IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 3, June 2026, pp. 2357~2368
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i3.pp2357-2368

Accurate detection of Alzheimer's disease using machine learning model on magnetic resonance imaging data

Rahman Md Mojnur[1], Md. Tanjil Sarker[2], Hezerul Abdul Karim[2], Fahmid Al Farid[2], Aziah Ali[1], Wan-Noorshahida Mohd-Isa[1]
[1]Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia
[2]Faculty of Artificial Intelligence and Engineering, Multimedia University, Cyberjaya, Malaysia

Article Info

Article history:

Received Jul 9, 2024
Revised Feb 23, 2026
Accepted Apr 21, 2026

Keywords:

Alzheimer's disease
Convolutional neural networks
Deep learning
Machine learning
Magnetic resonance imaging data
Medical image analysis
Neuroimaging

ABSTRACT

The rapid identification and diagnosis of Alzheimer's disease (AD) is essential for initiating early intervention and effective treatment planning. Magnetic resonance imaging (MRI) provides valuable structural insights into the pathological alterations in the brain associated with AD. Early and accurate detection of AD is critical for initiating timely interventions. This study presents a classical machine learning (ML) approach for detecting AD using structured features extracted from MRI metadata, such as mini-mental state examination (MMSE) scores, brain volume metrics, and cognitive attributes. Unlike deep learning models that rely on raw imaging data, the interpretable framework offers reduced computational complexity and better alignment with real-world clinical constraints. Models such as random forest (RF) and extreme gradient boosting (XGBoost) achieved up to 85% accuracy, showing strong potential for deployment in resource-limited environments. The results demonstrate the potential of classical ML in supporting early AD diagnosis, particularly in low-resource clinical settings. Moreover, the proposed approach offers a computationally efficient and interpretable alternative to deep learning models, facilitating adoption in real-world healthcare environments.

Corresponding Author:

Wan-Noorshahida Mohd-Isa
Faculty of Computing and Informatics, Multimedia University
Cyberjaya 63100, Malaysia
Email: wan.noorshahida.isa@mmu.edu.my

1. INTRODUCTION

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by cognitive decline, memory loss, and a deterioration in the ability to perform everyday activities [1]. It is among the most prevalent forms of dementia, affecting approximately one in every nine individuals aged 65 and older. With the global population continuing to age, the incidence of AD is expected to rise significantly, posing growing challenges for healthcare systems worldwide [2]. Early and accurate diagnosis of AD is crucial for timely intervention, improved patient management, and the development of effective therapeutic strategies that may alleviate symptoms and slow disease progression [2]. Magnetic resonance imaging (MRI) has emerged as a vital tool in the diagnosis and monitoring of AD, offering high-resolution structural information that enables the visualization of neuroanatomical changes, such as hippocampal atrophy and cortical thinning [3]. However, manual interpretation of MRI scans remains time-consuming, subjective, and prone to inter-observer variability, thus motivating the need for automated diagnostic approaches. In recent years, machine learning (ML) has demonstrated remarkable success in medical image analysis, including the detection and classification of neurological diseases using MRI data [4]. While deep learning methods such

as convolutional neural networks (CNNs) have shown superior performance in extracting spatial features from raw imaging data [5], their practical implementation often demands large datasets, significant computational power, and specialized GPU-based hardware.

In contrast, classical ML algorithms such as random forest (RF), support vector machines (SVM), and extreme gradient boosting (XGBoost) can offer high accuracy, computational efficiency, and greater interpretability, especially when applied to structured clinical and imaging-derived metadata [6]. These attributes make classical ML models highly suitable for deployment in low-resource clinical environments where transparency and scalability are essential. This research aims to develop and evaluate a robust ML framework for the accurate detection of AD using clinically relevant features derived from MRI metadata [7]. Through comprehensive data preprocessing, statistical feature selection using analysis of variance (ANOVA)-based methods, and comparative analysis of multiple ML models, the study demonstrates the viability of using structured tabular data for AD classification. The proposed approach offers a practical and interpretable diagnostic aid to support early-stage AD detection in both research and clinical applications [8].

As shown in Table 1, recent studies emphasize deep learning approaches for MRI-based AD classification. However, most of them face limitations such as high computational demand, lack of interpretability, or limited modality input. In contrast, the present study addresses these gaps by applying classical ML models to structured clinical and neuroanatomical data, providing a lightweight and interpretable framework suitable for deployment in resource-limited environments. Future work will aim to integrate multimodal inputs and incorporate explainable AI techniques. Table 1 provides the comparison of recent studies.

Table 1. Comparison of recent studies

Reference	Research summary	Research gaps	Future work
[9]	Developed a hybrid 3D-CNN and attention mechanism for AD classification using raw MRI scans. Achieved high accuracy by leveraging spatial features and deep attention layers.	Model requires high computational power; lacks integration with structured clinical data; interpretability is limited.	Integrate multimodal data (images+clinical), improve model explainability using tools like Shapley additive explanations (SHAP).
[10]	Used a CNN-based model to classify MRI data into stages of cognitive decline. Emphasized early detection through image feature extraction.	Dataset imbalance not addressed; lacks comparison with classical ML methods.	Employ data balancing (e.g., synthetic minority over-sampling technique (SMOTE), and include baseline ML comparisons for robustness.
[11]	Proposed an ensemble of deep learning models for longitudinal AD prediction. Utilized time-series imaging data for disease progression modeling.	Methodology depends heavily on longitudinal data availability; not applicable in all settings.	Develop models that work with single-timepoint data; explore clinical-data fusion strategies.
[12]	Compared performance of CNN, SVM, and k-nearest neighbors (KNN) models using OASIS dataset for binary AD classification. Found CNN slightly outperformed SVM.	Limited model interpretability; only binary classification (no mild cognitive impairment (MCI) or early-stage granularity).	Extend to multi-class classification including MCI; add explainability layer to improve clinical trust.
[13]	Implemented VGG-16 and ResNet-50 architectures on AD image classification. Reported high accuracy using transfer learning on limited data.	Transfer learning from non-medical domains may introduce domain bias; no tabular features used.	Combine deep features with cognitive scores for multimodal learning; retrain on medical-specific data.

