BALKAN JOURNAL OF ELECTRICAL & COMPUTER ENGINEERING, Vol. 13, No. 2, June 2025
Research Article

Alzheimer’s Disease Diagnosis in MRI Images Using Transfer Learning Methods: Evaluation of Different Model Performances

Eren Gunduzvar, Abdulsamet Kayik and Mehmet Ali Altuncu

Abstract— **The most common type of dementia in older adults is Alzheimer's disease. Currently, there is no known cure for this illness. The progression of the disease can lead to loss of cognitive and physical abilities. In addition, the process of caring for patients causes both economic and psychological difficulties for their relatives. Therefore, early detection of Alzheimer's disease is vital. With an early diagnosis, patients' quality of life can be improved, and the progression of the disease can be slowed. Many clinical methods are used in the diagnosis of Alzheimer's disease. One of the most preferred of these methods is Magnetic Resonance Imaging (MRI). This study compares the performance of transfer learning-based Convolutional Neural Network (CNN) models, including VGG-19, ResNet-50, DenseNet-201, and InceptionV3. These models are used to classify Alzheimer's disease into four stages: Non Demented, Very Mild Demented, Mild Demented, and Moderate Demented. To compare the performance of the models, accuracy, precision, sensitivity, F1-score, and area under the curve (AUC) metrics were measured. A publicly available dataset of MRI images was used in the study. SMOTE (Synthetic Minority Over-sampling Technique) was applied to overcome the class imbalance in the dataset. Experimental results show that transfer learning-based CNN models are effective in classifying Alzheimer's stages. In particular, the DenseNet-201 model outperformed the other models with an accuracy of 96.52%.**

Index Terms— **Alzheimer, CNN architectures, MRI imaging, Transfer learning.**

I. INTRODUCTION

DEMENTIA, A type of disease caused by impairment in cognitive functions, leads to problems in skills such as memory formation, speech, thinking, judgment and behavior. According to the World Alzheimer Report 2023, the number of dementia cases, which reached 55 million in 2020, is estimated to reach 139 million in the 2050s as societies age. According to the same report, the cost of treating dementia patients, which was USD 1.3 trillion per year in 2019, is projected to increase to USD 2.8 trillion in 2030 [1].

In Alzheimer's disease (AD), symptoms develop gradually but progressively worsen, particularly affecting those over the age of 65. Since there is no definitive cure, it has become one of the most important diseases for the elderly. AD patients are in need of care, and caring for them is both costly and psychologically burdensome for families. Therefore, early diagnosis is of great importance to prevent the disease from progressing [2, 7]. Moreover, early diagnosis can prolong the survival of AD patients by an average of 3 years [3].

Depending on the level of brain damage and the patient's condition, AD is usually classified into four stages. These stages are defined to reflect the progression of the disease and the severity of cognitive impairments. The first is Mild Cognitive Impairment (MCI), when symptoms begin to appear, and the second is Mild Alzheimer's, when memory loss and cognitive impairments become more pronounced. In the third stage, Moderate Alzheimer's, the person may have difficulty recognizing himself or herself and the people around them, and may struggle to perform complex tasks. In the last stage, Severe Disability, abilities for daily living are severely impaired, and patients need full-time care to meet their basic needs [8].

In clinical practice, various imaging techniques are employed to identify the stages of Alzheimer's disease; these techniques help in understanding the effects, structure, and function of the disease. These methods include Computed Tomography (CT), Positron Emission Tomography (PET), MRI, and ultrasonography [4]. Particularly in the early stages of Alzheimer's disease, MRI is the most commonly used imaging modality because it can capture small structural changes in different brain regions more clearly [5].

Detecting the first stage of Alzheimer’s disease, MCI, enables the implementation of necessary measures to prevent the disease from progressing to other stages. While clinical assessments and expert evaluations are necessary to prevent the progression of Alzheimer’s disease, symptoms often need to become quite pronounced for experts to accurately identify the stages. Therefore, the presence of automated assistive detection systems in conjunction with expert evaluations is of vital

Eren Gündüzvar is with the Department of Computer Engineering University of Kocaeli, Kocaeli, Turkey,(e-mail: gunduzvareren@gmail.com).
https://orcid.org/0009-0009-2375-9080

Abdulsamet Kayik is with the Department of Computer Engineering University of Kocaeli, Kocaeli, Turkey, (e-mail: sametkayik@gmail.com).
https://orcid.org/0009-0002-3212-8618

Mehmet Ali Altuncu is with the Department of Computer Engineering University of Kocaeli, Kocaeli, Turkey, (e-mail: mehmetali.altuncu@kocaeli.edu.tr).
https://orcid.org/0000-0002-2948-3937

Manuscript received Aug 20, 2023; accepted Feb, 17, 2025.
DOI: 10.17694/bajece.1535631

importance [6]. A number of sophisticated and successful machine learning techniques have been developed in recent research to classify Alzheimer's disease stages, such as from MRI scans.

This study assessed how well CNN-based transfer learning models performed when used to categorize MRI images into the four stages of Alzheimer's disease. The examined models underwent experiments, and the output of each model was contrasted with each other. The rest of this essay is structured as follows: In Section II, the methods, and studies that are currently being used to identify AD stages are thoroughly reviewed. Section IV provides a detailed presentation of the experimental outcomes, whereas Section III outlines the suggested technique. The paper is concluded in Section V, also addressing future research directions.

II. RELATED WORKS

Various methods have been proposed for the classification of AD stages. This section reviews studies that used MRI. A model for multi-class AD diagnosis based on Linear Discriminant Analysis (LDA) was proposed by Lin et al. [7]. The MR images were initially adjusted based on the patient's age. Using the least absolute shrinkage and selection operator (LASSO) approach, features were selected in the second stage, and Principal Component Analysis (PCA) was used to reduce dimensionality. Finally, a decision tree based on an Extreme Learning Machine (ELM) was employed to perform multi-class categorization. The AD Neuroimaging Initiative (ADNI) dataset was used for the studies, and the proposed model outperformed an approach that relied only on raw characteristics.

