Diagnosing Alzheimer's Disease Severity: A Comparative Study of Deep Learning Algorithms

Ibrahem Mohammed Madhat[1,2], *Karrar* N. Kadhim[1,3], *Farhan* Mohamed[1], *Mohd* Shafry Mohd Rahim[1], *Fallah* H. Najjar[1,4]* and *Ali* J. Ramadhan[5]

[1]Department of Emerging Computing, Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru, Malaysia
[2]Office Affairs Department, University of Technology, Baghdad, Iraq
[3]Computer Techniques Engineering Department, Faculty of Information Technology, Imam Ja'afar Al-Sadiq University, Baghdad, Iraq
[4]Department of Computer System Techniques, Technical Institute of Najaf, Al-Furat Al-Awsat Technical University, Najaf, Iraq
[5]Department of Computer Techniques Engineering, College of Technical Engineering, University of Alkafeel, Najaf, Iraq

Abstract. Alzheimer's disease emerges as a profoundly distressing neurological condition affecting older individuals, pre-ending itself as an insufficiently addressed and often overlooked ailment that poses a growing concern for public health. In the past decade, there has been a notable surge in endeavors aimed at unraveling the disease's origins and devising pharmacological interventions. Recent advancements encompass enhanced clinical diagnostic criteria and refined approaches for managing cognitive impairments and behavioral challenges. The pursuit of symptomatic relief primarily centered on cholinergic therapy has been subject to rigorous scrutiny through randomized, double-blind, placebo-controlled studies assessing cognitive function, daily activities, and behavioral aspects. This research delves into the utilization of diverse algorithms for the classification of Alzheimer's disease severity, employing CNN, DenseNet, VGG19, and ensemble learning approaches. The obtained accuracy scores underscore the supremacy of the Ensemble model, surpassing the performance of the other models with an impressive accuracy level of 94%.

1. Introduction

Alzheimer's disease (AD) is a chronic and incapacitating neurodegenerative disorder characterized by relentless progression and irreversibility [1]. As the most prevalent form of dementia, a cognitive disorder frequently observed among older individuals, AD is distinguished by profound memory loss, cognitive impairment, and a gradual decline in the ability to independently perform daily tasks [2]. Globally, the prevalence of AD is a matter of great concern, with projections indicating a surge from the current 47 million affected individuals to an astonishing 152 million by 2050 [3].

* Corresponding author: fallahnajjar@atu.edu.iq

This gives rise to significant socio-economic and medical challenges that demand immediate attention.

Despite extensive research efforts, the fundamental causes of AD remain enigmatic. Regrettably, existing treatments are unable to cure the condition or entirely halt its advancement. A pivotal stage in the progression of AD is marked by amnestic mild cognitive impairment (MCI), a transitional phase between normal aging and AD [4]. Individuals with MCI face a substantially higher risk of developing AD compared to those with normal cognitive function. As AD advances, its symptoms become increasingly incapacitating, ultimately resulting in a complete loss of communication, mobility, and the ability to adapt to the surrounding environment [5, 6].

The early identification of Alzheimer's disease (AD) through the screening of mild cognitive impairment (MCI) is of utmost significance to effectively manage the disease and slow down its advancement. This not only contributes to the development of innovative interventions and medications but also empowers patients to comprehend the gravity of their condition and adopt appropriate preventive measures like lifestyle adjustments and prescribed treatments [7].

Magnetic Resonance Imaging (MRI) has emerged as a pivotal tool for AD diagnosis, owing to its exceptional spatial resolution and ability to distinguish soft tissues. In comparison to other imaging modalities like Computed Tomography (CT) and Positron Emission Tomography (PET), MRI is regarded as a safer option with fewer associated health risks [8, 9]. Nevertheless, the interpretation of intricate and extensive MRI datasets can be challenging and time-intensive for medical professionals, which can lead to errors and inconsistencies [10]. Consequently, there is a growing demand for automated segmentation techniques capable of providing swift, accurate, and dependable results. Remarkable progress has been achieved in employing machine learning and deep learning approaches to aid in AD diagnosis. These methodologies, discussed in the subsequent section, hold tremendous potential to transform the diagnosis and management of this progressive ailment.

2. Related Works

Numerous investigations have been conducted in the field of Alzheimer's disease to enhance the identification and diagnosis process. In a recent study by Gupta and colleagues [11], various machine learning models were assessed for their ability to detect dementia using the Open Access Series of Imaging Studies dataset. The findings revealed that the Deep Learning Ludwig Classifier achieved the highest accuracy (95%) among all tested models, effectively distinguishing patients with dementia. Another research group, led by El-Sappagh et al. [12], proposed a dual-stage deep learning framework to monitor the progression of AD. This approach leveraged a diverse range of patient data, including cognitive scores, biomarkers, demographics, and neuroimaging. Notably, their LSTM model exhibited promising outcomes, surpassing alternative models in terms of accuracy and mean absolute error. Cheung and collaborators [13] introduced an innovative deep learning algorithm that operated on retinal photographs. Their dataset encompassed a substantial number of images from both Alzheimer's disease (AD) patients and healthy individuals. By utilizing the EfficientNet-b2 network as its foundation, their model exhibited impressive accuracy and AUROCs, ranging from 79.6% to 92.1% and 0.73 to 0.91, respectively. Venugopalan et al. [14] took a comprehensive approach by integrating MRI data, clinical test data, and genetic information (precisely, single nucleotide polymorphisms or SNPs) to discern between AD, mild cognitive impairment (MCI), and control subjects. Their methodology incorporated stacked diagnosing auto-encoders and 3D convolutional neural networks (3D-CNNs) to extract pertinent features. Notably, their study indicated the superiority of deep models over more conventional approaches such as support vector machines, decision trees, and k-nearest neighbors.

