aurum

AURUM SAĞLIK BİLİMLERİ DERGİSİ
AURUM JOURNAL OF HEALTH SCIENCES

Cilt 6, Sayı 2 | Yaz 2024
Volume 6, No 2 | Summer 2024

TRANSFER LEARNING IN SEVERITY CLASSIFICATION IN ALZHEIMER'S : A BENCHMARK COMPARATIVE STUDY ON DEEP NEURAL NETWORKS

Seda KIRTAY[1], Muhammed Tayyip KOÇAK[2]
[1]Electrical And Electronics Engineering, Marmara University, Turkey, seda.kirtay@marmara.edu.tr
(https://orcid.org/0000-0003-2415-9131)
[2]Software Engineering, Istanbul Health and Technology University, Turkey, muhammed.kocak@istun.edu.tr
(https://orcid.org/0000-0003-2276-2658)

Received: 05.09.2023
Accepted: 27.05.2024
Published: 31.07.2024
*Corresponding author

Research Article
pp.91-108

Abstract

Alzheimer's disease has become a condition of the brain that progresses over time and impacts a significant number of individuals worldwide. Early diagnosis, timely intervention and management of this disease process are very important in Alzheimer's disease. With regard to this study, we propose a transfer learning based early detection approach for Alzheimer's disease using Moderate Demented, Mild Demented, No Demented and Very Mild Demented classification sets. The proposed approach utilizes transfer learning based on the use of a deep neural network model that has been trained to extract features from brain imaging data. To evaluate the performance in transfer learning, a dataset of 6,400 images from brain MRI scans is augmented using data augmentation techniques and used in various convolutional neural network models the like VGG-19, Resnet-50, DenseNet-121, Inception-V3, VGG-16. The results are planned to show that these models achieve high sensitivity, specificity and high accuracy in detecting early signs of Alzheimer's disease. The study also emphasizes these advantages of using transfer methods of learning for early Alzheimer's detection by comparing it with various other deep learning models. The findings of this research suggest that transfer learning-based approaches can aid in the early detection of Alzheimer's disease., which affects millions of people, and offer a practical solution to classify cognitive impairment. With the proposed approach, it is shown that by helping clinicians to detect individuals at risk of Alzheimer's at an early stage, it will be possible to provide timely intervention and, in fact, better patient care. In terms of more effective applicability in clinical applications, the proposed approach can be applied to different and larger datasets and populations to make improvements and provide convenience to clinicians and patients. The best success rate of the models we used is achieved on the VGG19, RESNET50 KNN model with 99 percent.

Keywords: Deep learning, transfer learning, alzheimer, disease classification

1. Introduction

Alzheimer's disease, which is incurable neurodegenerative deterioration that is ongoing condition that impairs memory, cognition, and conduct. It is actually especially widespread cause of dementia in elderly adults and has a substantial impact on both individuals and society. According to Gaugler et al (Alzheimer's Association, 2016), there are an estimated There are currently 5.8 million Alzheimer's disease patients in the US, and this figure is projected to increase to 13.8 million by 2050. Alzheimer's disease prevalence is also rising globally, posing immense challenges to society (Ribe & Lovestone, 2016; Qiu, Kivipelto, & Von Strauss, 2022).

Alzheimer's disease affects not only those affected by it, but also their caregivers and society as a whole. Caregivers of Alzheimer's disease patients experience significant burden and stress, which can result in adverse health outcomes (Hayajneh & Shehadeh, 2014; Scott, 2013). Estimates range from $290 billion to $1 trillion annually in the United States alone (Alzheimer's Association, 2016) for the cost of Alzheimer's disease care. This cost encompasses both direct and indirect costs, such as decreased productivity and caregiver burden.

The consequences of Alzheimer's on society extend beyond monetary costs. Additionally, it has substantial social and affective effects on individuals, families, and communities. Alzheimer's disease patients and those who care for them may experience social isolation, stigma, and diminished quality of life as a result of the disease (Alzheimer's Association, 2016; Ribe & Lovestone, 2016; Qiu, Kivipelto, & Von Strauss, 2022). In addition, the rising Dementia and Alzheimer's disease have a high prevalence has increased the need for resources and assistance for those impacted persons and the relatives of those individuals.

The influence of Alzheimer's on individuals, caregivers, and society is significant. The increasing prevalence of the disease and its associated costs and burden emphasize the need for continued research and assistance for those affected and their families. It is necessary to conduct a thorough evaluation of treatment efficacy (Winblad et al., 2001) by evaluating the effects of treatment on persons suffering from Alzheimer's disease, in addition to the effects their condition has on caregivers and society as a whole.

Early and accurate Alzheimer's disease diagnosis is essential for effective treatment and care. As to the Alzheimer's Organization, early identification helps those suffering from Alzheimer's disease along with their loved ones to access support services, plan for the future, and make educated decisions regarding treatment and care. Individuals with Alzheimer's disease who are diagnosed early have the opportunity to participate in clinical trials and research studies, which can contribute to the development of novel treatments and therapies. Importantly, getting a diagnosis of Alzheimer's disease early enables patients to begin therapy at the earliest possible point in the disease's progression. Alzheimer's disease currently lacks a definitive cure, but pharmacological and non-pharmacological interventions exist that can effectively alleviate symptoms and impede the advancement of the disease. The efficacy of these treatments is highest when they are applied at an early stage of the disease's progression, prior to the occurrence of substantial cerebral damage. In addition to the benefits for Alzheimer's disease patients, early diagnosis has significant advantages for caregivers and society as a whole. Early diagnosis enables caregivers to plan for the future and gain access to support services, thereby reducing caregiver burden and tension. Additionally, early diagnosis permits the allocation of resources and funding for research and support services, which can help reduce the economic and social burden of Alzheimer's disease on society (Alzheimer's Association, 2016; Ribe & Lovestone, 2016; Qiu, Kivipelto, & Von Strauss, 2022; Hayajneh & Shehadeh, 2014; Scott, 2013; Winblad et al., 2001; Korolev, Safiullin, Belyaev, & Dodonova, 2017; Islam & Zhang, 2018).