The findings of this research have the potential to significantly impact clinical practice by providing clinicians with an efficient and objective tool for early detection and monitoring of AD. By harnessing the power of ML and MRI technology, we strive to advance the field of neuroimaging and contribute to improved patient care in the management of this debilitating condition. The rest of the paper is organized as follows: section 2 discusses the dataset and pre-processing. Model development and hypertuning is described in sections 3. Performance comparison is shown in the section 4. In conclusion, final observations and recommendations for future work are presented in section 5.

2. DATASET AND PRE-PROCESSING

Alzheimer's disease neuroimaging initiative (ADNI) is a widely used dataset containing MRI scans, clinical, and genetic data from AD patients, individuals with MCI, and healthy controls [14]. By carefully selecting a suitable dataset and conducting thorough preprocessing, researchers can ensure the robustness and

reliability of the deep learning model developed for accurate detection of AD using MRI data. Additionally, documenting preprocessing steps is crucial for transparency and reproducibility in research.

2.1. Image dataset

The dataset used for this project was sourced and was procured by Singh on Kaggle [15]. The dataset is a collection of MRI images of people's brains. MRI is a method that utilizes strong magnetic fields and radio waves to produce images of the internal structure of the body. MRI is a commonly used imaging technique for diagnosing neurological conditions [16].

The dataset has a total of 6,400 of high-quality jpeg files, separated into 4 folders, each containing different classes of MRI images. Table 2 is shown the dataset for different classes of MRI images. The images in the dataset provide insight into the brain patterns of Alzheimer's patients and can be used to develop and evaluate ML algorithms for the early detection and diagnosis of AD.

Table 2. Dataset for different classes of MRI images

Folder name	Description
Normal	This class contains images of healthy brains with no signs of AD.
Very mild dementia	This class contains images of brains affected by very mild dementia, which is a mild form of cognitive decline.
Mild cognitive impairment	This class contains images of brains with mild cognitive impairment, which is often a precursor to AD.
AD	This class contains images of brains affected by AD.

2.2. Table dataset

In another dataset used for this project was also sourced from Kaggle by Choi [17]. The dataset is a csv file containing information about patients diagnosed with AD. The dataset relates to the OASIS project at https://sites.wustl.edu/oasisbrains/. A project that aims to provide neuroimage datasets for free, in the hopes of improving the research on neurodegenerative diseases. Table 3 is shown the dataset information about patients diagnosed with AD. The dataset could be used to find, study and understand how internal and external factors play a role in the development of AD.

Table 3. Dataset information about patients diagnosed with AD

Item	Description
Subject ID	A unique identifier for each patient in the study.
MR Delay	The time elapsed between the MRI scan and the patient ‘s assessment.
MRI ID	A unique identifier for each MRI scan.
Visit	The number of visits to the study center.
M/F	The gender of the patient.
Hand	The handedness of the patient, left-handed or right-handed.
Age	The age of the patient.
EDUC	The level of education of the patient.
Socio-economic status (SES)	The SES of the patient.
Mini-mental state examination (MMSE)	The MMSE score, a measure of cognitive function.
Clinical dementia rating (CDR)	The CDR score, a scale for evaluating the severity of dementia.
Estimated total intracranial volume (eTIV)	The eTIV, a measure of the size of the brain.
Normalized whole brain volume (nWBV)	The nWBV, a measure of brain volume relative to overall size.
Atlas scaling factor (ASF)	The ASF, a value used to correct for differences in brain size across subjects.
Group	Classification for whether the patient has dementia, recovered from it or has never had it.

2.3. Image pre-processing

Pre-processing for the image dataset is done using the ImageDataGenerator class from Keras [18]. The ImageDataGenerator generates new data from pre-existing ones. This increases the number of images that can be used to train our deep learning model. The parameters and explanation for the ImageDataGenerator is shown the Table 4.

It is important to note that SMOTE was applied exclusively to the tabular dataset to address class imbalance between 'demented' and 'nondemented' classes. While image preprocessing such as horizontal flipping, rescaling, and brightness augmentation was conducted using ImageDataGenerator, no model was trained on this image data due to computational limitations [19]. Therefore, SMOTE was not applied to high-dimensional image data, as doing so can produce unrealistic and structurally invalid synthetic samples.

Table 4. Parameters and explanation for the ImageDataGenerator

Item	Description
rescale	rescale =1/255. The pixel values of the images are scaled down from 0-255 to 0-1. This normalizes the values within each image as well as making it faster to process images.
brightness_range	brightness_range = [0.8, 1.2]. New images are generated with 80% to 120% brightness levels of the original dataset.
zoom_range	zoom_range = [.99, 1.01]. New images are generated that came from zooming in 99% and zooming out 101% of the original dataset.
data_format	data_format = “channel_last”. Images are shaped in the format of (height, width, channels).
fill_mode	fill_mode = "constant". Areas that are out of bounds of the image are filled with 0s or black.
horizontal_flip	horizontal_flip = true. Randomly flips the original images horizontally to produce new images.

2.4. Table pre-processing

In the table dataset, only columns SES and MMSE had empty values. These empty values are then replaced with their respective median values. Mean is very sensitive to outliers, but median is robust, hence it is much better to replace the empty values with median when the distribution is skewed [20].

Columns that are removed from the dataset are subject ID, MRI ID, Visit, and CDR. Subject ID and MRI ID are just unique identifiers, Visit is the number of visits the patient made with a doctor. CDR was removed because it represented the same information as the group feature [21]. Decided to utilize the group feature over CDR since it is a categorical variable, making it easier to handle in the later stages of model development. The categories of patients are depicted in Figure 1. There are 3 categories of patients, patients with dementia (demented), patients that had mild but later develop worse conditions (converted), and patients that never had dementia (nondemented). Because the converted group also experiences dementia, the group was re-categorized as demented as well.