Acharya et al. [8] proposed transfer learning models for predicting AD stages. In this study, the proposed modified AlexNet architecture was compared with traditional CNN, VGG-16, and ResNet-50 models, and it was reported to achieve higher accuracy, F-score, Recall, and Precision on the Kaggle MRI dataset [9].

The stages of Alzheimer's disease were also categorized by Shamrat et al. [10] using transfer learning models. The ADNI dataset's image quality was first enhanced using a histogram equalization-based approach, and the dataset's class imbalance was eliminated using data augmentation approaches. After training five CNN models (VGG-16, MobileNetV2, AlexNet, ResNet-50, and InceptionV3), the InceptionV3 model achieved the highest accuracy. In the final phase of the study, the InceptionV3 model, which had the best accuracy, was modified, resulting in an improvement in accuracy beyond that of the classical approach.

A modified version of the VGG-16 model was presented by Mehmood et al. [11] to categorize AD patients into four stages. The present study used the Open Access Series of Imaging Studies (OASIS) dataset and applied data augmentation techniques. The suggested model, which has two modified VGG-16 layers operating in parallel, includes three batch normalization layers, three Gaussian noise layers, five max-pooling layers, and 14 convolutional layers. This approach resulted in a significant improvement in the accuracy of multi-class AD classification.

Using a portion of the ADNI dataset, Nawaz et al. [12] presented a unique CNN architecture for the classification of AD phases. The suggested model outperformed the conventional AlexNet and VGG-16 models in terms of accuracy, when tested on a dataset that was produced from a subset of the ADNI images.

Fu'adah et al. [13] attempted to detect AD stages using a CNN model based on AlexNet architecture. The dataset used in this study also comprised MRI scans. The study compared different learning rates using Adam optimization, and the best accuracy and loss parameters were obtained with a learning rate of 0.0001.

A CNN model was proposed by Ajagbe et al. [14] to categorize AD stages from MRI images. The accuracy area under curve (AUC), F1 score, precision, recall and computation time of the suggested model were evaluated using the Kaggle MRI dataset and compared with the results obtained with the VGG-16 and VGG-19 models. The proposed model demonstrated better performance in terms of the F1 score, recall, and computational time.

Rao et al. [15] applied machine learning techniques to classify the AD stages. The proposed method generally comprises data preprocessing, feature extraction and selection, and classification stages. Correlation Matrices and Exhaustive Feature Selection were employed for feature selection, and the Support Vector Machine (SVM) and Multilayer Perceptron (MLP) methods were used for classification. The proposed model was compared with the classical AlexNet method and yielded better results in terms of accuracy, F1 score, precision, and recall metrics.

Ramzan et al. [16] proposed a hybrid method that combines residual neural networks (RNN) and transfer learning (ResNet-18) to classify the six stages of AD. The proposed model was tested on the ADNI dataset and was reported to achieve better results than some previous studies.

Nawaz et al. [17] also proposed a hybrid method. In this study, the AlexNet model was used for feature extraction, and K-Nearest Neighbors (KNN), SVM, and Random Forest (RF) methods were used for classification. The model's performance was evaluated on the OASIS dataset, and the highest accuracy was achieved using the SVM method during classification.

Savaş [18] compared the performance of 29 different pretrained models for detecting AD stages using MRI images from the ADNI dataset. The performances of the models were evaluated using accuracy, precision, sensitivity, and specificity metrics, and the EfficientNet versions demonstrated the best performance in terms of these parameters.

Degadwala et al. [19] proposed another CNN-based model to better capture specific features in AD dataset images and enhance classification performance. The performance of the proposed model was compared with those of the AlexNet, VGG-16, and ResNet-50 models, and it was reported to have better performance than all of these methods.

In terms of accuracy and loss parameters, Mirchandani et al. [20] evaluated the performance of three CNN-based algorithms for multi-class classification of AD stages (four stages): AlexNet, Faster R-CNN, and YOLOv4. In this study, data augmentation techniques were applied to address the class imbalance in the Kaggle MRI dataset. The results demonstrate that AlexNet and YOLOv4 achieved the highest accuracy.

A hybrid CNN approach based on the ResNet-50 architecture was proposed by Yildirim and Çinar [21] for the classification of four distinct AD phases. The study used the Kaggle MRI dataset [3], and the performance of the proposed model was compared to that of the classical AlexNet, ResNet-50, DenseNet-201 and VGG-16 methods. It was reported that the developed hybrid model improved accuracy by 3% compared to traditional CNN architectures.

Esam and Mohammed [33] also proposed an original CNN architecture for the classification of four different AD stages. In the pre-processing stage of the study, images were converted to 150-x-150 pixels and data augmentation was applied using the SMOTE technique. The accuracy of the proposed model was compared with pre-trained CNN models. As a result of the evaluations, the original CNN model exhibited superior performance.

Arafa et al. [34] proposed a method consisting of data pre-processing, data augmentation, cross-validation and classification stages for the detection of AD. In the classification stage, they first performed AD detection with an original CNN architecture. In the second method, they tried to improve the performance of the model by testing the pre-trained CNN model VGG-16 with different optimizers. The results show that the original CNN architecture provides higher classification accuracy than the optimized VGG-16 model.

Singh and Kumar [35] compared the performance of seven pre-trained CNN models for the detection of six different stages of AD. In order to improve the images used in the study, methods such as image reorientation, shadow correction, and segmentation were applied. In the study, the performance of the models was evaluated with metrics such as accuracy, AUC, loss values, F1 score, and it was stated that the best performance was obtained with the EfficientNet0 model.

Khalid et al. [36] developed three different feature extraction methods for the detection of AD stages and compared their performance. In the first method, features of MRI images were separately obtained using the GoogLeNet and DenseNet-121 models were used. In the second method, features obtained using the same models were combined, and the dimensions of the features were reduced using the Principal Component Analysis (PCA) algorithm. In the third method, the features obtained with the GoogLeNet and DenseNet-121 models were combined with the hand-crafted features obtained with the Discrete Wavelet Transform (DWT), Local Binary Pattern (LBP) and Gray Level Equivalence Matrix (GLCM) methods. The features obtained with the three different methods were classified, using the Feed Forward Neural Network (FFNN). It was stated that the highest accuracy was obtained from FFNN classification by combining the features obtained with the DenseNet-121 model with the hand-crafted features.