Brain MRI imaging have become a valuable classification tool for Alzheimer's disease. MRI imaging can detect structural changes in the brain, such as hippocampal atrophy, associated with Alzheimer's disease (Platero, Lin, & Tobar, 2019). Several studies (Korolev, Safiullin, Belyaev, & Dodonova, 2017; Platero, Lin, & Tobar, 2019; Zhan et al., 2015) have demonstrated the efficacy of using MRI imaging and machine learning algorithms to classify Alzheimer's disease. These studies demonstrate that techniques of deep learning, that include convolutional neural networks, can classify comparing Alzheimer's disease to mild cognitive impairment and normal controls with high precision (Korolev, Safiullin, Belyaev, & Dodonova, 2017; Islam & Zhang, 2018). MRI scans can also be used to monitor disease progression and identify biomarkers for Alzheimer's disease early detection (Zhan et al., 2015).

Early and accurate Alzheimer's disease diagnosis is essential for effective treatment and care. Using brain MRI scans as a classification tool for Alzheimer's disease can aid in early disease detection and monitoring. This can lead to enhanced patient outcomes and quality of life for Alzheimer's disease patients and their caregivers.

In conclusion, brain MRI imaging have become a valuable classification instrument for Alzheimer's disease. Alzheimer's disease can be distinguished from moderate cognitive impairment and normal controls with a high degree of precision using machine learning algorithms and deep learning techniques. MRI scans can also be used to monitor disease progression and identify biomarkers for Alzheimer's disease early detection. The use of brain MRI imaging as a classification tool for Alzheimer's disease can aid in early detection and monitoring of the disease, leading to enhanced patient outcomes and quality of life for Alzheimer's disease patients and their caregivers.

2. Conceptual Framework and Literature

2.1. Transfer Learning

Transfer learning is a prevalent paradigm in deep learning in which models trained on standard datasets can be efficiently adapted to subsequent tasks. It has been utilized in numerous disciplines, technologies such as natural language processing, voice recognition, computer vision etc are examples. Transfer learning enables the reuse of a previously trained model for a new problem, which is extremely useful in data science given that the majority of real-world problems lack the millions of data points required to train these complex models (He, Zhou, Ma, Berg-Kirkpatrick, & Neubig, 2021).

2.2. MRI Image Conversion to Feature Vectors

MRI is a potent diagnostic instrument which used to be obtain images of the body's internal organs. An MRI creates images of organs, tissues, and bones using a strong magnetic field and radio radiation. It is a non-intrusive harmless procedure that produces high-resolution images regarding the internal structures of the human body's anatomy. MRI is a commonly employed diagnostic tool for a range of medical conditions, encompassing injuries to the brain and spinal cord., malignancies, joint and bone issues, and cardiovascular diseases. Magnetic resonance imaging (MRI) has been shown to be a valuable tool for healthcare professionals in facilitating precise diagnoses and devising efficacious treatment strategies for their patients (Wang et al., 2017; Varuna Shree & Kumar, 2018; Vlaardingerbroek & Boer, 2013).

MRI images Converting into feature vectors enables a vast array of sophisticated analysis and insights. By utilizing methods like preprocessing, feature extraction,ROI selection, normalization,vector representation, and optional feature selection, clinicians and researchers may extract pertinent data from MRI scans and accelerate breakthroughs in fields such as medical imaging, disease diagnosis, and drug development (Zulkoffli & Shariff, 2019; Higaki, Nakamura, Tatsugami, Nakaura, & Awai, 2019).

There have been many studies on the conversion of MRI images into feature vectors. Moreover, these citations illustrate how MRI images can be converted into feature vectors using different machine learning methods. Based on the feature vector of MRI images, (Sreeja & Mubarak, 2022) designed a Hounsfield unit (HU) conversion model using a fuzzy clustering algorithm for image segmentation and an adaptive neuro-fuzzy algorithm. Using MRI data and support vector machines (SVMs), (Xu et al., 2021) predicted the survival of glioblastoma patients. The creation of a feature vector was documented in reference (Kavitha, Shini, & Priya, 2022), which involved the utilization of diverse statistical texture analyses on brain lesions that were extracted from an MRI image. A study on pediatric neuro-oncology conducted at multiple centers employed a support vector machine (SVM) that was trained using 3D textural attributes derived from conventional MRI, as documented in reference (Fetit et al., 2018). The authors employed morphometric similarity mapping to derive various anatomical indices from individual cortical regions. Subsequently, they calculated the Pearson product moment correlation coefficient between each pair of MRI feature vectors for the respective regions. (Tian et al., 2020) In their study, the authors utilized a discrete wavelet packet transform in conjunction with Tsallis entropy and a generalized eigenvalue proximal support vector machine (GEPSVM) to develop an innovative computer-aided diagnosis (CAD) system for the differentiation of abnormal brains from healthy brains in MRI scans. This approach represents a novel contribution to the field of medical imaging analysis. The article referenced as (Sun, Morris, & Babyn, 2009) presents an extension of the optimal linear transformation (OLT) method, originally developed for magnetic resonance imaging (MRI), to functional magnetic resonance imaging (fMRI). The purpose of this extension is to improve the performance of conventional fMRI analysis techniques in identifying stimulation. The authors of reference (Chan, Chen, Wang, & King, 2019) employed a learning-by-example methodology to perform voxel segmentation and classification in in vivo cancer imaging. Specifically, a support vector machine (SVM) was trained to classify voxels based on the labels obtained from the clustering phase. The aforementioned studies showcase the capacity of machine learning methodologies in transforming MRI images into feature vectors that can be effectively utilized in diverse applications such as disease diagnosis and treatment approaches.

2.3. Machine Learning Algorithm

On the basis of a set of features or characteristics, algorithms that use machine learning for classification utilize them to foresee the class or grouping of a given input. These algorithms are extensively utilized in numerous disciplines, which include natural language processing, image detection and speech recognition, and the detection of fraud. Among the most prevalent machine learning algorithms for classification are statistical technique; Random Forest, Naive Bayes, K-Nearest Neighbors and Support Vector Machines (Mahesh, 2020a; Ayodele, 2010; Singh, Thakur, & Sharma, 2016).

