A COMPARATIVE ANALYSIS AND PERFORMANCE EVALUATION OF MACHINE LEARNING AND DEEP LEARNING MODELS FOR EARLY DIAGNOSIS OF ALZHEIMER'S DISEASE
A PROJECT REPORT

submitted by

GOVIND M NAIR
KH.PS.P2ASD23013

under the supervision of

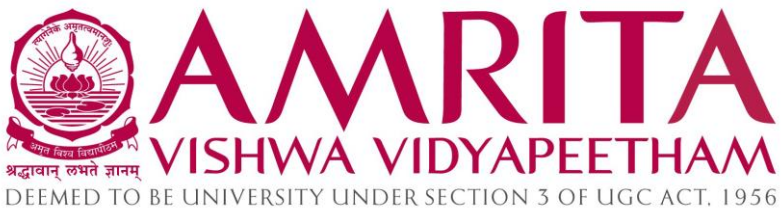
DR. T. SENTHILKUMAR
Assistant Professor
Department of Mathematics
School of Physical Sciences

in partial fulfilment of the requirement of

AMRITA VISHWA VIDYAPEETHAM

for the award of the degree of

MASTER OF SCIENCES
IN
APPLIED STATISTICS AND DATA ANALYTICS



AMRITA VISHWA VIDYAPEETHAM, KOCHI CAMPUS

MAY 2025

AMRITA VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY UNDER SECTION 3 OF UGC ACT, 1956

BONAFIDE CERTIFICATE

This is to certify that the dissertation report entitled **"A COMPARATIVE ANALYSIS AND PERFORMANCE EVALUATION OF MACHINE LEARNING AND DEEP LEARNING MODELS FOR EARLY DIAGNOSIS OF ALZHEIMER'S DISEASE"** submitted by **Govind M Nair, Reg. No: KH.PS.P2ASD23013,** for the award of the degree of **MASTER OF SCIENCES in APPLIED STATISTICS AND DATA ANALYTICS**, is a bonafide record of the work carried out by her under my guidance and supervision at the Department of Mathematics, School of Physical Sciences, Amrita Vishwa Vidyapeetham, Kochi Campus.

Dr. T. Senthilkumar
Project guide
Assistant Professor
Department of Mathematics
School of Physical Sciences
Amrita Vihwa Vidyapeetham,
Kochi Campus

Mr. Sreekumar K
Head of the Department
Assistant Professor (Sr.Gr.)
Department of Mathematics
School of Physical Sciences
Amrita Vishwa Vidyapeetham,
Kochi Campus

The project was evaluated by as on:

Internal Examiner

External Examiner

DECLARATION

I affirm that the project work entitled "**A COMPARATIVE ANALYSIS AND PERFORMANCE EVALUATION OF MACHINE LEARNING AND DEEP LEARNING MODELS FOR EARLY DIAGNOSIS OF ALZHEIMER'S DISEASE**" being submitted in partial fulfilment for the award of the **DEGREE OF MASTER OF SCIENCES in APPLIED STATISTICS AND DATA ANALYTICS** is the original work carried out by me. It has not formed the part of any other project work/internship submitted for award of any degree or diploma, either in this or any other University.

Place: Kochi **GOVIND M NAIR**

Date: **KH.PS.P2ASD23013**

DEDICATION

To

My parents, all my teachers and the eternal God. Thank you for believing in me and encouraging me throughout.

CONTENTS

ACKNOWLEDGEMENT

No endeavour is ever accomplished alone, and I take this moment to sincerely thank everyone who guided and supported me throughout this project journey.

First and foremost, I offer my heartfelt gratitude to the Almighty for blessing me with the strength, perseverance, and clarity needed to complete this work successfully.

I am deeply thankful to **Dr. U. Krishnakumar**, Director of Amrita Vishwa Vidyapeetham, Kochi Campus, for granting me the opportunity to undertake this project.

I would like to express my deep gratitude to **Mr. Sreekumar K**, Head of the Department of Mathematics, School of Physical Sciences, Amrita Vishwa Vidyapeetham, Kochi Campus for his valuable guidance and for the support he rendered on me.

