International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 11-No.2 (2025) pp. 1707-1715
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Comparative Evaluation of Feature Selection Techniques and Machine Learning Algorithms for Alzheimer's Disease Staging

Gayathri L[1*], Muralidhara B L[2], Rajesh B[3]

[1]Department of Computer Science and Applications, Bangalore University, Karnataka, India
* **Corresponding Author Email:** gayathri.lingaraju@gmail.com - **ORCID:** 0009-0003-4553-3468

[2]Department of Computer Science and Applications, Bangalore University, Karnataka, India
Email: murali@bub.ernet.in - **ORCID:** 0000-0001-7994-8694

[3]Mahindra University, School of Management, India,
Email: rajesh.balarama@mahindrauniversity.edu.in - **ORCID:** 0000-0001-7011-5338

Article Info:

DOI: 10.22399/ijcesen.1077
Received : 22 January 2025
Accepted : 05 March 2025

Keywords :

Alzheimer's disease,
Feature selection,
Machine learning,
Feature extraction,
Texture analysis.

Abstract:

Dementia encompasses a range of brain disorders characterized by cognitive decline, with memory loss as a hallmark symptom. Alzheimer's disease (AD), the most common form of dementia, progressively affects cognitive functions, leading to severe memory loss. Early and accurate detection of AD is essential for timely intervention, preventing further neuronal damage, and improving patient outcomes. This study employs machine learning (ML) techniques, feature selection methods, and texture analysis to enhance AD diagnosis. By systematically evaluating various feature selection techniques and Principal Component Analysis (PCA) in conjunction with multiple ML algorithms, the study identifies the most effective approach for classifying AD stages. The integration of texture-based features with ML models demonstrates a significant improvement in distinguishing Cognitive Normal, Mild Cognitive Impairment, and AD stages. These findings highlight the clinical significance of combining feature selection and texture analysis with ML for early AD diagnosis, facilitating more precise disease classification and contributing to personalized treatment strategies.

1. Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder defined by the gradual deterioration of brain cells and their connections. This decline leads to memory loss, cognitive impairment, and challenges in performing daily tasks. The disease's progression involves shrinking of the brain, particularly the hippocampus, and enlargement of fluid-filled spaces within the brain [1]. AD poses a major public health crisis globally, affecting millions of people and imposing substantial economic and social burdens [2]. The number of people living with dementia is alarmingly high, with over 50 million affected worldwide. This figure is expected to nearly triple by 2050, reaching approximately 152 million individuals [3]. India is also struggling with a rising tide of dementia. With an estimated 6.1 million elderly citizens, approximately 3.7% or 46,000 individuals are currently living with Alzheimer's [4]. Notably, the widespread occurrence of dementia is disproportionately higher among women and in rural areas [5], underscoring the critical need for effective prevention, diagnosis, and treatment strategies. AD is typically categorized into three stages: Cognitive Normal (CN), Mild Cognitive Impairment (MCI), and AD patients. Accurate diagnosis, particularly during the early stages, remains a challenge in healthcare.

In medical imaging, texture features play a vital role in capturing patterns within Magnetic Resonance Imaging (MRI) that indicates the presence and progression of AD. Texture features quantify variations in pixel intensity, capturing details about the structural properties of brain tissue. These features can highlight differences between healthy and diseased brain regions, providing valuable input for classification models.

Feature selection techniques are utilized to boost the efficiency and accuracy of classification models. These methods, help in determining the most

influential features from a large dataset, thereby reducing dimensionality and improving model performance. Additionally, Principal Component Analysis (PCA) is used as a dimensionality reduction technique that reduces data dimensions while preserving variance.

Machine Learning (ML) has become a potent instrument in disease classification, offering robust algorithms capable of learning from complex datasets [6]. Integrating feature selection and dimensionality reduction techniques with ML models enhances the classification performance of AD stages, ultimately contributing to improved patient outcomes.

This study evaluates the effectiveness of various feature selection methods alongside PCA in combination with multiple ML algorithms to identify the optimal approach for classifying AD stages. To enhance efficiency, only half of the most important features were selected based on each feature selection method, which aids in reducing dimensionality, minimizing computational complexity, and eliminating redundant or less relevant features. This comparative analysis provides valuable insights into the most effective feature selection methods for AD classification.

