Turkish Journal of Engineering
https://dergipark.org.tr/en/pub/tuje
e-ISSN 2587-1366

Deep learning based classification for alzheimer's disease detection using MRI images

Seda Nur Polater*[1], Onur Sevli [1]

[1]*Burdur Mehmet Akif Ersoy University, Faculty of Engineering and Architecture, Department of Computer Engineering, Türkiye, sedanurpolater7@gmail.com; onursevli@mehmetakif.edu.tr*

Cite this study: Polater, S. N., & Sevli, O. (2024). Deep Learning Based Classification for Alzheimer's Disease Detection Using MRI Images. Turkish Journal of Engineering, 8 (4), 729-740.

https://doi.org/10.31127/tuje.1434866

Keywords
Alzheimer's Diagnosis
Deep Learning
Classification

Research/Review Article

Received: 10.02.2024
Revised: 06.03.2024
Revised: 10.03.2024
Accepted: 11.03.2024
Published: 31.10.2024

Abstract

Alzheimer's disease is a common type of dementia that can cause serious problems in cognitive functions and activities of daily living. Although there is no definitive cure for Alzheimer's disease today, early diagnosis is important to slow down the adverse conditions that may arise and to improve the quality of life. As a result of the development of artificial intelligence technologies and their consistent application in different fields, machine learning techniques have the potential to play an important role in the detection of Alzheimer's disease. In particular, deep learning-based methods, which have the ability to automatically extract patterns from complex patterns, are promising in this field. Recent studies show that the use of deep learning models for Alzheimer's detection on images is becoming widespread. In addition to contributing to the early diagnosis of the disease, these models also show potential in detecting different stages of the disease by analyzing the symptoms in magnetic resonance images. These developments enable the development of more effective treatment methods for patients. However, more studies are needed to evaluate the efficacy and safety of these technologies in clinical applications. In this study, classification studies were performed using MobileNetV2, InceptionV3, Xception, Vgg16 and Vgg19 models for the diagnosis of the disease on a publicly shared Alzheimer's dataset consisting of 6400 different samples and 4 different classes. An accuracy of 99.92% was calculated for the MobileNetV2 model. The performances of the models used in this study were compared with similar studies in the literature and their performances were reported in terms of different metrics. Among the five different models used, the highest accuracy value of 99.92% was obtained with MobileNetV2. It was concluded that the architectures used in the experimental studies produced generally better results than similar studies in the literature.

1. Introduction

With aging, cognitive decline may occur in individuals and the likelihood of diseases such as Dementia increases. In individuals aged 65 and over, Alzheimer's Disease (AD) accounts for the majority of dementia cases. AD causes serious loss of cognitive abilities and disrupts daily life activities. Alzheimer's is a neurodegenerative disorder for which there is currently no definitive cure [1]. AD is characterized by cognitive and behavioral problems such as speech difficulties, memory problems, comprehension difficulties and attention deficit. As the disease progresses, it can make it difficult to maintain independent living. AD is divided into 4 stages: very mild dementia, mild dementia, moderate dementia and no dementia. The clinical diagnosis of AD is usually based on tests, which can be costly and difficult to administer. Structural and chemical changes in the brain are observed to determine the difference between the healthy brain and the brain with AD. Traditional diagnostic methods can pose various difficulties in practice and accurate diagnosis requires expertise [2]. Artificial intelligence techniques, which have gained popularity in recent years, offer successful solutions in clinical studies. Machine learning techniques, a sub-discipline of artificial intelligence, offer an alternative and successful approach for the detection and classification of AD on medical imaging data. Deep learning methods that make successful inferences on image data enable computer-aided early diagnosis and these systems can help experts in the field in disease detection [3].

Deep learning can be an effective tool for diagnosing the symptoms of Alzheimer's Disease (AD) and diagnosing it at an early stage. This method enables computers to understand the content of data and learn models by using the multi-layered structures of artificial neural networks inspired by the nervous systems of living things. The depth of the network refers to the number of layers it contains [4]. While classical artificial neural networks or convolutional neural networks usually consist of a few layers, deep neural networks can have hundreds of layers. Deep learning can be effective in diagnosing AD by using it to analyze biomarkers and different imaging techniques. These techniques examine large datasets to predict the symptoms of the disease and identify patterns in disease-related data. Research on Alzheimer's Disease shows that deep learning models are still an emerging field [5]. In recent years, doctors have been using brain Magnetic Resonance Imaging (MRI) data for early diagnosis of Alzheimer's Disease. Researchers have developed various computer-aided diagnostic systems for accurate disease detection. Rule-based expert systems were used for this purpose from the 1970s to the 1990s and supervised models from the 1990s onwards. However, the use of supervised systems usually requires intensive involvement of human experts, which is costly in terms of time, money and effort. Recently, the development of deep learning models has offered the possibility to extract features directly from images without the involvement of human experts. In this context, researchers have focused on using deep learning models for the diagnosis of Alzheimer's Disease [6]. Deep learning models have been successfully applied for different medical image analyses and have achieved significant results in various disease detection and classification in the fields of organ and substructure segmentation, pathology. However, so far, research on deep learning models for Alzheimer's Disease diagnosis is limited.

There are various studies in the literature on deep learning-based detection of FH. Aydın et al. [7] achieved an accuracy of 88% in their study using the 3D CNN method. In another study, Muhamed et al. [8] used ConvNext method and achieved 99.5% accuracy. In the study of Sharma et al. [9] using the Inception model, an accuracy of 94.92% was obtained. Zena et al. [10] obtained an accuracy of 97.625% in their study using the VGG16 model. In another study by Liu et al. [11], an accuracy of 96.25% was obtained with the Resnet+ method. In a study by Singh et al. [12] using CNN and Softmax model, 98.59% accuracy was achieved. Shu et al. [13] using supervised and unsupervised adversarial learning method achieved 92% accuracy. In the study of A.ER, S. Varma [14] using GLCM method, an accuracy of 75.71% was obtained. Islam and Zhang using Proposed ensembled model [15] and Resnet18 [16], achieved 93.18% and 86.03% accuracy respectively.

In this study, classification studies were carried out with MobileNetV2, InceptionV3, Xception, Vgg16 and Vgg19 models on a dataset consisting of brain MRI images containing 4 different classes and 6400 samples. This study is expected to be useful in diagnosing Alzheimer's disease, which is difficult and costly to diagnose early, and in determining the stages of the disease. This study aims to evaluate the effectiveness of these models in the diagnosis and stage determination processes of Alzheimer's disease.

2. Material and method

2.1. Dataset

In this study, a publicly available dataset was used of 6400 brain magnetic resonance (MR) images in JPEG format created by Dubey [17]. The dataset contains brain MRI images of four different case types: Non Demented (NOD), Very Mild Dementia (VMD), Mildly Demented (MID) and Moderate Dementia (MOD). Sample images of different class types in the dataset are given in Figure 1.

The number of samples in each class in the dataset is given in Table 1 and the proportional distribution of the classes in the dataset is given in Figure 2.