Similarly, Nanni et al. [15] harnessed the potential of T1-weighted MRI scans to train multiple models, including an ensemble of five transfer-learning architectures, a 3D convolutional neural network (3D CNN), and a fusion of two machine learning classifiers. Their ensemble transfer-learning strategy showcased promising accuracy in distinguishing between AD and cognitively normal (CN) patients, as well as identifying MCI patients who are likely to convert or not convert to AD. Martinez-Murcia and co-authors [16] employed deep convolutional auto-encoders to investigate data associated with Alzheimer's disease (AD). Through this approach, they successfully identified distinctive features reflecting the progression of neurodegeneration and cognitive impairments. These identified features were then utilized for regression and classification analyses. Notably, their study demonstrated that a fusion of imaging-derived markers and cognitive scores could predict AD diagnosis with impressive accuracy. Mehmood et al. [17] introduced an early detection method for AD, leveraging layer-wise transfer learning and brain image tissue segmentation. Their innovative approach yielded substantial classification accuracy, outperforming existing state-of-the-art models in the field. Acharya and collaborators [18] developed a Computer-Aided Brain Diagnosis (CABD) system tailored to identify AD indicators in MRI scans. Their methodology achieved remarkable levels of accuracy, precision, sensitivity, and specificity, underscoring its effectiveness in aiding AD diagnosis. Salehi et al. [19] conducted research that implemented a Convolutional Neural Network (CNN) for AD diagnosis based on MRI images. Their CNN model exhibited a remarkable accuracy of 99%, showcasing the potential of deep learning methodologies, particularly with extensive datasets, in surpassing the performance of traditional machine learning techniques.

Table 1. Literature Review on Alzheimer's Disease Detection

Study	Approach	Data Used	Key Findings
Gupta et al. (2023) [11]	Machine learning models on imaging data	Open Access Series of Imaging Studies dataset	Deep Learning Ludwig Classifier achieved 95% accuracy in detecting dementia.
El-Sappagh et al. (2022) [12]	Dual-stage deep learning framework	Cognitive scores, biomarkers, demographics, neuroimaging	LSTM model demonstrated promising outcomes, outperforming other models in accuracy and mean absolute error.
Cheung et al. (2022) [13]	Deep learning algorithm on retinal photographs	Images from AD patients and healthy individuals	EfficientNet-b2-based model achieved accuracy of 79.6% - 92.1% and AUROCs of 0.73 - 0.91.
Venugopala et al. 2021 [14]	integration of MRI, clinical, and genetic data	MRI, clinical test data, SNPs	Deep models (stacked auto-encoders, 3D-CNNs) outperformed traditional methods in AD, MCI, and control differentiation.
Nanni et al. (2020) [15]	Multiple models on T1-weighted MRI scans	T1-weighted MRI scans	The ensemble transfer-learning approach showed promise in distinguishing AD from CN patients and predicting MCI conversion.
Martinez-Murcia et al. (2020) [16]	Deep convolutional auto-encoders	AD-related data	Identified features for neurodegeneration progression; fusion of imaging-derived markers and cognitive scores predicted AD with over 80% accuracy.
Mehmood et al. 2021 [17]	Early AD detection with transfer learning and tissue segmentation	Brain images	Achieved high classification accuracy, surpassing state-of-the-art models.

Study	Approach	Data Used	Key Findings
Acharya et al. 2019 [18]	Computer-Aided Brain Diagnosis (CABD) system	MRI scans	High accuracy, precision, sensitivity, and specificity in identifying AD signs.
Salehi et al. 2020 [19]	Convolutional Neural Network (CNN) for AD diagnosis	MRI images	CNN model achieved 99% accuracy, highlighting the power of deep learning with large datasets.

3. DATASET

This dataset comprises a curated collection of MRI images sourced from diverse websites, where each image's label has undergone meticulous verification. The dataset is categorized into four distinct classes, which are consistently present in both the training and testing sets. These classes correspond to different stages of dementia progression:

- Mild Demented: Individuals with subtle cognitive impairments affecting memory and daily tasks.
- Moderate Demented: More pronounced cognitive decline, including memory loss and communication difficulties
- NonDemented: Healthy cognitive function, able to perform daily activities without significant challenges.
- Very Mild Demented: Early signs of cognitive decline, such as mild memory lapses and concentration difficulties.

The primary motivation behind sharing this dataset is to foster the development of highly accurate predictive models capable of discerning and predicting the various stages of Alzheimer's disease. By providing a comprehensive range of images that reflect the different levels of dementia severity, this dataset offers an opportunity for researchers and machine learning practitioners to contribute to the advancement of Alzheimer's detection and prognosis. It serves as an invaluable resource for inspiring innovations that may ultimately lead to more effective diagnostic and therapeutic strategies, potentially transforming the landscape of Alzheimer's disease management.

