El Cezeri
Journal of Science and Engineering
ISSN: 2148-3736

Research Article

Classification of Dementia Levels by Using Different Convolutional Neural Network Architectures

Iclal Cetin Tas[1a], Murat Simsek[2b]
[1] Electrical-Electronics Engineering Department, Ostim Technical University, Ankara, Türkiye
[2] Artificial Intelligence Engineering Department, Ostim Technical University, Ankara, Türkiye
icetintas@baskent.edu.tr

DOI : 10.31202/ecjse.1512362
Received: 08.07.2024 Accepted: 24.12.2024
How to cite this article:
Iclal Cetin Tas, " Classification of Dementia Levels by Using Different Convolutional Neural Network Architectures", El-Cezeri Journal of Science and Engineering, Vol: 12, Iss: 1, (2025), pp.(74-85).
ORCID: [a]0000-0002-1101-9773, [b] 0000-0002-8648-3693.

Abstract : Dementia or Alzheimer is a disease that causes symptoms such as forgetfulness and loss of physical ability, which will add to the individual's life in later stages, along with morphological changes in the brain. Unfortunately, a definitive treatment for these diseases has not yet been found. However, it is aimed at slowing down the progression of the disease to ensure that the patient is less affected by these adverse conditions and to protect living standards with early diagnosis of the disease. In addition, a complete diagnosis of the disease requires a series of tests and a tiring diagnostic phase to be evaluated by an experienced specialist. High-resolution magnetic resonance imaging is used to make this determination. This study tries to determine the stage of the disease or whether the individual is healthy by using MR.MR images of individuals in 4 stages of the disease, one of which is a healthy individual, were described as a classification problem and tried to be solved using VGG, Resnet, and Mobilenet architectures. Over %95 success has been achieved by supporting the proposed architecture with feature analysis and classical architectures.

Keywords : Alzheimer, dementia levels, CNN, SMOTE, classification.

1 Introduction

Nowadays, many diseases seriously affect daily life and quality of life for a long time. Dementia types of especially Alzheimer's is one of the most important ones. Generally, there may be age-related memory problems called dementia. To summarize the relationship between dementia and Alzheimer's, dementia is a general term for a set of diseases characterized by cognitive decline, and then Alzheimer's disease is the most common type of dementia under this generalization. According to the World Health Organization, %60- %80 of dementia cases result in Alzheimer's. Alzheimer's disease (AD), which is the most common type of dementia, causes memory loss in the brain and disruption of daily life, especially in elderly individuals. AD is a common type of dementia and neurological disease in which the steps in the progression process that destroy brain cells are critical. This disease causes a decrease in thinking, memory, and behavioral functions, and symptoms appear gradually with age. The transition between the stages of the disease can take a long time [1]. The disease has profound physical and psychological effects on individuals, their families, and their social environments [2]–[5]. As population growth slows down and the elderly population gradually increases throughout the world, especially in developed countries, the number of Alzheimer's patients is increasing, and it is predicted that it will increase annually. Fig. 1 and 2 also contain information on the two patient populations. Although many clinical studies continue to be conducted worldwide, no treatment has yet been provided to stop the disease [6]. Fig. 1 shows the distribution of patients by age. We observe that the incidence of the disease increases with advancing age. It should not be forgotten that one of the determining criteria here is that not all individuals live to the age of 85 and above.

When the graph in Fig. 2 is examined, unfortunately, it is predicted that the number of patients will increase every year. Many reasons leading to this result are mentioned in the literature. Cerebral vascular occlusions, brain infections, vitamin deficiencies, excessive alcohol use, brain tumors, active ingredients of some drugs, and metabolic or psychological problems can be listed. Under current conditions, it is not possible to eliminate the disease through the treatment process. As with many diseases, early diagnosis is essential in dementia and Alzheimer's. Currently, only the rate of progression can be slowed down, and patients' relative quality of life can be kept constant. With early diagnosis and starting treatment in the early stages of the disease, significant progress can be made before permanent damage occurs in the brain. Analysis of MRI images is widely used

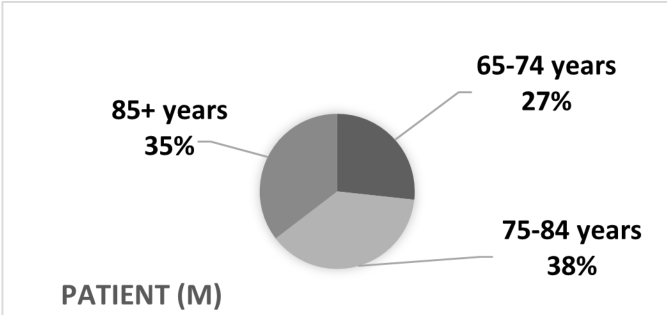

Figure 1: Distribution of patients by age [7]

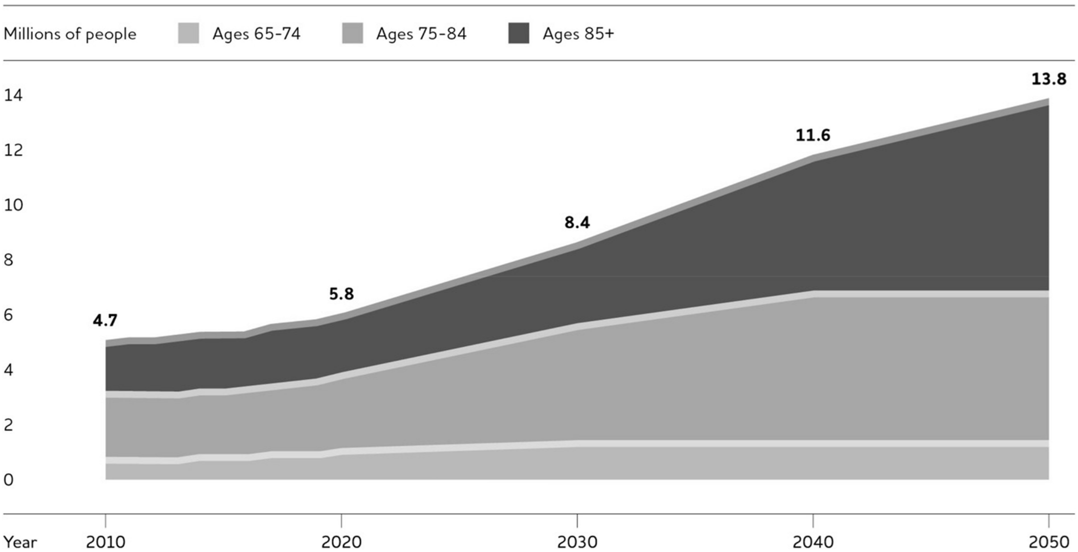

Figure 2: Distribution of patients by age in years [7]

in diagnosing AD [8]. With these analyses, it is possible to determine and classify the stages of the disease. Short information on symptoms and durations depending on the stages of the disease is shown in Fig. 3.