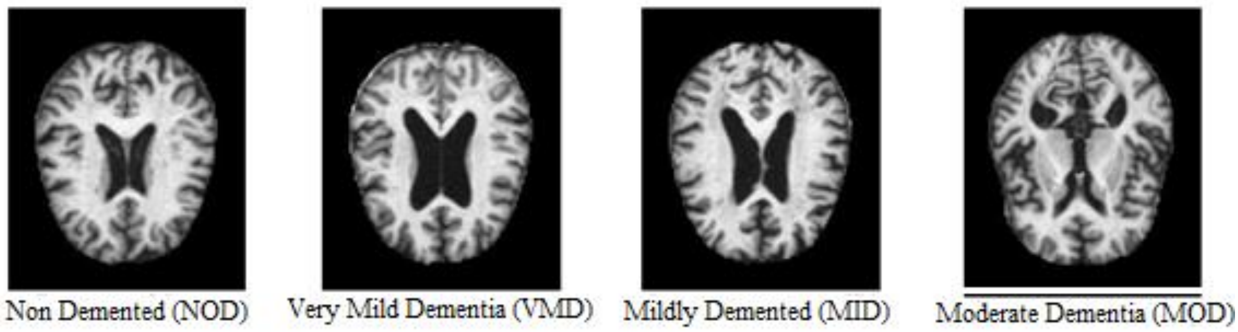


Figure 1. Example images of Alzheimer's dataset.

Table 1. Classes in the dataset and the number of samples in each class.

Class	Sample Count
Non Demented (NOD)	3200
Very Mild Dementia (VMD)	2240
Mildly Demented (MID)	896
Moderate Dementia (MOD)	64
Total number of samples	**6400**

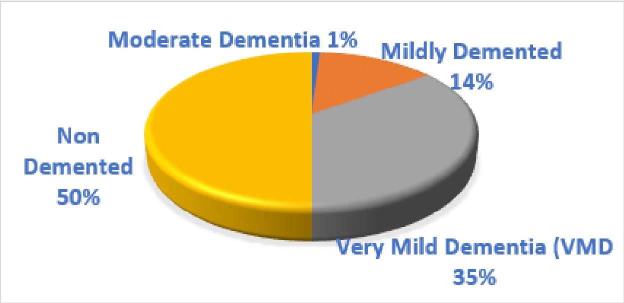


Figure 2. Class distributions of the dataset.

As can be seen in Figure 2, the dataset used has an unbalanced structure in terms of class distributions. The imbalance in the class distribution may cause the models used to learn the dominant class more while failing to learn the classes with fewer examples. To avoid this situation, alternative methods such as data augmentation are used to create class balance during classification.

2.2. Convolutional neural networks

Convolutional Neural Networks (CNNs) are a type of deep neural network that is widely used, especially in tasks focused on image analysis. CNN architecture is given in Figure 3. Although the proposed CNN-based method outperforms the compared methods and has promising results, it requires longer running time on the original images compared to a normal CNN, since each sub-band of the wavelet transform is fed into separate CNNs and the methodology is not parallelized [18].

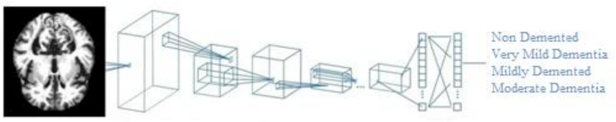


Figure 3. CNN Architecture.

As can be seen in Figure 3, CNN consists of a convolution layer, a pooling layer and an optional dropout layer. The convolution layer creates feature maps using convolution filters that transform the images. The stride determines how many steps the filter matrix is moved over the input pixels. In each layer, an appropriate activation function is used to improve the performance of the CNN. Pixel padding is applied after the convolution process to control the size difference between the input and output matrices and allows for more accurate analysis of the image. Pooling is an important component that allows CNN to perform dimension reduction on large images by reducing the number of parameters. It can be applied in three optional ways: maximum, minimum and averaging. Dropout is preferred in complex networks in order to avoid overfitting. Overlearning is when the network overfits to the data and memorizes it. The dropout layer temporarily turns off some nodes during training to prevent memorization. With CNN, the feature matrices obtained from the input image are flattened and transferred to the classification layer through the fully connected layer.

Depending on the problem to be addressed, CNN architectures are constructed. This process is usually done experimentally. In addition, some state-of-the-art models, which have shown superior performance in the ImageNet competition [19] on a huge dataset consisting of 1.2 million samples and 1000 classes, can exhibit high performance when used in different studies. In this study, MobileNetV2, InceptionV3, Xception, Vgg16 and Vgg19 architectures were used for classification.

2.2.1. MobileNetV2

MobileNetV1, proposed by Google in 2017, is a deeply separable neural network with low memory requirements and designed for use on mobile or embedded devices. MobileNetV2 is an improved version of MobileNetV1, built by introducing inverted residuals and bottlenecks. This version obtains more features by expanding the channel with convolution operations in 1x1 format and uses 3x3 depth convolution to obtain features. Point-to-point convolutions of size 1x1 are used to compress the channel numbers. The activation functions in the inverted residual structures not only accelerate learning but also increase the stability of the model. The architecture of the model consists of 53 layers and the input layer is 224x224x3 in size. The MobileNetV2 model offers a number of advantages compared to other models [20]. It is lightweight and fast, allowing it to operate effectively in resource-limited environments, and it adapts to memory constraints with low memory consumption. Moreover, its suitability for transfer learning and the variety of input resolutions make it adaptable to different application requirements. However, it also has some disadvantages. Less precision means that the model can operate with low accuracy compared to larger and more complex models. Difficulty in fine-tuning indicates that the model may be difficult to adapt to complex tasks. Furthermore, scope limitations indicate that MobileNetV2 may have performance limitations for large and complex tasks.

2.2.2. InceptionV3

InceptionV3 has a complex architecture developed by Google. It is preferred for large data sets and complex classification tasks. It contains parallel structures. Parallel structures that process multiple convolution filter sizes at the same time offer the ability to identify and combine attributes at different scales. It is predominantly used on larger datasets and requires higher computational power, making it generally preferred for large-scale tasks. The InceptionV3 model has significant advantages compared to other models [21]. Its high accuracy and low error rate, transfer learning capability and scalable architecture improve the overall performance of the model. However, disadvantages such as its high computational demand, complex architecture, memory consumption and long training time can limit its use and require careful management of hardware resources.

2.2.3. Xception

Xception is an improved version of Inception and takes the "Depthwise Separable Convolution" structure even further. It uses more deeply separated processing for convolution layers [22]. By focusing on processing each pixel separately, it achieves more efficient representation. It gives more effective results with fewer parameters and higher representation capability. It is generally preferred for visual recognition, classification and detailed feature extraction tasks. The Xception model has superior features compared to other models. High performance, lightweight and fast model with low parameter usage, transfer learning ability, adaptability and generalizability are among the advantages of the model. However, the model also has some weaknesses. Disadvantages such as high computational power requirement, long training time, recall to hardware limitations and difficulty in fine-tuning limit the use of the model.