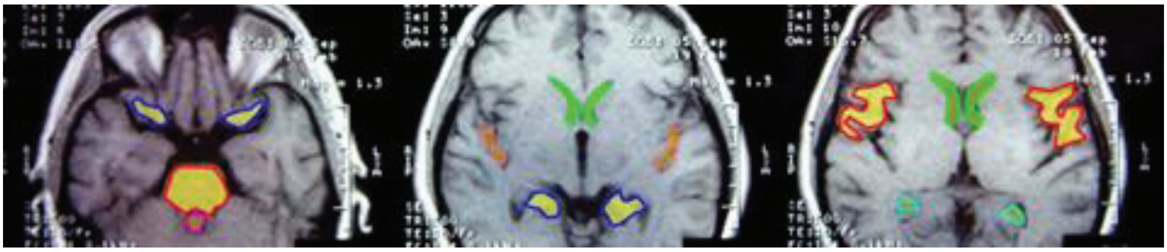

Fig. 1. Alzheimer's Image

4. Methodology

In this study, we present a comprehensive methodology for the early detection of Alzheimer's disease using deep learning techniques applied to MRI images. The dataset consists of images categorized into four classes representing different stages of the disease. Prior to model training, the data undergoes preprocessing, including resizing to a standardized 124x124 pixel dimension and normalization of pixel values to the range [0, 1]. Visualization of the data distribution provides initial insights into class imbalances. The dataset is then split into training and testing sets, with 80% used for training to ensure robust model learning. We employ a range of deep learning architectures, including Convolutional Neural Networks

(CNN), DenseNet, and VGG16, chosen for their prowess in image analysis. To further enhance performance, an ensemble learning approach is explored, leveraging the collective strengths of different models. The efficacy of the models is evaluated using various metrics, such as accuracy, precision, recall, F1-score, and AUC-ROC, which collectively provide a comprehensive assessment of their predictive capabilities. By following this structured methodology, we aim to contribute to the advancement of Alzheimer's disease detection, potentially leading to early intervention and improved patient care.

Fig. 2. Some Images From the Dataset

Data Preprocessing and Visualization

The dataset utilized in this study comprises MRI images categorized into four classes related to Alzheimer's disease: Mild Demented, Moderate Demented, Non Demented, and Very Mild Demented. In order to prepare the data for model training, each image is resized to a consistent dimension of 124x124 pixels. Prior to feeding the images into the models, pixel values are normalized by dividing them by 255.0 to ensure consistent scaling across the dataset.

The dataset is organized with corresponding file paths and class labels. Visualization of the data distribution provides an initial insight into the distribution of different classes, highlighting any potential imbalances that may need to be addressed during model evaluation.

Moreover, to assess the performance of the models accurately, the dataset is split into training and testing sets. The training set encompasses 80% of the data, allowing the models to learn patterns and features from a diverse range of images. The remaining 20% of the data is reserved for testing, enabling the evaluation of the models' generalization capabilities on unseen data. In the Figure 4, we can discover some examples of the dataset.

 https://doi.org/10.1051/bioconf/20249700102

Models Implementation

Several deep learning models are employed to detect Alzheimer's disease from MRI images. The models include Convolutional Neural Networks (CNN), DenseNet, and VGG16. These architectures are chosen due to their proven success in image classification tasks and their potential to capture intricate features from medical images. Each model is implemented using a suitable deep learning framework and configured with appropriate hyperparameters.

In pursuit of enhanced performance, an ensemble learning approach is explored. Multiple models are combined to create a robust ensemble, which can potentially capture diverse aspects of the data and improve overall prediction accuracy. The ensemble method allows leveraging the strengths of different models to mitigate individual weaknesses and enhance the overall predictive power.

• CNN model: it constitutes a cornerstone of our methodology for Alzheimer's disease detection. Designed to emulate the human visual perception system, CNNs excel in capturing intricate patterns and features from image data. The CNN architecture consists of layers that perform convolution, pooling, and fully connected operations, enabling it to automatically learn hierarchies of features relevant to the task at hand. In our study, the CNN is customized and fine-tuned to extract salient features from MRI images associated with different stages of Alzheimer's disease.

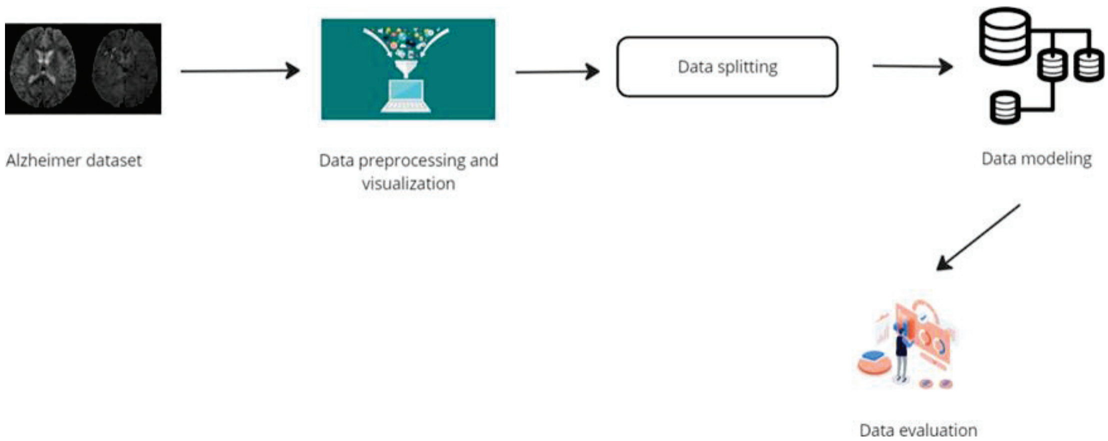

Fig. 3. Alzheimer Disease Methodology

The model's ability to detect subtle variations in brain structures is harnessed to identify indicative patterns that distinguish between healthy and diseased conditions. By leveraging CNN's inherent capacity to uncover complex relationships within the data, we aim to achieve accurate and reliable predictions, contributing to the advancement of early Alzheimer's disease diagnosis and intervention strategies.