Studies based on traditional machine learning and, especially in recent years, deep learning-based techniques have focused on developing models for detecting physical, anatomical, and functional disorders due to types of dementia and Alzheimer's disease in the human brain [10]–[16].

The studies started by making binary classification, which is the basis of classification. Binary classification is a method that produces one of two outputs for input data. Binary classification is divided into positive class and negative class. These are "1" and "0". For this reason, the classification process was carried out by grouping the data groups labeled as very mild demented, mild demented, and moderate demented among the classes in the data set as non-demented, which is at risk of disease, and the data as healthy. There are approaches applied for similar datasets in the literature [3], [17], [18].

In their approach, Nguyen et al. aimed to investigate the ability to detect AD during the first visit of patients with suspected Alzheimer's disease. For this reason, they stated that all the data used for the test included only the initial and first visit scans. They used the Extreme Gradient Boosting method with 5-fold cross-validation. They achieved an average AUC of $\%100$ during training and $\%96$ in testing. They evaluated machine learning methods from a temporal perspective. They tried to prioritize the prediction of the 3D-ResNet model through the heat map [19].

Venugopalan et al. study showed that deep models outperformed shallow models, including support vector machines, decision trees, random forests, and k-nearest neighbors, by using the AD neuroimaging initiative (ADNI) dataset. Integrating multimodal data outperforms single-mode models in terms of performance evaluation criteria (accuracy, precision, recall, and average F1 scores). It is seen that approximately $\%88$ success was achieved in the analyses made with the proposed method [20]

Ahmed et al. examined both the left and right hippocampus regions on MRI images. They analyzed feature extraction and softmax cross-entropy in convolution neural network (CNN) structures in their study. The analyses used the Gwangju Alzheimer and Related Dementia (GARD) cohort dataset from the National Dementia Research Center (GARD) in Gwangju, South Korea. The results obtained achieved an accuracy of $\%88$ [10].

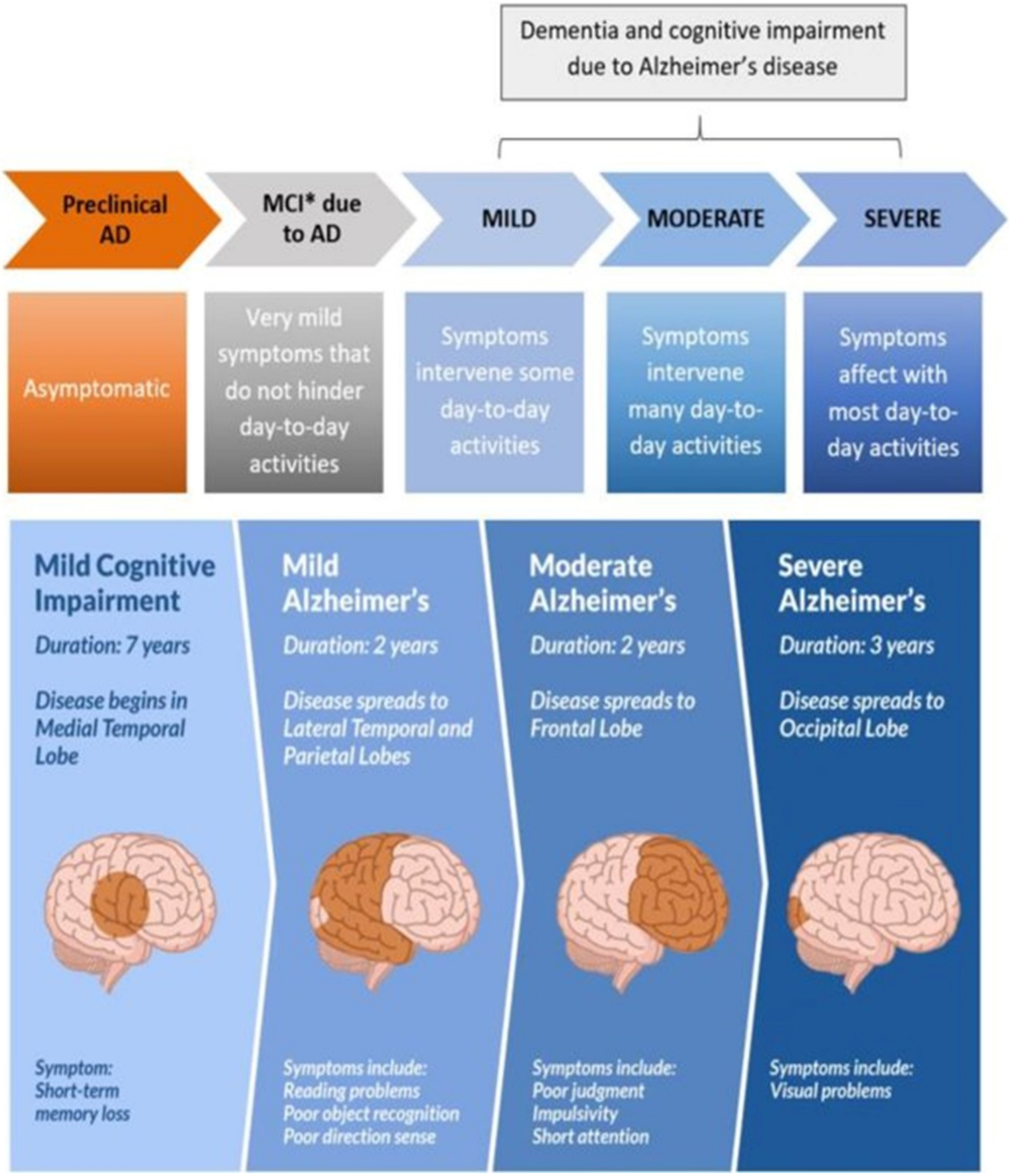

Figure 3: Changes caused by Alzheimer's disease in the brain and its [9]

In this study, dementia levels, including Alzheimer's, were classified with different CNN architectures using MRI data. The study is constructed as follows: Section II presents information about the dataset containing MRI images frequently used in the diagnosis of dementia and Alzheimer's disease and the CNN methods applied to it. Section III includes the data obtained from the analysis and discusses the data in question. The section mentions conclusions and predictions for future studies, which are given in Section IV.

2 Materials

The dataset used in the study was obtained from Kaggle [21]. The dataset contains 6400 MRI images. These images contain images of patient groups belonging to 4 different classes. Sample images are shown in Fig. 4. In all steps carried out within the scope of this study, a computer with an Intel i5 processor (2.5 GHz Turbo), four cores and 8 MB memory was used. Software development was done using the Python programming language. All software operations performed in this study used PyCharm 022.2.2 (Professional Edition). Python is a dedicated Python Integrated Development Environment that provides essential tools in various areas. (IDE). Python-based deep learning tools also offer various advantages in biomedical image analysis. These tools offer a powerful ability to understand, analyze, and extract features from complex biomedical data sets. These tools can analyze data from medical imaging devices, classify diseases, and recognize critical anatomical structures. Python-based deep learning tools offer a robust set of tools to obtain more effective, faster, and accurate results in biomedical image analysis.

