MR Image Feature Analysis for Alzheimer's Disease Detection Using Machine Learning Approaches

D. S. A. Aashiqur Reza , Sadia Afrin , Md. Ahsan Ullah , Sourav Kumar Kha , Sadia Chowdhury Toma , Raju Roy , Lasker Ershad Ali*

Mathematics Discipline, Khulna University, 9208 Khulna, Bangladesh

* Correspondence: Lasker Ershad Ali (ershad@math.ku.ac.bd)

Abstract: Alzheimer's disease (AD), a progressive neurological disorder, predominantly impacts cognitive functions, manifesting as memory loss and deteriorating thinking abilities. Recognized as the primary form of dementia, this affliction subtly commences within brain cells and gradually aggravates over time. In 2023, dementia's financial burden for elderly adults aged 65 and older was projected to reach $345 billion, encompassing health care, long-term care, and hospice services. Alarmingly, Alzheimer's disease claims one in three seniors, outnumbering combined fatalities from breast and prostate cancer. Currently, the diagnostic landscape for Alzheimer's lacks definitive tests, and diagnoses based purely on biological definitions have been observed to possess low predictive accuracy. In the presented study, a diagnostic methodology has been proposed using machine learning models that harness image features derived from brain MRI scans. Specifically, nine salient image features, grounded in color, texture, shape, and orientation, were extracted for the study. Four classifiers — Naïve-Bayes, Logistic regression, XGBoost, and AdaBoost — were employed, as the challenge presented a binary classification scenario. A grid search parameter optimization technique was employed to fine-tune model configurations, ensuring optimal predictive outcomes. Conducted experiments utilizing the Kaggle dataset, and for each model, parameters were rigorously optimized. The XGBoost classifier demonstrated superior performance, achieving a test accuracy of 92%, while Naïve Bayes, Logistic Regression, and AdaBoost registered accuracies of 63%, 70%, and 72%, respectively. Relative to contemporary methods, the proposed diagnostic approach exhibits commendable accuracy in predicting AD. If AI-based predictive diagnostics for AD are realized using the strategies delineated in this study, significant benefits may be anticipated for healthcare practitioners.

Keywords: Alzheimer's disease (AD), Machine learning, Image classification, Disease detection, XGBoost, MRI, Feature extraction

1 Introduction

AD has been identified as a predominant aging-related illness. Increases in average life expectancy in recent years have amplified the economic and societal implications of AD [1]. As advancements in treatments for central nervous system (CNS) disorders emerge, the significance of aiding medical practitioners in establishing precise early diagnostic distinctions has been accentuated [1]. Over the next forty years, a tripling in AD prevalence is anticipated, escalating from 27 to 106 million. During this period, it is projected that 1 in 85 individuals worldwide will be afflicted by the disorder. Notably, even a minor one-year delay in the onset and progression of the disease could potentially reduce the number of cases by 9 million [2].

AD has been noted as the leading cause of memory impairment in both senile and presenile individuals, with the severity of the condition amplifying as individuals age [3]. Neuroimaging investigations have become instrumental in the early diagnosis of AD, acquiring valuable data and insights from multiple sources, including structural MRI, functional MRI (fMRI), and blood perfusion information [4]. In 1980, clinical criteria for AD diagnosis were established using a binary evaluation methodology by the National Institute of Neurologic and Communicative Disorders and Stroke in conjunction with the AD and Related Disorders Association [5].

MR brain imaging, in evaluating degeneration in grey and white matter tissues, has frequently been employed in AD diagnosis [6]. The preliminary diagnostic steps often involve differentiating between normal aging-related physical and cognitive changes and symptoms suggestive of Mild Cognitive Impairment. An in-depth review of the MRI, complemented by a clinical interview with the affected individual, can facilitate the confirmation of memory loss or AD diagnosis [7]. Contemporary research endeavors have focused on the development of deep learning-based diagnostics for AD, exploring MRIs using robust classification models [8]. Integration of state-of-the-art computational machine learning algorithms with advanced neuroimaging methodologies has enabled the identification of structural and molecular biomarkers associated with AD [9]. AD, characterized by a progressive loss of cognitive capabilities, currently lacks effective or curative treatments [10]. Given the escalating number of AD cases, early diagnosis has become paramount. One challenge in image signal processing is the vast number of features, which contributes to extended execution times. Thus, pivotal image features have been extracted, aiming to reduce this execution duration by limiting feature numbers.

The study at hand seeks to detect AD through brain MRIs via machine learning methodologies. Its significance is rooted in its potential contribution to the early detection and diagnosis of AD, thereby paving the way for effective management, treatment planning, and preparatory care for patients and their families. Features like image intensity, texture, shape, eccentricity, orientation, entropy, coarseness, energy, homogeneity, dissimilarity, standard deviation, variance, and skewness of image intensities have been harnessed for the feature extraction phase. Utilizing these features, AD detection is pursued through machine learning models such as Naïve-Bayes, Logistic Regression, XGBoosting, and AdaBoosting. It is hoped that these findings might facilitate early AD diagnosis, potentially enhancing life expectancy for affected individuals.

The subsequent section (section 2) encompasses an extended literature review. The materials and methodologies adopted in this investigation are detailed in section 3. The findings of this research are elucidated in section 4, culminating in a comprehensive conclusion in section 5.