The Random Forest algorithm is a type of collective learning approach that involves the creation of numerous decision trees during the training phase. The output of this algorithm is determined by the mode of the classes in the case of classification or the mean prediction in the case of regression, as determined by the individual trees (Gray et al., 2013). The decision tree is a graphical tool that represents decisions and their corresponding outcomes in a tree-like structure. The nodes depicted in the aforementioned graph signify an occurrence or determination, whereas the boundaries denote the norms or circumstances governing the decision. Each arboreal structure is comprised of interconnected nodes and branches. According to reference (Mahesh, 2020b), nodes in a classification group represent attributes, and each branch represents a potential value for the corresponding node. The Support Vector Machines (SVM) technique is a supervised machine learning approach that can be utilized for classification or regression tasks. Support Vector Machines (SVMs) are capable of generating a hyperplane or a series of hyperplanes that are suitable for classification in a space with a large number of dimensions (Awad & Khanna, 2015). The Naive Bayes algorithm is a statistical technique that utilizes Bayes' theorem to estimate the likelihood of a specific category based on the existence or non-existence of specific features (Saritas & Yasar, 2019). The K-Nearest Neighbors (KNN) algorithm is utilized for classification of a new data point, which relies on a group of its K-Nearest Neighbors in the area of features (Guo et al., 2003). The algorithms have versatile applications in classification tasks, and the selection of an appropriate algorithm is contingent upon the problem's characteristics and the data at hand.

3. Materials Method

3.1. Data Set

Alzheimer's disease is a progressive neurodegenerative disorder affecting millions of people worldwide. Early diagnosis, timely intervention, and management of this disease process are very important in Alzheimer's disease. In this study, we propose a transfer learning-based early detection approach for Alzheimer's disease using the moderate demented, mild demented, no demented, and very mild demented classification sets. In this context, the dataset obtained through Kaggle website consists of 6400 images in total. In the data set we obtained, the images in the train and test sections were combined and applied on the basis of random selection. The distribution of the images in the dataset is shown in Table 1.

Table 1. Dataset distribution of images by alzheimer class.

Data set percentage distribution rates			
Non-Demented	Very Mild Demented	Mild Demented	Moderate Demented
50%	35%	14%	1%

In the course of the initial development of machine learning models, the data set is split into the training data set and the test data set, and this is a vital concern for the process of learning as well as the evaluation of the model.

A training set is a segment of data that the machine learning algorithm is going to utilize within the training process. The machine learning algorithm learns relationships and patterns according to the data in the set used for training. The training set is the data collection utilized to develop and optimize the model. The set of tests is a distinct data set utilized to evaluate the efficacy of the model. This dataset is used to ascertain how the model performs on actual world data. The test set is used to evaluate the accuracy, precision, or other performance metrics of the model. The allocation of the data set is usually conducted in a randomized fashion. A certain percentage of the data set (typically 70%–80%) is utilized as the training set, while the remaining percentage (usually 20%–30%) is allocated as the test set. This procedure of separation provides a distinct data set that is used to evaluate the extension ability of the model. In short, if the data set is being utilized, it is split into two distinct portions: the "training set" and the "test set". The training set is used to train and optimize the model's parameters, whereas the test set is the data set assigned to assess the performance of the predictive algorithm.

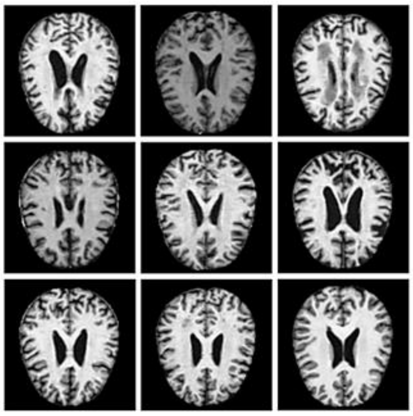


Figure 1. MRI images randomly selected from the dataset.

Figure 1 depicts a set of MRI images arbitrarily selected from the Alzheimer dataset. The images is jpeg format and represent various segments or views of the brain obtained via magnetic resonance imaging (MRI) technology. These images are essential for identifying multiple neurological disorders while examining the structures of the brain. The randomly chosen group of MRI images guarantees an assortment of the dataset and assists with training and evaluating machine learning algorithms for tasks that include image categorization, division, or identifying abnormalities in brain scans. In this investigation, a selection of brain scan images was used to classify Alzheimer's disease. A dataset was utilized. The dataset includes a total of 6,400 unique images. Each image depicts a distinct person at various phases of Alzheimer's disease progression and is labelled with one of four categories:

Mild Dementia: 896 images in this category show people with mild dementia.

Moderately Dementia: The category with the fewest images, 64, represents individuals with moderate dementia.

Non-Dementia: This category contains 3,200 images depicting individuals without dementia symptoms. This category consists of 2,240 individuals with a very moderate form of dementia. This consists of images.

Compared to the Mildly Demented and Moderately Demented categories, we observe that the dataset's distribution is asymmetrical. It should be noted that more images were discovered in the Non-Dementia and Very Mild Dementia categories. This imbalance can be identified during classification model training and evaluation. can result in some complications. Understanding the distribution of images within various categories is crucial for evaluating the representativeness and correctly interpreting classification results.

3.2. Performance metrics

3.2.1. F1 Score

The F1-score is a performance evaluation metric utilized in the field of machine learning to quantify the balance between precision and recall. The measure in question is the harmonic mean of the precision and recall metrics of the model. This metric is deemed to be balanced as it takes into account both precision and recall. The F1-score is calculated according to the following formula:

Precision refers to the proportion of accurate positive forecasts produced by the model, while recall pertains to the proportion of true positive cases in the dataset. The F1-score is a useful metric in the context of imbalanced datasets, where one class has a substantially larger number of instances than the other. Under such circumstances, the dependability of the model's performance as an indicator of accuracy may be compromised due to potential bias towards the majority class. The F1-score is deemed to be a more precise measure of a model's performance as it takes into account both precision and recall (Huang et al., 2015).

$$F1 - score = 2 * (precision * recall) / (precision + recall) \quad (1)$$

Equations 1 shows that F1 score formulas. The score for F1 is typically applied to binary classification tasks, but it can also be applied to multi-class classification tasks by calculating the F1-score for each class and then averaging them. In conclusion, the F1-score is a valuable performance evaluation metric in machine learning that provides a balanced evaluation of the model's precision and recall. It is especially valuable for unbalanced datasets and can be applied to multiclass classification tasks.

3.2.2. Recall

In machine learning, recall is a performance evaluation metric that measures the proportion of true positives among all positive instances in the dataset. Also called sensitivity or true positive rate. Calculating recall is as follows: Equation 2:

$$Recall = true\ positives / (true\ positives + false\ negatives) \quad (2)$$

During which true positives are the number of instances correctly classified as positive and false negatives are the number of instances incorrectly classified as negative when they are actually positive. Recall is especially useful in situations where the identification of positive instances is crucial, such as medical diagnosis and fraud detection. In such situations, a high recall is desired, as it indicates that the model correctly identifies the majority of positive instances despite producing some false positives.