2. Literature review

Computer-aided diagnosis (CAD) systems have become indispensable tools for assisting physicians in recent years. Numerous studies have focused on developing CAD systems to aid in the detection of AD stages. This section reviews existing research that employed traditional ML techniques, texture analysis and feature selection methods, for AD diagnosis.

Several studies have explored ML approaches and feature selection techniques for AD classification, emphasizing the importance of selecting optimal features to enhance predictive performance. Alshamlan et al. [7], compared Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR) with feature selection methods such as Minimum Redundancy Maximum Relevance (mRMR) and Mutual Information (MI), finding that LR combined with mRMR achieved the highest accuracy. Arjaria et al. [8], investigated various ML algorithms, including SGD, k-NN, Decision Tree, AdaBoost, Neural Network, and Naïve Bayes, alongside feature selection and dimensionality reduction techniques like Information Gain, Gini Index, Chi-Squared, and PCA to optimize classification performance. Similarly, Uddin et al. [9], developed an ML model integrating GaussianNB, Decision Tree, Random Forest, XGBoost, Voting Classifier, and GradientBoost, leveraging clinical, demographic, and brain imaging features for AD classification, with the Voting Classifier combined with the Select K Best feature selection algorithm achieving the highest validation accuracy. Further studies have refined feature selection techniques to improve classification performance. Gu et al. [10], designed a pipeline integrating supervised and unsupervised feature selection methods, validated through a bootstrap sampling-based workflow on the ADNI dataset. Their findings demonstrated that incorporating stability with discriminability significantly enhances AI model performance. Yue et al. [11], developed an explainable prediction model using ensemble learning and feature selection on longitudinal aging study data from China, evaluating five feature selection techniques and nine ML classifiers. Their model identified critical predictive features for AD and MCI, achieving high accuracy, sensitivity, and specificity. Collectively, these studies highlight the effectiveness of combining advanced feature selection and classification techniques to enhance ML-based AD diagnosis. Few studies have explored ML approaches for AD classification using different feature extraction techniques. In study [12], authors Reddy and Nagireddy (2022), combined multiple features, including GLCM, 3D SIFT, HOG-TOP, and Complete LBP of Sign and Magnitude-Three Orthogonal Planes (CLBPSM-TOP), to classify subjects into CN, MCI, and AD using the OASIS dataset. Their ensemble approach, integrating SVM and k-NearestNeighbors (KNN), outperformed individual classifiers. In [13], AlSaeed & Omar (2022), investigated CNN-based feature extraction for automated AD classification using MRI images, analyzing the impact of fully connected layers and evaluating performance with Softmax, SVM, and RF classifiers. The study by Ahmadi et al. [14], employs ML techniques to assess the severity of AD using MRI images, with the Kaggle dataset categorized into four severity levels. A hybrid approach combining 12 feature extraction methods is utilized for diagnosis, followed by feature reduction using PCA. Six traditional ML classifiers such as, Decision Tree, KNN, Linear Discriminant Analysis, Naïve Bayes, SVM, and ensemble learning are applied to classify disease severity. Optimization is performed during training to enhance classifier performance. Additionally, a Convolutional Neural Network (CNN) model is trained using the extracted features, demonstrating superior accuracy in predicting the disease compared to traditional ML algorithms.

The literature survey reveals a notable lack of focus on texture features integrated with feature selection methods in conjunction with the performance of ML algorithms for AD diagnosis. While numerous

studies explore ML techniques and feature extraction individually, the synergistic impact of texture analysis, feature selection, and ML remains underexplored. Addressing this gap is crucial, as texture features can provide valuable insights into structural brain changes associated with AD, and optimal feature selection methods can enhance the discriminative power of ML models. This study aims to bridge this gap by systematically evaluating the role of texture features, feature selection and ML algorithms in improving AD stage classification.

3. Methodology

The following pseudocode provides an overview of the methodology employed in this study, encompassing the key steps of preprocessing MRI images, extracting texture-based features, selecting the most relevant features, and training ML models to classify AD stages. This structured representation highlights the workflow's key components and ensures replicability of the proposed approach.