2 Related Works

The literature encompasses various approaches for the prediction and detection of AD, primarily bifurcating into deep learning and machine learning methodologies.

In recent years, advancements have been observed in the deployment of convolutional neural networks (CNNs) for AD detection. An identification system for AD, encompassing both binary and multi-class tasks, was proposed by Maqsood et al. [11] in 2019, wherein a pre-trained CNN, augmented with transfer learning, was validated using data from brain MRI images. Parallel research by Tambe et al. [12] explored multiple deep-learning models, with the VGG19 model exhibiting an accuracy of 91.38%. Additionally, a machine learning methodology predicated on GoogleNet, VGG, and Alexnet was constructed by Muhammed Raees and Thomas [13]. Such patterns of utilizing CNN-based transfer learning for both binary and multiclass classifications of AD were also noticed in works by Ghaffari et al. [14], where ResNet101, Xception, and InceptionV3 were employed. Substantial performance metrics were reported by Mamun et al. [15], where CNN outperformed other models when applied to a dataset of 6219 MRI images for AD detection. Beyond MRI data, EEG recordings have also been harnessed. Ieracitano et al. [16] introduced a data-driven strategy, utilizing CNN to discern binary and multiclass images, achieving accuracies of 89.8% and 83.3% respectively. Complementing this, Liu et al. [17] leveraged a 3D ShuffleNet model based on ResNet and DenseNet within a CNN architecture. Song et al. [18] conducted an exhaustive analysis, and findings suggested RF, MLP, and CNN achieved accuracies of 90.2%, 89.6%, and 90.5%, respectively. It is worth noting that while these deep learning algorithms can process large and intricate datasets, occasionally outperforming traditional machine learning algorithms, there remains the possibility of overfitting.

Conversely, traditional machine-learning approaches have been pursued extensively for AD prediction. For instance, Achilleos et al. [19] employed decision trees (DT) and random forests (RF) for the classification of normal control (NC) cases of AD. Kadhim et al. [20] conducted a comprehensive review, emphasizing the positive results achieved by algorithms such as SVM, U-Net Architecture 2.5D, and ResNet. Savaş [21] embarked on a journey to discern the early stages of AD, classifying them using brain MRIs. In a more recent study by Bigham et al. [22], an automated system was developed to classify AD patients, leveraging diffusion tensor imaging (DTI) and MATLAB for the analysis. Gupta et al. [23] focused on an ensemble of machine learning algorithms, integrating features from voxel-based morphometry (VBM), cortical and subcortical volumetric features, and hippocampal volumetric features. Decision trees, particularly pruned variants like J48, were applied by Battineni et al. [24] to prognosticate late-life AD. Baglat et al. [25] tested an assortment of machine learning methodologies on T1-weighted MRI data, with RF and AdaBoost classifiers yielding an accuracy of 86%. Tuan et al. [26] proposed a deep learning model tailored for 3D brain MR image segmentation.

It is discernible from the literature that while deep learning models often proffer higher accuracy rates, they come with the trade-offs of increased computational complexity, interpretability challenges, and resource-intensive training processes. On the other hand, machine learning models, owing to their simplicity and well-defined mathematical

underpinnings, often exhibit swifter training durations and reduced computational demands. Therefore, it becomes pivotal to balance these trade-offs based on the specific requirements and constraints of a study. The following section (Section 3: Materials and Methods) delves into the machine learning methodologies adopted in this research, elucidating ten feature extraction techniques complemented by four classification models.

3 Methodology

The objective of this study was to predict AD using four distinct machine learning classification models, drawing upon nine specific features extracted from brain MR images.

Upon the careful acquisition of data, key image features were quantified. Intensity, shape, entropy, eccentricity, energy, coarseness, homogeneity, and dissimilarity were computed for both the training and testing sets. These computations were facilitated using formulas described in Section 3, executed within a Python environment supported by the OpenCV library. The criticality of feature selection in image classification became evident as intensity, representative of pixel brightness, and shape, delineating object structure, were assessed. Additionally, texture metrics such as entropy, which quantifies randomness, and dissimilarity, evaluating pixel value variations, were integrated. The amalgamation of these diverse features provided a comprehensive portrayal of visual attributes, enhancing the model's prowess in distinguishing and classifying image objects.

Post feature extraction, the machine learning models were trained using the identified features. Parameter optimization was undertaken to ascertain optimal results. While model performance is often evaluated using accuracy, this metric may not suffice in scenarios with imbalanced class distributions or varying error costs. As such, sensitivity, specificity, and AUC values were also considered pivotal performance indicators in this analysis.