However, it should be noted that a high recall rate may result in a significant number of false positives, which may pose challenges in specific use cases. Achieving a more thorough assessment of the model's performance necessitates the consideration of additional performance evaluation metrics, such as precision and F1-score, in conjunction with recall. Recall is a valuable metric for evaluating performance in machine learning, as it quantifies the ratio of true positives to genuine positive instances. The utilization of this technique proves to be particularly advantageous in situations where the accurate identification of affirmative occurrences holds paramount importance. However, it is recommended to supplement this metric with other evaluation criteria to obtain a more all-encompassing assessment of the model's performance, as suggested by reference (Acevedo, Jiménez-Valverde, Lobo, & Real, 2012).

3.2.3. Precision

In machine learning, recall is a performance evaluation metric that measures the percentage of true positives among all positive instances in the dataset. Also called sensitivity or true positive rate. Recall is especially useful in situations where the identification of positive instances is crucial, such as medical diagnosis and fraud detection. Calculating precision is as follows: Equation 3:

$$Precision = True\ positives / (True\ positives + False\ positives) \qquad (3)$$

An elevated recall shows that the predictive model correctly identifies the majority of positive instances in the dataset, whereas a low recall demonstrates that many positive instances are being missed (Sujatha & Mahalakshmi, 2020).

3.2.4. Confusuion Matrix

The confusion matrix is employed as a means of evaluating the effectiveness of a machine learning algorithm. The confusion matrix is a quantitative depiction of the correspondence between anticipated and observed classifications. In the matrix, the actual class is denoted by each row, while the predicted class is denoted by each column. The matrix displays the count of correct and incorrect forecasts produced by the model for every category.

3.3. Procedure

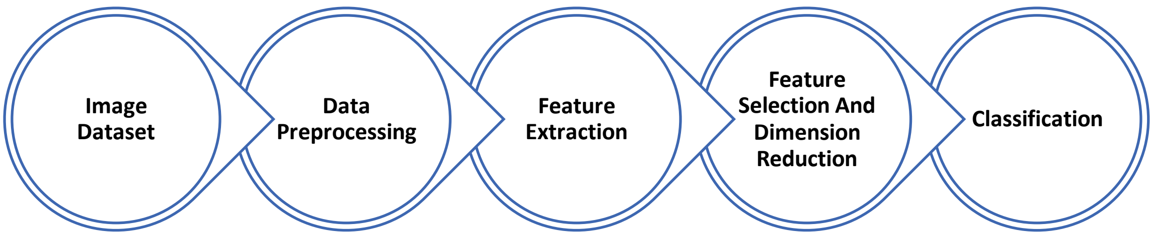


Figure 2. Procedure for alzheimer classification.

Figure 2 shows a process step consisting of the paths we have followed. These steps are, respectively:

Data preprocessing: Several pre-processing steps were undertaken to enhance the quality and compatibility of the data prior to classifying the images. Included in these stages are the following: All images were resized to a standard 224x224 pixel resolution. Resizing guarantees compatibility with the pre-trained models employed for feature extraction.

Normalisation: To facilitate optimal model performance and convergence, the pixel values of the images were normalised to a specific range (e.g., [0, 1]).

Using Pretrained Models for Feature Extraction: Using several pre-trained deep learning models, transfer learning was used to extract informative features from brain scan images. For feature extraction, the VGG19, ResNet50, InceptionV3, DenseNet121, and EfficientNet models were utilized. The pre-trained model was used to extract meaningful features from each image. The images were processed by the model, and the output of the final convolutional layer was used as a representation of feature vectors. This process transformed high-dimensional images into concise and informative representations of their features.

Data Partitioning: The dataset was separated into training and test sets so that classification models could be accurately evaluated. A common division ratio of 80:20 was employed, with 80% of the data reserved for model training and 20% for testing and performance evaluation. In order to assure impartial evaluations, stratified sampling was used to divide the data. By conserving the original class distribution in both the training and test sets, this technique avoids imbalance issues during model training and evaluation.

Selection of Features and Dimensionality Reduction: To reduce the dimensionality of the extracted features and mitigate the risk of overfitting, dimension reduction techniques such as Principal Component Analysis (PCA) were utilized. The training set was subjected to PCA by selecting the most pertinent components that encapsulate the majority of the data's variance. Depending on the investigation, the number of selected components can fluctuate; different values were tested, and for this study, 2000 components were chosen.

Using specified features from each pre-trained model, K-Nearest Neighbours (KNN), Support Vector Machines (SVM), Gaussian Naive Bayes (GNB), Decision Tree Classifier, Random Forest classifiers were trained and assessed. For each classifier, specific steps were followed:

Training data consisting of selected features obtained after dimensionality reduction were used to train each classifier.

- The classifiers were adapted to the training data by learning the underlying patterns and relationships between features and their corresponding Alzheimer's disease classes.

- Once the classifiers were trained, they were evaluated using the previously separated test data. The test data was fed to the trained classifiers to predict class labels.

- Performance metrics such as accuracy, precision, recall and F1-score were calculated to evaluate the classification performance of each model.

- Furthermore, confusion matrices were constructed to visualise the distribution of predicted classes and to identify misclassifications or biases in the models' predictions.

4. Result and Discussion

In this study, the KNN, SVM, Decisiontree, Randomforest, and GaussianNB algorithms are used to evaluate the efficacy of various deep learning models such as ResNet50, VGG, DenseNet121, EfficientNet, and InceptionNet; the evaluation metrics and confusion matrices are depicted in Figures 3-7.

Several evaluation metrics, including accuracy, F1 score, recall, and precision, are used to determine the efficacy of our proposed method. Confusion matrices were generated to obtain insight into the model's predictive capabilities and to identify potential misclassification areas.

Figure 3 illustrates the confusion matrices and evaluation metrics generated by the ResNet50 architecture and KNN algorithm. The confusion matrices demonstrate the efficacy of the model by comparing the distribution of predicted classes to the distribution of actual classes, with diagonal elements representing instances that were correctly classified. It is observed that its rate of accuracy is 0.97.