I extend my sincere appreciation to **Dr. T. Senthilkumar**, Asst. Professor, Mathematics, School of Physical Sciences, Amrita Vishwa Vidyapeetham, Kochi Campus, for his continuous support, valuable insights, and encouragement, which were instrumental in shaping this work.

My thanks are also due to the staff of the ICTS Department, School of Physical Sciences, for their timely technical assistance and guidance.

Last but not least, I am truly grateful to my friends and family, whose unwavering support and motivation helped me stay strong and focused even during the most challenging phases of this journey.

Govind M Nair

ABSTRACT

Alzheimer's disease (AD) is an emerging global healthcare concern, contributing significantly to neurological deterioration and impacting millions worldwide. Early diagnosis of Alzheimer's Disease (AD) is essential for effective patient management and treatment planning. The application of Machine Learning (ML) and Deep Learning (DL) techniques has shown substantial promise in predicting AD, leveraging large sets of structured clinical data and MRI scans to aid early diagnosis. This research presents a comparative evaluation of two computational approaches—Support Vector Machine (SVM) with Histogram of Oriented Gradients (HOG) as a ML model, and DenseNet121 as a DL model—for classifying MRI images into four cognitive categories: Mild Impairment, Moderate Impairment, No Impairment, and Very Mild Impairment. The models are trained and tested on a combined dataset composed of grayscale brain MRI scans, and their performance is assessed using key classification metrics, including accuracy, precision, recall, and F1-score. The findings advocate for the careful selection of diagnostic models based on dataset characteristics, model interpretability, and available computational resources, contributing to the broader efforts in improving automated medical diagnosis systems.

CHAPTER 1

1 INTRODUCTION

Alzheimer's Disease (AD) is a progressive and irreversible neurodegenerative disorder that gradually impairs cognitive functions such as memory, reasoning, and behavior, ultimately leading to severe dependency and death. Globally, over 55 million people live with dementia, and AD accounts for approximately 60–70% of these cases [1]. The disease is primarily associated with the buildup of beta-amyloid plaques and neurofibrillary tangles in the brain, resulting in neuronal loss and cortical shrinkage [2]. Despite its growing prevalence, AD often remains undiagnosed until significant cognitive decline occurs. Therefore, early detection is critical to slow disease progression and improve the quality of life for patients and caregivers. MRI imaging has become a preferred diagnostic tool due to its non-invasive nature and ability to capture brain abnormalities such as hippocampal atrophy and ventricular enlargement in early-stage AD [3,4].

1.1 BACKGROUND OF THE STUDY

Artificial intelligence (AI) and neuroimaging have emerged as promising tools in the early detection of AD. ML models such as Support Vector Machines (SVM) and Decision Trees have been used with feature extraction methods like Histogram of Oriented Gradients (HOG) to detect dementia-related patterns in MRI scans [5]. ML approaches especially when paired with handcrafted feature extraction like Histogram of Oriented Gradients (HOG), have shown effectiveness in analysing brain texture and shape patterns.
In contrast, DL architectures like DenseNet121 can automatically learn spatial features from raw grayscale MRI images without manual intervention. Several recent studies have demonstrated that CNN-based models outperform classical ML algorithms, particularly in high-dimensional, image-based classification problems by autonomously learning spatial features from raw images, bypassing the need for manual preprocessing [6,7]. However, comparative studies that evaluate ML and DL approaches under uniform settings are still scarce, prompting further investigation. The integration of AI into medical imaging holds promise not only for enhancing diagnostic precision but also for making AD screening more accessible and efficient [8].

1.2 MOTIVATION OF THE STUDY

While both Machine Learning (ML) and Deep Learning (DL) approaches have been widely applied to Alzheimer's Disease (AD) detection, there is a noticeable lack of research that directly compares the two using the same dataset, identical preprocessing steps, and uniform performance evaluation metrics. Most existing studies tend to focus on comparisons within the same methodological category—ML vs. ML or DL vs. DL—leaving a significant gap in head-to-head analyses between traditional and modern AI approaches.