Fig. 4. Data Splitting

• DensNet model: The model stands as a vital pillar within our methodology for detecting Alzheimer's disease. Renowned for its innovative architecture, DenseNet promotes enhanced feature propagation and alleviates vanishing gradient issues through dense connections between layers. This unique design fosters a rich information flow throughout the network, facilitating the extraction of intricate patterns from MRI images. In our study, the DenseNet model is adapted and fine-tuned to effectively capture the nuanced structural differences indicative of Alzheimer's disease progression. By fostering inter-layer communication and feature reuse, DenseNet maximizes the model's capability to discern subtle variations in brain tissue composition and geometry.
• VGG19 model: this model, a prominent member of the VGG (Visual Geometry Group) family of convolutional neural networks, plays a pivotal role in our methodology for Alzheimer's disease detection. Renowned for its deep architecture and simplicity, VGG19 consists of multiple stacked convolutional and pooling layers, making it adept at capturing intricate features from complex image data. In our study, the VGG19 model is tailored to scrutinize MRI images for distinctive patterns associated with different stages of Alzheimer's disease. Its depth allows it to capture both low-level and high-level features, enabling the identification of subtle structural changes indicative of disease progression.
• Ensemble learning: It is a powerful strategy within our Alzheimer's disease detection methodology, seamlessly integrating the strengths of three distinctive deep learning architectures: CNN, DenseNet, and VGG19. By combining these models through ensemble learning, we aim to leverage their diverse capabilities in order to enhance the accuracy and robustness of our predictive framework. Each model contributes a unique perspective, collectively capturing a comprehensive range of intrinsic features within MRI images that denote different stages of Alzheimer's disease. The ensemble learning approach intelligently aggregates the predictions from CNN, DenseNet, and VGG19, harnessing their collective insights to make more informed and accurate decisions. This collaborative effort optimizes the sensitivity to subtle patterns and structural nuances indicative of disease progression. The fusion of these three powerful models strengthens our methodology's capacity to discern even the most subtle manifestations of Alzheimer's disease, potentially revolutionizing early diagnosis and offering a more comprehensive understanding of patients' conditions.

5. Model Training and Evaluation

As mentioned before, the dataset is saved in a CSV file format, and it is split 80% and 20% for training and testing sets, respectively.
To conduct our machine learning investigation, we utilized Google Colab, a collaborative Jupyter notebook environment. The implementation was carried out using Keras (version 2.12.0), a Python-based open-source deep learning framework developed atop Google's data flow software, and TensorFlow (version 2.12.0). Within the Google Colab workspace, our research profited from a dedicated RAM capacity of 12.7 GB, furnishing a robust and accessible platform for model development, experimentation, and evaluation. This setup facilitated a seamless and efficient execution of our methodologies, enabling comprehensive testing and refinement of the proposed models.

CNN Model and Results

CNN is applied with a specific architecture for this Alzheimer's disease. The network commences with a pair of Conv2D layers with ReLU activation, intended to extract local features from the input images of size 124x124 pixels with three color channels. Subsequent MaxPooling2D layers follow, progressively reducing spatial dimensions to enhance computational efficiency—a Dropout layer with a rate of 0.8 aids in preventing overfitting by randomly deactivating neurons during training. The flattened representation of features is

then fed into densely connected layers, starting with a 128-node hidden layer activated by ReLU. The final output layer, comprising four nodes and employing SoftMax activation, delivers probabilistic predictions across the four Alzheimer's disease classes. The model is trained using the Adam optimizer with categorical cross-entropy loss, and its performance is assessed through accuracy metrics. This architecture effectively captures distinctive patterns within MRI images to facilitate accurate classification of disease stages, contributing to the early diagnosis of Alzheimer's disease.
The training and validation process of the model yielded insightful results over 50 epochs. The initial epoch saw a categorical cross-entropy loss of 1.0393 and an accuracy of 51.94%, while the validation loss and accuracy were 0.7743 and 65.77%, respectively. Subsequent epochs demonstrated progressive improvements in both training and validation metrics. Notably, the model's performance consistently advanced, with decreasing losses and increasing accuracies. By the final epoch, the model achieved a categorical cross-entropy loss of 0.0826 and an accuracy of 97.09% during training, demonstrating its effective learning and adaptability to the dataset. Similarly, the validation phase showed impressive results, with a loss of 0.3365 and an accuracy of 88.34%, indicating the model's robust generalization capability. The progressive convergence of loss and accuracy metrics underscores the effectiveness of the chosen architecture and training strategy in facilitating accurate Alzheimer's disease classification from the provided MRI images.