The overarching process of the study is illustrated in Figure 1: Systematic Flowchart of the Proposed Approach. Detailed discussions on feature extraction and classification strategies are presented in the ensuing sections.
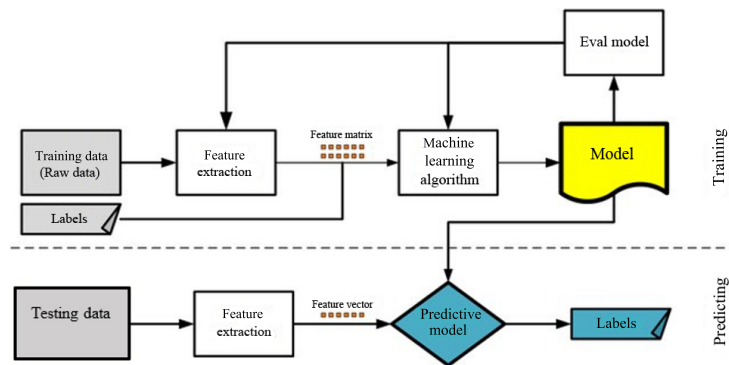

Figure 1. The systematic flowchart of the proposed approach

3.1 Image Feature Extraction

Feature extraction is understood as the transformation of raw, unprocessed data into numerical attributes that can be processed, ensuring the retention of the original dataset's pertinent information. Image features, in this context, signify visual patterns or characteristics extractable from an image, presenting its content in a more concise and interpretable manner. These extracted features encapsulate pivotal information, from shapes and textures to edges and other visually relevant properties. Historically, attributes such as shape, texture, pixel intensity, homogeneity, and dissimilarity have been identified as the most pertinent features in an image. It has been observed that leveraging these specific image attributes results in enhanced performance. In this study, the features of pixel intensity, shape, entropy, orientation, eccentricity, energy, coarseness, homogeneity, and pixel dissimilarity were examined, as delineated in the following subsections.

3.1.1 Pixel intensity

Pixel intensity describes the level of brightness or darkness exhibited by image pixels. In digital imaging, a numerical value is attributed to each pixel, representing its intensity. Pixels with higher luminosity are associated with elevated intensity values, whereas their darker counterparts exhibit diminished values. This elemental property has been harnessed in diverse applications spanning edge detection, contrast augmentation, and image segmentation. Not just mean intensity, but also variance, standard deviation, and skewness of the intensity have been considered as features in this investigation. The average intensity of an image is computationally determined using the equation:

$$Intensity = \frac{1}{m} \sum I(x, y) \tag{1}$$

where, *m* represents the total number of pixels, and *I(x, y)* denotes the intensity at each specific image point.

3.1.2 Shape features

The concept of shape in imagery pertains to the spatial layout and geometric properties inherent to objects or regions encapsulated within. This encompasses the contour, boundary delineations, and the general structure of the visual content. Shape analysis, a method of extracting and chronicling geometric object properties, finds utility in various applications such as object recognition, character recognition, and medical image diagnostics. The geometric features of objects embedded in images are deciphered to gain invaluable insights, underscoring the significance of shape analysis in the broader realms of image processing and computer vision.

3.1.3 Image entropy

Entropy, in the image domain, delineates the unpredictability quotient inherent to an image. In the realm of image coding, entropy serves as a baseline, dictating the average coding length, quantified in bits per pixel, achievable by optimal coding techniques without information compromise. Its value oscillates not just with image focus levels but also entropy. The computation of pixel value entropy within a 2-dimensional expanse centered at coordinates *(i, j)* reveals the image's entropy. A multifaceted image, boasting a diverse range of pixel values, is indicated by a higher entropy number, whereas simplicity and uniformity are represented by a diminished entropy figure. The entropy of an image is mathematically defined by:

$$\eta = \sum p * \log(p) \tag{2}$$

where, *p* denotes the normalized histogram counts corresponding to a specific grey level of a pixel.

3.1.4 Patterns local orientations

Orientation, a salient feature of edges and patterns exhibiting a marked direction, provides insights into intricate image facets. Local orientation, beyond merely deciphering the directionality of a pattern, is an indispensable component of motion analysis, involving the scrutiny of an image to deduce the orientation of an embedded object based on its angular disposition.

3.1.5 Eccentricity of an image region

Eccentricity, in the context of a connected graph, is defined as the shortest path length between one vertex and any other vertex. Within the realm of ellipses, the ratio of the length of the major axis to the distance between its foci determines the eccentricity, producing values ranging between 0 and 1, with degenerate cases at both extremities. Specifically, an ellipse possessing an eccentricity of 0 is identified as a perfect circle, while an eccentricity of 1 delineates a line segment. Eccentricity measures the discrepancy between the predicted central position of an object in the image and its actual location. It has been postulated that for precise measurements, potential eccentricities ought to be rectified. The intersection of a double-napped cone, related to a conic section, and a plane offers a method to ascertain eccentricity. In this study, the mean eccentricity of the image was derived as a feature by calculating the eccentricity of each region, utilizing the formula:

$$e = 1 - \frac{b^2}{a^2} \tag{3}$$

where, *a* represents the length of the major axis and *b* symbolizes the minor axis length of the region.

3.1.6 Energy features

The term 'energy' in texture attributes encapsulates the collective magnitude or intensity of the pixel values within an image, signifying the image's intricacy spectrum. A subdued energy signifies an image leaning towards smoothness and reduced texture, while an elevated energy resonates with an image possessing richer details and textures. The energy of an image is mathematically deduced using:

$$E = \sum_{i=1}^{N} \sum_{j=1}^{N} p(i,j)^2 \tag{4}$$

where, $p(i, j)$ denotes the (i, j)th element of the normalized Gray-Level Co-occurrence Matrix (GLCM).