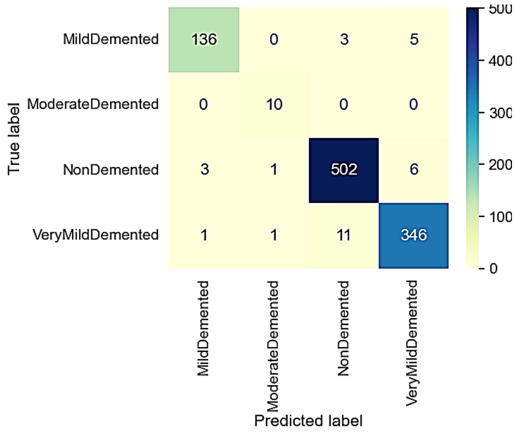
The accuracy, F1 score, and recall metrics for each ResNet50 architecture SVM model are displayed in Figure 4. These metrics assess classification performance quantitatively. The accuracy metric represents the overall correct classification rate, whereas the F1 score evaluates the model's ability to correctly classify positive and negative examples by incorporating both precision and recall. The recall metric measures the model's ability to accurately recognize all positive examples. By comparing these metrics across various models, we can determine which architecture provides the highest levels of precision and diagnostic capability. It is evident that its accuracy rate is 0.98. It is observed that its rate of accuracy is 0.98.

Figure 5 demonstrates the accuracy, F1 score, recall, and confusion matrix for the ResNet50 architecture GaussianNB model. It has been determined that its accuracy rate is 0.61.

The accuracy, F1 score, and recall measurements, as well as the confusion matrix, for the ResNet50 architecture Decisiontree model are displayed in Figure 6. It is observed that its accuracy is 0.69.

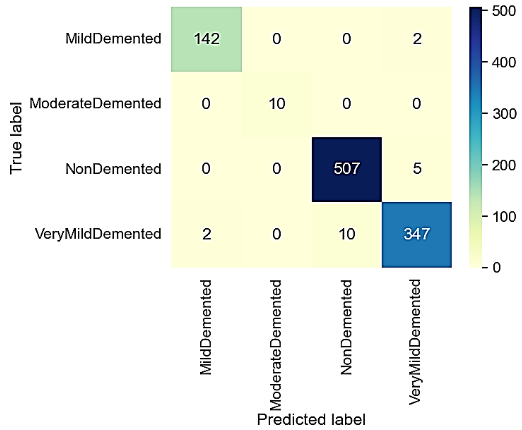
Figure 7 depicts the accuracy, F1 score, and recall measurements, as well as the confusion matrix, for the Randomforest model based on the ResNet50 architecture. It is observed that its rate of accuracy is 0.88.

The evaluation metrics and visualisations presented in this study provide a comprehensive analysis of the performance of ResNet50, VGG, DenseNet121, EfficientNet, and InceptionNet architectures in the classification of Alzheimer's disease severity. These findings contribute to our comprehension of the strengths and limitations of each model, aiding in the development of more accurate and reliable Alzheimer's disease diagnostic instruments.



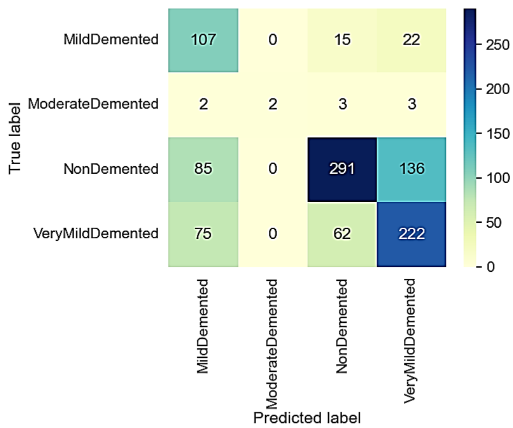
	precision	recall	f1-score	support
MildDemented	0.97	0.94	0.96	144
ModerateDemented	0.83	1.00	0.91	10
NonDemented	0.97	0.98	0.98	512
VeryMildDemented	0.97	0.96	0.97	359
accuracy			0.97	1025
macro avg	0.94	0.97	0.95	1025
weighted avg	0.97	0.97	0.97	1025

Figure 3. Resnet 50 KNN algorithm confusion matrix and performance metric.



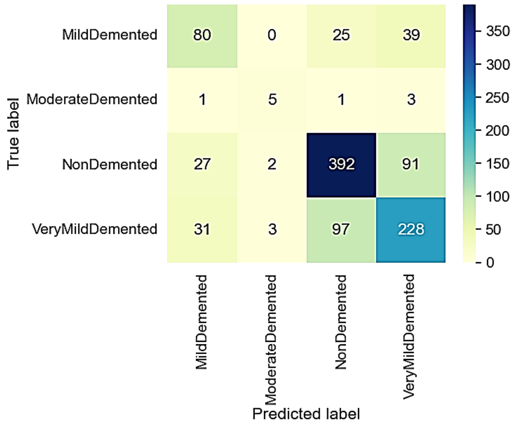
	precision	recall	f1-score	support
MildDemented	0.99	0.99	0.99	144
ModerateDemented	1.00	1.00	1.00	10
NonDemented	0.98	0.99	0.99	512
VeryMildDemented	0.98	0.97	0.97	359
accuracy			0.98	1025
macro avg	0.99	0.99	0.99	1025
weighted avg	0.98	0.98	0.98	1025

Figure 4. Resnet 50 SVM algorithm confusion matrix and performance metric.



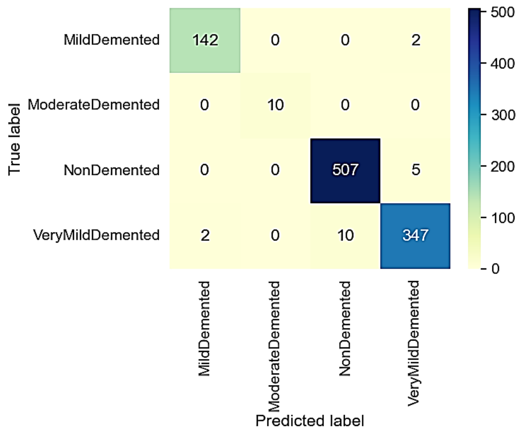
	precision	recall	f1-score	support
MildDemented	0.40	0.74	0.52	144
ModerateDemented	1.00	0.20	0.33	10
NonDemented	0.78	0.57	0.66	512
VeryMildDemented	0.58	0.62	0.60	359
accuracy			0.61	1025
macro avg	0.69	0.53	0.53	1025
weighted avg	0.66	0.61	0.61	1025