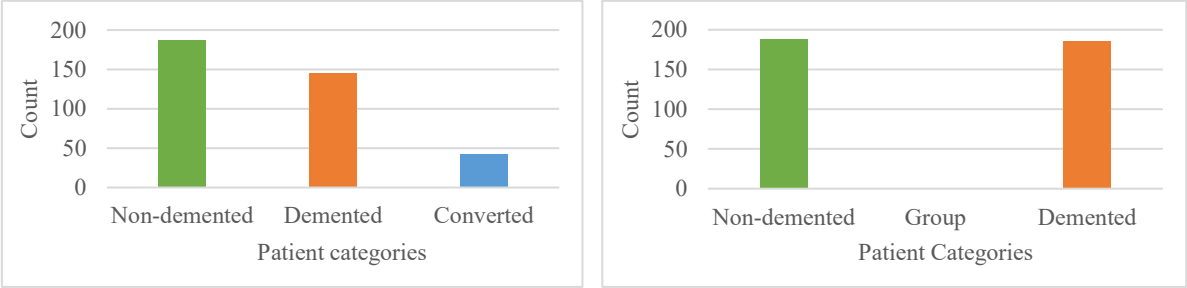

Figure 1. Distribution of patients across dementia categories

2.5. Table exploratory data analysis

Exploratory data analysis (EDA) is done on the table dataset to find the patterns and relations between the columns of the dataset [22]. From Figure 2, for both dementia and non-dementia patients, there is a wide spread in brain size distribution. However, for patients that recovered from dementia the available stable funding (ASF) values ranged mostly around 1-1.4. An assumption could be made that dementia does not affect the brain size [23].

From the Figure 3, The study cohort included a demographic breakdown by age, gender, and dementia status. Participant data was collected and categorized into five distinct age groups, ranging from 60-85 years, to analyze the distribution of males and females. The largest concentration of subjects was observed within the 71-75 age group, with subsequent analyses examining the correlation between gender and dementia prevalence across the entire sample.

From Figure 4, further examination of dementia severity using the CDR score revealed a distinct pattern between genders. As shown in the violin plot, the distribution of CDR scores for females was more dispersed and extended to higher values compared to males. This wider range suggests a greater variability in the severity of dementia among female subjects. The CDR data reinforces the observation of a more pronounced impact of the condition within the female demographic of the study group. The only strong positive correlations from the dataset are between the number of visits (Visit) and the time interval between the examination and MRI scan in months (MR delay) [24]. The only strong negative correlation is between the brain’s volume (eTIV) and the proportion of brain volume occupied by Alzheimer's-specific neurodegeneration (ASF). Figure 5 is shown the assessing relationships between clinical measures using correlation analysis.

2.6. Table encoding

Categorical columns are encoded to integers. For the group column, demented is set as 1 while nondemented is set as 0. For the M/F column, which is the gender column, M is set as 1 while F is set as 0. Finally, for the hand column, R is set as 0.

Figure 2. Brain size (ASF) distribution across dementia, non-dementia, and recovered groups