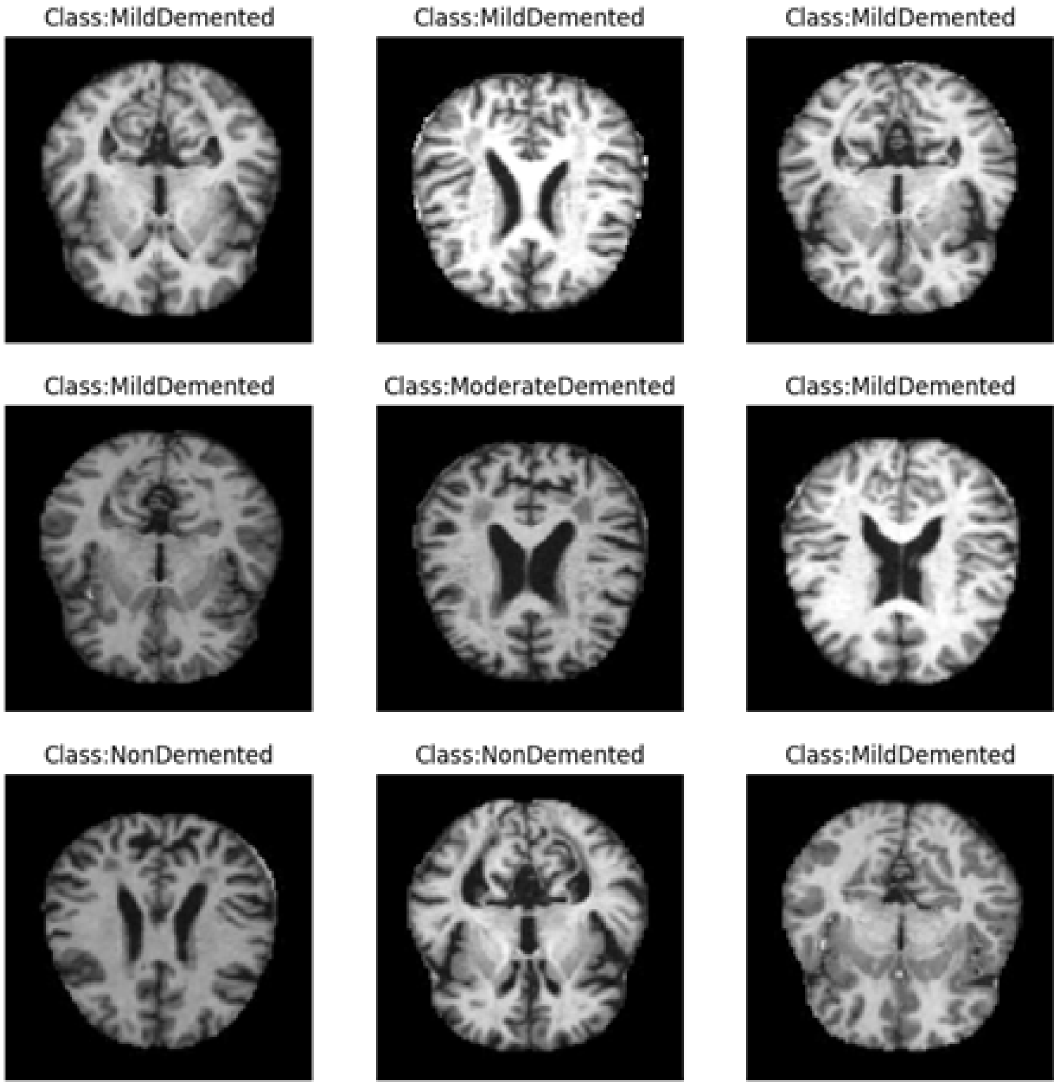

Figure 4: Sample images for different classes

Fig. 5 shows the block diagram of the methodology applied for the multiclass classification of dementia disease. Sub-steps for each step are included in the diagram.

3 Methods

3.1 Convolutional Neural Networks

CNNs are deep learning models that have been successfully used in visual data analysis tasks such as computer vision, image recognition, and processing. They have been shown to be very effective in detecting patterns and features in images, especially in studies. The layers of CNN and their properties are summarized below.

• Convolutional Layers: Convolutional layers are the basic components that help detect features in the input data (for example, edges, shapes, patterns in images). They perform convolution on the input using filters or kernels. This allows specific patterns and features to be identified. Each convolution layer can contain multiple filters, each used to identify different features. The convolution operation transforms the data into smaller and particularly more representative feature maps.

• Pooling Layers: Pooling layers shrink and summarize the feature maps produced by the convolution layers. They usually work with operations such as maximum pooling or average pooling. Reducing the size of feature maps is important to reduce computational cost and sensitivity to translations.

• Fully Connected Layers: Fully connected layers are the traditional structures found at the end of the CNN. These layers take a flattened version of the feature maps and are often used for output tasks such as classification or regression.

These layers help to learn higher-level representations of features. CNN's main purpose is to recognize complex features in images or visual data and perform certain tasks (e.g., object recognition, face recognition) using these features. The convolution

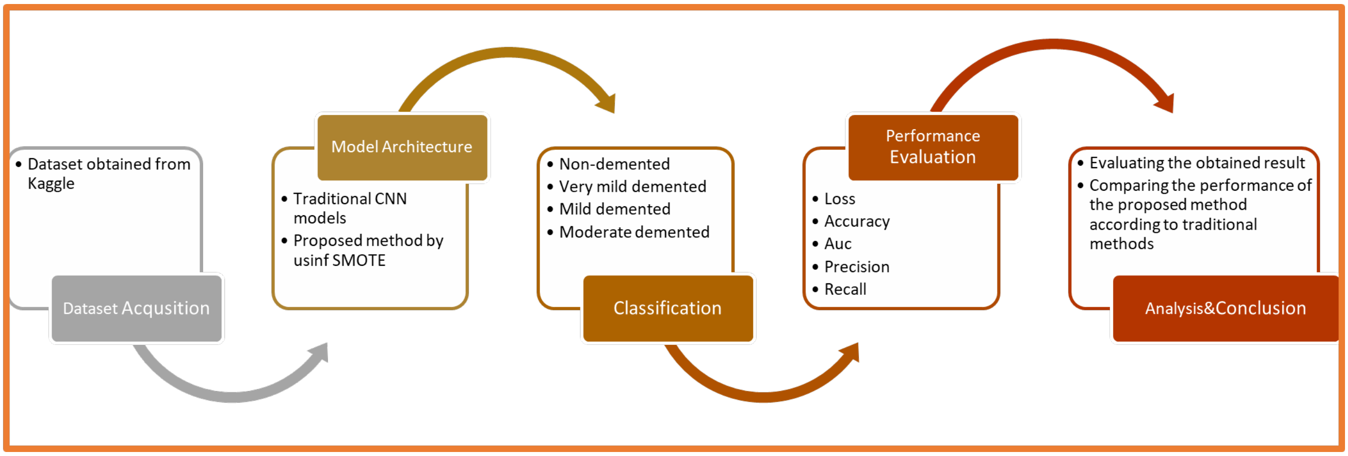

Figure 5: This study flowchart for multi-class classification analysis

and pooling layers help to learn these features in a hierarchical way, while the fully connected layers translate these features into task results. An attractive alternative to training from scratch is fine-tuning a deep network (especially a CNN architecture) via transfer learning. Through transfer learning, these trained networks can be used with smaller datasets by fine-tuning only the fully connected final layers of the CNN. Studies have proven that transfer learning is successful in applications with medical images [22]–[24].