III. METHODOLOGY

The proposed method for classifying the AD stages is illustrated in Fig. 1. The SMOTE technique was initially applied to address class imbalance in a publicly available dataset. The dataset was then divided into 80% training data and 20% test data. During the classification phase, four transfer learning methods—VGG-19, ResNet-50, InceptionV3, and DenseNet-201—were employed. These models were used to classify AD stages into four categories.

.

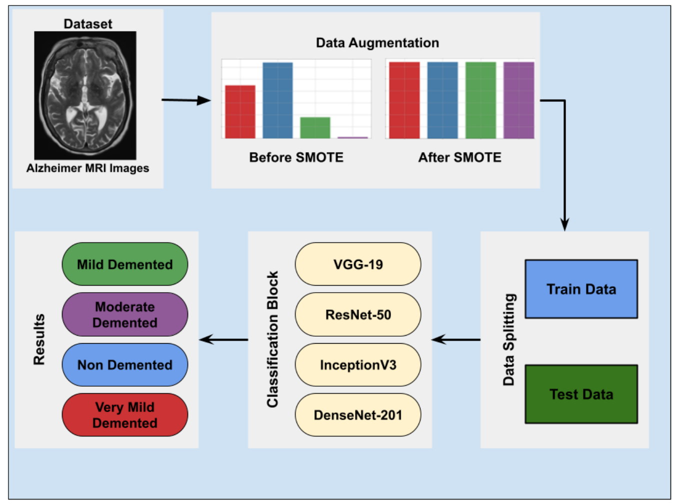

Fig.1. Architecture of the proposed system

A. VGG-19

The University of Oxford’s Visual Geometry Group (VGG) created the VGG-19 model, a CNN architecture trained on millions of images from the ImageNet collection (Fig. 2). Three fully connected layers and sixteen convolutional layers comprise VGG-19. Each of the five blocks that comprised the convolutional layers contained 3×3 filters. The max-pooling layer lowers the dimensionality and computing cost of the model, which follows each convolutional block. The nonlinear activation function is the rectified linear unit (ReLU). The last layer, which is a fully connected layer, performs classification using the softmax activation function [22, 23].

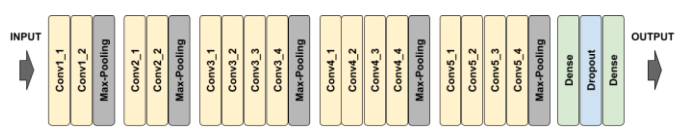

Fig.2. The architecture of the VGG-19 model [24]

B. ResNet-50

He et al. created the Residual Network (ResNet) architecture in 2015 to address the vanishing gradient problem, which occurs when deep networks have more layers [25]. The existence of structures known as "residual blocks" which appear every few layers is the primary way that ResNet differs from conventional CNN systems. With the help of skip connections, these blocks directly add input to the output layer, which improves the learning efficiency of the network. This method solves problems like disappearing gradients and makes training deeper networks easier. Depending on the number of layers, there are variations in the ResNet design with 50, 101, and 152 layers [26, 27]. These versions are frequently used in various applications, such as location, object identification, and image categorization. ResNet-50 is a model that was trained using millions of images from the ImageNet database, just like VGG-19. The 50 layers of the ResNet-50 architecture, which include pooling, activation functions, and convolutional layers, are shown in Fig. 3.

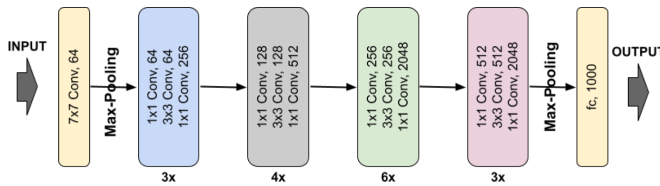

Fig.3. The architecture of the ResNet-50 model [28]

C. DenseNet-201

A variation of the DenseNet architecture called DenseNet-201 has dense blocks that comprise several convolutional layers connected directly to one another. 1x1 convolutions make up the Transition layers between Dense blocks, which lessen feature maps and maximize computational load [29]. Fig. 4 depicts the DenseNet-201 model's overall network structure.

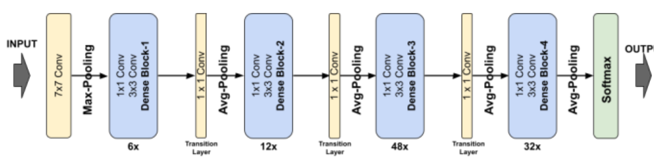

Fig.4. The architecture of the DenseNet-201 model [30]

D. InceptionV3

InceptionV3, an evolution of GoogLeNet (InceptionV1), offers significant improvements in terms of both performance and computational efficiency. This architecture features “Inception Blocks,” which enable the model to learn features at various scales by simultaneously performing convolutions and max-pooling operations of different sizes. Additionally, similar to the 1x1 convolutions used in the DenseNet-201 model, InceptionV3 employs 1x1 convolutions. These convolutions reduce the size of the feature maps, thereby reducing the computational load of the model and facilitating a faster learning process [31].

IV. EXPERIMENTAL RESULTS

In this study, the performance of popular transfer learning models, including VGG-19, ResNet-50, DenseNet-201, and InceptionV3, was analyzed for the classification of AD stages. The experiments were conducted using the TensorFlow framework, which is commonly used in deep learning applications. The dataset was divided into 80% training and 20% testing, for model training and evaluation. The training and testing data were randomly selected using Python's scikit-learn library. To ensure a fair comparison of model performance, the same data were used during both training and testing phases for each model.

A. Dataset Description

This study utilized the open-source Alzheimer’s Dataset (4 Classes of Images) [9], which is available on Kaggle. The dataset contains 6,400 MRI images in JPEG format, each with dimensions of 176x208 pixels. The dataset is organized into four classes and the distribution of each class is given in Table I.