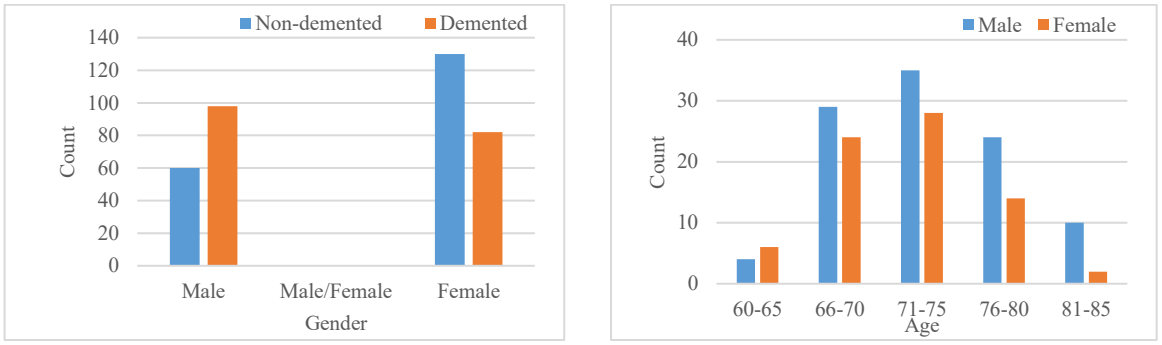

Figure 3. Exploratory data analysis between male and female with age group

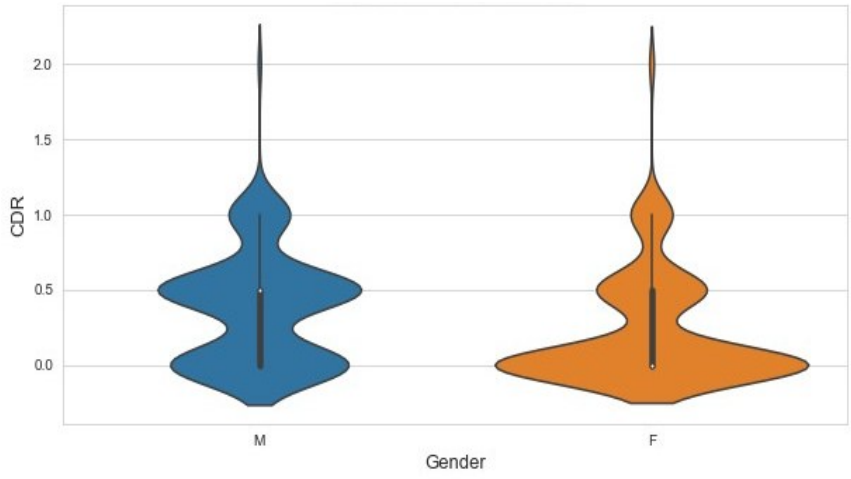

Figure 4. Violin plots of CDR by gender

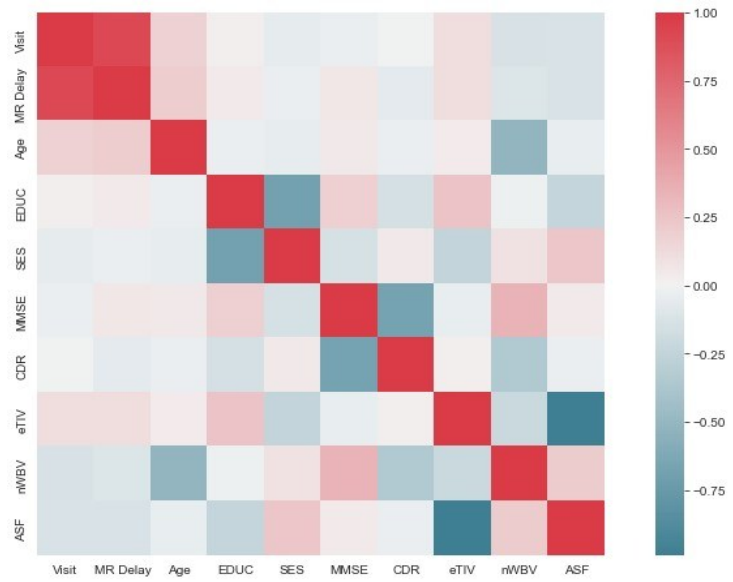

Figure 5. Assessing relationships between clinical measures using correlation analysis

3. MODEL DEVELOPMENT AND HYPERTUNING

All model development was focused on tabular data derived from clinical and structural MRI features. These features include MMSE scores, eTIV, nWBV, and other cognitive and anatomical indicators extracted from the OASIS dataset. This approach allowed for faster experimentation and greater interpretability, particularly beneficial in resource-constrained environments. The table dataset is separated into 2 parts X and Y. Y contains the group column, which acts as our classifier and X contains the rest of the columns in the table dataset. X is the features that might determine whether a person has dementia or not.

3.1. Feature selection for machine learning models

The feature selection for the table dataset is performed using the SelectKBest from the Sklearn library. The SelectKBest uses the ANOVA F-value to rank the features [25], [26]. The ANOVA method measures the difference in means between two or more features and tests if this difference is significant. A large F-value indicates that the model provides a good fit to the data and the differences between the features are significant. The ANOVA-based feature selection shown in Figure 6. As revealed in the feature selection analysis, MMSE, nWBV, and gender (M/F) were among the most discriminative features, reinforcing their clinical relevance in AD diagnosis. From this we extracted the top 7 most relevant features, which are MMSE, nWBV, M/F, EDUC, MR Delay, eTIV, and SED. These 7 features are the ones that will be used later in our ML models.

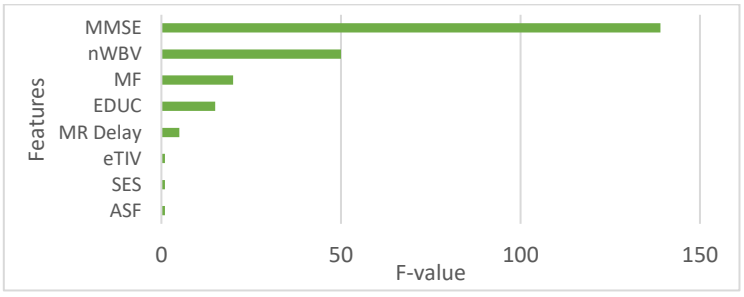

Figure 6. ANOVA F-value scores of selected features for the ML model

3.2. Data preparation

The values for each column are normalized using the StandardScaler class from Sklearn. StandardScaler normalizes the data by subtracting it with the mean value and scaling it to unit variance [27]. The formula for StandardScaler is as in (1).

$$z = \frac{x-\mu}{\sigma} \quad (1)$$

Where μ is the mean and σ is the standard deviation. The dataset is then divided into 2 sets. 80% of the dataset is reserved as training data, while the remaining 20% is reserved for testing the model.

3.3. Extreme gradient boosting classifier

Developing a classification model using the XGBClassifier class from the XGBoost library. XGBClassifier is an implementation of the XGBoost algorithm for binary and multi-class classification [28]. Similarly, to XGBoost, it uses gradient boosting (GB) on decision trees (DTs) to produce accurate predictions for classification problems. The XGBClassifier model is developed using the default parameters of the class. By using the default parameters, the model is able to achieve decent results. The model has an accuracy, average precision, average recall, and average F1-score of 72%. The only lacking performance comes from the recall score for the demented group. The performance metric has a value of 65%, the only metric of the model to have a value less than 70%.

For hyperparameter tuning, GridSearchCV class is utilized from the Sklearn library. GridSearchCV takes in a ML model, a parameter grid to search over, and a scoring method to evaluate the performance of the model on the validation set. In this example GridSearchCV is performed on the XGBClassifier parameters using accuracy as its metric of performance. Using cross-validation, the XGBClassifier model with default parameters had an average accuracy of 83.2%. By tuning the parameters using GridSearchCV, the cross-validation accuracy of the model had increased to 85%.

3.4. Random forest

Developing a classification model using the RF classifier class from the sklearn library. RF classifier is an ensemble learning method that trains multiple DTs and aggregates [29] their outputs to make a final prediction. Each DT is trained on a random subset of the data, then the final prediction is made by taking the average of the predictions of all trees. The RF classifier model is developed using the default parameters of the class. By using the default parameters, the model is able to achieve decent results. The model has an accuracy, average precision, average recall, and average F1-score of 72%. There are no outlier results in the performance metrics. For hyperparameter tuning, GridSearchCV is utilized class from the Sklearn library. In this example GridSearchCV is performed on the RF classifier parameters using accuracy as its metric of performance. Using cross-validation, the RF classifier model with default parameters had an average accuracy of 82.9%. By tuning the parameters using GridSearchCV, the cross-validation accuracy of the model had increased to 84.9%.

3.5. Decision tree

Developing a classification model using the DT classifier class from the Sklearn library. DT classifier works by recursively dividing the dataset into smaller and smaller subsets based on a feature until all the samples in a particular subset belong to the same class [30]. The final result of the model will have a tree-like structure with its root node as the original dataset, branches corresponding to features, and leaf nodes representing the final prediction made by the model. The DT classifier model is developed using the default parameters of the class. By using the default parameters, the model is able to achieve decent results. The model has an accuracy, average precision, average recall, and average F1-score of 79%. There are no outlier results in the performance metrics. Using cross-validation, the DT classifier model with default parameters had an average accuracy of 72.6%. By tuning parameters using GridSearchCV, cross-validation accuracy of the model had increased to 78.2%.