3.1 Dataset

For this experiment, 1,026 3D brain 1.5T T1-weighted structural MRI scans were selected, including balanced 402 images of CN, AD, and MCI from ADNI database (http://adni.loni.usc.edu/) [15]. The slice thickness of MRI was 1.2 millimeters. The rows, columns, and slices have dimensions of 256 × 256 × 166, with voxels measuring 1.0 × 1.0 × 1.2 mm3. Figure 1 represents different stages of AD in coronal view.

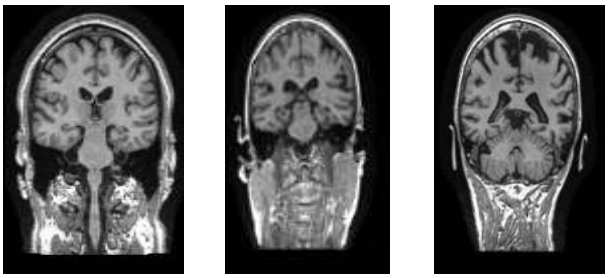

Figure 1. *Sample MRI images representing stages (shown from left to right): CN, MCI, and AD.*

Pseudocode for AD Stage Classification Workflow

```
BEGIN

Step 1: Load and preprocess MRI images
    LOAD 3D MRI images
    FOR each image:
        APPLY skull removal
        CO-REGISTER images to a standard anatomical template
        NORMALIZE voxel intensities

Step 2: Feature extraction
    INITIALIZE feature vector as empty

    FOR each 3D MRI image:
        COMPUTE GLCM features: energy, homogeneity, contrast, entropy, correlation, dissimilarity
        COMPUTE GLHA features: mean, variance, skewness, kurtosis
        COMPUTE NGTDM features: coarseness, contrast, busyness, complexity, strength
        COMBINE GLCM, GLHA, NGTDM features into feature vector

    STORE all feature vectors in database
Step 3: Feature selection
    APPLY feature selection methods:
      ANOVA, Chi-Square, Mutual Information Classifier (MICIF), PCA and ML algorithms
    SELECT top 50% of ranked features for each method
    STORE selected features

Step 4: Train machine learning models
    SPLIT dataset into training (80%) and testing (20%)
    INITIALIZE classifiers: KNN, DTC, GNB, SVM, MLP, RFC, GBC, XGB, ADB, ETC, CBC

    FOR each feature selection method:
        FOR each classifier:
            TRAIN classifier on training set using selected features
            TEST classifier on testing set
        RECORD accuracy, ROC-AUC, precision, recall, F1-score

Step 5: Evaluate performance
    COMPARE metrics across classifiers and feature selection methods
    IDENTIFY best-performing combination of feature selection method and classifier

Step 6: Output results
    DISPLAY top-performing model and feature selection combination

END
```

3.2 Preprocessing

The dataset acquired from ADNI underwent initial preprocessing steps during its acquisition, specifically addressing Gradwarp, B1 non-uniformity, and N3biasfield correction. Gradwarp utilizes an algorithm to enhance image appearance by eliminating distortion caused by gradient nonlinearity, which is particularly significant for non-linear gradient models. This correction improves image geometry, enhancing the utility of image information for analysis. B1 non-uniformity correction is used to rectify color and intensity distortions in the image, often caused by improper radio frequency transmission. Additionally, N3 bias field correction is applied to address intensity distortions resulting from dielectric effects during acquisition [16]. N3 bias field-corrected images underwent skull stripping using the deepbrain [17] method in Python. The extracted skull is linearly registered against 1mm MNI152-space using FSL-FLIRT [18].

3.3 Feature Extraction

Feature extraction techniques are crucial for capturing the essential characteristics of images, specifically in case of medical imaging for disease classification. In this work, three prominent texture analysis techniques are employed: GLCM, Gray Level Histogram Analysis (GLHA), and Neighboring Gray-Tone Difference Matrix (NGTDM).