2.2.4. Vgg16 ve Vgg19

VGG16 and VGG19 are models with the same architecture but different depths. They contain 16 and 19 layers respectively. Their structure is quite simple and is based on successive layers of convolution and pooling. Therefore, they have an understandable and simple structure. They offer an understandable structure for those who are new to working with deep networks. They are also common for transfer learning purposes because they are presented as pre-trained models on a general dataset. The VGG16 model is a deep learning model with strong advantages and drawbacks to be aware of. The advantages of this model include its simple and straightforward architecture, which has gained popularity among users. This feature provides a significant advantage in terms of understanding and making sense of the model's meaning. Furthermore, VGG16's efficient feature extraction capability enables it to successfully identify and extract features at different scales, which allows it to be widely used in various visual tasks. The transfer learning capability is an important advantage, as pre-trained models can be successfully adapted to other tasks. However, the disadvantages of the VGG16 model should not be overlooked [23]. The high number of parameters and the consequent weight of the model leads to a long training time and the need for more computational power. Memory consumption is also a major drawback and increases due to the large number of parameters and confusion. It also tends to show poor performance on tasks such as classifying small objects. The VGG19 model is a powerful deep learning model with a number of advantages and disadvantages [24]. The advantages of the model include high accuracy, transfer learning capability, efficient feature extraction and wide application. High accuracy emphasizes the model's ability to provide reliable results in classification tasks. Transfer learning capability means that pre-trained models can be successfully adapted to different tasks. Efficient feature extraction refers to the model's ability to successfully identify features at different scales. The wide range of applications emphasizes the ability of the VGG19 model to be used effectively in a variety of tasks. However, the VGG19 model also faces disadvantages. The computational power requirement increases due to the large number of parameters, which requires a more powerful computing infrastructure.

3. Experimental study and results

In this study, MobileNetV2, InceptionV3, Xception, Vgg16 and Vgg19 models were used to classify a dataset of 6400 brain MRI images in four different case types. Due to the unbalanced nature of the dataset, data augmentation was applied during the classification process. Data augmentation is a technique that increases the number of instances of the dataset and achieves positive results. At the same time, data augmentation helps to reduce overlearning by balancing the data distribution. Data augmentation aims to create new instances by applying various manipulations to the available data. These manipulations include adding noise to the image, rotating the image at different angles, tilting and bending the image. Table 2 shows the number of samples of the augmented data in the dataset after data augmentation.

Table 2. Classes in the augmented dataset and the number of instances in each class.

Class	Data Count
Non Demented (NOD)	9600
Very Mild Dementia (VMD)	8960
Mildly Demented (MID)	8960
Moderate Dementia (MOD)	6464
Dataset Total	**33.984**

The 5 different CNN architectures used during the feature extraction of the images used similar classification layers. A summary of the common classification layer used for each model is given in Table 3.

Table 3. The classification layer summary.

Layer Type	Output Shape	Parameters
Dense (Dense)	(None, 256)	2769152
Dense_1 (Dense)	(None, 128)	32896
Dropout Dropout)	(None, 128)	0
Dense_2 (Dense)	(None, 4)	516

Total params: 2,802,564

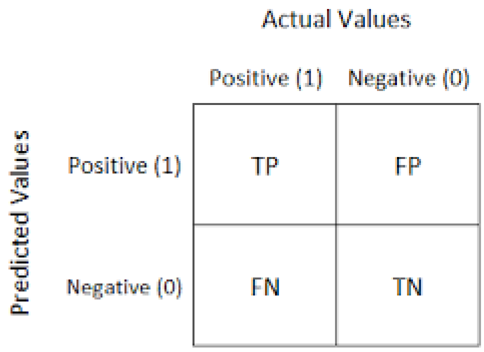


Figure 4. General structure of confusion matrix.

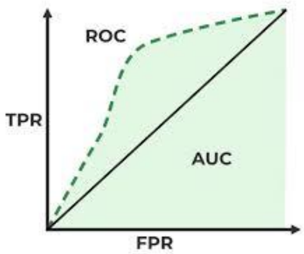
The confusion matrix is a matrix that evaluates the performance of a classification model. This matrix shows the model's true and false classifications, which allows to analyze the performance in detail. It basically contains four terms, namely True Positive, True Negative, False Positive and False Negative. These terms indicate which classes the model predicts correctly or incorrectly and help calculate performance metrics. A class that is actually positive in the confusion matrix is called True Positive (TP) if the model predicts it as positive, and False Negative (FN) if the model predicts it as negative. When a class that is actually negative is predicted positive by the model, it is called False Positive (FP), and when it is predicted negative by the model, it is called True Negative (TN). Other metrics calculated based on these values and their descriptions are given in Table 4.

Graphs showing the changes in accuracy and error values during model training also provide information about model performance. The accuracy graph is expected to show an increasing trend during the training process, while the error graph is expected to show a decreasing trend, indicating that the model is learning the data better and better.

The AUC change graph obtained as a result of the classification processes performed with the MobileNetV2 model is given in Figure 5.

The confusion matrix obtained in the test process after training the MobileNetV2 model is given in Figure 6.

Table 4. Performance Metrics.

Metric	Formula	Description
Precision	TP/(TP+FP)	It measures the proportion of correct positive predictions of a classification model out of the total positive predictions.
Recall	TP/(TP+FN)	It measures the success of a classification model in detecting true positives.
F1-Score	2 * (Precision * Recall)/(Precision+Recall)	It is a performance measure that represents a balanced combination of precision and recall metrics of a classification model.
Accuracy	(TP+TN)/(TP+TN +FP+FN)	It is a value that expresses the ratio of correct predictions of a classification model to the total number of samples.
ROC Curve and AUC		The ROC curve is a graph used to evaluate the performance of a classification model. AUC refers to the area under the ROC curve.

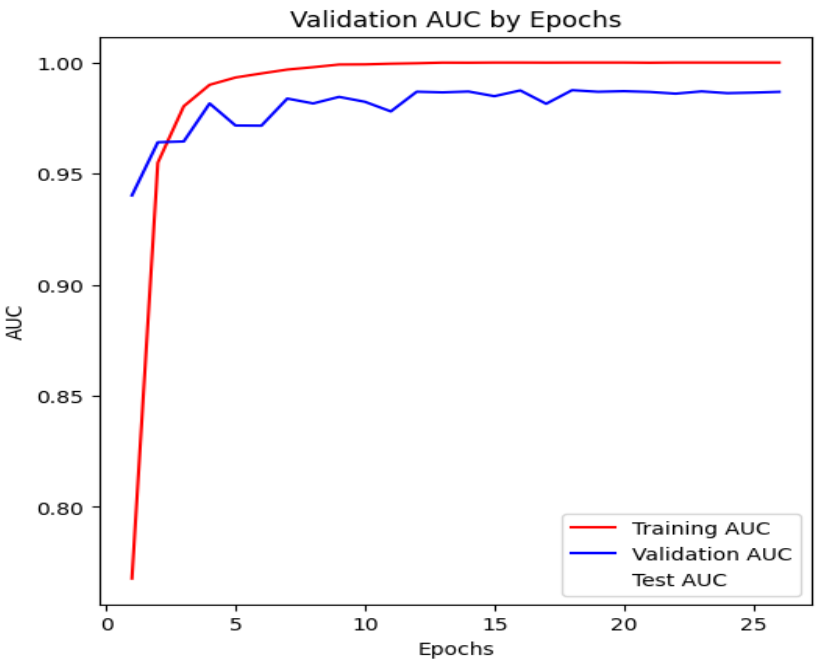


Figure 5. MobileNetV2 AUC graph.