3.6. K-nearest neighbor

Developing a classification model using the K-neighbors classifier class from the Sklearn library. K-neighbors classifier finds the class of an unseen sample data by finding the KNN to it in the training dataset and voting for the class that appears the most among these KNN [31]. The K-neighbors classifier model is developed using the default parameters of the class. By using the default parameters, the model is able to achieve subpar results compared to the previous models. The model has an accuracy, average precision and average recall of 67%. The recall for the demented group is considerably low, at 59% compared to the nondemented group at 74%. Using cross-validation, the K-neighbors classifier model with default parameters had an average accuracy of 71.8%. By tuning the parameters using GridSearchCV, the cross-validation accuracy of the model had increased to 74.1%.

3.7. Support vector machine

Developing a classification model using the support vector classification (SVC) class from the Sklearn library. SVC works by finding the hyperplane in high dimensional space that best separates the data

into classes. The hyperplane is chosen in a way that maximizes the distance between the hyperplane and the closest data points from each class, called support vectors. During the classification process, new data points are mapped into the high dimensional space and then classified based on which side of the hyperplane they fall on. The SVC model is developed using the default parameters of the class. By using the default parameters, the model is able to achieve decent results. The model has an accuracy, average recall, and average F1-score of 76%. The recall for the demented group is considerably low, at 68% compared to the nondemented group at 84%.

For hyperparameter tuning, GridSearchCV class is utilized from the sklearn library. In this example GridSearchCV is performed on the SVC parameters using accuracy as its metric of performance. Using cross-validation, the SVC model with default parameters had an average accuracy of 76.9%. By tuning the parameters using GridSearchCV, the cross-validation accuracy of the model had increased to 78.5%.

3.8. Multilayer perceptron

Developing a classification model using the multilayer perceptron (MLP) classifier class from the Sklearn library. MLPClassifier is composed of multiple layers of interconnected nodes, called neurons, which process and transmit information. Each neuron in the MLP receives input from the neurons in the previous layer, processes the information using an activation function, and transmits the output to the next layer. This step repeats until the output layer where the model will assign a class to the data. The MLPClassifier model is developed using the default parameters of the class. The following are the default parameters of MLPClassifier.

By using the default parameters, the model is able to achieve above average results. The model has an accuracy, average recall, average precision, and average F1-score of 72%. There are no outlier results in the performance metrics. For hyperparameter tuning, GridSearchCV class is utilized from the sklearn library. In this example GridSearchCV is performed on the MLPClassifier parameters using accuracy as its metric of performance. Using cross-validation, the MLPClassifier model with default parameters had an average accuracy of 76.2%. By tuning the parameters using GridSearchCV, the cross-validation accuracy of the model had increased to 79.5%.

3.9. Stochastic gradient descent

Developing a classification model using the stochastic gradient descent (SGD) classifier class from the Sklearn library. SGDClassifier uses SGD algorithm for training the model. The basic idea of SGD is to approximate the gradient of the loss function with respect to the model's weights using one training sample. Then, the model's weights are updated in the direction of the negative gradient to reduce the loss. The SGDClassifier model is developed using the default parameters of the class.

By using the default parameters, the model is able to achieve really good results. The model has an accuracy, average recall, and average F1-score of 84%. The recall value for the nondemented group is and the precision value for the demented group are both 100%. Using cross-validation, the SGDClassifier model with default parameters had an average accuracy of 77.2%. By tuning the parameters using GridSearchCV, the cross-validation accuracy of the model had increased to 80%.

3.10. Stochastic gradient descent and logistic regression

By using the default parameters, the model is able to achieve really good results. The model has an accuracy, average recall, and average F1-score of 81%. There are no outlier results in the performance metrics. In this example GridSearchCV is performed on the GB classifier parameters using accuracy as its metric of performance. Using cross-validation, the GB classifier model with default parameters had an average accuracy of 83.2%. By tuning the parameters using GridSearchCV, the cross-validation accuracy of the model had decreased to 83%. GridSearchCV is performed on the logistic regression parameters using accuracy as its metric of performance. Using cross-validation, the GB classifier model with default parameters had an average accuracy of 79.6%. By tuning parameters using GridSearchCV, cross-validation accuracy of the model remained the same at 79.6%.

4. PERFORMANCE COMPARISON

To assess the effectiveness of the proposed ML models for AD classification, a comprehensive evaluation was conducted using four standard performance metrics: accuracy, precision, recall, and F1-score. Accuracy represents the overall proportion of correct predictions, while precision quantifies how many of the predicted positive cases are actually positive. Recall, or sensitivity, measures the model's ability to identify true positive cases, which is especially important in clinical applications such as AD detection. The F1-score, the harmonic mean of precision and recall, provides a balanced metric for assessing both correctness and completeness of the classification performance. Table 5 is the performance metrics of ML models.

Table 5. Performance metrics of ML models

Model	Accuracy (default) (%)	Accuracy (tuned) (%)	F1-score (tuned)	Recall (demented) (%)	Notes
XGBoost	83.2	85.0	0.85	84.1	Best tuned model overall
RF	82.9	84.9	0.84	82.0	High recall and interpretability
SVM	76.9	78.5	0.76	68.0	Lower recall on demented class
MLPClassifier	76.2	79.5	0.79	74.3	Good balance, slow convergence
SGDClassifier	77.2	80.0	0.81	81.0	Very fast training
KNN	71.8	74.1	0.72	59.0	Weakest recall performance
DT	72.6	72.2↓	0.70	65.0	Slight performance degradation
GB	81.3	83.0	0.83	80.0	Good generalist performer

The performance of all classifiers was evaluated using 5-fold cross-validation, and hyperparameter tuning was performed using GridSearchCV to identify optimal configurations. Across all models tested, XGBoost demonstrated the highest classification performance after tuning, achieving an accuracy of 85.0% and an F1-score of 0.85. Most notably, XGBoost attained a recall of 84.1% for the demented class, indicating a strong ability to correctly identify patients with AD. RF also performed robustly, with a tuned accuracy of 84.9%, an F1-score of 0.84, and a demented-class recall of 82.0%. Both models benefitted significantly from ensemble learning techniques, which reduce overfitting and improve generalization. Other models such as the SVM and MLP classifiers showed moderate performance improvements after tuning, achieving recall values of 68.0% and 74.3% respectively. The SGDClassifier offered a lightweight yet effective alternative, with a tuned recall of 81.0%. In contrast, the KNN model exhibited the weakest recall for the demented class at 59.0%, indicating poor sensitivity and limiting its clinical applicability. The DT classifier slightly degraded in performance post-tuning, likely due to overfitting specific parameter combinations, underscoring the superiority of ensemble methods for this task.

