early Alzheimer's Disease: Clinical Practice in 2021*. *3*(8), 371–386.

Rajayyan, S., & Mustafa, S. M. M. (2023). Prediction of dementia using machine learning model and performance improvement with cuckoo algorithm. *International Journal of Electrical and Computer Engineering*, *13*(4), 4623–4632. https://doi.org/10.11591/ijece.v13i4.pp4623-4632

Salami, F., Bozorgi-Amiri, A., Hassan, G. M., Tavakkoli-Moghaddam, R., & Datta, A. (2022). Designing a clinical decision support system for Alzheimer's diagnosis on OASIS-3 data set. *Biomedical Signal Processing and Control*, *74*(September 2021), 1–7. https://doi.org/10.1016/j.bspc.2022.103527

Salhi, S., Kora, Y., Ham, G., Zadeh, H., Id, H., & Simon, C. (2023). Network analysis of the human structural connectome including the brainstem. *PLoS ONE*, *18*(4), 1–20. https://doi.org/10.1371/journal.pone.0272688

Saratxaga, C. L., Moya, I., Picón, A., Acosta, M., Moreno-Fernandez-de-leceta, A., Garrote, E., & Bereciartua-Perez, A. (2021). Mri deep learning-based solution for alzheimer's disease prediction. *Journal of Personalized Medicine*, *11*(9). https://doi.org/10.3390/jpm11090902

Scheltens, P., Strooper, B. De, Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C. E., … Flier, W. M. Van Der. (2022). *Alzheimer ' s disease*. *397*(10284), 1577–1590. https://doi.org/10.1016/S0140-6736(20)32205-4.Alzheimer

Shrivastava, R. K., Singh, S. P., & Kaur, G. (2023). shrivastava.pdf. In D. Koundal, D. K. Jain, Y. Guo, A. S. Ashour, & A. Zaguia (Eds.), *Data Analysis for Neurodegenerative Disorders. Cognitive Technologies.* (pp. 111–126). https://doi.org/10.1007/978-981-99-2154-6_6