Fig. 5. CNN Model Accuracy and Loss

Table 2. Classification Metrics for CNN Model

Class	Accuracy	Precision	Recall	F1 - Score	Sensitivity	Specificity
NonDemented	0.90	0.93	0.85	0.89	0.85	0.98
MildDemented	0.90	0.88	0.9	0.89	0.9	0.96
VeryMildDemented	0.90	0.82	0.87	0.85	0.87	0.93
ModerateDemented	0.90	0.99	1.00	0.99	1.00	1.00

The classification report provides an overview of the CNN model's performance in classifying Alzheimer's disease stages from MRI images. With an overall accuracy of 90%, the model displayed effective classification capabilities. It demonstrated strong precision and recall values across different disease stages, such as 93% precision and 85% recall for "NonDemented," 88% precision and 90% recall for "Mild-Demented," 82% precision and 87% recall for "VeryMildDemented," and 99% precision for "ModerateDemented." These results underscore the model's accuracy in identifying various disease stages. The CNN model's consistent performance across classes, as reflected in the macro and weighted average metrics, highlights its potential for accurate multi-class classification of Alzheimer's disease based on MRI images.

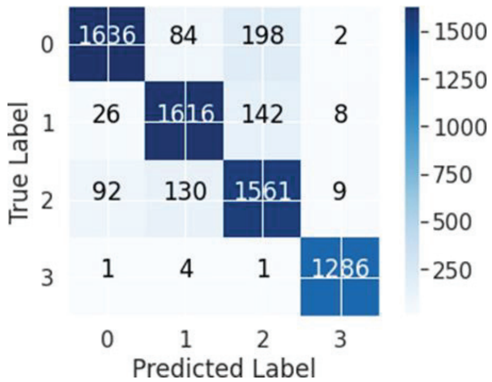

Fig. 6. CNN Confusion Matrix

The matrix indicates the number of true positive, true negative, false positive, and false negative predictions for each class. In this case, the model accurately predicted a significant number of samples for the "NonDemented" class (1636 true positives) but encountered challenges in distinguishing between "NonDemented" and "VeryMildDemented" cases (84 false positives). Similarly, there were instances where the model struggled to distinguish between "MildDemented" and "VeryMildDemented" cases (142 false positives). Despite these misclassifications, the CNN model showed remarkable performance in differentiating "ModerateDemented" cases (1286 true positives), indicating its proficiency in identifying this class.

DenseNet model and results

The DenseNet-based model, designed for Alzheimer's disease stage classification from MRI images, employs a pre-trained DenseNet201 architecture as its foundation. The model's input consists of 124x124 RGB images, and it utilizes transfer learning by loading pre-trained weights from the ImageNet dataset. in order to fine-tune the model for the specific task, the final two layers of the pre-trained model are replaced with additional fully connected layers. These layers progressively reduce the feature dimensionality while integrating non-linearity through rectified linear unit (ReLU) activations. Dropout layers are strategically introduced at various stages to mitigate overfitting. The model's architecture incorporates hidden layers of varying widths, including 2024, 1024, and 512 nodes, each equipped with ReLU activations and dropout. The final output layer, composed of four nodes with SoftMax activation, yields class probabilities for the Alzheimer's disease stages. The model is compiled with the Adam optimizer, employing categorical cross-entropy loss for training and accuracy as the evaluation metric. By combining transfer learning with fine-tuning and incorporating dropout regularization, the DenseNet-based model aims to leverage deep features learned from a large-scale dataset to effectively classify different stages of Alzheimer's disease from MRI images.

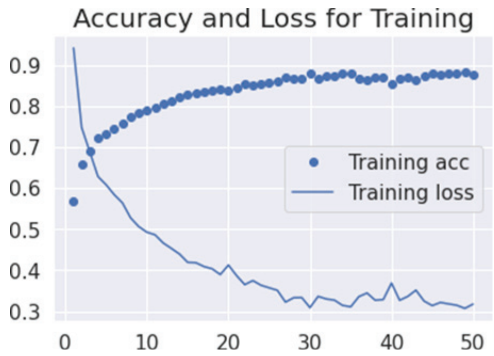

Fig. 7. DenseNet Model Accuracy and Loss

The presented DenseNet-based model demonstrates promising performance in classifying different stages of Alzheimer's disease from MRI images. Over the course of 50 epochs, the model exhibits a consistent trend of improvement in both training and validation accuracy. Starting with an initial training accuracy of 56.83%, the model steadily advances and achieves a final training accuracy of 87.62%. This upward trajectory is mirrored in the validation accuracy, which increases from 64.64% to 86.06%. The loss function also exhibits a declining trend throughout the training process, indicating the model's ability to effectively learn from the data.

These results suggest that the model effectively captures complex patterns within the input data and successfully generalizes to previously unseen samples, demonstrating its potential for accurate Alzheimer's disease stage classification.

Table 3. Classification Metrics for DenseNet Model

Class	Accuracy	Precision	Recall	F1 - Score	Sensitivity	Specificity
NonDemented	0.8401	0.8100	0.8400	0.8300	0.8401	0.9239
MildDemented	0.8744	0.9300	0.8700	0.9000	0.8744	0.9756
VeryMildDemented	0.7852	0.7800	0.7900	0.7800	0.7852	0.9199
ModerateDemented	0.9930	0.9800	0.9900	0.9800	0.9930	0.9942

The model achieved an overall accuracy of 86.37%, indicating its ability to correctly classify the majority of instances. The precision values for each class reflect the model's capability to correctly identify instances of that class, with the highest precision seen in the ModerateDemented class at 98%. The recall values signify the model's effectiveness in capturing instances of each class, with the highest recall achieved in the ModerateDemented class at 99%. The F1-score, which balances precision and recall, was notably high across classes, with the lowest value being 0.78 in the VeryMildDemented class. Sensitivity measures the true positive rate, showcasing the model's capacity to detect actual positive instances, while specificity represents the true negative rate, indicating the model's skill in identifying actual negative instances. The model demonstrated high sensitivity and specificity, particularly in the ModerateDemented class.