TABLE I
DISTRIBUTION OF CLASSES IN THE KAGGLE MRI DATASET

Class Name	Number of samples
Mild Demented	896
Moderate Demented	64
Non Demented	3200
Very Mild Demented	2240

As shown in Table I, the distribution among classes is imbalanced. To address this imbalance, the SMOTE technique [32] was applied to equalize the number of images in each class to 3200.

B. Evaluation metrics

Four distinct criteria were used to assess the performance of the proposed transfer learning models: accuracy, precision, sensitivity, and F1 score. Four basic parameters were used to calculate these metrics: False Positive (FP), True Negative (TN), True Positive (TP), and False Negative (FN). Equations (1-4) include the formulas required to compute the metrics. In addition, the loss value, which measures the difference between the model's predictions and actual values and the AUC (Area Under the Curve) metric indicating how well the model differentiates between classes at various thresholds, are reported for each transfer learning model individually.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = \frac{2 * Precision * Sensitivity}{Precision + Sensitivity} \quad (4)$$

The hyperparameters used during the training of VGG-19, ResNet-50, DenseNet-201, and InceptionV3 models are detailed in Table II. The Adam algorithm was preferred in model optimization; learning rate was fixed at 0.001; mini-batch size was determined to be 32; and the number of training iterations (epochs) was configured as 100.

TABLE II
HYPERPARAMETERS USED IN MODEL TRAINING

Hyperparameter Name	Value
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Maksimum Epoch	100

C. *Results obtained with the VGG-19 model*

The accuracy and loss function values of the VGG-19 model during the training process are shown in Figure 5. Upon examining the accuracy graph history in Figure 5, the training accuracy value increased steadily as the number of epochs increased, while validation accuracy fluctuated but stabilized towards the end of training. This shows that the model achieved convergence without overfitting. When the history of loss graph is examined, the training loss decreased over time and reached low levels, which showed that the model's errors are gradually decreasing.

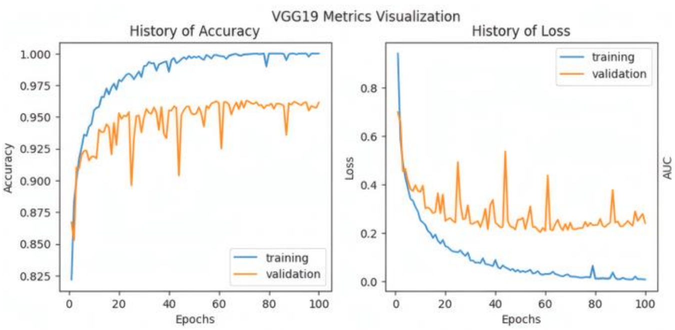

Fig.5. Accuracy and loss functions during VGG-19 training

The confusion matrix obtained using the test data after training the VGG-19 model is shown in Fig. 6.

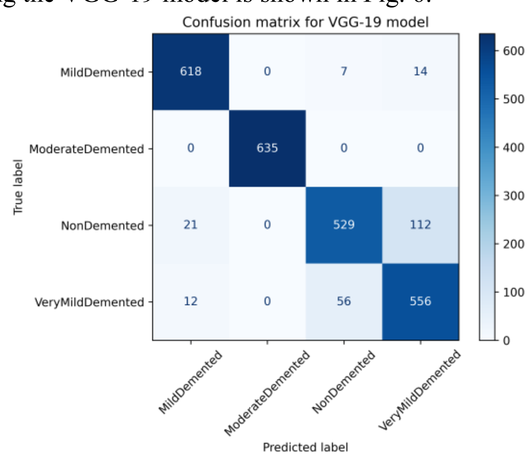

Fig.6. Confusion matrix for the VGG-19 model

With the VGG-19 model, 91.32% overall accuracy was obtained. Other performance metrics of the model are presented in Table III. According to the table, the 'Moderate Demented' class achieved the highest performance with 100% Precision, Sensitivity, and F1-Score. On the other hand, the 'Very Mild Demented' class showed the lowest performance with Precision (81.52%), Sensitivity (89.1%), and F1-Score (85.15%). This shows that the model has greater difficulty in distinguishing the class.

TABLE III
EVALUATION METRICS FOR VGG-19 MODELS

Class	Precision (%)	Sensitivity (%)	F1-Score (%)
Mild Demented	94.93	96.71	95.81
Moderate Demented	100	100	100
Non Demented	89.36	79.91	84.37
Very Mild Demented	81.52	89.1	85.15
Weighted average	**91.48**	**91.33**	**91.29**

The ROC curve for the VGG-19 model is shown in Figure 7. When the ROC curve is examined, it is seen that the lowest performance among the four classes is in the 'Very Mild Demented' class with 0.97. However, the AUC value for the 'Moderate Demented' and 'Mild Demented' classes was calculated as 1.0.

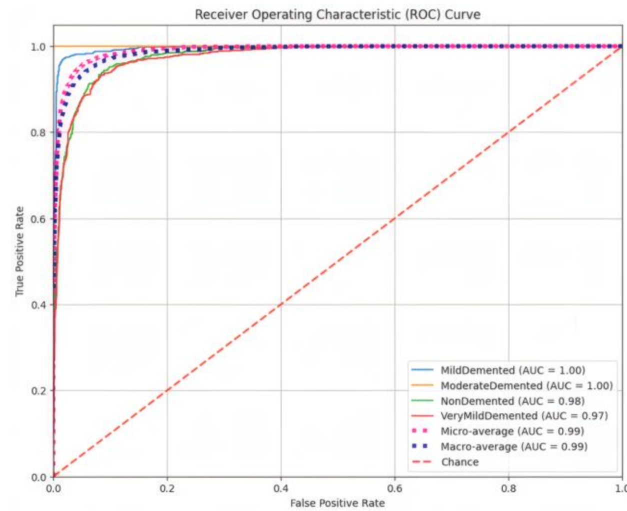

Fig.7. ROC curve of the VGG-19 model

D. *Results obtained with the ResNet-50 model*

The accuracy and loss function values of the ResNet-50 model during the training process are shown in Figure 8. When the 'history of accuracy' graph in Figure 8 is examined, it shows that the model generalized successfully on the validation set and that there was no overfitting. The history of loss graph shows that the validation loss is lower than the training loss.