Figure 5. Resnet 50 GaussianNB algorithm confusion matrix and performance metric.



	precision	recall	f1-score	support
MildDemented	0.58	0.56	0.57	144
ModerateDemented	0.50	0.50	0.50	10
NonDemented	0.76	0.77	0.76	512
VeryMildDemented	0.63	0.64	0.63	359
accuracy			0.69	1025
macro avg	0.62	0.61	0.62	1025
weighted avg	0.69	0.69	0.69	1025

Figure 6. Resnet 50 Decision tree algorithm confusion matrix and performance metric.



	precision	recall	f1-score	support
MildDemented	0.99	0.99	0.99	144
ModerateDemented	1.00	1.00	1.00	10
NonDemented	0.98	0.99	0.99	512
VeryMildDemented	0.98	0.97	0.97	359
accuracy			0.98	1025
macro avg	0.99	0.99	0.99	1025
weighted avg	0.98	0.98	0.98	1025

Figure 7. Resnet 50 random forest algorithm confusion matrix and performance metric.

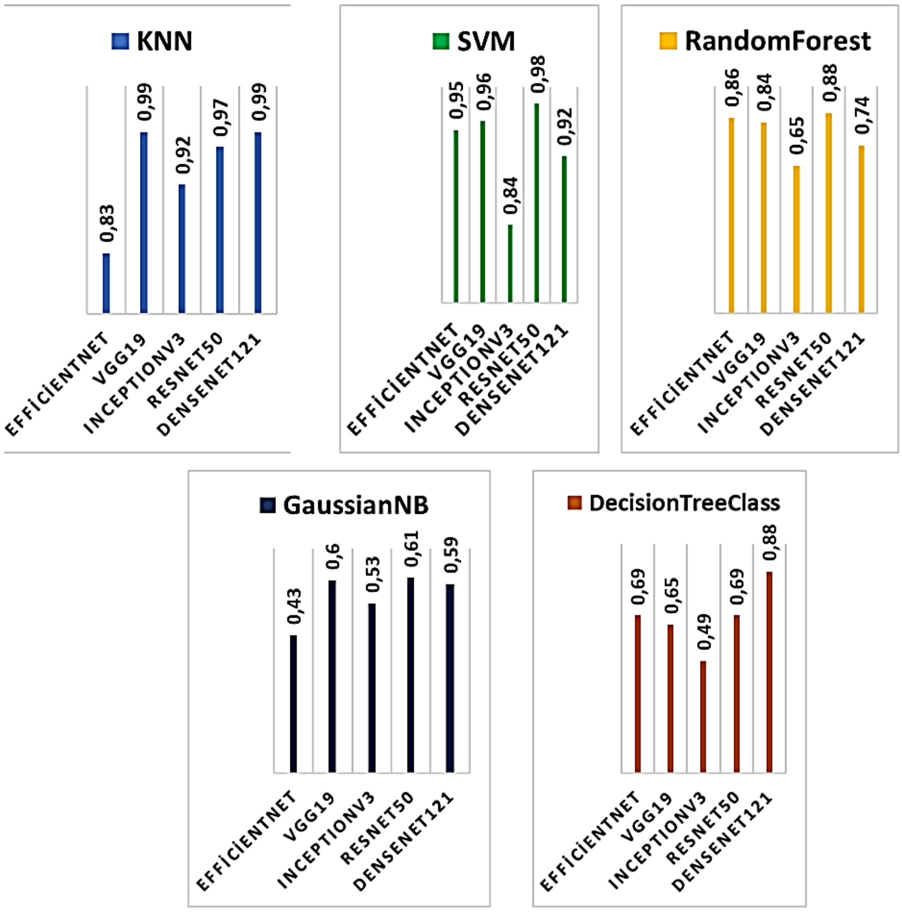


Figure 8. Algorithm models accuracy comparision.

Figure 8 shows the performance of different deep learning models such as ResNet50, VGG, DenseNet121, EfficientNet, and InceptionNet with accuracy ratios obtained using KNN, SVM, decision tree, random forest, and GaussianNB algorithms. The best accuracy ratios for the KNN model are densenet121 and resnet50. For the SVM model, the best accuracy ratio is RESNET50 with 0.98. For Random Forest, Resnet50 was 0.88 with accuracy, while for GaussianNB and Decision Tree, it was RESNET50 with 0.61 and 0.69, respectively.

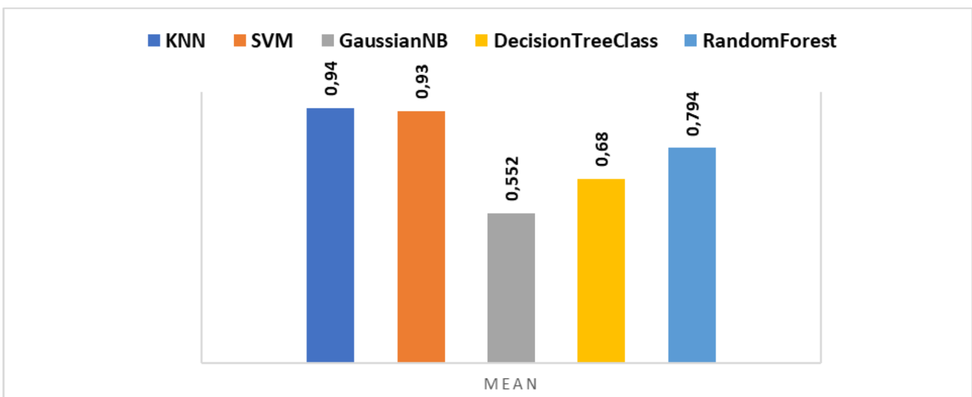


Figure 9. Algorithm models accuracy means.