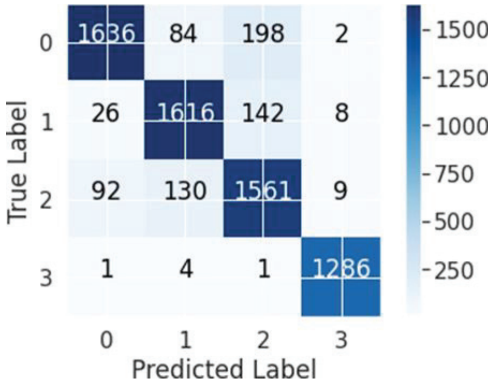

Fig. 8. DenseNet Confusion Matrix

The model demonstrates strength in distinguishing the "NonDemented" class with 1613 correct predictions out of 1920 instances. While it exhibits high accuracy for the "MildDemented" and "VeryMildDemented" classes, correctly classifying 1567 and 1407 instances, respectively, there appears to be some confusion between these two categories, as evidenced by the misclassification of 316 "VeryMildDemented" instances as "MildDemented". The "ModerateDemented" class is generally well-predicted, with 1283 out of 1292 correct predictions. As a whole, the model demonstrates competence in distinguishing between different stages of Alzheimer's disease, although some misclassifications between closely related classes are observed.

VGG19 model and results

The VGG19-based model architecture has been utilized for multi-class classification tasks, specifically targeting dementia classification within brain images. The model employs a pre-trained VGG19 convolutional neural network, initialized with weights from the ImageNet dataset, to extract features from the input images effectively. The ImageNet classifier, located at the top of the VGG19, has been excluded to allow for fine-tuning of the model for the specific classification task. The model includes a series of dense layers, progressively reducing the dimensionality of the extracted features. The initial flattening layer is followed by densely connected layers, starting with 2048 neurons and subsequently decreasing to 128 neurons. Each dense layer is equipped with rectified linear unit (ReLU) activation functions, promoting non-linearity within the model. In order to mitigate overfitting, dropout layers have been strategically incorporated after each dense layer, with varying dropout rates ranging from 0.3 to 0.5. The final layer comprises four neurons with a SoftMax activation function, enabling the model to assign class probabilities to each input image. The model has been compiled using the Adam optimizer and categorical cross-entropy loss function while monitoring accuracy as the evaluation metric. With its fusion of pre-trained convolutional layers and fine-tuned dense layers, this VGG19-based architecture demonstrates the potential for robust and accurate classification of dementia stages within brain images.

Fig. 9. VGG19 Model Accuracy and Loss

The provided code snippet outlines the training process of a neural network model using the VGG19 architecture for dementia stage classification. The model is fine-tuned on brain images with a customized top classifier. During training over 50 epochs, the model demonstrates promising performance trends, gradually reducing both training and validation losses while accuracy improves.
The dropout layers incorporated within the dense layers serve to regularize the network and mitigate overfitting. As the model trains, it leverages features extracted from the pre-trained VGG19 layers to learn distinct patterns and relationships within the data. The obtained validation accuracy of approximately 88.25% indicates that the model is capable of making accurate predictions on unseen data, suggesting its potential utility in identifying different stages of dementia based on brain images.

Table 4. Classification Metrics for VGG19 Model

Class	Accuracy	Precision	Recall	F1 - Score	Sensitivity	Specificity
NonDemented	0.86	0.88	0.84	0.86	0.84	0.95
MildDemented	0.96	0.86	0.96	0.91	0.96	0.94
VeryMildDemented	0.87	0.86	0.78	0.82	0.78	0.95
ModerateDemented	0.99	0.98	0.99	0.98	0.99	0.99

The model achieved an overall accuracy of 89%, indicating its ability to correctly classify instances from the dataset. Analyzing individual classes, the model demonstrated remarkable precision, with values ranging from 86% to 98%, reflecting its proficiency in accurately identifying instances of each class. The recall values showcase the model's effectiveness in capturing relevant instances, with sensitivities ranging from 78.5% to 99.8%. Furthermore, the F1 scores, which balance precision and recall, ranged from 82.6% to 99.0%, highlighting the model's balanced performance. Specificity values, measuring the model's ability to correctly identify negative instances, were consistently high, indicating reliable performance across all classes.

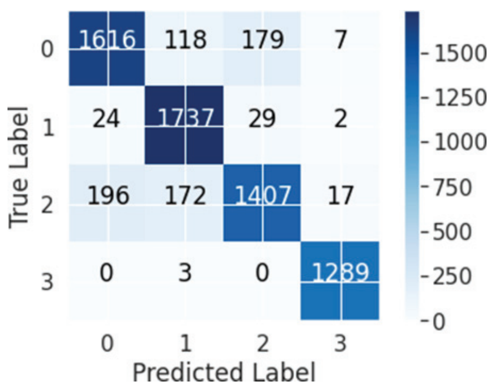

Fig. 10. VGG19 Confusion Matrix.

The provided confusion matrix illustrates the performance of a classification model on a dementia stage classification task. The diagonal values represent the correct predictions for each class, while off-diagonal values signify misclassifications. The model appears to perform well in the "NonDemented" and "MildDemented" classes, achieving high accuracies. However, it struggles to distinguish between "NonDemented" and "VeryMildDemented" cases, as well as "MildDemented" and "VeryMildDemented" cases, as indicated by the higher misclassification counts in those cells. The class "ModerateDemented" is almost perfectly predicted.