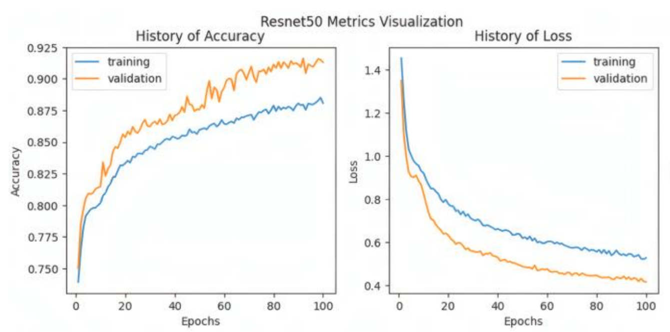

Fig.8. Accuracy and loss functions during ResNet-50 training

The confusion matrix obtained using the test data after training the ResNet-50 model is shown in Fig. 9.

With the ResNet-50 model, 80.62% overall accuracy was obtained. Other performance metrics of the model are presented in Table IV. According to Table IV, the 'Moderate Demented' class showed the highest performance with 99.84% Precision, 100% Sensitivity, and 99.92% F1-Score. On the other hand, the 'Very Mild Demented' class exhibited the lowest performance with Precision (61.35%), Sensitivity (62.82%) and F1-Score (62.07%). This shows that the model does not adequately distinguish the 'Very Mild Demented' class. In addition, although the Precision (71.18%) is low for the 'Non Demented' class, it was observed that the false negative rate was low because the Sensitivity value was high at 97.52%.

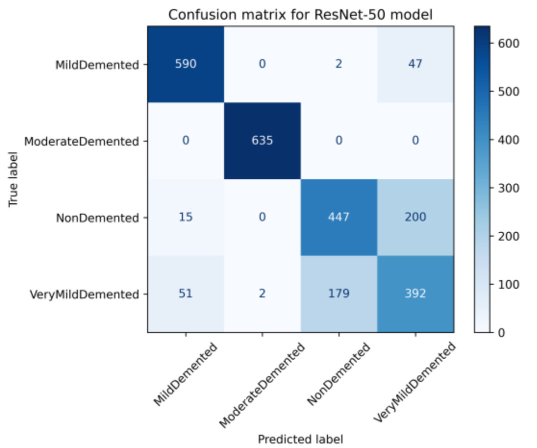

Fig.9. Confusion matrix for the ResNet-50 model

TABLE IV
EVALUATION METRICS FOR THE RESNET-50 MODEL

Class	Precision (%)	Sensitivity (%)	F1-Score (%)
Mild Demented	89.94	92.33	91.12
Moderate Demented	99.84	100	99.92
Non Demented	71.18	97.52	69.3
Very Mild Demented	61.35	62.82	62.07
Weighted average	**80.54**	**80.62**	**80.56**

The ROC curve for the ResNet-50 model is presented in Figure 10. When the ROC curve is examined, it is seen that the lowest performance among the four classes is in the 'Very Mild Demented' class with an AUC value of 88%. The AUC for the 'Moderate Demented' class is calculated as 100%. The results show that the ResNet-50 model needs improvements, especially in the 'VeryMildDemented' and 'NonDemented' classes.

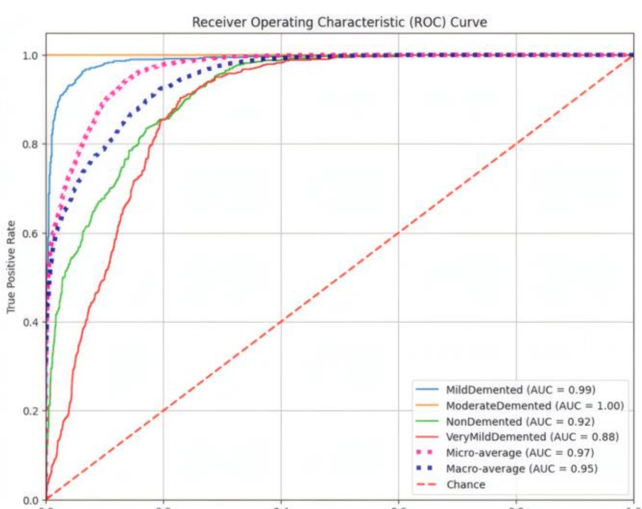

Fig.10. ROC curve of ResNet-50

E. *Results Obtained with the DenseNet-201 model*

The accuracy and loss function values of the DenseNet-201 model during the training process are shown in Figure 11. When the 'history of accuracy' graph in Figure 11 is examined, the difference between the model's training and validation accuracy is small, and stable convergence is achieved. The "history of loss" graph shows that the fluctuations in the validation loss are small and the model progresses stably during the training process.

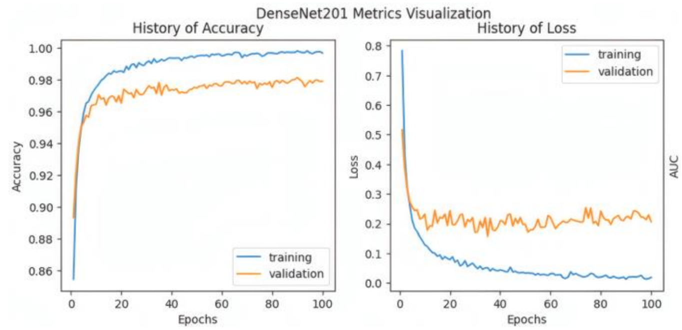

Fig.11. Accuracy and loss functions during the DenseNet-201 training

The confusion matrix obtained using the test data after training the DenseNet-201 model is shown in Fig. 12.