The performance of traditional approaches was measured using the pre-trained network architectures used in the study. Visual geometry group (VGG) architecture based Vgg16, Vgg19, residual network (Resnet) based Resnet50, Resnet101 and mobile network (Mobilenet) based Mobilenet, Mobilenetv2 architectures were used for this study's analysis. VGG, ResNet, and MobileNet are three important convolutional neural networks that are considered important building blocks in the field of deep learning. They have architectures suitable for different tasks. This text will examine the common and different aspects of these three architectures, focusing on their advantages and application areas.: VGG, ResNet, and MobileNet share convolutional neural network (CNN) principles. This provides specifically designed building blocks for visual recognition, object detection, and classification tasks. Transfer Learning Ability: All three models have a common feature in that they can share pre-trained weights and are suitable for transfer learning applications. This allows them to be used effectively in tasks with limited data.

VGG generally has a simple structure consisting of deep and consecutive layers. ResNet contains blocks spliced together to solve the vanishing gradient problem that occurs in deep networks. MobileNet, on the other hand, offers a lightweight and fast architecture for use in mobile and embedded systems. VGG generally has more parameters and higher computational power. ResNet requires fewer parameters than VGG due to its block structure, which is designed to work more effectively.

MobileNet, on the other hand, is optimized especially for devices with low computing power and storage space. It is designed to provide high performance on mobile devices and embedded systems. The other two generally require larger computational resources and, therefore, have broader application areas.

3.2 Proposed Methods

Table 1 contains the layer information of the proposed CNN architecture. After preprocessing, the input dataset will pass the convolution, dropout, and maximum 2D pooling layers. In the convolution layer, which gives CNN architectures its name, there are several filters (or kernels) whose parameters must be examined as it progresses. The first and second convolution layers consist of 16 filters with a kernel size of 128*128. After the next layer, we apply dropout layers in this model to prevent all neurons from converging toward the same target [25]. We periodically utilize dropout layers to reduce overfitting and increase generalization error in the entire deep neural network with different architectures. Dropout layers are preferred because their generalization performance in many datasets outperforms neural networks that do not use dropouts [26].

The CNN architecture proposed in the study was supported by the Synthetic Minority Over-sampling Technique (SMOTE) approach. Although deep learning is a powerful tool for training complex model structures on large data sets, it may present some challenges, such as unbalanced class distributions. One of several techniques developed to overcome these difficulties is called SMOTE. SMOTE alleviates the problem of class imbalance by creating synthetic instances to empower the minority class. When combined with deep learning, the positive features of SMOTE come to the fore. This approach can help the model generalize better, better represent minority class samples, and avoid overfitting. It can also optimize the performance of deep learning models by increasing their learning ability, allowing the model to learn rare cases in the minority class better. Therefore, the SMOTE approach in deep learning can be considered an effective strategy to combat class imbalance and improve the model's overall performance.

Table 1: Proposed model CNN architecture

Model: cnn model Layer (type)	Output size	Parameter numbers
conv2d (Conv2D)	[128 128 16]	448
conv2d_1 (Conv2D)	[128 128 16]	2320
max_pooling2d (MaxPooling2D)	[64 64 16]	0
sequential (Sequential)	[32 32 16]	14016
sequential_1 (Sequential)	[16 16 64]	55680
sequential_2 (Sequential)	[8 8 128]	221952
dropout (Dropout)	[8 8 128]	0
sequential_3 (Sequential)	[4 4 256]	886272
dropout_1 (Dropout)	[4 4 256]	0
flatten (Flatten)	4096	0
sequential_4 (Sequential)	512	2099712
sequential_5 (Sequential)	128	66176
sequential_6 (Sequential)	64	8512
dense_3 (Dense)	4	260

In this study, the hyperparameters were tuned to optimize the performance of the convolutional neural network (CNN) architectures. The learning rate, batch size, and optimizer were systematically adjusted based on performance metrics observed during validation. The Adam optimizer was selected for its efficiency in converging the model during training. A learning rate of 0.001 was chosen after testing multiple configurations, ensuring a balance between convergence speed and performance. The batch size was set to 32, which allowed for efficient use of computational resources while maintaining the stability of the gradient updates. These hyperparameters were fine-tuned through iterative testing to maximize classification accuracy and minimize loss across training and validation datasets. The CNN model was constructed with multiple convolutional layers followed by max-pooling layers to extract spatial features from the input data progressively. The model architecture includes two initial convolutional layers with 16 filters and 3x3 kernel sizes, followed by convolutional blocks with 32, 64, 128, and 256 filters. Each convolutional block is followed by a max-pooling layer to reduce the dimensionality of the feature maps and improve computational efficiency. Dropout layers with a 0.2 dropout rate were employed after the deeper convolutional blocks to prevent overfitting. The fully connected layers included 512, 128, and 64 units, which progressively reduced the dimensionality of the feature vector before the final classification layer. The final output layer, with a softmax activation function, consisted of 4 units corresponding to the four classes in the dataset. This architecture was selected to balance computational efficiency with the need for deep feature extraction and classification accuracy.

4 Results and Discussion

The dataset contains 6400 images in total. Images belong to four different classes: 'NonDemented,' 'VeryMildDemented,' 'MildDemented,' and 'ModerateDemented.' The image dimensions were rescaled to 128x128. The parameter numbers used for parameter analysis of the images are given in Table 2.

Table 2: Parameter numbers

Model: Parameter Type	Number
Non-trainable	2368
Trainable	3352980
Total	3355348

When using multi-class classification performance of dementia disease, comparisons were made on five criteria: accuracy rate, the area under the curve, loss, precision, and recall. All tables and visualizations are presented in a way to emphasize these features. Loss is a metric that measures how far a model's predictions are from the actual values during training. The loss function is used to set the parameters of the model. The model tries to minimize the outcome of this function. Common loss functions use cross-entropy calculation. In multiple classification problems, cross-entropy measures the probabilities between multiple classes. Each class has a probability estimate, and the sum of these estimates must be 1.

Accuracy defines the ratio of correctly classified samples to the total number of samples. It is usually expressed as a percentage (%). However, accuracy may be an inadequate performance measure in dataset situations with unbalanced class distribution.

AUC generally refers to the area under the ROC (Receiver Operating Characteristic) curve. This curve shows the change of false positive rate (FPR) with response value, while it shows the change of true positive rate (TPR). AUC takes a value between 0 and 1. An AUC value closer to 1 indicates better classification performance of the model.