In the clinical context, recall is particularly important, as it reflects the model's ability to identify patients who are actually affected by AD. False negatives cases where patients with AD are incorrectly classified as healthy can lead to delayed intervention and worsen patient outcomes. Therefore, high recall in the demented class is prioritized over precision in this study. Ensemble models such as XGBoost and RF are thus more suitable for early-stage AD diagnosis, where sensitivity is vital for patient care. While the current evaluation framework provides a strong initial understanding of model performance, future work will expand upon these results by incorporating additional evaluation tools. In particular, receiver operating characteristic (ROC) curves and their corresponding area under the curve (AUC) values will be utilized to quantify model discrimination ability across classification thresholds. While accuracy and recall are useful metrics, the ROC-AUC will provide a more nuanced assessment of classifier discrimination ability, especially under imbalanced class conditions typical of medical datasets. Confusion matrices will also be generated to visually represent class-specific prediction outcomes and misclassification patterns. Furthermore, precision-recall curves will be employed to evaluate performance in imbalanced data scenarios where the demented class may be underrepresented. These additional analyses will offer a more nuanced understanding of the strengths and weaknesses of each model and will enhance the overall robustness and clinical interpretability of the proposed diagnostic framework.

This research focused on structured tabular data derived from MRI metadata and clinical assessments, such as MMSE scores, brain volume metrics (e.g., eTIV and nWBV), and demographic features. These features are clinically meaningful and have been shown to be effective indicators of cognitive decline and neurodegeneration. The use of structured data allowed the application of classical ML models such as RF, XGBoost, and SVMs, which are computationally efficient, easier to interpret, and more suitable for deployment in real-world low-resource settings. Additionally, structured data-based models often require smaller training datasets and can still yield competitive performance in disease classification tasks.

Looking forward, future work will address the limitations of this study by incorporating multimodal approaches that combine both imaging data and structured clinical features. A multimodal ML architecture could involve parallel processing of MRI images using CNNs and integration with tabular features via fully connected networks or ensemble fusion methods. This would enable the model to benefit from both spatial imaging patterns and domain-specific clinical insights. Furthermore, to overcome hardware limitations, future experiments will be conducted using cloud-based platforms such as Google Colab Pro, Amazon web services (AWS), or institutional high-performance computing (HPC) environments. These platforms provide access to GPU resources and scalable infrastructure, allowing for the training and evaluation of more complex deep learning architectures, including CNNs, autoencoders, and hybrid models. Through these enhancements, the proposed diagnostic framework can be expanded to deliver higher accuracy, richer interpretability, and broader applicability across diverse clinical contexts.

5. CONCLUSION

This study demonstrated that classical ML algorithms, applied to structured clinical and neuroanatomical features from MRI metadata, can effectively detect AD with high accuracy. The XGBoost and RF classifiers achieved classification accuracies of 85.0% and 84.9%, respectively, with recall scores exceeding 82% for the demented class. These results highlight the potential of using lightweight and interpretable models for early-stage AD detection, particularly in low-resource clinical settings. Unlike deep learning models, which require large-scale datasets and extensive computing power, our approach leverages easily obtainable tabular data, offering faster and more transparent decision-making. Future work will focus on integrating multimodal learning strategies and further enhancing model interpretability through explainable AI techniques.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Research Management Center, Multimedia University, Malaysia, for supporting the article processing charge (APC) of this publication.

FUNDING INFORMATION

The authors declare that no funding has been received for this research work.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Rahman Md Mojnur	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			
Md. Tanjil Sarker	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			
Hezerul Abdul Karim		✓		✓	✓	✓	✓			✓		✓	✓	
Fahmid Al Farid			✓		✓	✓		✓		✓	✓			
Aziah Ali		✓		✓	✓		✓			✓		✓	✓	
Wan-Noorshahida Mohd-Isa				✓	✓		✓			✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest regarding the publication of this paper.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author and the first author upon reasonable request.

REFERENCES

[1] M. Nasb, W. Tao, and N. Chen, "Alzheimer's disease puzzle: delving into pathogenesis hypotheses," *Aging and Disease*, vol. 15, no. 1, pp. 43–73, 2024, doi: 10.14336/AD.2023.0608.

[2] J. L. Liss *et al.*, "Practical recommendations for timely, accurate diagnosis of symptomatic Alzheimer's disease (MCI and dementia) in primary care: a review and synthesis," *Journal of Internal Medicine*, vol. 290, no. 2, pp. 310–334, 2021, doi: 10.1111/joim.13244.

[3] E. S. Finn, R. A. Poldrack, and J. M. Shine, "Functional neuroimaging as a catalyst for integrated neuroscience," *Nature*, vol. 623, no. 7986, pp. 263–273, 2023, doi: 10.1038/s41586-023-06670-9.

[4] A. Bari *et al.*, "Advancements in multi-view human activity recognition for ambient assisted living," in *2024 Multimedia University Engineering Conference (MECON)*, Cyberjaya, Malaysia, 2024, pp. 1-6, doi: 10.1109/MECON62796.2024.10776109.

[5] S. Iqbal, A. N. Qureshi, J. Li, and T. Mahmood, "On the analyses of medical images using traditional machine learning techniques and convolutional neural networks," *Archives of Computational Methods in Engineering*, vol. 30, no. 5, pp. 3173–3233, 2023, doi: 10.1007/s11831-023-09899-9.