This study is driven by the need to bridge that gap and provide a clear understanding of the practical trade-offs involved in choosing between ML and DL techniques for Alzheimer's diagnosis using MRI images. While DL models such as DenseNet121 have demonstrated superior accuracy in image classification tasks, they require substantial computational resources and are often less interpretable. Conversely, ML models like Support Vector Machines (SVMs), when combined with feature engineering methods such as Histogram of Oriented Gradients (HOG), offer lower computational overhead and greater interpretability, albeit sometimes at the cost of predictive performance.

By comparing a ML model (SVM + HOG) with a CNN model (DenseNet121) under identical experimental conditions, this study aims to deliver actionable insights into the practical viability of each method. Key factors such as dataset size, grayscale modality, model complexity, and hardware availability are considered alongside performance metrics including accuracy, precision, recall, and F1-score. Ultimately, this comparative analysis contributes to the development of scalable and accessible diagnostic tools for early-stage Alzheimer's, particularly in settings with limited computational infrastructure.

1.3 AIM OF THE STUDY

The primary aim of this study is to perform a comparative analysis of machine learning and deep learning approaches for early detection of Alzheimer's Disease. Specifically, it evaluates the performance of

- Support Vector Machine (SVM) trained on Histogram of Oriented Gradients (HOG) for Machine Learning (ML)
- DenseNet121 for Deep Learning (DL)

using identical datasets and evaluation metrics — including accuracy, precision, recall, and F1-score. Beyond performance, the study also examines model complexity, computational requirements, and adaptability to medical imaging tasks. This work ultimately seeks to recommend the most suitable approach based on practical and technical criteria, contributing to the advancement of automated, reliable Alzheimer's diagnostic systems.

CHAPTER 2

2 METHODOLOGY

2.1 DATA COLLECTION

For both the machine learning (ML) and deep learning (DL) approaches, a publicly available MRI imaging dataset was used, sourced from Kaggle. The dataset comprises grayscale axial MRI scans of the human brain, categorized into four distinct classes representing the stages of Alzheimer's disease:

- No Impairment
- Very Mild Impairment
- Mild Impairment
- Moderate Impairment

The dataset is organized into two separate folders: train and test. The training set contains a total of 10,240 images, with 2,560 images per class, ensuring a balanced distribution across the four categories. In contrast, the test set consists of 1,279 images, distributed as follows:

- No Impairment: 640 images
- Very Mild Impairment: 448 images
- Mild Impairment: 179 images
- Moderate Impairment: 12 images

This imbalance in the test set poses a challenge for classification performance, especially for the underrepresented "Moderate Impairment" class. Nonetheless, the balanced training data helps ensure fair learning across all categories. The dataset serves as the foundation for classifying Alzheimer's disease stages based on structural brain differences observed in the MRI scans.

2.2 PRE-PROCESSING

In the machine learning workflow, each MRI image was initially read using skimage.io.imread. If the image was not already in grayscale, it was explicitly converted. All images were resized to a standard 128×128 resolution to ensure uniformity across the dataset. From these resized grayscale images, texture-based features were extracted using the Histogram of Oriented

Gradients (HOG) technique. The HOG extraction parameters were carefully selected to capture local structural variations commonly associated with Alzheimer's disease. Specifically, features were computed with 9 gradient orientation bins, a cell size of 8×8 pixels, and a block size comprising 2×2 cells. L2-Hys normalization was applied at the block level to ensure contrast invariance and stability across lighting variations. These configurations resulted in robust one-dimensional descriptors for each image, encapsulating critical edge and shape information relevant to neurological abnormalities. The resulting feature matrix, along with corresponding labels, was used to train a linear Support Vector Classifier (LinearSVC). The HOG-based approach enabled the model to distinguish subtle textural differences across categories, which is essential in identifying early-stage dementia patterns from structural brain scans.

For the deep learning model, a systematic preprocessing pipeline was implemented using the ImageDataGenerator utility from Keras to prepare grayscale MRI images. Each image was resized uniformly to 128×128 pixels to ensure consistency in input dimensions. Pixel intensity values, originally in the range of 0–255, were normalized to the [0, 1] range by dividing by 255. This normalization facilitated smoother gradient updates during model training. To enhance the model's generalization capability and reduce overfitting, extensive data augmentation techniques were applied. These included random rotations of up to 25 degrees, horizontal shifts and vertical shifts up to 15% of the image size, and zoom operations up to 20%. Horizontal flipping was also enabled to simulate real-world variance in brain scan orientations. Any missing or newly exposed pixel regions resulting from transformations were filled using the nearest-neighbor method, preserving edge continuity. The dataset directory containing the images was divided into training and validation sets using the validation_split parameter with an 80:20 ratio. Although the MRI images were grayscale, the chosen architecture, DenseNet121, is designed for 3-channel (RGB) inputs. To address this, each grayscale channel was duplicated across three channels using a Keras Concatenate layer, ensuring full compatibility with the DenseNet framework while retaining the original image information.