3.1.7 Image coarseness features

Image coarseness characterizes the magnitude or dimensionality of the predominant textures within an image, shedding light on the granularity or fidelity of texture patterns. Images leaning towards fineness possess diminutive, less conspicuous textures, in contrast to coarse images characterized by pronounced, substantial textures. The coarseness of an image is evaluated through:

$$Coarseness = \frac{1}{Q}\sum_{d=1}^{Q}\frac{1}{2(L-d)}\sum_{x=1}^{L-d}\sum_{y=1}^{L}|I(x,y)-I(x+d,y)| \tag{5}$$

where, Q refers to the number of directional considerations, L represents the image size, and $I(x, y)$ is the intensity at the specified pixel location.

3.1.8 Pixels homogeneity

Pixel homogeneity alludes to a measure denoting the uniformity or similarity among pixel values in an image. A criterion for its evaluation is how congruently pixel values align with the median value of their proximal neighborhood. Enhanced homogeneity implies that neighboring pixels possess analogous values, resulting in an image that appears more coherent. The homogeneity of an image is computed as:

$$H = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{1+|i-j|}p(i,j) \tag{6}$$

3.1.9 Image pixels dissimilarity features

Conversely, dissimilarity functions as a counterpoint to homogeneity, gauging the disparity or variation in pixel values within an image. It signals the presence of contrasting textures or patterns, emphasizing distinctions between adjacent pixels. A heightened dissimilarity is emblematic of an image teeming with variations. The dissimilarity metric is derived using:

$$Diss = \sum_{i=1}^{N}\sum_{j=1}^{N}|i-j|p(i,j) \tag{7}$$

3.2 Classification Models

Classification pertains to the task of predicting a class label for a given input data sample. In this study, four distinct classification models were utilized for the analysis of extracted features, aiming to predict AD. These models include logistic regression, naïve bayes, XGBoost, and AdaBoost.

3.2.1 Logistic regression model

Logistic regression, often employed in binary classification scenarios, is designed to predict the probability of an instance aligning with one of two categorical outcomes. The output is usually bound between two classes, such as true/false or yes/no. Using the logistic function, predicted values are constrained between 0 and 1. A decision boundary at 0.5 is typically established; values below this threshold are assigned to one class, while values above belong to the other. The logistic function is represented as:

$$P(I) = \frac{Exp(c+dI)}{1+Exp(c+dl)} \tag{8}$$

where, $P(I)$ denotes the probability associated with the image pixels, while c and d represent model parameters.

3.2.2 Naïve bayes classifier

The Naïve Bayes algorithm, despite its simplicity, is a robust mechanism for binary classification. Drawing from the Bayes theorem, it operates under the assumption of conditional independence of features, given the class label. Probabilities of an instance falling under each class are computed, with the class of highest probability chosen as the final prediction. It is particularly adept at handling text and categorical data. The Bayes theorem, in its most rudimentary form, is expressed as:

$$P(A \mid B) = \frac{P(B \mid A) * P(A)}{P(B)} \tag{9}$$

where, $P(A \mid B)$ is the posterior probability of interest, while $P(A)$ denotes the prior probability or the marginal likelihood of occurrence.

3.2.3 AdaBoost classifier

AdaBoost, or Adaptive Boosting, emerges as a recognized ensemble learning strategy for classification tasks. By aggregating a multitude of weak learners, commonly decision trees, a robust classifier is constructed. Training instances are assigned weights, which are then adjusted based on the performance of preceding models. Emphasis is particularly laid on misclassified instances in subsequent iterations. The outputs from these weak learners are combined through weighted voting to deliver the final classification.

3.2.4 XGBoost classifier

XGBoost, an abbreviation for Extreme Gradient Boosting, is an advanced gradient boosting machine learning algorithm that has shown proficiency in various tasks such as classification, regression, and ranking. This technique integrates a series of weak learners, predominantly decision trees, to formulate a potent prediction model. The gradient descent optimization technique is employed by XGBoost to minimize the associated loss function. Regularization algorithms are incorporated to counteract overfitting, enabling the model to efficiently manage high-dimensional datasets. In this investigation, the decision to employ these specific classifiers was informed by their individual merits. XGBoost, with its optimized speed and accuracy, and AdaBoost, celebrated for boosting decision trees' performance on binary classifications, were deemed suitable. The logistic regression model, owing to its simplicity and effectiveness with linearly separable classes, and the memory-efficient Naïve Bayes Classifier were also integrated into this research framework.

4 Results and Discussions

The experimental setup, encompassing the detailed description of the database utilized and the results achieved using various statistical measures, is presented in this section. Brain MR image data, crucial for the detection of AD, were sourced from Kaggle [27]. Within the original dataset, four distinct classes were identified: Mild impairment, Moderate impairment, No impairment, and Very low impairment. For the purpose of this research, these classes were amalgamated into two broader categories: Impairment and Non-impairment. The dataset comprised 5121 training images and 1279 test images. Of the training set, 2561 images were categorized under impairment, and 2560 under non-impairment. Conversely, the test set included 639 impairment images and 640 non-impairment images.

Each image was processed individually, and the features delineated in the methodology section were computed utilizing the Python programming language. Subsequent to this feature computation, two datasets, corresponding to the training and test sets, were established and archived in CSV format. Both datasets encompassed features elaborated upon in the image feature extraction section, coupled with a binary response variable: 0 indicating non-impairment and 1 indicating impairment.