Precision refers to the ratio of samples predicted as positive to those that are positive. It aims to reduce the number of false positives, which are cases where true negatives are incorrectly predicted as positives. Precision measures how much of the samples classified as true positives are correctly predicted as positive. It aims to ensure that the model does not miss all instances of positives. False negatives are cases where true positives are incorrectly predicted as negatives.

The proposed model compares transfer learning by employing pre-trained CNN architectures, specifically VGG16, VGG19, ResNet50, and MobileNet. During transfer learning, the convolutional layers of the pre-trained models were frozen, preserving the weights learned from extensive datasets like ImageNet. Only the fully connected layers at the end of the model were fine-tuned to adapt to the Alzheimer's dataset. This approach capitalizes on the general feature extraction capabilities of the pre-trained models while allowing for specialization in the final layers. The fine-tuning process involved adjusting the weights of the last few layers to better represent the characteristics of the MRI images used for dementia classification, enhancing the model's ability to differentiate between the various stages of the disease.

In the first experiment, classification was applied using the data of non-demented and three-stage demented individuals. Pre-trained network architectures were used for this analysis. Results are obtained using this method, which is shown in Table 3..

Table 3: Traditional pre-trained networks classification result for four classes.

	Vgg16	Vgg19	Resnet 50	Resnet 101	Mobilenet	Mobilenetv2
Loss	0.7175	0.7000	0.8709	0.8767	1.2799	1.8565
Accuracy	0.6873	0.6919	0.6231	0.6067	0.6719	0.6489
AUC	0.9065	0.9121	0.8630	0.8626	0.8798	0.8481
Precision	0.7312	0.7193	0.6811	0.6493	0.6830	0.6551
Recall	0.6145	0.6411	0.4926	0.5152	0.6740	0.6474

VGG19 architecture has a very deep network structure and a wide learning capacity. This increases the network's ability to learn more complex features and relationships. When dealing with a complex and multidimensional problem such as Alzheimer's disease, this depth appears to allow the extraction of high-level features and these features to classify disease levels more accurately. VGG19 architecture can provide better results than other pre-trained network architectures as it can better extract feature maps using smaller filter sizes and consecutive convolution layers, ensuring that the features derived from previous layers represent lower-level and general features. In addition, with the transfer learning advantage, it can be said that VGG19 is an architecture with better generalization ability since it has been trained on a large dataset before. The VGG19 architecture includes various convolution layers and fully connected layers, resulting in more parameters in the model's learning process. This allows the model to gain more flexibility and better adapt to the data set. These aspects can explain why the approaches performed can perform better.

The model, which started with Conv2D layers, captured the spatial relationships in the input data. In these layers, feature maps were created through filters and essential patterns and building blocks in the data were detected. Conv2D layers performed deep feature extraction using different filter numbers and kernel sizes. After these convolution layers, the MaxPooling2D layer was added. The MaxPooling2D layer reduced the computational load of the model by performing dimensionality reduction and selecting the most significant information in the feature maps. With this layer, the complexity of the model was kept under control, and overfitting was prevented during the learning process. In addition, multiple Sequential layers were used in the model to efficiently organize the layers and optimize their interactions with each other. Sequential layers were formed by the combination of layers added in a certain order, and this structure was intended to increase the modularity and reusability of the model. Dropout layers were strategically placed to prevent over-learning of the model and to increase its overall performance. In these layers, certain neurons were randomly disabled during training, making the model more robust and generalizable. A flatten layer was used to combine and flatten the features. Multidimensional feature maps obtained from the flatten layer and convolution layers were converted to a one-dimensional vector and transferred to fully connected layers. The model classification process was performed with Dense layers. In Dense layers, learned features were used to increase the classification performance and it was concluded whether there was Alzheimer's in the output layer.

Callback mechanisms such as early stopping and learning rate reduction were incorporated into the model's training process to prevent overfitting and ensure optimal training performance. Early stopping was used to monitor the validation accuracy, halting the training if no significant improvement was observed after a patience threshold of five epochs. This strategy helped mitigate overfitting by ensuring that the model did not continue training beyond the point of diminishing returns. Additionally, a dynamic learning rate adjustment mechanism was implemented, reducing the learning rate when the validation accuracy plateaued. These callbacks not only improved the training efficiency but also ensured that the model converged to an optimal solution without unnecessary iterations, thereby enhancing the overall performance and generalizability of the model.

Callbacks were used in the model implemented using the proposed architecture. Callbacks are functions that are called when certain events or conditions occur during training. They can perform a few tasks, such as controlling the model's training, preventing overfitting, adjusting the training pace, or performing different functions. Early stopping and learning rate adjustment methods were used in this study. Early stopping is a standard callback used to prevent the model from being overfitting. If a particular metric (for example, accuracy) does not improve during training, it can automatically stop training. Learning rate is a vital hyperparameter that determines the training speed. Adjusting the learning rate during training enables faster or slower learning. In this way, an attempt was made to reduce the possibility that the results obtained from the model would be misleading. The callback parameters used in the study are shown in Table 4.

Fig. 6 visualizes the model's success rates during the training process and its performance on the validation set. The achievements without the SMOTE method support the robustness and general applicability of the deep learning-based dementia classification model, highlighting the model's ability to deal with the minority class in the dataset.

Table 4: Callback parameters values

Callback parameters	Value
Epoch Number	50
Monitor	Accuracy
Min_delta	0.01
Patience	5
Mode	“max”