2.3 TOOLS AND TECHNOLOGIES

a) Python

Python served as the primary language for coding all components of the project—from data preprocessing and feature engineering to model training and evaluation.

b) TensorFlow and Keras

The deep learning component of this project was implemented using TensorFlow, with Keras as its high-level interface. These frameworks were essential for designing and training the DenseNet121 convolutional neural network. Key functionalities used include:

- ImageDataGenerator for real-time data augmentation and grayscale image handling.

- Functional API layers like Input, Concatenate, and pre-configured DenseNet121.
- Training optimizations via callbacks such as EarlyStopping and ReduceLROnPlateau

c) Scikit-Learn (sklearn)

Scikit-learn was employed for building and evaluating the machine learning. Key utilities included:

- LinearSVC for classification.
- Handling imbalanced data using class weights.
- Generating metrics like accuracy, precision, recall, F1-score, and the confusion matrix.

d) Scikit-Image (skimage)

This library was crucial for preprocessing the MRI. It enabled:

- Reading and resizing images using imread and resize.
- Converting images to grayscale with rgb2gray.
- Extracting features using HOG with configurable parameters for orientation, cell size, and block size.

e) NumPy

NumPy was used extensively for array operations, image data handling, and preparing input features for both ML and DL models.

2.4 DATA ANALYSIS TECHNIQUES

a) Support Vector Machine + Histogram of Oriented Gradients (HOG)

Support Vector Machine (SVM) is a powerful supervised learning algorithm typically used for classification tasks. SVM works by finding the optimal hyperplane that maximizes the margin between different classes in the feature space. The data points that are closest to this hyperplane, known as support vectors, are critical in determining the position and orientation of the hyperplane. If the data is not linearly separable, SVM utilizes kernel functions to map the data into a higher-dimensional space where a linear separation is possible. The Linear SVM is particularly efficient when dealing with linearly separable data and high-dimensional spaces, making it a good fit for structured datasets like patient information in medical classifications.
In this project, SVM was used to classify features extracted from MRI images using Histogram of Oriented Gradients (HOG). The HOG features capture the local image gradients and texture patterns, which are essential in differentiating between varying stages of dementia. The SVM

model was trained on these features, and performance metrics such as accuracy, precision, recall, and F1-score were evaluated to assess its effectiveness in classifying the images into different dementia categories. The Linear kernel was chosen for this task due to its simplicity and its balance of computational efficiency, robustness in high-dimensional feature space, reduced risk of overfitting on a limited dataset, and improved interpretability—making it a practical and effective choice for early-stage Alzheimer's classification using grayscale MRI scans. The SVM model performed well, especially in distinguishing between non-demented and mildly impaired subjects, demonstrating the model's potential in medical image classification tasks.

b) DenseNet121

DenseNet121 is a deep learning architecture that falls under the category of Convolutional Neural Networks (CNNs), which are widely used for image classification tasks. DenseNet121 is particularly notable for its dense connections, where each layer is connected to every other layer in a feed-forward fashion. This architecture allows for more efficient gradient flow and feature reuse across layers, making it effective for tasks that require learning hierarchical features from visual data, such as medical image analysis. The DenseNet121 model is typically used in combination with global pooling layers, batch normalization, and dense layers to provide robust feature extraction and classification capabilities.