[6] N. Kagiyama *et al.*, "PRIME 2.0: an update to the proposed requirements for cardiovascular imaging-related machine learning evaluation checklist," *JACC: Cardiovascular Imaging*, vol. 19, no. 2, pp. 225–251, Feb. 2026, doi: 10.1016/j.jcmg.2025.08.004.
[7] K. Jomova, S. Y. Alomar, S. H. Alwasel, E. Nepovimova, K. Kuca, and M. Valko, "Several lines of antioxidant defense against oxidative stress: antioxidant enzymes, nanomaterials with multiple enzyme-mimicking activities, and low-molecular-weight antioxidants," *Archives of Toxicology*, vol. 98, no. 5, pp. 1323–1367, 2024, doi: 10.1007/s00204-024-03696-4.
[8] A. P. Porsteinsson, R. S. Isaacson, S. Knox, M. N. Sabbagh, and I. Rubino, "Diagnosis of early Alzheimer's disease: clinical practice in 2021," *Journal of Prevention of Alzheimer's Disease*, vol. 8, no. 3, pp. 371–386, 2021, doi: 10.14283/jpad.2021.23.
[9] C. Özdemir, "Adapting transfer learning models to dataset through pruning and Avg-TopK pooling," *Neural Computing and Applications*, vol. 36, no. 11, pp. 6257–6270, 2024, doi: 10.1007/s00521-024-09484-6.
[10] C. Özdemir, "Classification of brain tumors from MR images using a new CNN architecture," *Traitement du Signal*, vol. 40, no. 2, pp. 611–618, 2023, doi: 10.18280/ts.400219.
[11] Y. Dogan, C. Özdemir, and Y. Kaya, "Enhancing CNN model classification performance through RGB angle rotation method," *Neural Computing and Applications*, vol. 36, no. 32, pp. 20259–20276, 2024, doi: 10.1007/s00521-024-10232-z.
[12] C. Özdemir, "Designing effective models for COVID-19 diagnosis through transfer learning and interlayer visualization," *Balkan Journal of Electrical and Computer Engineering*, vol. 11, no. 4, pp. 340–345, 2023, doi: 10.17694/bajece.1274253.
[13] C. Özdemir, M. A. Gedik, H. Küçüker, and Y. Kaya, "Gender identification from left hand-wrist X-ray images with a hybrid deep learning method," *Konya Journal of Engineering Sciences*, vol. 11, no. 4, pp. 1118–1133, 2023, doi: 10.36306/konjes.1294139.
[14] S. Grueso and R. V. Sobera, "Machine learning methods for predicting progression from mild cognitive impairment to Alzheimer's disease dementia: a systematic review," *Alzheimer's Research and Therapy*, vol. 13, no. 1, 2021, doi: 10.1186/s13195-021-00900-w.
[15] P. P. Singh, "Alzheimers-dataset-4-class-of-images," *kaggle.com*. 2019. Accessed: Mar. 28, 2024. [Online]. Available: https://www.kaggle.com/datasets/preetpalsingh25/alzheimers-dataset-4-class-of-images
[16] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00444-8.
[17] H. Choi, "Detecting early Alzheimer's," *kaggle.com*. 2018. Accessed: Mar. 28, 2024. [Online]. Available: https://www.kaggle.com/code/hyunseokc/detecting-early-alzheimer-s/input?select=oasis_longitudinal.csv
[18] S. E. Sorour, A. A. Wafa, A. A. Abohany, and R. M. Hussien, "A deep learning system for detecting cardiomegaly disease based on CXR image," *International Journal of Intelligent Systems*, vol. 2024, no. 1, 2024, doi: 10.1155/2024/8997093.
[19] A. S. Tarawneh, A. B. A. Hassanat, K. Almohammadi, D. Chetverikov, and C. Bellinger, "SMOTEFUNA: synthetic minority over-sampling technique based on furthest neighbour algorithm," *IEEE Access*, vol. 8, pp. 59069–59082, 2020, doi: 10.1109/ACCESS.2020.2983003.
[20] R. Foorthuis, "On the nature and types of anomalies: a review of deviations in data," *International Journal of Data Science and Analytics*, vol. 12, no. 4, pp. 297–331, 2021, doi: 10.1007/s41060-021-00265-1.
[21] J. A. Ruffolo, L.-S. Chu, S. P. Mahajan, and J. J. Gray, "Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies," *Nature Communications*, vol. 14, no. 1, Apr. 2023, doi: 10.1038/s41467-023-38063-x.
[22] S. K. Mukhiya and U. Ahmed, *Hands-on exploratory data analysis with Python: perform EDA techniques to understand, summarize, and investigate your data*. Birmingham, United Kingdom: Packt Publishing, 2020.
[23] M. T. Khatun, M. G. Sadeque, M. N. Akter, M. T. Sarker, D. Kundu, and A. H. M. I. Ferdous, "Doctor house: a telemedicine health service for people in remote areas of Bangladesh," *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Chittagong, Bangladesh, 2025, pp. 1-6, doi: 10.1109/ECCE64574.2025.11013894.
[24] S. P. Trivedi, S. Corderman, E. Berlinberg, A. Schoenthaler, and L. I. Horwitz, "Assessment of patient education delivered at time of hospital discharge," *JAMA Internal Medicine*, vol. 183, no. 5, pp. 417–423, 2023, doi: 10.1001/jamainternmed.2023.0070.
[25] M. A. Tariq, "A study on comparative analysis of feature selection algorithms for students grades prediction," *Journal of Information and Organizational Sciences*, vol. 48, no. 1, pp. 133–147, 2024, doi: 10.31341/jios.48.1.7.
[26] M. R. Barman, A. A. Biswas, M. Sultana, A. Rajbongshi, M. S. Zulfiker, and T. Tabassum, "Automated classification of diseased cauliflower: a feature-driven machine learning approach," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 22, no. 4, pp. 887–896, 2024, doi: 10.12928/TELKOMNIKA.v22i4.25812.
[27] A. Cuadros, D. L. Torre, and W. Ticona, "Machine learning predictive model for thyroid disease detection," in *Software Engineering Methods Design and Application*, Cham: Springer, 2024, pp. 267–282, doi: 10.1007/978-3-031-70285-3_20.
[28] M. S. Khan *et al.*, "Comparison of multiclass classification techniques using dry bean dataset," *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 6–20, 2023, doi: 10.1016/j.ijcce.2023.01.002.
[29] P. V'kovski, A. Kratzel, S. Steiner, H. Stalder, and V. Thiel, "Coronavirus biology and replication: implications for SARS-CoV-2," *Nature Reviews Microbiology*, vol. 19, no. 3, pp. 155–170, 2021, doi: 10.1038/s41579-020-00468-6.
[30] M. A. Jahin, M. S. H. Shovon, M. F. Mridha, M. R. Islam, and Y. Watanobe, "A hybrid transformer and attention based recurrent neural network for robust and interpretable sentiment analysis of tweets," *Scientific Reports*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-76079-5.
[31] S. K. Gunda, "Fault prediction unveiled: analyzing the effectiveness of random forest, logistic regression, and kneighbors," in *2nd International Conference on Self Sustainable Artificial Intelligence Systems, ICSSAS 2024*, 2024, pp. 107–113, doi: 10.1109/ICSSAS64001.2024.10760620.