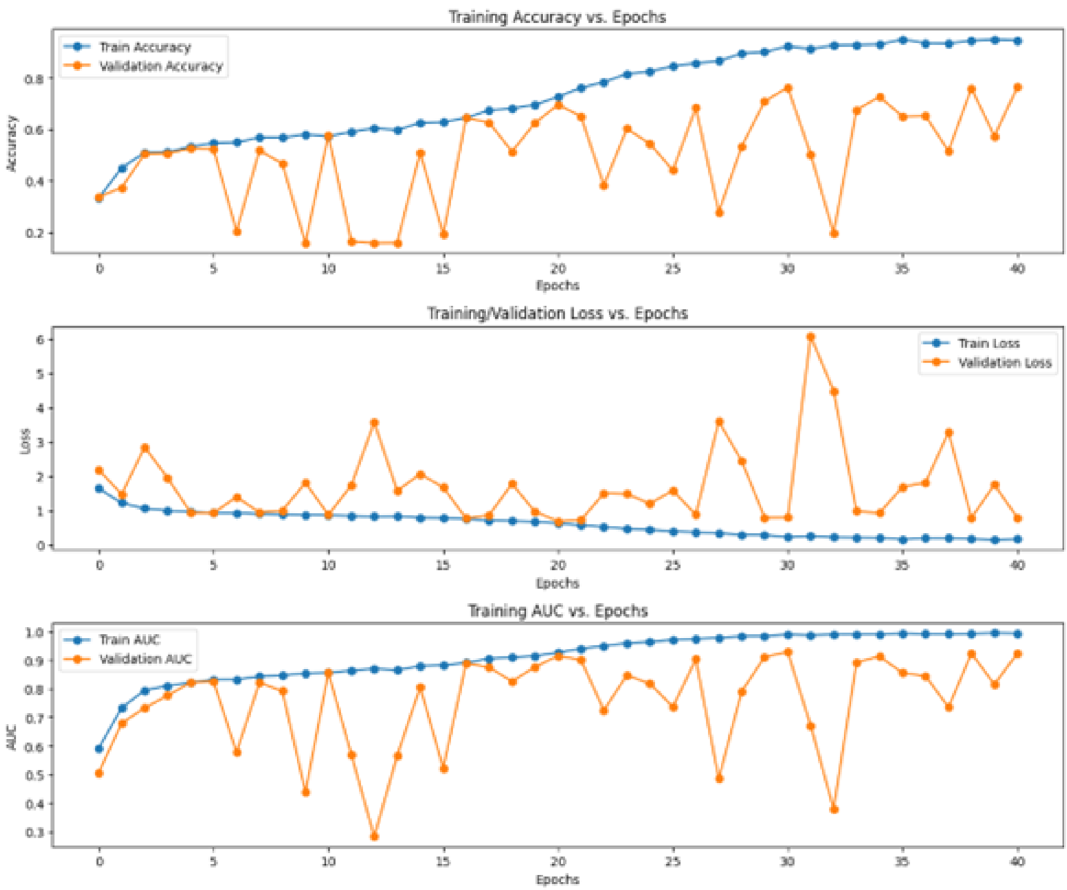

Figure 6: Training without SMOTE algorithm for four classes

The last step of the proposed model is aimed at preventing the imbalance between the data by using the SMOTE algorithm and reducing the error rates by increasing the inter-class predictive ability of the model. SMOTE algorithm was applied to eliminate data imbalance. The purpose of using this algorithm is to eliminate the imbalance of the dataset by ensuring that all classes contain equal numbers of data. With this approach, the number of data was 12800. The graphics are given in Fig. 7 in the analysis, and the test size was determined to be 0.2.

SMOTE was applied to address the inherent class imbalance in the dataset. SMOTE creates synthetic samples for the minority classes, thus increasing their representation within the dataset. This method was critical in enhancing the model’s ability to generalize to minority classes, such as "Moderate Demented" and "Very Mild Demented," which were underrepresented in the original dataset. The SMOTE algorithm was applied before model training, and its effects were evident in the improved classification metrics, particularly in precision and recall for the minority classes. By balancing the dataset, the model could learn more robust feature representations for all classes, ultimately leading to a more reliable classification performance across the board.

Since the transition between phases of the disease and knowing which stage the patient is at the time of diagnosis are essential, the classification was first made for very mildly demented, mildly demented, and moderately demented classes. The results obtained for the four-class and three-class classification problems are shown in Table 5. and Table 6., respectively.

However, when the results obtained from three-class and four-class analyses are evaluated together, classification ability decreases as the number of classes increases and the nature of the added data changes. The values obtained in the four-class results are worse for all criteria than the three. Information about the literature studies is shown using the same dataset in Table 7.

The contrastive learning method used by Shu et al. [27] provided an accuracy rate of $\%92$. This rate is considerably higher

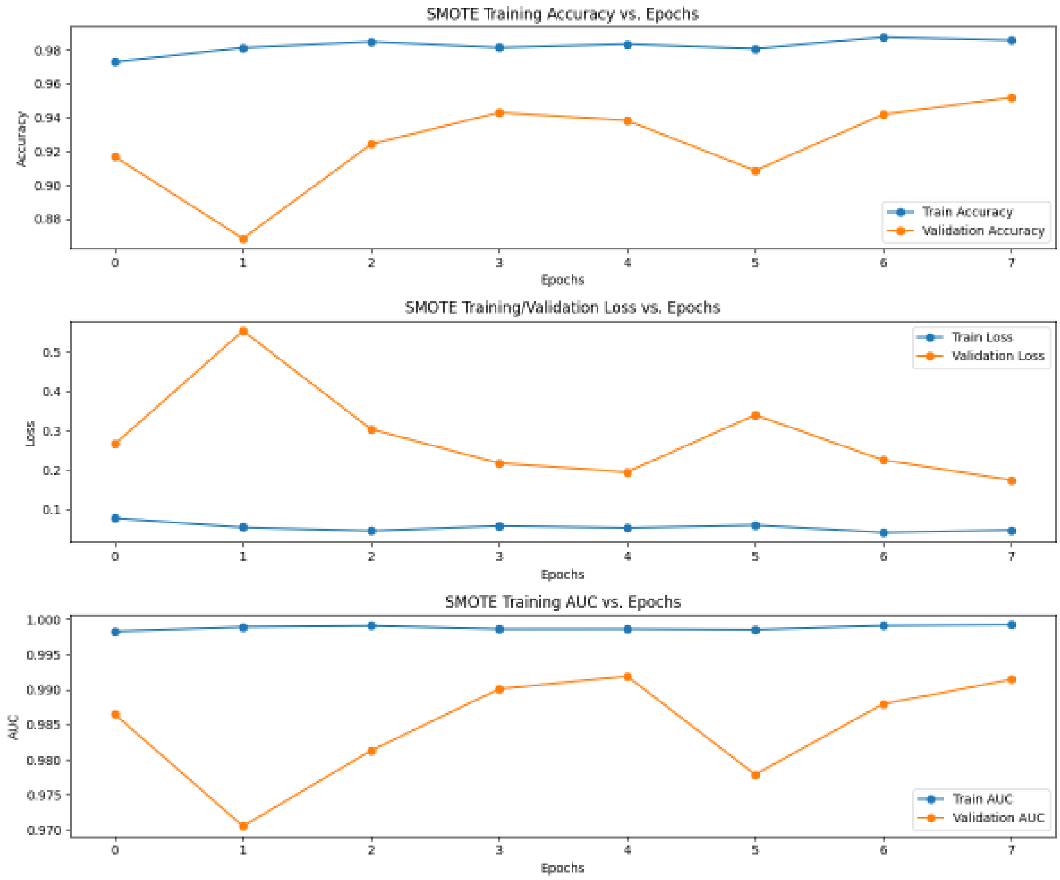

Figure 7: Training without SMOTE algorithm for four classes

Table 5: Results for multiclass classification for four classes

	Without SMOTE Model	SMOTE Model
Loss	0.7870	0.175
Accuracy	0.7641	0.9520
AUC	0.9236	0.9914
Precision	0.7674	0.9519
Recall	0.7602	0.9512

Table 6: Results for multiclass classification for three classes

	Without SMOTE Model	SMOTE Model
Loss	0.7684	0.9341
Accuracy	0.7011	0.7500
AUC	0.8309	0.8998
Precision	0.7011	0.7547
Recall	0.7011	0.7485

than the 70.$\%$30 accuracy rate obtained by Mggdadi et al. [28] using the VGG16-based 2D CNN. In the study conducted by Ajagbe et al. [29] , $\%71.02$ and $\%77.66$ accuracy rates were achieved with the VGG16 and VGG19 models, respectively, which shows that different CNN configurations can create significant differences in terms of performance. In this study, it is seen that similar results are obtained when the traditional methods in question are applied.