In this project, DenseNet121 was used to classify grayscale MRI images into four categories representing different stages of Alzheimer's disease. The DenseNet121 architecture was selected due to its efficient feature reuse, superior performance on subtle grayscale brain patterns, and ability to generalize well despite limited data, making it an ideal fit for early-stage dementia diagnosis. The model was trained from scratch, without relying on pre-trained weights, to ensure that it learned domain-specific features relevant to Alzheimer's diagnosis. A custom classification head was added, consisting of global average pooling, batch normalization, and a dense layer with ReLU activation, followed by a SoftMax layer to output class probabilities. Training was guided using early stopping, learning rate reduction, and model checkpointing to prevent overfitting and ensure optimal convergence. Performance was evaluated using accuracy, precision, recall, and F1-score, demonstrating the model's strong capability in identifying early and mild stages of impairment. DenseNet121 proved to be a powerful deep learning approach for Alzheimer’s classification, showing high potential for application in medical image analysis tasks.

CHAPTER 3

3 RESULTS AND DISCUSSION

3.1 MODEL BUILDING AND EVALUATION

For the Machine Learning (ML) approach, the dataset was pre-processed by extracting Histogram of Oriented Gradients (HOG) features from grayscale MRI images. These images were uniformly resized to 128×128 pixels before feature extraction. The resulting HOG descriptors captured edge and texture patterns critical for distinguishing between stages of cognitive impairment. These features were then used as input to train a Linear Support Vector Machine (SVM) classifier using a linear kernel. The model was trained on data categorized into four dementia classes, with class labels serving as the target output. Although no explicit validation set was used during training, the model's performance was later assessed on a separate test set using standard evaluation metrics such as accuracy, precision, recall, and F1-score.

For the Deep Learning (DL) approach, a DenseNet121-based convolutional neural network (CNN) was developed to classify grayscale MRI images into the same four diagnostic categories. The architecture was constructed using the DenseNet121 backbone without pre-trained weights, followed by a custom classification head consisting of a global average pooling layer, batch normalization, and a dense layer with ReLU activation. The final output layer used a SoftMax activation function to compute class probabilities. Since DenseNet121 expects three-channel input, the grayscale MRI images were replicated across channels using a concatenation layer. All images were normalized and resized to 128×128 pixels before training. The deep learning model was compiled using the Adam optimizer and categorical cross-entropy as the loss function. To improve generalization and prevent overfitting, extensive data augmentation was applied during training, including random rotations, translations, zooming, and horizontal flipping. An 80:20 split was used to separate training and validation data, and callbacks such as EarlyStopping, ReduceLROnPlateau, and ModelCheckpoint were employed to monitor validation loss, adjust learning rates, and preserve the best-performing model. Class weights were also computed and applied to address imbalance in the training dataset. Both models were trained and evaluated on the same dataset to ensure a fair and consistent comparison in terms of classification performance and robustness.

The performance of both the models was evaluated using standard classification metrics to ensure a comprehensive assessment across all dementia classes. The primary metrics used were **Accuracy**, **Precision**, **Recall**, and **F1-Score**—each offering different perspectives on the model's performance.

- Accuracy which measures the overall proportion of correctly classified instances.

- Precision which measures how many of the predicted positive cases were actually correct, making it useful in evaluating the model's ability to avoid false positives—particularly important when distinguishing between similar dementia classes.

- Recall which evaluates how well the model captures actual positive cases, indicating its ability to detect impaired individuals.

- The F1-Score which is the harmonic mean of Precision and Recall, providing a balance between the two.

To evaluate class-wise performance, each of these metrics was computed for all four classes in the dataset—Mild, Moderate, Very Mild, and No Impairment This multi-metric evaluation helped ensure that the comparison between machine learning and deep learning models was fair, especially in the presence of class imbalance and varying prediction confidence levels.

3.2 INTERPRETATION OF RESULTS

The DenseNet121 model achieved an overall accuracy of **86%**, with strong performance across all evaluation metrics. It demonstrated particularly high recall for the **Mild Impairment (93%)**, **Moderate Impairment (100%)**, and **No Impairment (91%)** classes, indicating reliable detection of these categories. Although the recall for the **Very Mild Impairment** class was slightly lower at **75%**, the precision and F1-scores remained high across the board.