Following the application of the stacked ensemble technique, its efficacy was ascertained using a labeled training set. The ensuing performance was quantified via the generation of a confusion matrix, with overarching metrics of accuracy, sensitivity, and specificity being derived for the experiment. Accuracy, indicative of the proportion of correctly predicted values, was ascertained using the formula:

$$Accuracy = \frac{TP + FP}{TP + FP + TN + FN} \tag{10}$$

where, *TP* and *TN* represent the counts of true positives and true negatives, respectively. *FP* denotes false positives, while *FN* corresponds to false negatives. Sensitivity, which represents the proficiency of the model in correctly identifying positive cases, is calculated as:

$$Sensitivity = \frac{TP}{TP + FN} \tag{11}$$

Conversely, specificity, reflecting the model's capability in accurately identifying negative cases, is computed as:

$$Specificity = \frac{TN}{TN + FN} \tag{12}$$

Another performance metric, the receiver operating characteristic curve (ROC), provides insights into the performance of binary classifiers by capturing the entire area under this two-dimensional curve. The area under the ROC curve (AUC) serves as an indicator of the likelihood that the model would rank a randomly selected positive instance higher than a randomly selected negative one.

Table 1. Parameter settings for the classifiers

Model	Parameter Setup	Range of Parameters
Naiive-Bayes	Tuning not required	None
Logistic Regression	Regularization strength $= 10$, Regularization $= L2$	Regularization strength $= 0.01, 0.1, 1, 10$
XGBoosting	Estimators $= 100$, Max depth $= 5$, Learning rate $= 0.3$	Estimators $= 50, 100, 200$. Max Depth $= 3, 5, 7$. Learning rate $= 0.01, 0.03, 0.05, 0.1, 1$ None
AdaBoosting	Learning rate $= 0.1$, n_estimators $= 100$	Learning rate $= 0.01, 0.1, 0.2$ Estimator $= 50, 100, 200$

In the context of AD detection, predictive models were constructed employing the Naïve-Bayes classifier, logistic regression, XGBoost, and AdaBoost. To optimize the model's performance, parameter tuning was undertaken. The grid search method, coupled with a 5-fold cross-validation technique, was deployed to discern the optimal parameter values. After extensive iterations with diverse parameter settings for each model, the configurations presented in Table 1 were identified as the most efficacious.

The optimal parameter setting was applied to the training data, resulting in the subsequent training of the model. The test data were then employed to calculate accuracy, among other statistical values. The predictive capability of the models, as well as their respective accuracies, were determined using measures derived from the confusion matrix, juxtaposed with prediction results. These findings are elucidated in Table 2.

Table 2. Performance of the classifiers

Model	Accuracy	Sensitivity	Specificity	AUC
Naïve-Bayes	0.63	0.76	0.50	0.73
Logistic Regression	0.70	0.71	0.70	0.74
AdaBoost	0.72	0.64	0.80	0.78
XGBoost	**0.92**	**0.90**	**0.94**	**0.98**

As evidenced by Table 2, accuracy rates observed for Naive Bayes, Logistic Regression, AdaBoost, and XGBoost were 63%, 70%, 92%, and 72%, respectively. Among the classifiers investigated in this study, XGBoost displayed superior overall accuracy. The results suggest that the XGBoost model offers enhanced performance in predicting AD relative to the other models under consideration. Moreover, in comparison to the other three classifiers, XGBoost also exhibited the highest sensitivity, specificity, and AUC scores, noted at 90%, 94%, and 98% respectively. ROC curves were delineated, and the AUC was calculated to provide a deeper understanding of the prediction model, as depicted in Figure 2.
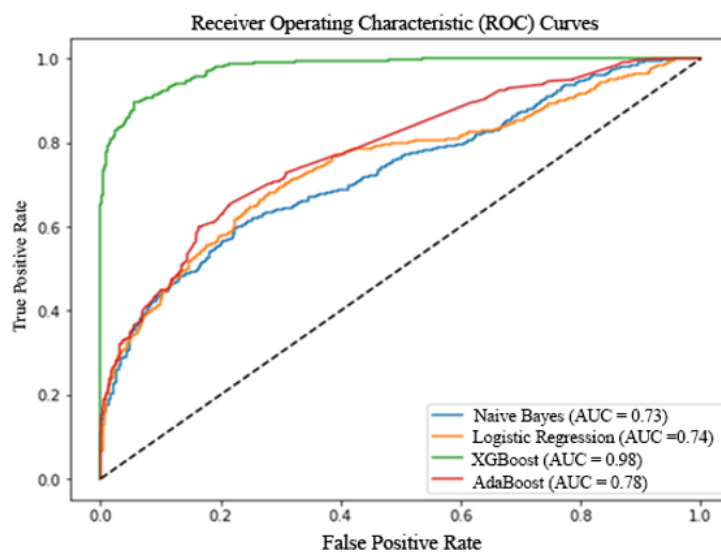

Figure 2. ROC curve for different classifiers

From an analysis of Figure 2, it can be discerned that the XGBoost model's ROC curve, represented by the green trajectory, is more fluent than those of the other three classifiers. The accompanying AUC values are annotated within the plot legend of Figure 2. AUC values ranging between 0.7 and 0.8 are deemed commendable, those between 0.8 and 0.9 are considered exceptional, and values exceeding 0.9 are characterized as outstanding. An AUC of 0.5 typically implies a lack of distinction. Given that the XGBoost classifier yielded an AUC of 0.98, it can be inferred that the proposed approach delivers significant results. The relative effectiveness of the four classifiers, as indicated by accuracy, sensitivity, specificity, and AUC values, is further portrayed in Figure 3. A perusal of Figure 3 reaffirms the superior performance of XGBoost across all metrics.