The DEMNET model proposed by Murugan et al. [30] reveals that special network configurations can be effective. In another study, conducted with AlexNet and ResNet-based models, the AlexNet + SVM combination stood out with an accuracy rate of $\%94.80$ [31]. It shows that integrating traditional machine learning algorithms such as SVM with deep learning models can improve performance. The hybrid CNN model proposed by Techa et al. [32] and including DenseNet196, VGG16 and ResNet50, achieved an accuracy rate of $\%89$. Sharma et al. [34] emphasize that transfer learning can be a powerful tool in Alzheimer's detection with $\%94.92$ accuracy using the Transfer learning and Inception model. When the results are examined for proposed method, it is shown that the model's errors during training have significantly decreased, and it has undergone a better learning process. While the accuracy rate of the model before SMOTE was applied was $\%76.41$, this rate increased to $\%95.20$ after SMOTE was applied. This increase shows that the SMOTE method has significantly increased the model's overall

Table 7: Literature summary for the same dataset

Reference	Method	Year	Accuracy (%)
Shu et al. [27]	Contrastive learning	2018	92.00
Mggdadi et al. [28]	2D CNN	2021	67.50
	VGG16	2021	70.30
Ajagbe et al. [29]	CNN	2021	71.02
“	VGG16	2021	77.04
“	VGG19	2021	77.66
Murugan et al. [30]	DEMNET(Dementia Networks)	2021	95.23
Mohammed et al. [31]	AlexNet	2021	92.20
	ResNet	2021	93.10
	AlexNet+SVM	2021	94.80
	ResNet-50+SVM	2021	94.10
Techa et al. [32]	A proposed convolution neural network (included DenseNet196, VGG16 and ResNet50)	2022	89.00
Sharma et al. [33]	Transfer learning, SVM, and permutation based machine learning	2022	91.75
Sharma et al. [34]	Transfer based Inception model	2022	94.92
Hussain et al. [35]	Random Forest	2023	91.25
	SVM	2023	80.70
	CNN	2023	93.96
Proposed method*	With SMOTE (three classes)	2024	75.00
“	With SMOTE (four classes)	2024	95.20

performance. The AUC value was $\%92.36$ before SMOTE was applied, while it became $\%99.14$ after SMOTE was applied. This shows that the classification ability of the model has been significantly improved with SMOTE and provides more reliable results. It shows that the model’s ability to catch true positives has increased and produces fewer false negative results. This study aims to contribute to the literature by including a more comprehensive classification framework targeting the stages of dementia. While focusing on the binary or quadruple-class classification of Alzheimer’s diagnosis using machine learning and deep learning models, this study applied a three- and four-class classification to guide decision-makers in making decisions about stage transitions and to provide a similar contribution to the initial diagnosis. Although the proposed method shows lower accuracy in the three-class classification than the four-class one, it highlights the importance of correctly defining different disease stages. This finer level of detail can provide valuable clinical insights not emphasized in previous studies that focused mainly on broader classifications. It offers potential benefits for more detailed diagnostic processes to improve patient care.

In this part, let’s briefly summarize the restrictive reasons and performance criteria. A research limitation is that the data set used in the analysis cannot be tested on real data. In addition to this situation, the long duration of the analyses can be considered another limiting factor.

5 Conclusions

This research presented multiple classifications of medical images of Alzheimer’s disease with the proposed CNN, VGG16, VGG19, Resnet50, Resnet101, Mobilenet, and Mobilenetv2 and demonstrated that deep convolutional neural network approaches for multiple classifications are possible.

This study aims to contribute to developing effective treatment strategies for the current stage by focusing on the classification of different dementia stages, including Alzheimer’s disease, by using CNN architectures, providing early diagnosis and stage determination of the disease. MRI is the most critical imaging method that contributes to this process. The results show that new approaches reinforced CNN architectures as a powerful tool for diagnosing and classifying dementia levels. In this study, MRI images containing three and four different classes were classified using different deep learning architectures. The performance of the obtained classification results was compared through metrics. The study obtained the highest classification performance using the proposed method. The proposed method achieved the best performance regarding accuracy, area under the curve, loss, recall, and precision. VGG-19 closely followed it, while Resnet 50 had a lower performance.

When the results obtained for the CNN architecture were compared, it was seen that the dimensionality reduction and feature acquisition methods applied in the study were effective in detecting dementia levels through MRI images. In subsequent studies, the performances of different CNN architectures and the features obtained from these architectures can be evaluated in classical classifiers to detect dementia levels. Performance evaluations of hybrid models can be made by combining new architectures with classical methods. It is envisaged that this study will provide a basis for future studies based on image analysis and that approaches can be used to reduce the mentioned limitations.

Ethical Declarations

The dataset used in the study is a dataset that has been previously used in the literature and is published publicly. Access link:https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images

Authors' Contributions

Iclal Cetin Tas: Conceptualization, Methodology, Writing - original draft, Data, Software. Murat Simsek: Data curation, Software, Writing - original draft.

Competing Interests

The authors have no conflict of interest to report.

References

[1] S. Al-Shoukry, T. H. Rassem, and N. M. Makbol, "Alzheimer's diseases detection by using deep learning algorithms: A mini-review," *IEEE Access*, vol. 8, pp. 77 131–77 141, 2020.

[2] 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 *Proceedings - International Conference on Smart Electronics and Communication, ICOSEC 2020*, 2020.

[3] W. Salehi, P. Baglat, G. Gupta, S. B. Khan, A. Almusharraf, A. Alqahtani, and A. Kumar, "An Approach to Binary Classification of Alzheimer's Disease Using LSTM," *Bioengineering 2023, Vol. 10, Page 950*, vol. 10, no. 8, p. 950, aug 2023.

[4] E. Hanbay and A. Ari, "Özel Blok Yapıları Kullanarak Tasarlanan Derin Öğrenme Mimarileri ile Alzheimer Hastalık Tespiti," *Firat Universitesi Muhendislik Bilimleri Dergisi*, vol. 35, no. 2, pp. 745–752, sep 2023.

[5] K. Aderghal, A. Khvostikov, A. Krylov, J. Benois-Pineau, K. Afdel, and G. Catheline, "Classification of Alzheimer Disease on Imaging Modalities with Deep CNNs Using Cross-Modal Transfer Learning," *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, vol. 2018-June, pp. 345–350, jul 2018.

[6] M. Ü. ÖZİÇ and S. ÖZŞEN, "3B Alzheimer MR Görüntülerinin Hacimsel Kayıp Bölgelerindeki Voksel Değerleri Kullanılarak Sınıflandırılması," *El-Cezeri Fen ve Mühendislik Dergisi*, 2020.