Class	Precision	Recall	F1-Score	Support
Mild Impairment	0.85	0.93	0.89	179
Moderate Impairment	1.00	1.00	1.00	12
No Impairment	0.86	0.91	0.88	640
Very Mild Impairment	0.85	0.75	0.80	448

The machine learning model (SVM + HOG) obtained a slightly lower overall accuracy of **81%**. While the model performed reasonably well for the **No Impairment** and **Moderate Impairment** classes (F1-scores of 0.85 and 0.80 respectively), its performance on the **Mild Impairment** and **Very Mild Impairment** categories was relatively lower. Specifically, the recall for **Mild Impairment** was **72%**, suggesting some difficulty in correctly identifying all true cases from this class.

Class	Precision	Recall	F1-Score	Support
Mild Impairment	0.82	0.72	0.77	179
Moderate Impairment	0.77	0.83	0.80	12
No Impairment	0.84	0.85	0.85	640
Very Mild Impairment	0.76	0.78	0.77	448

The DenseNet121 model outperformed the SVM + HOG classifier across all metrics, particularly in the **Moderate Impairment** category where it achieved perfect scores. The CNN also maintained balanced performance across all classes, making it more suitable for detecting varying stages of cognitive decline. These results suggest that deep learning approaches are more effective for handling raw medical imaging data, as they can capture complex spatial patterns that traditional hand-crafted features may miss.

Overall, both models demonstrated meaningful classification abilities, but the DenseNet121 showed higher accuracy and stronger generalization across multiple impairment stages, making it a more reliable approach for Alzheimer's disease prediction using MRI images.

3.3 DISCUSSION

This study presents a performance-based comparison between a traditional Machine Learning model and a Deep Learning architecture for the classification of Alzheimer's Disease stages using grayscale MRI images. The Machine Learning pipeline utilized a Support Vector Machine (SVM) trained on HOG (Histogram of Oriented Gradients) features extracted from the input MRI scans. The model achieved an overall accuracy of 81%, with particularly strong performance on the "No Impairment" and "Moderate Impairment" classes, recording F1-scores of 0.85 and 0.80 respectively. However, the SVM exhibited lower recall on "Mild Impairment" (72%), indicating limitations in identifying early-stage cognitive decline. In contrast, the DenseNet121 trained directly on the grayscale MRI images achieved a higher overall accuracy of 86%, with near-perfect precision and recall for the "Moderate Impairment" class (both 1.00) and strong generalization across all classes. Its macro-averaged F1-score of 0.89 and a weighted average F1-score of 0.85 suggest superior discriminative ability, particularly in capturing subtle spatial and structural patterns in neuroimaging data. Notably, although the test dataset was imbalanced, the CNN model—trained on a well-balanced dataset and evaluated using class-wise metrics—still demonstrated consistent performance across all impairment categories. The comparative analysis indicates that while SVM with hand-crafted features offers reasonable classification, deep learning models like DenseNet121 are better suited for complex visual pattern recognition in MRI-based dementia detection. This reinforces the growing relevance of CNN-based architectures in medical image analysis and highlights their potential in supporting early and accurate diagnosis of Alzheimer's Disease.

CHAPTER 4

4 CONCLUSION

This study presented a comparative analysis of a traditional machine learning approach using SVM with HOG features and a deep learning approach using DenseNet121 for the classification of Alzheimer's disease stages from grayscale MRI images. The deep learning model demonstrated a clear advantage in terms of overall predictive performance and robustness across all classes, effectively capturing structural variations that are often subtle in medical imaging. Its ability to generalize well even under class imbalance conditions highlights its strength in handling complex neuroimaging data. In contrast, the SVM model, though less powerful in terms of accuracy, maintained value through its simplicity and interpretability, offering clearer insights into decision boundaries — a feature still highly valued in clinical contexts where transparency is essential.

The study also underscored important trade-offs between performance, interpretability, and resource efficiency in the design of automated diagnostic systems. While DenseNet121 delivered higher reliability and precision, its computational demands and black-box nature may pose challenges in real-world deployment, especially in under-resourced medical settings. On the other hand, the SVM model with handcrafted features remains a practical choice when infrastructure or model transparency is a constraint. These findings suggest that model selection in healthcare AI should not rely solely on predictive performance but must also consider the operational context, data availability, and clinical usability. Together, these insights contribute to a more informed and strategic integration of AI in Alzheimer's diagnosis and broader medical imaging applications.