The marked performance of XGBoost over its counterparts, namely Naive Bayes, Logistic Regression, and AdaBoost, can be attributed to its inherent ability to process complex non-linear relationships within data sets. By progressively constructing an ensemble of decision trees designed to rectify errors engendered by preceding models, XGBoost adeptly discerns intricate patterns. The regularization techniques employed by XGBoost, including depth limitations and minimal child weight, have been identified as instrumental in mitigating overfitting. Furthermore, its

inherent support for handling missing data streamlines the preprocessing phase. The gradient boosting framework is noted to iteratively refine model weights, bolstering accuracy by focusing on challenging data instances. The flexibility endowed by XGBoost's hyperparameters renders it a formidable tool for achieving elevated predictive accuracy.

A direct comparison between the results gleaned from the proposed XGBoost classification model and existing state-of-the-art techniques proves challenging due to the disparate nature of the databases. However, previous studies, such as those conducted by Mehmood et al. [28], Odusam et al. [29], Venugopalan et al. [30], Pradhan et al. [31], Razavi et al. [32], and Islam and Zhang [33], have reported accuracies of 98%, 99%, 89%, 88%, 94.5%, and 88%, respectively, using diverse datasets from the AD Neuroimaging Initiative (ADNI) database. Remarkably, the model proposed in this study achieved a 92% accuracy with XGBoost by leveraging pertinent image features from a Kaggle dataset. This performance stands as competitive, even when juxtaposed with preceding research endeavors.
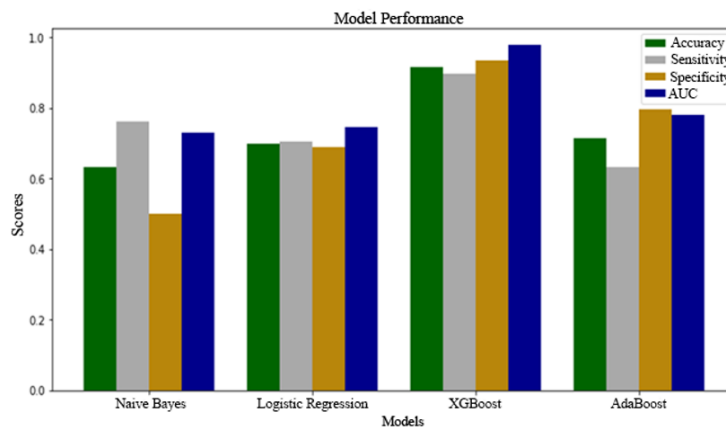

Figure 3. Performance of the classifiers

5 Conclusions

The principal objective of this study was to classify individuals into either impairment or non-impairment groups based on brain MR images. To achieve this distinction, brain MR images were amassed, and machine learning methodologies were employed. The results and corresponding performance metrics ascertain that an accuracy of 92% can be attained utilizing the proposed approach, effectively classifying individuals into the aforementioned groups. Furthermore, it was observed that the discrepancy between sensitivity and specificity is minimal, signifying the model's adeptness at unbiased classification, rather than gravitating towards prevalent responses. The research established that the proposed XGBoost machine learning model holds the potential to execute classifications with a precision rivalling that of deep learning models. Machine learning models, being less intricate in structure and more straightforward in implementation than deep learning models, necessitate fewer resources and a reduced duration for training and testing. Consequently, the adaptation of machine learning models in industry practices might lead to substantial savings in terms of time, computational memory, and financial costs.

However, as with most methodologies, avenues for refinement persist. Features employed in this study were chosen arbitrarily. A more judicious and methodical selection could potentially amplify the accuracy of the results. An impressive 92% accuracy in Alzheimer's disease prediction using XGBoost underscores its potential for future therapeutic implications. Such a precise model could serve as an invaluable clinical decision support tool, potentially aiding medical practitioners in the early identification and intervention of the disease. This model might facilitate personalized healthcare trajectories and risk-informed treatments by pinpointing those at elevated risk. The insights furnished by the model concerning predictive attributes could further enrich our understanding of Alzheimer's disease progression.

Nonetheless, hurdles impede the seamless translation of this research into practice. Challenges including data quality, privacy concerns, clinical receptiveness, model validation, model interpretability, evolving disease patterns, regulatory nuances, and constrained resources have been identified. Overcoming these obstacles necessitates a holistic approach encompassing diverse data procurement, ethical data stewardship, collaborative clinician engagement, cross-population validation, innovative interpretation techniques, consistent model updates, regulatory adherence, and strategic resource allocation. Together, these steps can pave the way for the effective integration of predictive models into the Alzheimer's disease clinical landscape. It was also noted that no pre-processing techniques were applied to the acquired images. Incorporating pre-processing prior to feature extraction might bolster the model's predictive prowess.