GLCM is a technique that analyzes the spatial correlation between pixels in an image. It quantifies the frequency of pixel pairs with defined intensities and spatial arrangements. By examining these co-occurrence patterns, GLCM extracts essential texture features such as energy, homogeneity, contrast, entropy, correlation, and dissimilarity. These parameters offer insights into image texture characteristics, including uniformity, similarity, intensity variation, randomness, pixel interdependence, and local image variation [19,20]. The mathematical formulas for GLCM features are presented below.

$$Energy = \sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j)^2 \quad (1)$$

$$Homogeneity = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{P(i,j)}{1+|i-j|} \quad (2)$$

$$\text{Contrast} = \sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j).(i-j)^2 \quad (3)$$

$$\text{Entropy} = -\sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j)\log P(i,j) \quad (4)$$

$$\text{Correlation} = \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i\sigma_j} \quad (5)$$

$$\text{Dissimilarity} = \sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j)\,|i,j| \quad (6)$$

where, P(i,j): Probability of pixel pairs with intensity levels i and j occurring in the image. μ_i,μ_j: Mean intensity of rows and columns, $\sigma_i\sigma_j$: Standard deviation of rows and columns.

GLHA analyzes the spread of pixel intensities within an image. By plotting the frequency of occurrence of different gray levels, a histogram provides valuable understanding of the image's characteristics. From histogram, core statistical metrics such as mean, variance, skewness, and kurtosis are calculated to quantify image properties like overall brightness, contrast, asymmetry, and peakedness [21]. GLHA features are calculated using the formulas presented in Equations (7) to (11).

$$Mean\ (\mu) = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (7)$$

$$Variance(\sigma^2) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \quad (8)$$

$$Skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^3}{\sigma^3} \quad (9)$$

$$Kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^4}{\sigma^4} \quad (10)$$

$$Entropy = -\sum p(x_i)\log p(x_i) \quad (11)$$

where, x_i: Intensity of pixel i, N: Total number of pixels. p(x_i): probability of intensity x_i.

NGTDM is a texture analysis method that focuses on the contrast between a pixel and its neighbors. It quantifies the differences in gray levels between a pixel and the average gray level of its surrounding pixels within a defined region. Key features extracted from NGTDM include coarseness, which measures texture granularity; contrast, reflecting local intensity variations; busyness, indicating the rate of gray level change; complexity, representing image intricacy based on gray level fluctuations; and strength, quantifying the distinctiveness of textures within the image [22]. Equations (12) to (16) define the mathematical calculation of NGTDM features.

$$Coarseness = \frac{1}{\sum_{i=1}^{N} P(i)}\sum_{i=1}^{N}\frac{P(i)}{1+S(i)} \quad (12)$$

$$Contrast = \sum_{i=1}^{N} P(i)(i-\bar{I})^2 \quad (13)$$

$$Busyness = \frac{\sum_{i=1}^{N} P(i)S(i)}{\sum_{i=1}^{N} P(i)\,|i-\bar{I}|} \quad (14)$$

$$Complexity = \sum_{i=1}^{N} P(i)(i-\bar{I})^2 S(i) \quad (15)$$

$$Strength = \frac{\sum_{i=1}^{N} P(i)(s(i)+S(i+1))^2}{\sum_{i=1}^{N} P(i)S(i)} \quad (16)$$

where, S(i): Sum of absolute differences.

The texture-features derived from GLCM, GLHA, and NGTDM are integrated to enhance the analysis of medical images, providing a deeper understanding of their structural and textural properties and leading to improved diagnostic accuracy.

3.4 Feature Selection

Feature-selection is essential in reducing dimensionality, improving model performance, and enhancing interpretability in ML tasks. This research employed multiple feature selection techniques, including Analysis of Variance (ANOVA), Chi-Square, Mutual Information Classifier (MICIF), PCA, and various embedded ML models.