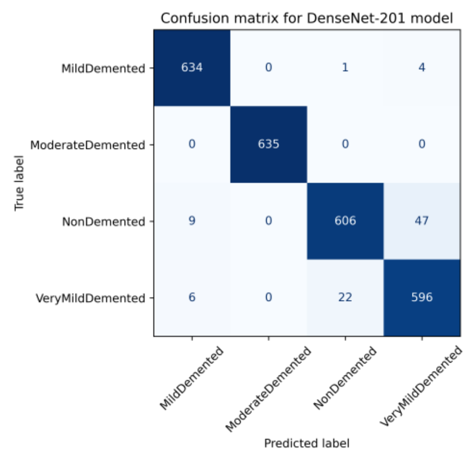

Fig.12. Confusion matrix for the DenseNet-201 model

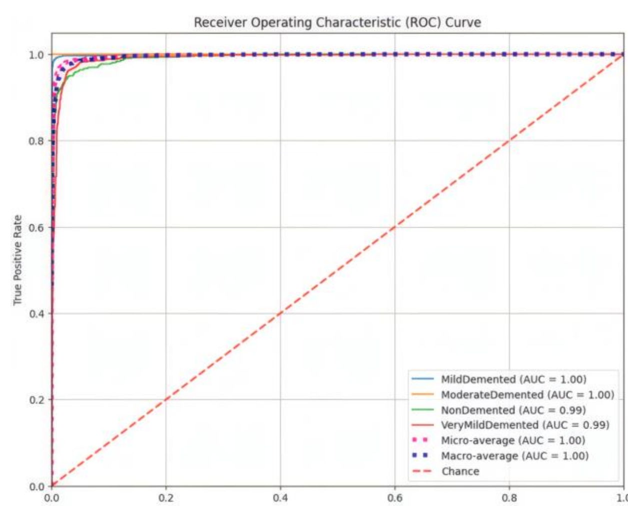

Fig.13. ROC curve of the DenseNet-201 model

With the DenseNet-201 model, 96.52% overall accuracy was achieved. Other performance criteria of the model are presented in Table V. When the table is examined, although there are low scores especially in the 'Non Demented' and 'Very Mild

Demented' classes, the overall model performance is highly balanced.

TABLE V
EVALUATION METRICS FOR THE DENSENET-201 MODEL

Class	Precision (%)	Sensitivity (%)	F1-Score (%)
Mild Demented	97.69	99.22	98.45
Moderate Demented	100	100	100
Non Demented	96.34	91.54	93.88
Very Mild Demented	92.12	95.51	93.78
Weighted average	**96.56**	**96.52**	**96.51**

The ROC curve for the DenseNet-201 model is shown in Figure 13. When the ROC curve is examined, it is seen that it can distinguish all classes quite well.

F. *Results Obtained with the InceptionV3 model*

The accuracy and loss function values of the InceptionV3 model during the training process are shown in Figure 11. As in the DenseNet-201 model, the difference between the training and validation accuracy of the InceptionV3 model is minimal, which reveals that the learning process of the model is completed efficiently and progresses without overfitting.

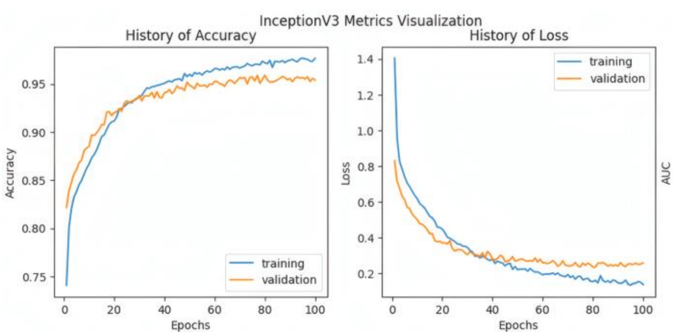

Fig.14. Accuracy and loss functions during the InceptionV3 training

The confusion matrix obtained using the test data after training the InceptionV3 model is shown in Fig. 15.

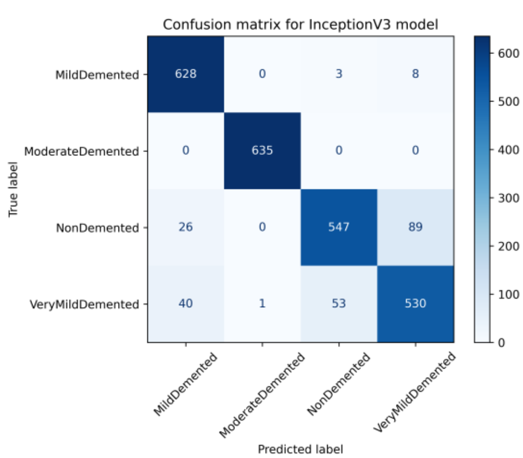

Fig.15. Confusion matrix for the InceptionV3 model

With the InceptionV3 model, 91.41% overall accuracy was achieved. Other performance criteria of the model are presented in Table VI. Upon examining the table, it is observed that although there are differences in the 'Non Demented' and 'Very Mild Demented' classes compared to other classes, the overall model performance is balanced with an F1-score of 91.32%.

TABLE VI
EVALUATION METRICS FOR INCEPTIONV3 MODELS

Class	Precision (%)	Sensitivity (%)	F1-Score (%)
Mild Demented	90.49	98.28	94.22
Moderate Demented	99.84	1.0	99.92
Non Demented	90.71	82.63	86.48
Very Mild Demented	84.53	84.94	84.73
Weighted average	**91.41**	**91.41**	**91.32**

The ROC curve for the InceptionV3 model is shown in Figure 16. When the ROC curve is examined, the AUC was 0.97 and above in all classes. These values show that the model is a strong classifier in distinguishing between all classes.