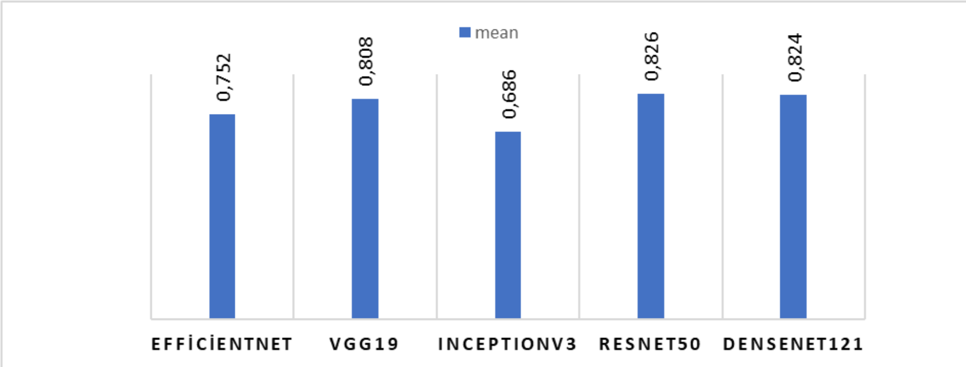


Figure 10. Architectures accuracy means.

Figure 9 shows the performance of ResNet50, VGG, DenseNet121, EfficientNet, and InceptionNet deep learning models with mean accuracy ratios obtained using KNN, SVM, decision tree, random forest, and GaussianNB algorithms. This figure shows the best mean accuracy is KNN algorithm as like show in the figure 0,94 accuracy.

Figure 10 shows the performance of with mean accuracy ratios obtained using KNN, SVM, decision tree, random forest, and GaussianNB algorithms with using ResNet50, VGG, DenseNet121, EfficientNet, and InceptionNet deep learning models. This figure shows the best mean accuracy is transfer learning algorithm is Resnet50 as like show in the figure 11 0,8226 accuracy.

Investigation on the same dataset in literature, Mggdadi et al. (Mggdadi et al., 2021) obtained a maximum accuracy of 49.9% with the Adam optimizer and 70.3% with the VGG16 model. Acharya et al. (Acharya, Mehta, & Singh, 2021) reported accuracy rates of 88.89% with CNN, 85.07% with VGG-16, 75.25% with ResNet50, and 95.70% with Modified AlexNet. Assmi et al. (Assmi, Elhabyb, Benba, & Jilbab, 2024) observed overall accuracy rates for models that VGG-19 (92.86%), VGG-16 (92.83%), Inception-V3 (91.04%), Xception (90.57%), ResNet-50 (85.99%), and DenseNet169 (88.64%).

5. Conclusion and Recommendations

Transfer learning is used in this investigation to derive features from brain MRI images. For Alzheimer's stage classification, these features were used to train and assess traditional machine learning algorithms such as Support Vector Machines (SVM) and K-Nearest Neighbors (KNN). Our objective was to assess the efficacy of transfer learning in conjunction with these algorithms for the classification of dementia using brain MRI data.

Utilizing brain MRI images to classify dementia stages, our experimental analysis revealed that transfer learning significantly enhances the performance of both SVM and KNN classifiers. The extracted features capture crucial patterns and information, allowing classifiers to achieve greater accuracy, precision, recall, and F1 scores than when using raw image data.

VGG19 and ResNet50 consistently outperformed the other architectures, obtaining superior precision, recall, and F1 scores across most classes. EfficientNetB7 and EfficientNet0 also demonstrated promising performance, particularly when paired with SVM and Random Forest classifiers. InceptionV3 demonstrated average performance, whereas GaussianNB struggled to provide precise classifications across all architectures. It is essential to note, however, that the performance of these models can be affected by the extent and quality of the dataset as well as the characteristics of the population under study. To determine the generalizability of these results, additional research and validation on larger and more diverse datasets are required.

In conclusion, our research emphasizes the promising potential of deep learning models for the classification of dementia using brain MRI images. The superior performance of these models opens up new avenues for the development of accurate and trustworthy dementia diagnostic instruments. Continued advances in deep learning techniques and access to larger datasets will enhance the accuracy and clinical applicability of these models.

References

Alzheimer's Association. (2016). 2016 Alzheimer's disease facts and figures. Alzheimer's & Dementia, 12(4), 459-509. https://doi.org/10.1016/j.jalz.2016.03.001

Ribe, E. M., & Lovestone, S. (2016). Insulin signalling in Alzheimer's disease and diabetes: from epidemiology to molecular links. Journal of internal medicine, 280(5), 430–442. https://doi.org/10.1111/joim.12534

Qiu, C., Kivipelto, M., & von Strauss, E. (2009). Epidemiology of Alzheimer's disease: occurrence, determinants, and strategies toward intervention. Dialogues in clinical neuroscience, 11(2), 111–128. https://doi.org/10.31887/DCNS.2009.11.2/cqiu

Hayajneh, F. A., & Shehadeh, A. (2014). The impact of adopting person-centred care approach for people with Alzheimer's on professional caregivers' burden: an interventional study. International journal of nursing practice, 20(4), 438–445. https://doi.org/10.1111/ijn.12251

Scott, C. B. (2013). Alzheimer’s Disease Caregiver Burden: Does Resilience Matter? Journal of Human Behavior in the Social Environment, 23(8), 879–892. https://doi.org/10.1080/10911359.2013.803451

Winblad, B., Brodaty, H., Gauthier, S., Morris, J. C., Orgogozo, J. M., Rockwood, K., Schneider, L., Takeda, M., Tariot, P., & Wilkinson, D. (2001). Pharmacotherapy of Alzheimer's disease: is there a need to redefine treatment success?. International journal of geriatric psychiatry, 16(7), 653–666. https://doi.org/10.1002/gps.496

Korolev, S., Safiullin, A., Belyaev, M., & Dodonova, Y. (2017). Residual and plain convolutional neural networks for 3D brain MRI classification. In 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017) (pp. 835-838). Melbourne, VIC, Australia: IEEE. https://doi.org/10.1109/ISBI.2017.7950647

Islam, J., & Zhang, Y. (2018). Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. Brain Informatics, 5(2). https://doi.org/10.1186/s40708-018-0080-3

Platero, C., Lin, L., & Tobar, M. C. (2019). Longitudinal Neuroimaging Hippocampal Markers for Diagnosing Alzheimer's Disease. Neuroinformatics, 17(1), 43–61. https://doi.org/10.1007/s12021-018-9380-2

Zhan, L., Liu, Y., Wang, Y., Zhou, J., Jahanshad, N., Ye, J., Thompson, P. M., & Alzheimer's Disease Neuroimaging Initiative (ADNI) (2015). Boosting brain connectome classification accuracy in Alzheimer's disease using higher-order singular value decomposition. Frontiers in neuroscience, 9, 257. https://doi.org/10.3389/fnins.2015.00257

He, J., Zhou, C., Ma, X., Berg-Kirkpatrick, T., & Neubig, G. (2021). Towards a unified view of parameter-efficient transfer learning. arXiv preprint arXiv:2110.04366. https://doi.org/10.48550/arXiv.2110.04366

Wang, Y., Song, Y., Xie, H., Li, W., Hu, B., & Yang, G. (2017). Reduction of Gibbs artifacts in magnetic resonance imaging based on Convolutional Neural Network. 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 1-5. doi: 10.1109/CISP-BMEI.2017.8302197.