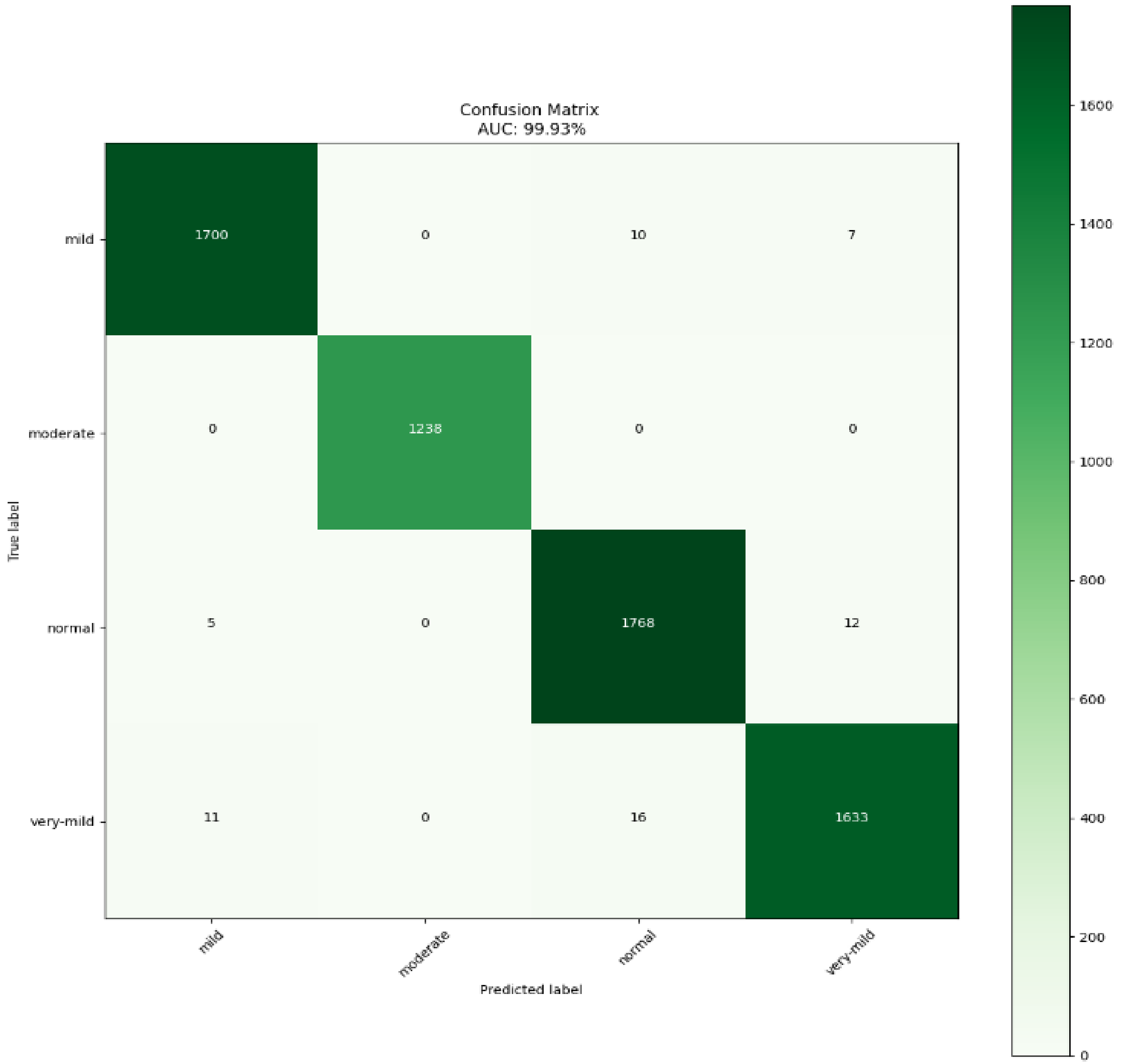


Figure 6. MobileNetV2 confusion matrix

The confusion matrix graph analysis of the MobileNetV2 model shows that it correctly classifies 1700, 1238, 1768 and 1633 samples in the mild, moderate, normal and very mild classes, respectively. The AUC value obtained as a result of these classifications was determined as 99.93%. The high AUC value obtained shows that the model distinguishes the classes effectively and the classification performance is quite high.

The performance metrics obtained based on the values in the confusion matrix are given in Table 5.

Table 5. MobileNetV2 Performance Metrics.

Metric	Obtained Values
Accuracy	0.9992
Precision	0.99
Recall	0.99
F1-Score	0.99
AUC	0.9993

In MobileNetV2 model, the accuracy value was 99.92%, precision value was 99%, recall value was 99%, f1-score value was 99% and AUC value was 99.93%.

The AUC change graph obtained as a result of the classification processes performed with the InceptionV3 model is given in Figure 7.

The confusion matrix obtained in the test process after training the InceptionV3 model is given in Figure 8.

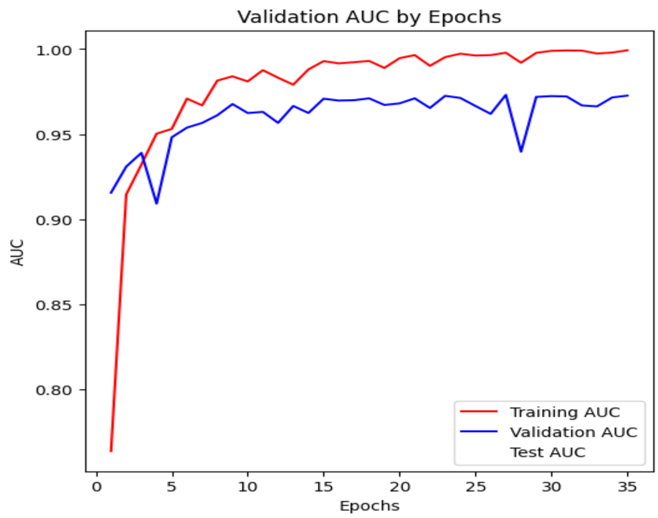


Figure 7. InceptionV3 AUC graph.