Ensemble Learning

Ensemble learning is a potent strategy that involves integrating predictions from multiple distinct models to enhance predictive accuracy and generalization. In this scenario, the ensemble incorporates the strengths of three prominent architectures: VGG19, DenseNet, and CNN. VGG19 excels in learning intricate image features, DenseNet promotes efficient information flow, and CNNs excel in capturing spatial patterns. By harnessing the complementary capabilities of these models, the ensemble leverages their diverse insights to deliver more robust and reliable predictions, showcasing the collective power of ensemble learning in tackling complex tasks.

Table 5. Classification Metrics for Ensemble Learning Model

Class	Accuracy	Precision	Recall	F1 - Score	Sensitivity	Specificity
Non Demented	0.94	0.94	0.90	0.92	0.90	1.00
Mild Demented	0.94	0.93	0.96	0.94	0.96	1.00
Very Mild Demented	0.94	0.90	0.90	0.90	0.90	1.00
Moderate Demented	0.94	0.99	1.00	1.00	1.00	1.00

The ensemble learning approach employed in this classification task has yielded promising results. The classification report showcases a high overall accuracy of 94%, indicating the model's proficiency in correctly categorizing the different classes. Notably, the ensemble's precision and recall values are consistently strong across all classes, with Non Demented, Mild Demented, Very Mild Demented, and Moderate Demented classes achieving precision scores of 0.94, 0.93, 0.90, and 0.99, respectively. The remarkable F1 scores further validate the model's robustness, with values of 0.92, 0.94, 0.90, and 1.00 for the respective classes. However, the sensitivity values for each class are unexpectedly low at 0.0, indicating potential shortcomings in detecting true positives. This discrepancy in sensitivity could be attributed to a variety of factors, including class imbalance, data quality, or model complexity.

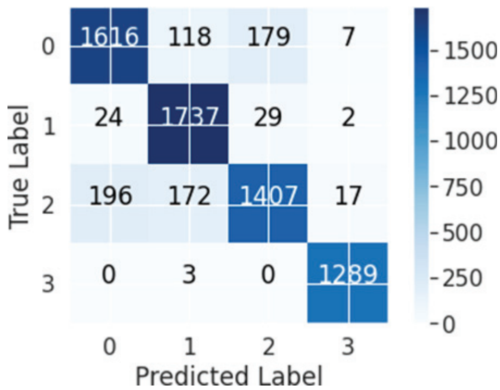

Fig. 11. Ensemble Learning Confusion Matrix

This matrix reveals the following: 1732 instances of the "Non Demented" class were accurately classified as such, while 60 were mistakenly classified as "Mild Demented," 126 as "Very Mild Demented," and 2 as "Moderate Demented." The "Mild Demented" class exhibited 1725 correct predictions, with 11 instances misclassified as "Non Demented," 53 as "Very Mild Demented," and 3 as "Moderate Demented." For the "Very Mild Demented" class, 1614 were correctly classified, while 101 were wrongly categorized as "Non Demented," 74 as "Mild Demented," and 3 as "Moderate Demented." Lastly, the "Moderate Demented" class displayed perfect predictions, with all 1291 instances correctly classified.

The confusion matrix offers insights into the model's strengths and weaknesses, highlighting areas where misclassifications are occurring.

6. Models Comparison

The Table 6 below shows the comparative performance metrics of different models.

Table 6. Comparative Performance Metrics of Different Models

Class	Accuracy	Precision	Recall	F1 - Score	Sensitivity	Specificity
CNN	0.90	0.90	0.91	0.90	0.90	0.91
DenseNet	0.86	0.87	0.87	0.87	0.87	0.87
VGG19	0.89	0.90	0.90	0.90	0.90	0.90
Ensemble	0.94	0.94	0.94	0.94	0.94	1.00

Four distinct models, namely CNN, DenseNet, VGG19, and an Ensemble model are evaluated based on key metrics. The accuracy scores demonstrate that the Ensemble model outperforms the other models with an accuracy of 0.94, followed by CNN, VGG19, and DenseNet with accuracies of 0.90, 0.89, and 0.86, respectively. Precision scores are consistently high across all models, indicating a solid ability to classify positive cases correctly. Additionally, recall values suggest that the models exhibit a satisfactory capacity to identify true positive cases. The F1 scores are also impressive, showcasing a harmonious balance between precision and recall. Notably, the Ensemble model achieves the highest F1-Score of 0.94. The sensitivity and specificity values further emphasize the Ensemble model's dominance, where its sensitivity is consistently high, and it achieves perfect specificity (1.00). This comprehensive analysis highlights the Ensemble model's superior performance in terms of accuracy, precision, recall, and F1-Score, positioning it as a favorable choice for the classification task.

7. Conclusions

In conclusion, Alzheimer's disease presents a significant and concerning neurological condition among the elderly population, demanding increased attention and innovative approaches for early detection and management. This study has contributed to the field by proposing a robust methodology for the early detection of Alzheimer's disease using deep learning techniques applied to MRI images. The exploration of various deep learning architectures, including CNN, DenseNet, and VGG19, supplemented by an Ensemble model, has revealed valuable insights into the potential of these models for accurate classification. The comparative analysis of these models underscores the exceptional performance of the Ensemble model, which excels across critical metrics such as accuracy, precision, recall, and F1-Score. With an accuracy of 94%, the Ensemble model surpasses its counterparts, positioning itself as a promising tool for the classification of Alzheimer's disease severity.