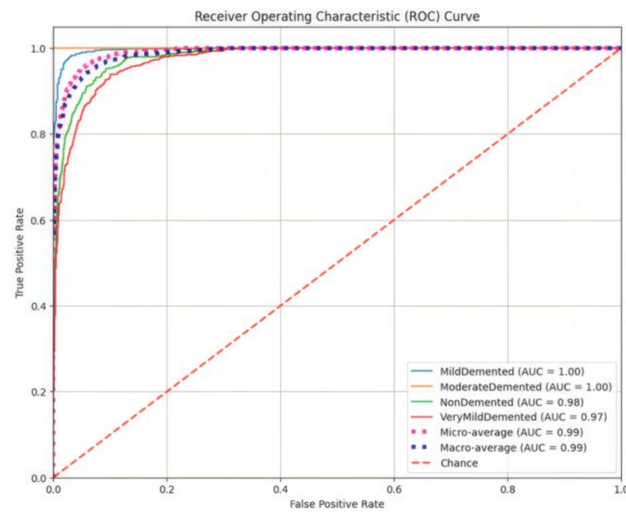

Fig.16. ROC curve of the InceptionV3 model

V. Conclusion

In this study, the performances of transfer learning-based CNN models are compared to classify four different stages of Alzheimer's disease via MRI images. In the experiments performed on VGG-19, ResNet-50, DenseNet-201, and InceptionV3 models, DenseNet-201 performed better than other models with Accuracy: 96.52%, Precision: 96.56%, Sensitivity: 96.52%, and F1-Score: 96.51%. However, the ResNet-50 model showed lower performance than DenseNet-201 with accuracy rates of 80.62%, InceptionV3 with 91.41%, and VGG-19 with 91.32%. The results show that transfer learning models can be used in the detection of AD stages. However, the errors observed in the early stages (Very Mild Demented) also suggest that the classification performance needs to be improved. In addition, the SMOTE technique was used in the study to eliminate the imbalance between the classes in the dataset. However, while SMOTE produces synthetic data, it can negatively affect the generalization performance of the model, as it cannot fully reflect the variations in real MRI images.

To eliminate these problems, it is aimed to use ensemble models created by combining different CNN architectures and to train these models on larger data sets. In addition, more realistic data augmentation can be achieved by GAN-based synthetic data generation instead of the SMOTE technique. It is anticipated that these approaches will provide more reliable and consistent results in the classification of the stages of Alzheimer's disease.

REFERENCES

[1] World Alzheimer Report 2023. Available: https://www.alzint.org/u/World-Alzheimer-Report-2023.pdf. Accessed 14 August 2024.

[2] F. Karakaya, C. Gurkan, A. Budak, and H. Karataş, "Classification and Segmentation of Alzheimer Disease in MRI Modality using the Deep Convolutional Neural Networks," Avrupa Bilim ve Teknoloji Dergisi, no. 40, pp. 99-105, 2022.

[3] M. Leela, K. Helenprabha, and L. Sharmila, "Prediction and classification of Alzheimer disease categories using integrated deep transfer learning approach," Measurement: Sensors, vol. 27, no. 100749, 2023.

[4] H. S. Suresha and S. S. Parthasarathy, "Alzheimer disease detection based on deep neural network with rectified Adam optimization technique using MRI analysis," in 2020 Third International Conference on Advances in Electronics, Computers and Communications (ICAECC), 2020, pp. 1-6.

[5] A. W. Salehi, P. Baglat, B. B. Sharma, G. Gupta, and A. Upadhya, "A CNN model: earlier diagnosis and classification of Alzheimer disease using MRI," in 2020 International Conference on Smart Electronics and Communication (ICOSEC), 2020, pp. 156-161.

[6] N. Raza, A. Naseer, M. Tamoor, and K. Zafar, "Alzheimer disease classification through transfer learning approach," Diagnostics, vol. 13, no. 4, p. 801, 2023.

[7] W. Lin, Q. Gao, M. Du, W. Chen, and T. Tong, "Multiclass diagnosis of stages of Alzheimer's disease using linear discriminant analysis scoring for multimodal data," Computers in Biology and Medicine, vol. 134, no. 104478, 2021.

[8] H. Acharya, R. Mehta, and D. K. Singh, "Alzheimer disease classification using transfer learning," in 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 1503-1508.

[9] Alzheimer's Dataset (4 class of Images). Available: https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images. Accessed 14 August 2024.

[10] F. J. M. Shamrat, S. Akter, S. Azam, A. Karim, P. Ghosh, Z. Tasnim, and K. Ahmed, "AlzheimerNet: An effective deep learning based proposition for alzheimer's disease stages classification from functional brain changes in magnetic resonance images," IEEE Access, vol. 11, pp. 16376-16395, 2023.

[11] A. Mehmood, M. Maqsood, M. Bashir, and Y. Shuyuan, "A deep Siamese convolution neural network for multi-class classification of Alzheimer disease," Brain Sciences, vol. 10, no. 2, p. 84, 2020.

[12] A. Nawaz, S. M. Anwar, R. Liaqat, J. Iqbal, U. Bagci, and M. Majid, "Deep convolutional neural network based classification of Alzheimer's disease using MRI data," in 2020 IEEE 23rd International Multitopic Conference (INMIC), 2020, pp. 1-6.

[13] Y. N. Fu'adah, I. Wijayanto, N. K. C. Pratiwi, F. F. Taliningsih, S. Rizal, and M. A. Pramudito, "Automated classification of Alzheimer's disease based on MRI image processing using convolutional neural network (CNN) with AlexNet architecture," in Journal of Physics: Conference Series, 2021, vol. 1844, no. 1, p. 012020.

[14] S. A. Ajagbe, K. A. Amuda, M. A. Oladipupo, F. A. Oluwaseyi, and K. I. Okesola, "Multi-classification of Alzheimer disease on magnetic resonance images (MRI) using deep convolutional neural network (DCNN) approaches," International Journal of Advanced Computer Research, vol. 11, no. 53, pp. 51, 2021.

[15] K. N. Rao, B. R. Gandhi, M. V. Rao, S. Javvadi, S. S. Vellela, and S. K. Basha, "Prediction and classification of Alzheimer's disease using machine learning techniques in 3D MR images," in 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), 2023, pp. 85-90.

[16] F. Ramzan, M. U. G. Khan, A. Rehmat, S. Iqbal, T. Saba, A. Rehman, and Z. Mehmood, "A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks," Journal of Medical Systems, vol. 44, pp. 1-16, 2020.