BIOGRAPHIES OF AUTHORS

Rahman Md Mojnur is a data science enthusiast from Bangladesh with a bachelor's degree in Computer Science (Hons) in software engineering from Multimedia University, Malaysia. He has developed expertise in software development, data analytics, and machine learning, using technologies like C++, Java, Python, JavaScript, MySQL, PHP, and MongoDB. He has worked on various projects, including a digital parking system, medical clinic management tools, and an Alzheimer's disease detection system using machine learning techniques. His research interests lie in the application of artificial intelligence and data science to solve real-world problems. Always eager to learn, he is passionate about exploring innovative solutions in AI and its impact on businesses and industries. He can be contacted at email: rahmanmdmojnur@gmail.com.

Md. Tanjil Sarker obtained his Ph.D. degree in Engineering in the Faculty of Engineering, Multimedia University, Malaysia in 2022. He has been a research scholar in the same institution for four years. He has completed his M.Sc. degree in Computer Science and Engineering from Jagannath University, Dhaka, Bangladesh in 2017. He received the B.Sc. degree in Electrical and Electronic Engineering and MBA degree in Human Resource Management from Bangladesh University, Dhaka, Bangladesh in 2013 and 2015, respectively. Now he is working as post-doctoral research fellow in the Faculty of Engineering, Multimedia University, Malaysia. He was a graduate student member in the IEEE student branch of the Malaysia section. He has conducted many research works in relevant fields. His research interest is in system identification, signal processing and control, power system analysis, high voltage engineering, renewable energy, and solar system. He can be contacted at email: tanjilbu@gmail.com.

Hezerul Abdul Karim obtained his B.Eng. degree in Electronics with Communications from University of Wales Swansea, United Kingdom, in 1998 and M.Eng. science degree from Multimedia University, Malaysia in 2003. He obtained his Ph.D. degree from Center for Communication Systems Research, University of Surrey, United Kingdom in 2008. He is an associate professor at Multimedia University since November 2015. He is also deputy dean of student affair and alumni at Faculty of Engineering, Multimedia University. He has been teaching multimedia and computing engineering subjects. His research interests include telemetry, 2D/3D image/video coding and transmission, error resilience and multiple description video coding, and deep learning in image and video. He is senior member of IEEE and currently the vice chair of IEEE signal processing society, Malaysia section. He can be contacted at email: hezerul@mmu.edu.my.

Fahmid Al Farid received his Doctor of Philosophy (Ph.D.) by research in Information Technology from Multimedia University, Cyberjaya, Malaysia in 2023. He received the B.S. degree in Computer Science and Engineering from University of Chittagong, Bangladesh, in 2010, and the M.S. degree from the Faculty of Computer Science and Electrical Engineering, University of Ulsan, South Korea, in 2015. From 2013 to 2014, he was a research assistant in Embedded System Laboratory, University for Ulsan, South Korea. In 2015, he was a research assistant in Ubiquitous Computing Technology Research Institute, Sungkyunkwan University, South Korea. He received Korean BK21 PLUS Scholarship, supported by Korean Government in MS (2012-2014). Now he is working as post-doctoral research fellow in the Faculty of Engineering, Multimedia University, Malaysia. He can be contacted at email: fahmid.farid@gmail.com.

Aziah Ali is a lecturer at the Faculty of Computing and Informatics, Multimedia University. She received her Ph.D. from the Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia. She teaches courses such as web programming, computer graphics, digital image processing, and multimedia fundamental courses for undergraduates, diploma, and foundation students. She is passionate about imparting wisdom and knowledge to students and ensuring the course objectives are achieved by the end of each course. She continuously pursues lifelong learning with the objective of upskilling herself with new technologies and building her research networks. Her research interests include image and signal processing focusing on the biomedical field. More specifically, her work examines how image and signal processing techniques could contribute towards better and more efficient ways of analysing and understanding medical data such as images and signals. She is now discovering new areas that can potentially benefit from image and signal processing, especially in industries. Her key area is in data analytics and she is proficient in MATLAB, Python, PowerBI, Google Looker Studio, SQL, PHP, and HTML. She has published articles on biomedical data processing in indexed journals and peer-reviewed conference proceedings. She can be contacted at email: aziah.ali@mmu.edu.my.

Wan-Noorshahida Mohd-Isa is an assistant professor at the Faculty of Computing and Informatics, Multimedia University, Cyberjaya. She holds bachelor's in Electrical Engineering from Vanderbilt University, a master's from the University of Southampton, and a Ph.D. from Multimedia University. Her research focuses on visual analytics, machine learning, and biometric analysis. She was a Laureate of the MyTIGER 2022 France-Malaysia Joint Research Programme. She has collaborated on biometric R&D with MIMOS and secured multiple research grants, including RM88,000 for stingless bee identification and a Telekom R&D grant for visual analytics in telecommunications. She played a key role in establishing MMU's center of excellence on big data and blockchain technology. Previously, she served as deputy dean (2016-2020) and is a MyTRIZ level 3 practitioner and IEEE member. She actively mentor's students and promotes STEM education in Malaysia. She can be contacted at email: wan.noorshahida.isa@mmu.edu.my.