Varuna Shree, N., & Kumar, T. N. R. (2018). Identification and classification of brain tumor MRI images with feature extraction using DWT and probabilistic neural network. Brain informatics, 5(1), 23–30. https://doi.org/10.1007/s40708-017-0075-5

Vlaardingerbroek, M. T., & Boer, J. A. (2013). Magnetic resonance imaging: theory and practice. Springer Science & Business Media.

Zulkoffli, Z., & Shariff, T. A. (2019). Detection of brain tumor and extraction of features in MRI images using K-means clustering and morphological operations. In 2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS) (pp. 1-5). Selangor, Malaysia: IEEE. https://doi.org/10.1109/I2CACIS.2019.8825094

Higaki, T., Nakamura, Y., Tatsugami, F., Nakaura, T., & Awai, K. (2019). Improvement of image quality at CT and MRI using deep learning. Japanese journal of radiology, 37(1), 73–80. https://doi.org/10.1007/s11604-018-0796-2

Sreeja, S., Mubarak, D.M.N. (2023). Pseudo-CT Generation from MRI Images for Bone Lesion Detection Using Deep Learning Approach. In: Ranganathan, G., Bestak, R., Fernando, X. (eds) Pervasive Computing and

Social Networking. Lecture Notes in Networks and Systems, vol 475. Springer, Singapore. https://doi.org/10.1007/978-981-19-2840-6_47

Xu, Y., He, X., Li, Y., Pang, P., Shu, Z., & Gong, X. (2021). The Nomogram of MRI-based Radiomics with Complementary Visual Features by Machine Learning Improves Stratification of Glioblastoma Patients: A Multicenter Study. Journal of magnetic resonance imaging : JMRI, 54(2), 571–583. https://doi.org/10.1002/jmri.27536

Kavitha, P. Subha, R. Priya, R. (2021). An Implementation Of Statistical Feature Algorithms For The Detection Of Brain Tumor. Journal of Journal of Cognitive Human-Computer Interaction, 1(2), 57 - 62. DOI: DOI: https://doi.org/10.54216/JCHCI.010202

Fetit, A. E., Novak, J., Rodriguez, D., Auer, D. P., Clark, C. A., Grundy, R. G., Peet, A. C., & Arvanitis, T. N. (2018). Radiomics in paediatric neuro-oncology: A multicentre study on MRI texture analysis. NMR in biomedicine, 31(1), 10.1002/nbm.3781. https://doi.org/10.1002/nbm.3781

Tian, T., Li, J., Zhang, G., Wang, J., Liu, D., Wan, C., Fang, J., Wu, D., Zhou, Y., & Zhu, W. (2020). Effects of childhood trauma experience and COMT Val158Met polymorphism on brain connectivity in a multimodal MRI study. Brain and behavior, 10(12), e01858. https://doi.org/10.1002/brb3.1858

Sun, F., Morris, D., & Babyn, P. (2009). The optimal linear transformation-based fMRI feature space analysis. Medical & biological engineering & computing, 47(11), 1119–1129. https://doi.org/10.1007/s11517-009-0504-6

Chan, H. P., Chen, W., Wang, L., & King, I. (2019). Neural keyphrase generation via reinforcement learning with adaptive rewards. *arXiv preprint arXiv:1906.04106*. https://doi.org/10.48550/arXiv.1906.04106

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR), 9*(1), 381-386.

Ayodele, T. O. (2010). Types of machine learning algorithms. In New Advances in Machine Learning (pp. 3-19). InTech.

Singh, A., Thakur, N., & Sharma, A. (2016). A review of supervised machine learning algorithms. 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), 1310-1315.

Gray, K. R., Aljabar, P., Heckemann, R. A., Hammers, A., Rueckert, D., & Alzheimer's Disease Neuroimaging Initiative (2013). Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. NeuroImage, 65, 167–175. https://doi.org/10.1016/j.neuroimage.2012.09.065

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR), 9*(1), 381-386.

Awad, M., Khanna, R., Awad, M., & Khanna, R. (2015). Support vector machines for classification. In Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers (pp. 39-66). Apress.

Saritas, M. M., & Yasar, A. (2019). Performance analysis of ANN and Naive Bayes classification algorithm for data classification. International journal of intelligent systems and applications in engineering, 7(2), 88-91.

Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K. (2003). KNN Model-Based Approach in Classification. In: Meersman, R., Tari, Z., Schmidt, D.C. (eds) On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. OTM 2003. Lecture Notes in Computer Science, vol 2888. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-39964-3_62

Huang, H., Xu, H., Wang, X., & Silamu, W. (2015). Maximum F1-score discriminative training criterion for automatic mispronunciation detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23*(4), 787-797.

Acevedo, P., Jiménez-Valverde, A., Lobo, J. M., & Real, R. (2012). Delimiting the geographical background in species distribution modelling. *Journal of Biogeography, 39*(8), 1383-1390.

Sujatha, P., & Mahalakshmi, K. (2020, November). Performance evaluation of supervised machine learning algorithms in prediction of heart disease. In *2020 IEEE International Conference for Innovation in Technology (INOCON)* (pp. 1-7). IEEE.

Mggdadi, E., Al-Aiad, A., Al-Ayyad, M. S., & Darabseh, A. (2021, May). Prediction Alzheimer's disease from MRI images using deep learning. In *2021 12th International Conference on Information and Communication Systems (ICICS)* (pp. 120-125). IEEE.

Acharya, H., Mehta, R., & Singh, D. K. (2021, April). Alzheimer disease classification using transfer learning. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1503-1508). IEEE.

Assmi, A., Elhabyb, K., Benba, A., & Jilbab, A. (2024). Alzheimer's disease classification: A comprehensive study. *Multimedia Tools and Applications, 83*(3), 4567-4590.