[17] H. Nawaz, M. Maqsood, S. Afzal, F. Aadil, I. Mehmood, and S. Rho, "A deep feature-based real-time system for Alzheimer disease stage detection," Multimedia Tools and Applications, vol. 80, pp. 35789-35807, 2021.

[18] S. Savaş, "Detecting the stages of Alzheimer's disease with pre-trained deep learning architectures," Arabian Journal for Science and Engineering, vol. 47, no. 2, pp. 2201-2218, 2022.

[19] S. Degadwala, D. Vyas, A. Jadeja, and D. D. Pandya, "Enhancing Alzheimer Stage Classification of MRI Images through Transfer Learning," in 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA), 2023, pp. 733-737.

[20] R. Mirchandani, C. Yoon, S. Prakash, A. Khaire, A. Naran, A. Nair, and S. Ganti. Comparing the Architecture and Performance of AlexNet Faster R-CNN and YOLOv4 in the Multiclass Classification of Alzheimer Brain MRI Scans. Available: https://ai-4-all.org/wp-content/uploads/2021/04/Comparing_the_Architecture_and_Performance_ofAlexNet__Faster_R_CNN__and_YOLOv4_in_theMulticlass_Classification_of_Alzheimer_Brain_MRIScans_Final.pdf. Accessed 14 August 2024.

[21] M. Yildirim and A. Cinar, "Classification of Alzheimer's Disease MRI Images with CNN Based Hybrid Method," Ingénierie des Systèmes d Inf., vol. 25, no. 4, pp. 413-418, 2020.

[22] A. Khattar and S. M. K. Quadri, "Generalization of convolutional network to domain adaptation network for classification of disaster images on twitter," Multimedia Tools and Applications, vol. 81, no. 21, pp. 30437-30464, 2022.

[23] V. Sudha and T. R. Ganeshbabu, "A Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Detection and Grading in Diabetic Retinopathy Based on Deep Learning," Computers, Materials & Continua, vol. 66, no. 1, 2021.

[24] J. Jaworek-Korjakowska, P. Kleczek, and M. Gorgon, "Melanoma thickness prediction based on convolutional neural network with VGG-19 model transfer learning," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019, pp. 0-0.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[26] M. Shafiq and Z. Gu, "Deep residual learning for image recognition: A survey," Applied Sciences, vol. 12, no. 18, p. 8972, 2022.

[27] Z. Qin, Q. Zeng, Y. Zong, and F. Xu, "Image inpainting based on deep learning: A review," Displays, vol. 69, p. 102028, 2021.

[28] S. Jahromi, M. N., P. Buch-Cardona, E. Avots, K. Nasrollahi, S. Escalera, T. B. Moeslund, and G. Anbarjafari, "Privacy-constrained biometric system for non-cooperative users," Entropy, vol. 21, no. 11, p. 1033, 2019.

[29] A. P. Syahputra, A. C. Siregar, and R. W. S. Insani, "Comparison of CNN models with transfer learning in the classification of insect pests," IJCCS (Indonesian Journal of Computing and Cybernetics Systems), vol. 17, no. 1, pp. 103-114, 2023.

[30] M. Bakr, S. Abdel-Gaber, M. Nasr, and M. Hazman, "DenseNet based model for plant diseases diagnosis," European Journal of Electrical Engineering and Computer Science, vol. 6, no. 5, pp. 1-9, 2022.

[31] S. Singh and R. Kumar, "Breast cancer detection from histopathology images with deep inception and residual blocks," Multimedia Tools and Applications, vol. 81, no. 4, pp. 5849-5865, 2022.

[32] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.

[33] S. Esam and A. Mohammed, "Alzheimer's disease classification for MRI images using Convolutional Neural Networks," in 2024 6th International Conference on Computing and Informatics, Mar. 2024, pp. 1-5.

[34] D. A. Arafa, H. E. D. Moustafa, H. A. Ali, A. M. Ali-Eldin, and S. F. Saraya, "A deep learning framework for early diagnosis of Alzheimer's disease on MRI images," Multimedia Tools and Applications, vol. 83, no. 2, pp. 3767-3799, 2024.

[35] A. Singh and R. Kumar, "Brain MRI image analysis for Alzheimer's disease (AD) prediction using deep learning approaches," SN Computer Science, vol. 5, no. 1, p. 160, 2024.

[36] A. Khalid, E. M. Senan, K. Al-Wagih, M. M. Ali Al-Azzam, and Z. M. Alkhraisha, "Automatic analysis of MRI images for early prediction of Alzheimer's disease stages based on hybrid features of CNN and handcrafted features," Diagnostics, vol. 13, no. 9, p. 1654, 2023.

BIOGRAPHIES

Eren GÜNDÜZVAR was born in Yalova, Turkey in 2001. He completed his B.S. in Computer Engineering at Kocaeli University in 2023. During his studies, he worked on projects and conducted research in areas such as Data Science, Full Stack Development, Mobile Development, Image Processing, and Artificial Intelligence. In 2022, he gained practical experience through internships at Döktaş and the Embedded Systems Laboratory at Kocaeli University, where he focused primarily on Full Stack Development. As of 2024, he is continuing his professional career as a FullStack Developer at DefineX Consulting, Technology, and Labs. His interests include web development, data science, image processing, and artificial intelligence.

Abdulsamet KAYIK was born in 2000 in Istanbul, Turkey. He completed his A.S. in Computer Programing at Marmara University in 2021 and began his studies in Computer Engineering at Kocaeli University in the same year. Throughout his academic career, he engaged with various aspects of software development.

In 2023, he gained practical experience through web development internships at Innova and Falla, where he developed frontend and backend development skills.

His interests include web development, image processing, and artificial intelligence.

Mehmet Ali ALTUNCU received his B.S. in Computer Engineering from the University of Sakarya in 2006, his M.S. in the same field from Kocaeli University in 2015, and his Ph. D. from Kocaeli University in 2021.

From 2013 to 2022, he worked as a Research Assistant at the Embedded Systems Laboratory at Kocaeli University. Since 2022, he has been serving as an Assistant Professor in the Department of Computer Engineering at Kocaeli University. His research interests include embedded systems, computer network security, pattern recognition, image processing, and machine learning.