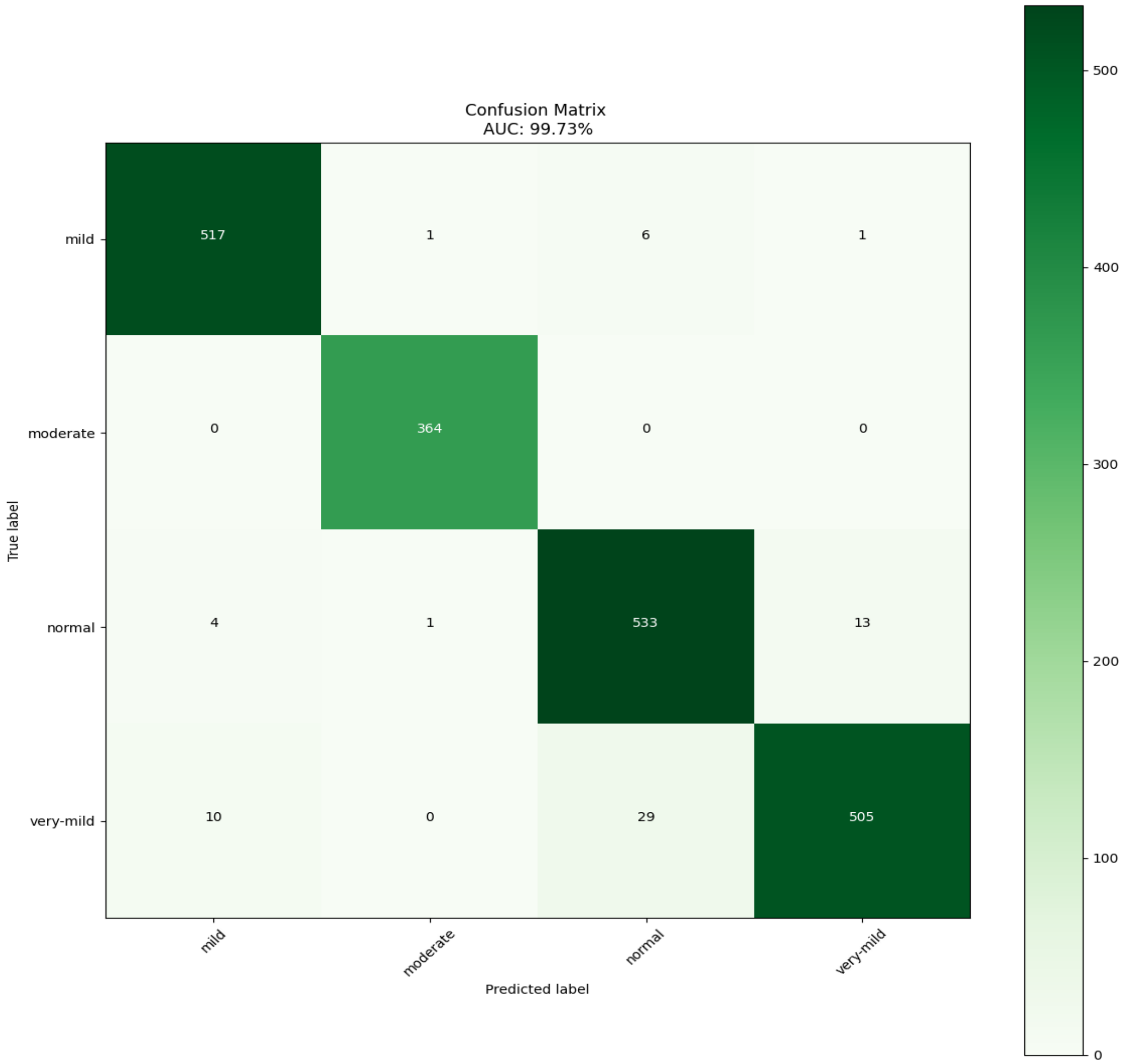


Figure 8. InceptionV3 confusion matrix.

The confusion matrix analysis of the InceptionV3 model shows that it correctly classifies 1517, 364, 533 and 505 samples in the mild, moderate, normal and very mild classes, respectively. The AUC value obtained as a result of these classifications was determined as 99.73%. The high accuracy and AUC of the InceptionV3 model indicate a strong classification capability.

The performance metrics derived from the values in the confusion matrix are given in Table 6.

Table 6. InceptionV3 Performance Metrics.

Metric	Obtained Values
Accuracy	0.9702
Precision	0.9675
Recall	0.97
F1-Score	0.97
AUC	0.9973

For the InceptionV3 model, accuracy was 97.02%, precision was 96.75%, recall was 97%, f1-score was 97% and AUC was 99.73%.

The AUC change graph obtained as a result of the classification processes performed with the Xception model is given in Figure 9.

The confusion matrix obtained in the test process after training the Xception model is given in Figure 10.

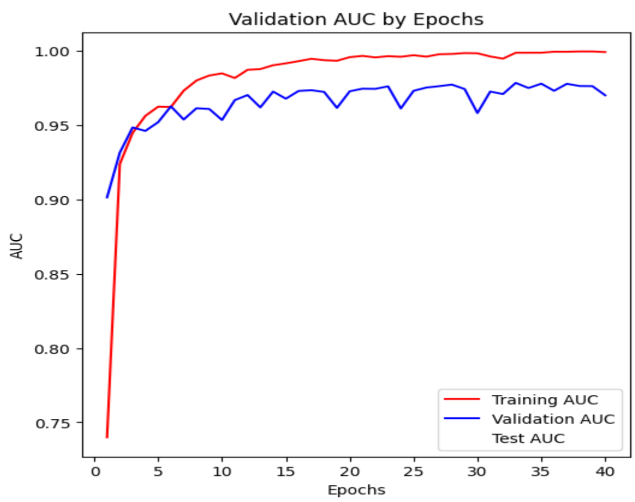


Figure 9. Xception AUC graph.

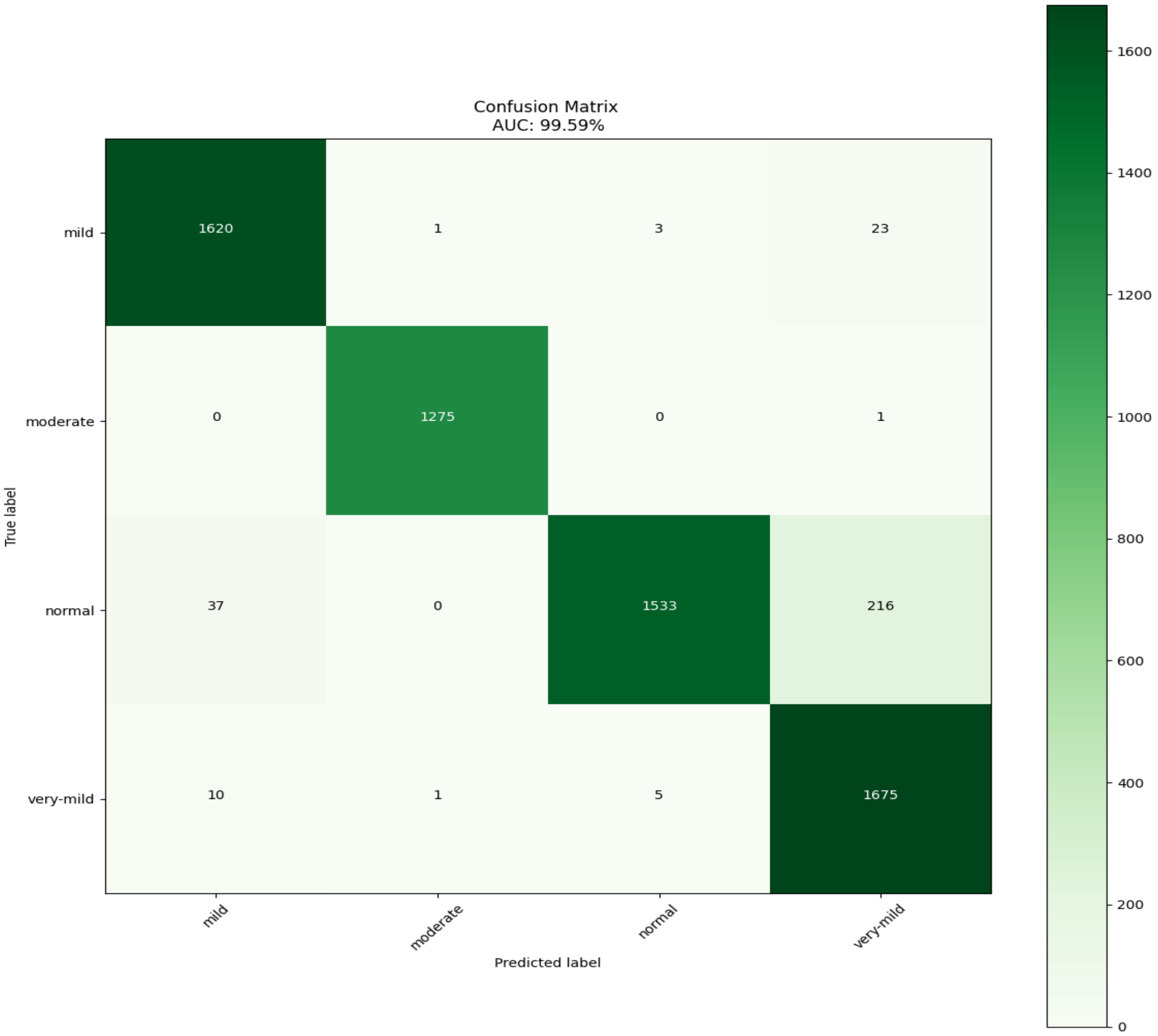


Figure 10. Xception confusion matrix.