Author Contributions

The contributions of each author is listed in the following statement: "Conceptualization, D. S. A. Aashiqur Reza; methodology, D. S. A. Aashiqur Reza, Sadia Afrin, Md. Ahsan Ullah; software, D. S. A. Aashiqur Reza; validation, Lasker Ershad Ali, Raju Roy; formal analysis, D. S. A. Aashiqur Reza; investigation, Lasker Ershad Ali and Raju Roy; writing- original draft preparation, D. S. A. Aashiqur Reza, Sadia Afrin , Md. Ahsan Ullah ,Sadia Chowdhury Toma, Sourav Kumar Kha.; writing- review and editing, Lasker Ershad Ali and Raju Roy; visualization, D. S. A. Aashiqur Reza; Overall supervision: Prof. Dr. Lasker Ershad Ali.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability

The data of AD supporting our research results are deposited in Kaggle [27].

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Y. Liu, L. Teverovskiy, O. Carmichael, R. Kikinis, M. Shenton, C. S. Carter, V. A. Stenger, S. Davis, H. Aizenstein, J. T. Becker, O. L. Lopez, and C. C. Meltzer, "Discriminative MR image feature analysis for automatic schizophrenia and AD classification," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2004: 7th International Conference.* Springer Berlin Heidelberg, 2004, pp. 393–401. https://doi.org/10.1007/978-3-540-30135-6_48

[2] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, "Forecasting the global burden of AD," *Alzheimers Dement*, vol. 3, no. 3, pp. 186–191, 2007. https://doi.org/10.1016/j.jalz.2007.04.381

[3] W. P. Santos, R. E. Souza, A. F. D. Silva, and P. B. Santos Filho, "Evaluation of AD by analysis of MR images using multilayer perceptrons and committee machines," *Comput. Med. Imaging Graph.*, vol. 32, no. 1, pp. 17–21, 2008. https://doi.org/10.1016/j.compmedimag.2007.08.004

[4] A. Pulido, A. Rueda, and E. Romero, "Classification of AD using regional saliency maps from brain MR volumes," in *Medical Imaging 2013: Computer-Aided Diagnosis.* Florida, United States, 2013, pp. 192–199. https://doi.org/10.1117/12.2007092

[5] C. Salvatore, A. Cerasa, P. Battista, M. C. Gilardi, A. Quattrone, and I. Castiglioni, "Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: A machine learning approach," *Front. Neurosci.*, vol. 9, p. 307, 2015. https://doi.org/10.3389/fnins.2015.00307

[6] F. J. Martinez-Murcia, J. M. Górriz, J. Ramírez, and A. Ortiz, "A spherical brain mapping of MR images for the detection of AD," *Curr. Alzheimer Res.*, vol. 13, no. 5, pp. 575–588, 2016. https://doi.org/10.2174/1567205013666160314145158

[7] K. M. Poloni and R. J. Ferrari, "Detection and classification of hippocampal structural changes in MR images as a biomarker for AD," in *Computational Science and Its Applications–ICCSA 2018: 18th International Conference.* Melbourne, VIC, Australia, 2018, pp. 406–422. https://doi.org/10.1007/978-3-319-95162-1_28

[8] Z. Cui, Z. Gao, J. Leng, T. Zhang, P. Quan, and W. Zhao, "AD diagnosis using enhanced inception network based on brain Magnetic Resonance Image," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).* San Diego, CA, USA, 2019, pp. 2324–2330. https://doi.org/10.1109/BIBM47256.2019.8983046

[9] Z. S. Aaraji, H. H. Abbas, and A. Asady, "ADs detection by using MRI brain images: A survey," *J. Posit. Sch. Psychol.*, vol. 6, no. 8, pp. 5616–5634, 2022. https://journalppw.com/index.php/jpsp/article/view/10824/6985

[10] K. Mahapatra, "Detection of AD using MRI scans based on inertia tensor and machine learning," 2023, arXiv preprint arXiv:2304.13314. https://arxiv.org/abs/2304.13314

[11] M. Maqsood, F. Nazir, U. Khan, F. Aadil, H. Jamal, I. Mehmood, and O. Song, "Transfer learning assisted classification and detection of AD stages using 3D MRI scans," *Sensors*, vol. 19, no. 11, p. 2645, 2019. https://doi.org/10.3390/s19112645

[12] P. Tambe, R. Saigaonkar, N. Devadiga, and P. H. Chitte, "Deep learning techniques for effective diagnosis of Alzheimer's disease using MRI images," in *ITM Web of Conferences*, vol. 40. EDP Sciences, 2021, p. 03021. https://doi.org/10.1051/itmconf/20214003021

[13] P. C. Muhammed Raees and V. Thomas, "Automated detection of AD using deep learning in MRI," *J. Phys.: Conf. Ser.*, vol. 1921, no. 1, p. 012024, 2021. https://doi.org/10.1088/1742-6596/1921/1/012024

[14] H. Ghaffari, H. Tavakoli, and G. Pirzad Jahromi, "Deep transfer learning–based fully automated detection and classification of AD on brain MRI," *Br. J. Radiol.*, vol. 95, no. 1136, p. 20211253, 2022. https://doi.org/10.1259/bjr.20211253

[15] M. A. A. Mamun, S. B. Shawkat, M. S. Ahammed, M. M. Uddin, M. I. Mahmud, and A. K. M. M. Islam, "Deep learning based model for AD detection using brain MRI images," in *2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. New York, USA, 2022, pp. 0510–0516. https://doi.org/10.1109/uemcon54665.2022.9965730