1. ANOVA: Is a statistical method used to identify notable differences between the means of multiple groups. In feature selection, ANOVA evaluates the significance of individual features by measuring the variance within groups to the variance between groups. Features with higher F-scores are considered more significant [23].
2. Chi-Square test: Determines whether categorical variables have a substantial relationship. In feature selection, it assesses the feature's independence from the target variable. Features with high Chi-Square values are more likely to be pertaining to classification tasks [24].
3. MICIF: MICIF evaluates the interdependence between attributes and target variable. Attributes with higher mutual information scores are deemed more predictive for classification [25].
4. PCA: A dimensionality-reduction technique that transforms high-dimension data into a low-dimensional space while preserving most of the data's variance [26]. It does this by generating new uncorrelated variables, known as principal components, which capture the maximum information from the raw data.
5. K-NearestNeighbors: KNN is a non-parametric method that evaluates feature importance by assessing how well features contribute to nearest neighbor classification accuracy [27].
6. DecisionTreeClassifier (DTC): DTC performs feature selection by prioritizing the most significant features at each node based on Gini impurity or information gain criteria. Features that lead to the high information gain are considered as most important [26].
7. ExtraTreesClassifier (ETC): ETC Combines multiple decision-trees and splits nodes based on random thresholds for feature values. The weight of each feature is derived from the reduction in impurity it achieves across all trees [28].
8. RandomForestClassifier (RFC): RFC [29], is an ensemble technique that constructs multiple decision trees. It computes feature importance by aggregating the decrease in impurity across all trees in forest, making it an effective tool for feature selection.
9. XGBoosting (XGB): Is a high-performance gradient boosting library that uses the Gradient-Boosting framework to create ML algorithms. Feature importance is measured by the gain, cover, and frequency of splits using each feature [30].

These feature selection methods are employed to enhance AD stage classification model performance by identifying and utilizing the most informative dataset features.

3.5 ML Algorithms

This section discusses the various ML algorithms employed to assess the classification task's performance based on different feature selection algorithms. A range of techniques, including instance-based, tree-based, Bayesian, margin-based learning, and neural network algorithms, were examined to evaluate their effectiveness in classifying data based on selected features.

1. Instance-Based Learning

KNN [27], is a simple, instance-based learning algorithm that operates on the principle of similarity. It classifies new instances by finding the predominant class among the k-nearestneighbors in the training set. In low-dimensional environments, this approach is helpful because to its simplicity and efficacy, but as dataset sizes and dimensions grow, it may become computationally costly.

2. Tree-based Methods

DTC is a hierarchical model that classifies data by sequentially splitting it based on feature values. Each internal node represents a decision based on a specific attribute, branches correspond to the possible outcomes of that decision, and leaf nodes represent class labels. The process continues untpure leaf nodes are reached, containing instances belonging to the same class [31]. RFC [29], is an ensemble method that builds multiple decision-trees and combines their predictions for classification or regression tasks. ETC, is similar to random forests but uses random splits of all observations at each node. It introduces more randomness to the tree-building process and can achieve better performance by reducing variance [28]. Gradient Boosting Classifier (GBC) (Friedman, 2001), sequentially builds an ensemble of trees, where each new tree focuses on correcting the errors of its predecessors. This method effectively handles complex datasets and is versatile for both classification and regression problems. XGB, is a powerful and efficient ML library that specializes in gradient boosting [30].

Designed for scalability and performance, it excels at creating complex models by combining numerous simpler models. Known for its speed and accuracy, XGB has gained popularity for various ML tasks and real-world applications. AdaBoost (Adaptive Boosting) [32], is an ensemble-learning method that sequentially integrates multiple weak classifiers to create a strong classifier. It assigns higher weights to misclassified instances in subsequent iterations, focusing the learning process on difficult examples. By iteratively improving the model's performance on challenging cases, AdaBoost effectively enhances overall classification accuracy. CatBoostClassifier (CBC) [33], is a high-performance gradient-boosting algorithm specifically designed to handle categorical features efficiently. It employs techniques like ordered boosting and categorical encoding to directly process categorical data without requiring extensive preprocessing. This enables CatBoost to achieve high performance on various datasets, particularly those with a significant number of categorical features.

3. Bayesian Methods

Gaussian Naive Bayes (GNB) [34], is based on Bayes' theorem with presupposition of independence between every pair of features. It's called 'naive' because it simplifies the calculation by assuming feature independence. GNB performs well in many real-world scenarios due to its simplicity and efficiency.

4. Margin-based learning

SVM [35], constructs a hyperplane or set of hyperplanes in a multi-dimensional space to separate different classes. To handle complex, non-linear patterns, SVMs employ kernel functions that implicitly project data into higher-dimensional spaces, allowing for more sophisticated decision boundaries.

5. Neural Networks

Multi-layer Perceptron (MLP) [36], is a feedforward artificial neural network with multiple interconnected layers of nodes. It processes input data through hidden layers, applying non-linear transformations to extract complex patterns. The output layer produces predictions or classifications. MLPs are trained with backpropagation, which iteratively adjusts weights to minimize prediction errors.