The confusion graph analysis of the Xception model showed that it correctly classified 1620 instances in the mild class, 1275 in the moderate class, 1533 in the normal class and 1675 in the very mild class. The AUC value obtained as a result of these classifications was 99.59%. The high AUC value indicates that the model successfully distinguishes the classes and the classification performance is quite high. The number of correct classifications obtained in the mild, moderate, normal and very mild classes emphasizes the model's ability to successfully recognize conditions of different severity levels.

Table 7. Xception Performance Metrics.

Metric	Obtained Values
Accuracy	0.9515
Precision	0.9575
Recall	0.9575
F1-Score	0.9575
AUC	0.9959

In the Xception model, the accuracy value was 95.15%, precision value was 95.75%, recall value was 95.75%, f1-score value was 95.75% and AUC value was 99.59%.

The AUC change graph obtained as a result of the classification processes performed with the VGG16 model is given in Figure 11.

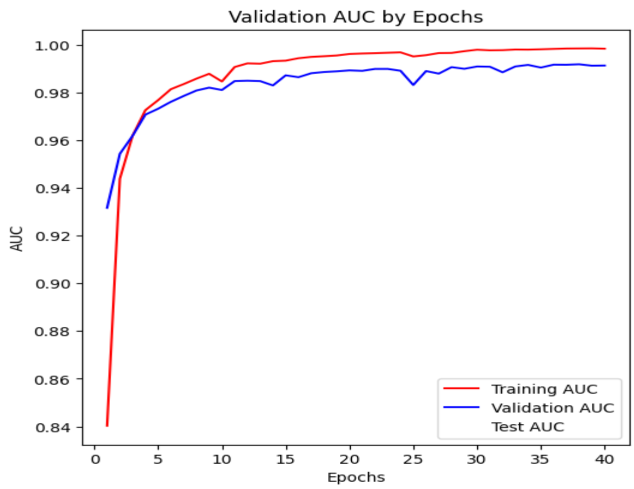


Figure 11. VGG16 AUC graph.

The confusion matrix obtained in the test process after training the VGG16 model is given in Figure 12.

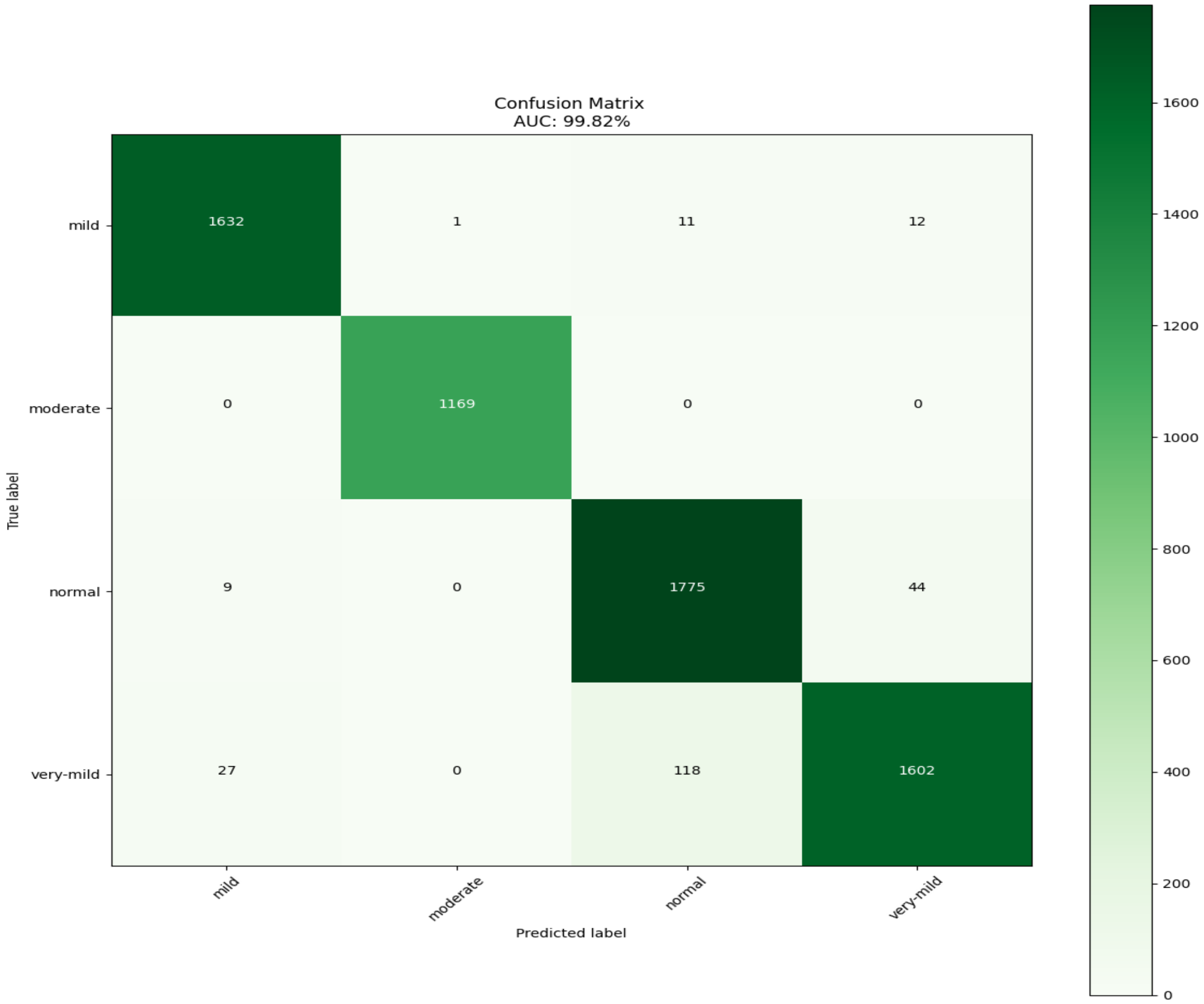


Figure 12. VGG16 confusion matrix.

When the confusion matrix of the VGG16 model is examined, correct classification was performed in 1632, 1169, 1775 and 1602 samples in the mild, moderate, normal and very mild classes, respectively. The AUC value obtained as a result of these classifications was determined as 99.82%. The high AUC value emphasizes that the model successfully distinguishes the classes and the classification performance is quite high. The number of correct classifications in the mild, moderate, normal and very mild classes demonstrates the model's ability to successfully recognize situations of different severity levels.

The performance metrics derived from the values in the confusion matrix are given in Table 8.

Table 8. VGG16 Performance Metric.

Metric	Obtained Values
Accuracy	0.9703
Precision	0.97
Recall	0.97
F1- Score	0.9675
AUC	0.9982

For the VGG16 model, accuracy was 97.03%, precision was 97%, recall was 97%, f1-score was 96.75% and AUC was 99.82%.