4.1 LIMITATIONS OF THE STUDY AND FUTURE WORK

Despite these results, the study is not without limitations. One of the primary constraints encountered in this study was the limited size and class imbalance of the available MRI dataset. In particular, the "Moderate Impairment" class contained significantly fewer samples compared to other categories, which may have biased the model's ability to generalize across all classes effectively. Although the training set was balanced through appropriate sampling strategies, the test set remained imbalanced, potentially affecting the reliability of evaluation metrics in real-world applications. Another challenge involved adapting grayscale MRI images for deep learning architectures such as DenseNet121, which are originally designed for three-channel (RGB) input. This necessitated additional preprocessing to replicate the grayscale channel across three dimensions, increasing memory usage and preprocessing complexity. Furthermore, deep learning models like DenseNet121 are computationally demanding and require dedicated GPU resources to train within a reasonable timeframe. In contrast, while traditional ML models such as SVMs are more lightweight and interpretable, they may struggle to extract high-level spatial features from raw medical images without deep hierarchical learning. This presents a trade-off between model simplicity and predictive performance.

Finally, although the evaluation was conducted using class-wise precision, recall, and F1-score to mitigate the effects of imbalance, the study was restricted to a single dataset without external validation. As a result, the generalizability of both ML and DL models to other clinical datasets

or real-world diagnostic environments remains an open question. Moreover, explainability—a critical factor in medical decision-making — was not addressed in depth for the deep learning model, limiting its immediate applicability in clinical workflows.

Future work could include incorporating multimodal approach and exploring advanced ensemble approaches. Future developments in AI-assisted diagnostic systems and the inclusion of explainable AI (XAI) methods—such as Grad-CAM or SHAP— can improve the disease prediction and management. Another potential direction is to enhance existing lightweight architectures such as MobileNet to improve their efficiency and generalization on limited datasets, thereby enabling accurate and scalable Alzheimer's disease prediction in low-resource clinical settings. Finally, to enhance generalizability, future work should use a larger and more diverse dataset, validate the models on external datasets from different institutions and scanners. This would test their robustness across varying patient populations, scan qualities, and clinical environments, further establishing the models' reliability for real-world application.

REFERENCES

[1] World Health Organization (2023). *Dementia*. Retrieved from: https://www.who.int/news-room/fact-sheets/detail/dementia

[2] Rajpa, Kanika, Rashmi Gourh, Parmeet Kaur, and Mehak Sharma. "Alzheimer's Disease Detection Using Machine Learning." *Available at SSRN 4666141* (2023).

[3] Saikia, Pallavi, and Sanjib Kumar Kalita. "Alzheimer disease detection using MRI: deep learning review." *SN Computer Science* 5, no. 5 (2024): 507.

[4] Bari Antor, Morshedul, AHM Shafayet Jamil, Maliha Mamtaz, Mohammad Monirujjaman Khan, Sultan Aljahdali, Manjit Kaur, Parminder Singh, and Mehedi Masud. "A comparative analysis of machine learning algorithms to predict Alzheimer's disease." *Journal of Healthcare Engineering* 2021, no. 1 (2021): 9917919.

[5] Nawaz, Ali, Syed Muhammad Anwar, Rehan Liaqat, Javid Iqbal, Ulas Bagci, and Muhammad Majid. "Deep convolutional neural network based classification of Alzheimer's disease using MRI data." In *2020 IEEE 23rd International Multitopic Conference (INMIC)*, pp. 1-6. IEEE, 2020.

[6] Arafa, Doaa Ahmed, Hossam El-Din Moustafa, Hesham A. Ali, Amr MT Ali-Eldin, and Sabry F. Saraya. "A deep learning framework for early diagnosis of Alzheimer's disease on MRI images." *Multimedia Tools and Applications* 83, no. 2 (2024): 3767-3799.

[7] Ghaffari, Hamed, Hassan Tavakoli, and Gila Pirzad Jahromi. "Deep transfer learning–based fully automated detection and classification of Alzheimer's disease on brain MRI." *The British journal of radiology* 95, no. 1136 (2022): 20211253.

[8] National Institute on Aging. (2022). *What Is Alzheimer's Disease?* Retrieved from: https://www.nia.nih.gov/health/what-alzheimers-disease