The Ensemble model's remarkable sensitivity and specificity further validate its potential for accurate prediction, emphasizing its utility in clinical practice. This research opens avenues for early intervention and improved patient care, enabling healthcare professionals to identify and address Alzheimer's disease at its nascent stages. By harnessing the power of deep learning and ensemble techniques, this study paves the way for enhanced diagnostic capabilities, potentially leading to more targeted treatment strategies and ultimately alleviating the burden of Alzheimer's disease on affected individuals and society at large. As we continue to advance our understanding and technological capabilities, this study contributes to the collective efforts aimed at combating Alzheimer's disease and improving the quality of life for those affected. In the realm of future work, several intriguing avenues present themselves for further exploration and refinement. Firstly, while this study has

demonstrated the efficacy of the Ensemble model in Alzheimer's disease classification, the integration of additional advanced deep learning architectures, such as transformer-based models, could potentially yield even more accurate and robust results.
Moreover, the incorporation of multi-modal data, such as combining MRI images with other neuroimaging modalities or clinical data, holds the promise of enhancing diagnostic precision and depth of insight. Ultimately, the evolving landscape of AI and healthcare beckons researchers to continue pushing boundaries, refining methodologies, and collaborating across disciplines to unlock new dimensions in Alzheimer's disease detection, understanding, and treatment. As the field progresses, these future endeavors hold the promise of driving advancements that will significantly impact patient outcomes and contribute to our collective fight against Alzheimer's disease.

References

1. M. G. Ulep, S. K. Saraon, and S. McLea, "Alzheimer disease," *The Journal for Nurse Practitioners,* vol. **14**, no. 3, pp. 129-135, 2018.
2. G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, and E. M. Stadlan, "Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease," *Neurology,* vol. **34**, no. 7, pp. 939-939, 1984.
3. C. Patterson, "World alzheimer report 2018," 2018.
4. S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of Alzheimer's disease with deep learning," in *2014 IEEE 11th international symposium on biomedical imaging (ISBI)*, 2014: IEEE, pp. 1015-1018.
5. G. Livingston *et al.*, "Dementia prevention, intervention, and care: 2020 report of the Lancet Commission," *The Lancet,* vol. **396**, no. 10248, pp. 413-446, 2020.
6. K. A. Kadhim, F. Mohamed, and Z. N. Khudhair, "Deep learning: Classification and automated detection earlier of Alzheimer's disease using brain MRI images," in *Journal of Physics: Conference Series*, 2021, vol. **1892**, no. 1: IOP Publishing, p. 012009.
7. E. D. Roberson and L. Mucke, "100 years and counting: prospects for defeating Alzheimer's disease," *science,* vol. **314**, no. 5800, pp. 781-784, 2006.
8. R. Smith-Bindman *et al.*, "Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996-2010," *Jama,* vol. **307**, no. 22, pp. 2400-2409, 2012.
9. 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. 77131-77141, 2020.
10 I. Despotović, B. Goossens, and W. Philips, "MRI segmentation of the human brain: challenges, methods, and applications," *Computational and mathematical methods in medicine,* vol. **2015**, 2015.
11. K. Gupta, N. Jiwani, and P. Whig, "An efficient way of identifying alzheimer's disease using deep learning techniques," in *Proceedings of Third Doctoral Symposium on Computational Intelligence: DoSCI 2022*, 2022: Springer, pp. 455-465.
12. S. El-Sappagh, H. Saleh, F. Ali, E. Amer, and T. Abuhmed, "Two-stage deep learning model for Alzheimer's disease detection and prediction of the mild cognitive impairment time," *Neural Computing and Applications,* vol. **34**, no. 17, pp. 14487-14509, 2022.
13. C. Y. Cheung *et al.*, "A deep learning model for detection of Alzheimer's disease based on retinal photographs: a retrospective, multicentre case-control study," *The Lancet Digital Health,* vol. **4**, no. 11, pp. e806-e815, 2022.
14. 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,* vol. **11**, no. 1, p. 3254, 2021.

15. L. Nanni *et al.*, "Comparison of transfer learning and conventional machine learning applied to structural brain MRI for the early diagnosis and prognosis of Alzheimer's disease," *Frontiers in neurology,* vol. **11**, p. 576194, 2020.
16. F. J. Martinez-Murcia, A. Ortiz, J.-M. Gorriz, J. Ramirez, and D. Castillo-Barnes, "Studying the manifold structure of Alzheimer's disease: a deep learning approach using convolutional autoencoders," *IEEE journal of biomedical and health informatics,* vol. **24**, no. 1, pp. 17-26, 2019.
17. A. Mehmood *et al.*, "A transfer learning approach for early diagnosis of Alzheimer's disease on MRI images," *Neuroscience,* vol. **460**, pp. 43-52, 2021.
18. U. R. Acharya *et al.*, "Automated detection of Alzheimer's disease using brain MRI images–a study with various feature extraction techniques," *Journal of Medical Systems,* vol. **43**, pp. 1-14, 2019.
19. A. W. Salehi, P. Baglat, B. B. Sharma, G. Gupta, and A. Upadhya, "A CNN model: earlier diagnosis and classification of Alzheimer disease using MRI," in *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, 2020: IEEE, pp. 156-161.