The validation curve showed that the training and cross validation scores are growing gradually, reflecting that model performed well [25]. The AUC change graph obtained as a result of the classification processes performed with the VGG19 model is given in Figure 13.

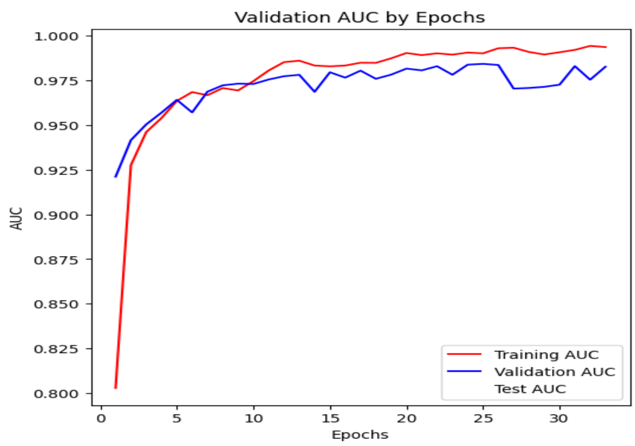


Figure 13. VGG19 AUC graph.

The confusion matrix obtained in the test process after training the VGG19 model is given in Figure 14.

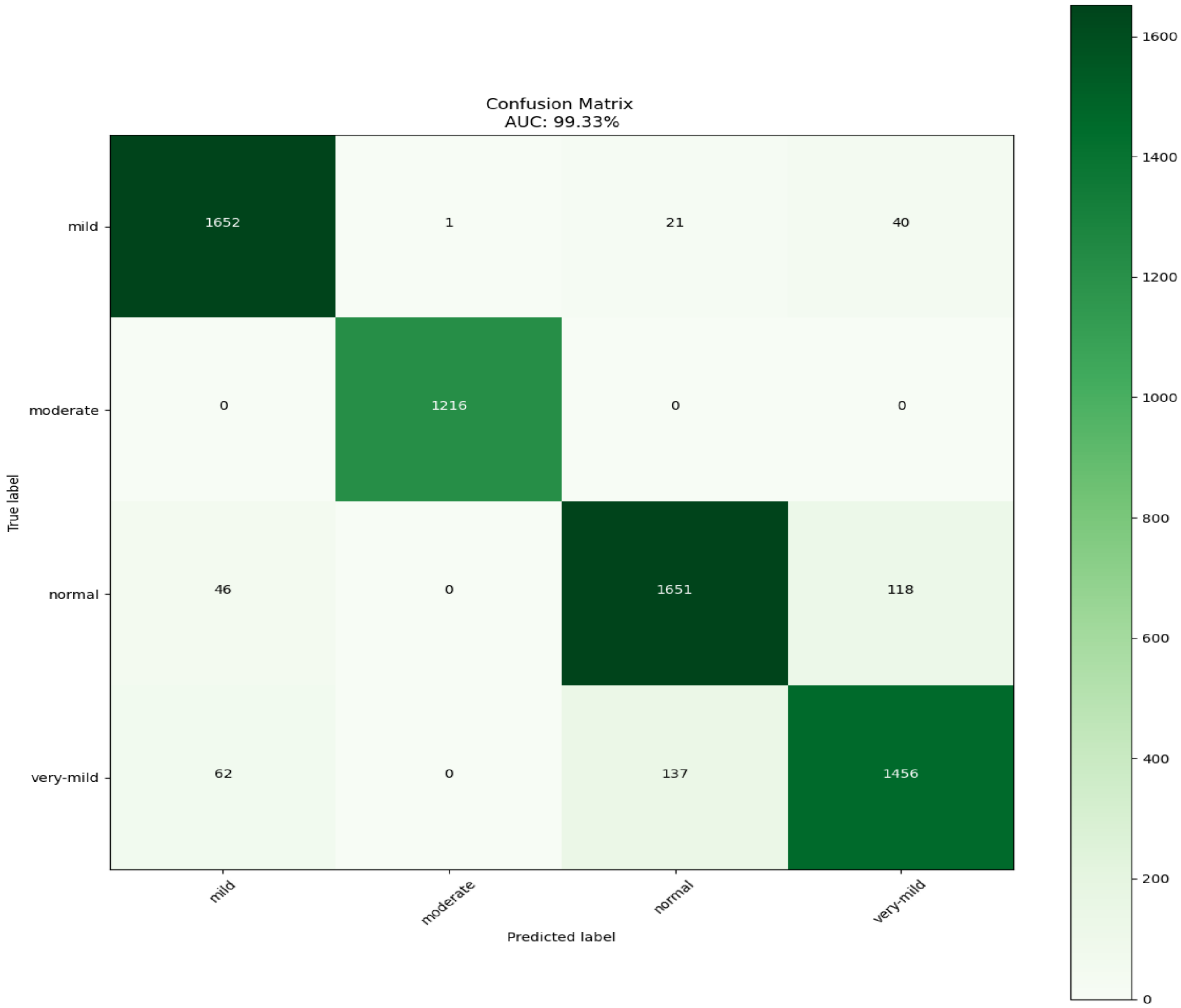


Figure 14. VGG19 confusion matrix.

When the confusion graph of the VGG19 model is examined, it is observed that it is correctly classified in 1652 instances in the light class, 1216 in the medium class, 1651 in the normal class and 1456 in the very light class. Moreover, the AUC value obtained as a result of these classifications was determined as 99.33%. The high AUC value and the number of correctly classified samples indicate that the model works successfully in general.

The performance metrics derived from the values in the confusion matrix are given in Table 9.

Table 9. VGG19 Performance Metric.

Metrik	Obtained Values
Accuracy	0.9311
Precision	0.9625
Recall	0.9625
F1-Score	0.9625
AUC	0.9933

For the VGG19 model, accuracy was 93.11%, precision was 96.25%, recall was 96.25%, f1-score was 96.25% and AUC was 99.33%.

The findings obtained with 5 different architectures used in classification are summarized in Table 10.

Table 10. VGG16 Performance Metric.

Model	Precision	Recall	F1-Score	Accuracy	AUC
MobileNetV2	0.99	0.99	0.99	0.9992	0.9993
InceptionV3	0.9675	0.97	0.97	0.9702	0.9973
Xception	0.9575	0.9575	0.9575	0.9515	0.9959
Vgg16	0.97	0.97	0.9675	0.9703	0.9982
Vgg19	0.9625	0.9625	0.9625	0.9311	0.9933

When the results in the table are evaluated, MobileNetV2 performs the best in the precision metric. It is followed by Vgg16, InceptionV3, Vgg19 and Xception models respectively. Regarding the recall metric, MobileNetV2 performs the best, followed by InceptionV3, Vgg16, Vgg19 and Xception. When evaluated on the F1-score metric, the MobileNetV2 model achieves the highest value, followed by InceptionV3, Vgg16, Vgg19 and Xception models. When analyzed according to the accuracy metric, the MobileNetV2 model shows the highest success, followed by Vgg16, InceptionV3, Xception and Vgg19 models. As a result of the evaluations made on all these metrics, it is seen that MobileNetV2 is the most successful model.

4. Discussion

Studies in the literature show that deep learning techniques have higher success rates than other traditional methods in diagnosing Alzheimer's disease, monitoring its course and understanding the factors associated with the disease. This shows that deep learning techniques offer significant potential for reducing the effects of the disease, improving treatment processes and better patient management.