[16] C. Ieracitano, N. Mammone, A. Bramanti, A. Hussain, and F. C. Morabito, "A convolutional neural network approach for classification of dementia stages based on 2D-spectral representation of EEG recordings," *Neurocomputing*, vol. 323, pp. 96–107, 2019. https://doi.org/10.1016/j.neucom.2018.09.071

[17] Z. Liu, H. Lu, X. Pan, M. Xu, R. Lan, and X. Luo, "Diagnosis of AD via an attention-based multi-scale convolutional neural network," *Knowl.-Based Syst.*, vol. 238, p. 107942, 2022. https://doi.org/10.1016/j.knosys.2021.107942

[18] M. Song, H. Jung, S. Lee, D. Kim, and M. Ahn, "Diagnostic classification and biomarker identification of AD with random forest algorithm," *Brain Sci.*, vol. 11, no. 4, p. 453, 2021. https://doi.org/10.3390/brainsci11040453

[19] K. G. Achilleos, S. Leandrou, N. Prentzas, P. A. Kyriacou, A. C. Kakas, and C. S. Pattichis, "Extracting explainable assessments of AD via machine learning on brain MRI imaging data," in *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*. Cincinnati, OH, USA, 2020, pp. 1036–1041. https://doi.org/10.1109/bibe50027.2020.00175

[20] K. A. Kadhim, F. Mohamed, and Z. N. Khudhair, "Deep learning: Classification and automated detection earlier of AD using brain MRI images," *J. Phys.: Conf. Ser.*, vol. 1892, no. 1, p. 012009, 2021. https://doi.org/10.1088/1742-6596/1892/1/012009

[21] S. Savaş, "Detecting the stages of AD with pre-trained deep learning architectures," *Arab. J. Sci. Eng.*, vol. 47, no. 2, pp. 2201–2218, 2022. https://doi.org/10.1007/s13369-021-06131-3

[22] B. Bigham, S. A. Zamanpour, and H. Zare, "Features of the superficial white matter as biomarkers for the detection of AD and mild cognitive impairment: A diffusion tensor imaging study," *Heliyon*, vol. 8, no. 1, p. e08725, 2022. https://doi.org/10.1016/j.heliyon.2022.e08725

[23] Y. Gupta, K. H. Lee, K. Y. Choi, J. J. Lee, B. C. Kim, and G. R. Kwon, "Early diagnosis of Alzheimer's disease using combined features from voxel-based morphometry and cortical, subcortical, and hippocampus regions of MRI T1 brain images," *PLoS One*, vol. 14, no. 10, p. e0222446, 2019. https://doi.org/10.1371/journal.pone.0222446

[24] G. Battineni, N. Chintalapudi, and F. Amenta, "Late-life AD detection using pruned decision trees," *Int. J. Brain Disord. Treat.*, vol. 6, p. 033, 2020. https://doi.org/10.23937/2469-5866/1410033

[25] P. Baglat, A. W. Salehi, A. Gupta, and G. Gupta, "Multiple machine learning models for detection of AD using OASIS dataset," in *Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation*. Tiruchirappalli, India, 2020, pp. 614–622. https://doi.org/10.1007/978-3-030-64849-7_54

[26] T. A. Tuan, T. B. Pham, J. Y. Kim, and J. M. R. Tavares, "Alzheimer's diagnosis using deep learning in segmenting and classifying 3D brain MR images," *Int. J. Neurosci.*, vol. 132, no. 7, pp. 689–698, 2022. https://doi.org/10.1080/00207454.2020.1835900

[27] "Dataset_Alzheimer," https://www.kaggle.com/datasets/yasserhessein/dataset-alzheimer/code?resource=download, accessed: 2023-07-22.

[28] A. Mehmood, S. Yang, Z. Feng, M. Wang, A. Ahmad, and R. Khan, "Transfer learning approach for early diagnosis of Alzheimer's disease on MRI images," *Neuroscience*, vol. 460, pp. 43–52, 2021. https://doi.org/10.1016/j.neuroscience.2021.01.002

[29] M. Odusami, R. Maskeliunas, R. Damaševičius, and T. Krilavičius, "Detection of early stage from functional brain changes in magnetic resonance images using a Finetuned ResNet18 network," *Diagnostics*, vol. 11, no. 6, pp. 1–16, 2021. https://doi.org/10.3390/diagnostics11061071

[30] J. Venugopalan, L. Tong, H. R. Hassanzadeh, and M. D. Wang, "Multimodal deep learning models for early detection of Alzheimer's disease stage," *Sci. Rep.*, vol. 11, no. 1, p. 3254, 2021. https://doi.org/10.1038/s41598-020-74399-w

[31] A. Pradhan, J. Gige, and M. Eliazer, "Detection of Alzheimer's Disease (AD) in MRI images using deep learning," *Int. J. Eng. Res. Technol.*, vol. 10, no. 3, pp. 580–585, 2021.

[32] F. Razavi, M. Tarokh, and M. Alborzi, "An intelligent Alzheimer's disease diagnosis method using unsupervised feature learning," *J. Big Data*, vol. 6, no. 32, pp. 1–16, 2019. https://doi.org/10.1186/s40537-019-0190-7

[33] J. Islam and Y. Zhang, "Early diagnosis of Alzheimer's disease: A neuroimaging study with deep learning architectures," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1881–1883. https://doi.org/10.1109/CVPRW.2018.00247