[7] "Alzheimer's Facts and Figures Report | Alzheimer's Association." [Online]. Available: https://www.alz.org/alzheimers-dementia/facts-figures

[8] Y. Eroglu, M. Yildirim, and A. Cinar, "mRMR-based hybrid convolutional neural network model for classification of Alzheimer's disease on brain magnetic resonance images," *International Journal of Imaging Systems and Technology*, vol. 32, no. 2, 2022.

[9] S. Dan, D. Sharma, K. Rastogi, Shaloo, H. Ojha, M. Pathak, and R. Singhal, "Therapeutic and diagnostic applications of nanocomposites in the treatment Alzheimer's disease studies," *Biointerface Research in Applied Chemistry*, vol. 12, no. 1, pp. 940–960, feb 2022.

[10] S. Ahmed, K. Y. Choi, J. J. Lee, B. C. Kim, G. R. Kwon, K. H. Lee, and H. Y. Jung, "Ensembles of Patch-Based Classifiers for Diagnosis of Alzheimer Diseases," *IEEE Access*, vol. 7, pp. 73 373–73 383, 2019.

[11] D. Shen, C. Y. Wee, D. Zhang, L. Zhou, and P. T. Yap, "Machine learning techniques for AD/MCI diagnosis and prognosis," *Intelligent Systems Reference Library*, vol. 56, pp. 147–179, 2014.

[12] Y. Wang, M. Liu, L. Guo, and D. Shen, "Kernel-based multi-task joint sparse classification for Alzheimer'S disease," *Proceedings - International Symposium on Biomedical Imaging*, pp. 1364–1367, 2013.

[13] J. Escudero, J. P. Zajicek, and E. Ifeachor, "Machine Learning classification of MRI features of Alzheimer's disease and mild cognitive impairment subjects to reduce the sample size in clinical trials," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 7957–7960, 2011.

[14] A. Ortiz, J. M. Górriz, J. Ramírez, and F. J. Martínez-Murcia, "LVQ-SVM based CAD tool applied to structural MRI for the diagnosis of the Alzheimer's disease," *Pattern Recognition Letters*, vol. 34, no. 14, pp. 1725–1733, oct 2013.

[15] S. T. Yang, J. D. Lee, T. C. Chang, C. H. Huang, J. J. Wang, W. C. Hsu, H. L. Chan, Y. Y. Wai, and K. Y. Li, "Discrimination between Alzheimer's disease and mild cognitive impairment using SOM and PSO-SVM," *Computational and Mathematical Methods in Medicine*, vol. 2013, 2013.

[16] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, and D. Rueckert, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *NeuroImage*, vol. 65, pp. 167–175, jan 2013.

[17] A. B. Tufail, Y. K. Ma, and Q. N. Zhang, "Binary Classification of Alzheimer's Disease Using sMRI Imaging Modality and Deep Learning," *Journal of Digital Imaging*, vol. 33, no. 5, pp. 1073–1090, oct 2020.

[18] R. Prajapati, U. Khatri, and G. R. Kwon, "An Efficient Deep Neural Network Binary Classifier for Alzheimer's Disease Classification," *3rd International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2021*, pp. 231–234, apr 2021.

[19] D. Nguyen, H. Nguyen, H. Ong, H. Le, H. Ha, N. T. Duc, and H. T. Ngo, "Ensemble learning using traditional machine learning and deep neural network for diagnosis of Alzheimer's disease," *IBRO Neuroscience Reports*, vol. 13, pp. 255–263, dec 2022.

[20] J. Venugopalan, L. Tong, H. R. Hassanzadeh, and M. D. Wang, "Multimodal deep learning models for early detection of Alzheimer's disease stage," *Scientific Reports 2021 11:1*, vol. 11, no. 1, pp. 1–13, feb 2021.

[21] "Alzheimer's Dataset (4 class of Images) | Kaggle." [Online]. Available: https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images

[22] M. Hon and N. M. Khan, "Towards Alzheimer's disease classification through transfer learning," *Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2017*, vol. 2017-January, pp. 1166–1169, dec 2017.

[23] J. Plested and T. Gedeon, "Deep transfer learning for image classification: a survey."

[24] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *Advances in Neural Information Processing Systems*, vol. 4, no. January, pp. 3320–3328, nov 2014.

[25] "Classification of Alzheimer ' s disease subjects from MRI using hippocampal visual features To cite this version : HAL Id : hal-00993379," *Computerized Medical Imaging and Graphics*, vol. 44, no. 1, 2015.

[26] J. Brownlee, "A Gentle Introduction to Dropout for Regularizing Deep Neural Networks," 2018.

[27] F. Shu and L. Tian, "Deep Learning Methods for Alzheimer's Disease Prediction Project Category: Computer Vision."

[28] E. Mggdadi, A. Al-Aiad, M. S. Al-Ayyad, and A. Darabseh, "Prediction Alzheimer's disease from MRI images using deep learning," in *2021 12th International Conference on Information and Communication Systems, ICICS 2021*, 2021.

[29] S. A. Ajagbe, K. A. Amuda, M. A. Oladipupo, O. F. AFE, 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, 2021.

[30] S. Murugan, C. Venkatesan, M. G. Sumithra, X. Z. Gao, B. Elakkiya, M. Akila, and S. Manoharan, "DEMNET: A Deep Learning Model for Early Diagnosis of Alzheimer Diseases and Dementia from MR Images," *IEEE Access*, vol. 9, pp. 90 319–90 329, 2021.

[31] B. A. Mohammed, E. M. Senan, T. H. Rassem, N. M. Makbol, A. A. Alanazi, Z. G. Al-Mekhlafi, T. S. Almurayziq, and F. A. Ghaleb, "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 2021, Vol. 10, Page 2860*, vol. 10, no. 22, p. 2860, nov 2021.

[32] C. Techa, M. Ridouani, L. Hassouni, and H. Anoun, "Alzheimer's Disease Multi-class Classification Model Based on CNN and StackNet Using Brain MRI Data," *Lecture Notes on Data Engineering and Communications Technologies*, vol. 152, pp. 248–259, 2023.

[33] S. Sharma, S. Gupta, D. Gupta, A. Altameem, A. K. J. Saudagar, R. C. Poonia, and S. R. Nayak, "HTLML: Hybrid AI Based Model for Detection of Alzheimer's Disease," *Diagnostics*, vol. 12, no. 8, 2022.

[34] S. Sharma, S. Gupta, D. Gupta, S. Juneja, A. Mahmoud, S. El–Sappagh, and K. S. Kwak, ''Transfer learning-based modified inception model for the diagnosis of Alzheimer's disease,'' *Frontiers in Computational Neuroscience*, vol. 16, 2022.

[35] M. G. Hussain and Y. Shiren, ''Identifying Alzheimer Disease Dementia Levels Using Machine Learning Methods,'' *Medical Research Archives*, vol. 11, no. 7.1, nov 2023.