Table 11. Comparison of literature search results.

References	Methods	Number of samples	Accuracy (%)
Aydın vd., 2023[7]	3D CNN	1439	88
Muhammed vd., 2023 [8]	ConvNext	6477	99,5
Sharma vd., 2022 [9]	Inception	6200	94,92
Zena vd., 2022 [10]	VGG16	-	97,625
Liu vd., 2022 [11]	ResNet+	2045	96,25
Singh vd., 2022 [12]	CNN+Softmax	-	98,59
Shu vd., 2018 [13]	Supervised and unsupervised adversarial learning	6400	92
A. ER, S. Varma, 2017 [14]	GLCM	-	75.71
Islam, J., & Zhang, Y., 2017 [15]	Proposed ensembled model	-	93.18
Islam, J., & Zhang, Y. 2017[16]	ResNet-18	2144	86.03
This study	MobileNetV2	33,984	99,92

According to the data presented in Table 11, this study highlights the 99.92% accuracy rate achieved with the MobileNetV2 architecture by evaluating previous works in the literature that address various classification problems. Among other references, prominent ones include the 99.5% accuracy of Muhammed et al and 98.59% accuracy of Singh et al. However, this study's high success rate using MobileNetV2 shows that an architecture optimized for mobile devices offers an effective solution to classification problems. Considering the generalization issues and lower accuracy rates in other references, the results of this study emphasize the superiority of MobileNetV2 in classification performance and suggest that it should be preferred in mobile applications.

5. Conclusion

This study examines deep learning models for the detection of Alzheimer's disease. A total of 40,400 MRI images were used and six different Convolutional Neural Network (CNN) architectures were analyzed. As a result of the experiments, the highest success rate was 99.92% using the MobileNetV2 architecture. These results can be considered as a potential Alzheimer's detection tool. A detailed examination of the data set and optimized parameters used in the training of the model can increase the success rate. Future studies aim to improve the generalization capability of the model and obtain broader, generalizable results. While this research offers a new perspective in the diagnosis of Alzheimer's disease, it can provide an important basis for future studies.

Author contributions

Seda Nur Polater: Data analysis, drafting the manuscript. Onur Sevli: Defining the methodology, evaluating the results and editing the draft.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Kırboğa, K. K., Tekin, B., & Demir, M. (2023). Applications of turmeric starch and curcumin. *Bayburt Üniversitesi Fen Bilimleri Dergisi*, *6*(1), 99-125.
2. Anadol, A. Z., & Tunaoğlu, F. S. (2023). Curriculum Development for Upper Endoscopy Training in Surgical Education. *Gazi Medical Journal*, *34*(1).
3. Santos, M. K., Ferreira, J. R., Wada, D. T., Tenório, A. P. M., Nogueira-Barbosa, M. H., & Marques, P. M. D. A. (2019). Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards precision medicine. *Radiologia brasileira*, *52*(06), 387-396.
4. Telgarsky, M. (2016, June). Benefits of depth in neural networks. In *Conference on learning theory* (pp. 1517-1539). PMLR.
5. Khojaste-Sarakhsi, M., Haghighi, S. S., Ghomi, S. F., & Marchiori, E. (2022). Deep learning for Alzheimer's disease diagnosis: A survey. *Artificial intelligence in medicine*, *130*, 102332.
6. Ebrahimighahnavieh, M. A., Luo, S., & Chiong, R. (2020). Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review. *Computer methods and programs in biomedicine*, *187*, 105242.
7. Aydın, S., Taşyürek, M., & Öztürk, C. (2023). MR Görüntülerinde Evrişimli Sinir Ağlar Kullanılarak Alzheimer Hastalık Tespiti. *Erciyes Üniversitesi Fen Bilimleri Enstitüsü Fen Bilimleri Dergisi*, *39*(3), 357-368.
8. Yüzgeç, E., & Talo, M. (2023). Alzheimer ve Parkinson Hastalıklarının Derin Öğrenme Teknikleri Kullanılarak Sınıflandırılması. *Fırat*

Üniversitesi Mühendislik Bilimleri Dergisi, *35*(2), 473-482.

9. Sharma, S., Gupta, S., Gupta, D., Juneja, S., Mahmoud, A., El–Sappagh, S., & Kwak, K. S. (2022). Transfer learning-based modified inception model for the diagnosis of Alzheimer's disease. *Frontiers in Computational Neuroscience*, *16*, 1000435.
10. Zena, J. I., Lucky, E., Ellaine, C. G., Edbert, I. S., & Suhartono, D. (2022, December). Deep Learning Approach based Classification of Alzheimer's Disease Using Brain MRI. In *2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* (pp. 397-402). IEEE.
11. Liu, S., Masurkar, A. V., Rusinek, H., Chen, J., Zhang, B., Zhu, W., ... & Razavian, N. (2022). Generalizable deep learning model for early Alzheimer's disease detection from structural MRIs. *Scientific reports*, *12*(1), 17106.
12. Singh, P., & Mishra, S. K. (2022, August). Alzheimer's Detection And Categorization using a Deep-Learning Approach. In *2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)* (pp. 727-734). IEEE.
13. Shu, F., & Tian, L. (2018). Deep Learning Methods for Alzheimer's Disease Prediction. *CS230: Deep Learn*, 1-10.
14. Er, A., Varma, S., & Paul, V. (2017). Classification of brain MR images using texture feature extraction. *International Journal of Computer Science and Engineering*, *5*(5), 1722-1729.
15. Islam, J., & Zhang, Y. (2017). A novel deep learning based multi-class classification method for Alzheimer's disease detection using brain MRI data. In *Brain Informatics: International Conference, BI 2017, Beijing, China, November 16-18, 2017, Proceedings* (pp. 213-222). Springer International Publishing.
16. Islam, J., & Zhang, Y. (2017). An ensemble of deep convolutional neural networks for Alzheimer's disease detection and classification. *arXiv preprint arXiv:1712.01675*.
17. Dubey, S. (2020). Alzheimer's Dataset four class of Images. *Kaggle. https://www. kaggle. com/tourist55/alzheimers-dataset-4-class-of-images/data. Accessed*, *1*.
18. Mogaraju, J. K. (2024). Machine learning empowered prediction of geolocation using groundwater quality variables over YSR district of India. *Turkish Journal of Engineering*, *8*(1), 31-45.
19. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.
20. Dong, K., Zhou, C., Ruan, Y., & Li, Y. (2020, December). MobileNetV2 model for image classification. In *2020 2nd International Conference on Information Technology and Computer Application (ITCA)* (pp. 476-480). IEEE.
21. Xia, X., Xu, C., & Nan, B. (2017, June). Inception-v3 for flower classification. In *2017 2nd international conference on image, vision and computing (ICIVC)* (pp. 783-787). IEEE.
22. Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1251-1258).
23. Salehi, A. W., Khan, S., Gupta, G., Alabduallah, B. I., Almjally, A., Alsolai, H., ... & Mellit, A. (2023). A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope. *Sustainability*, *15*(7), 5930.
24. Türkoğlu, M., & Hanbay, D. (2019). Plant disease and pest detection using deep learning-based features. *Turkish Journal of Electrical Engineering and Computer Sciences*, *27*(3), 1636-1651.
25. Aydın, V. A. (2024). Comparison of CNN-based methods for yoga pose classification. *Turkish Journal of Engineering*, *8*(1), 65-75.