4. Results and discussion

4.1 Evaluation Metrics

The ROC-AUC score is particularly valuable in medical image classification as it offers a comprehensive assessment of a model’s diagnostic performance, especially when handling imbalanced datasets. In medical applications such as disease classification, tumor detection, or disease severity prediction, relying solely on accuracy is insufficient, as misclassifications can lead to serious consequences. Therefore, this study considers the ROC-AUC score as the primary performance metric for evaluating ML models.

4.2 Performance Comparison

The selection of an appropriate ML algorithm and feature engineering techniques is crucial for achieving optimal performance in classification tasks. This section provides a comparative analysis of different ML algorithms applied to the AD dataset, incorporating the outcomes of various feature selection methods. The evaluation metrics are utilized to measure the effectiveness of each ML algorithm. Table 1, summarizes the evaluation metrics across different feature selection methods and ML algorithms, emphasizing their impact on AD stage classification performance. A comprehensive assessment of ML algorithms using different feature selection approaches provides several key insights:

Overall Performance: Overall, the models performed better with the feature selection methods that retained fifty percent of the most important features. This highlights the importance of selecting relevant features to enhance model performance.

Tree-Based Methods: Among the algorithms, tree-based methods such as RFC and ETC consistently outperformed other classifiers across most feature selection techniques. For instance, ETC achieved the highest score especially when paired with PCA and ETC feature selection methods. The strong performance of ETC can be attributed to its use of randomized split selection, which introduces diversity among trees, reduces overfitting, and improves generalization.

Gradient Boosting Methods: GBC and XGB also showed robust performance, particularly with feature-selection methods like MICIF and PCA. XGBoost, in particular, demonstrated superior ROC-AUC scores, suggesting its effectiveness in distinguishing between classes.

Impact of Feature Selection: Feature-selection methods such as ANOVA and MICIF provided a notable improvement in the performance of complex classifiers like GBC and XGB. The selection of the top 50% features seems to strike a balance between reducing dimensionality and retaining critical information.

Algorithm-Specific Observations: KNN, demonstrated worse performance across all parameters, indicating its sensitivity to irrelevant features and high dimensionality. SVM, performed

moderately well but lagged behind tree-based and ensemble methods. GNB, struggled with lower performance metrics, likely due to its assumption of feature independence which is not well-suited for this dataset. Ensemble methods such as ADB and CBC provided mixed results, with CBC performing better in feature-rich environments but ADB struggling across various feature selection methods. The analysis shows the efficiency of tree-based and ensemble learning methods in handling the selected features for AD classification. Feature selection has a vital role in enhancing model performance, with methods such as PCA and MICIF showing considerable advantages. A bar plot (Figure. 2) of AUC-ROC scores across different feature-selection methods and ML algorithms further illustrates this, highlighting that model like RFC, GBC, and XGBClassifier consistently achieving higher scores, especially when paired with PCA and MICIF. Alzheimer's disease is studied and reported in the literature [37-40].

5. Conclusion

This study has demonstrated the efficiency of various ML algorithms and feature-selection methods in classifying stages of AD. Tree-based and ensemble learning methods, particularly Random Forest, Gradient Boosting, and XGBoost, demonstrated superior results in classification of AD. The importance of feature selection was highlighted, with PCA and MICIF providing substantial benefits. While this study provides valuable insights, some limitations must be acknowledged. The dataset used might not fully represent the entire spectrum of AD classification. Including additional clinical and demographic information could potentially increase model performance. Additionally, exploring different hyperparameter tuning strategies for each algorithm may lead to further optimizations. Future research should explore on adopting deep learning technique, such as CNNs, to capture complex patterns in neuroimaging data. Additionally, investigating the impact of longitudinal data and multimodal information could give a more thorough insight of

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.

Table 1. *Score of different ML algorithms based on various feature selection methods.*

Feature selection method	ML algorithms										
	KNN	**DTC**	**GNB**	**SVM**	**MLP**	**RFC**	**GBC**	**XGB**	**ADB**	**ETC**	**CBC**
ANOVA	66.31	67.44	59.66	73.94	59.27	75.58	72.45	77.98	61.94	77.40	69.38
Chi2	60.71	52.62	55.01	69.11	59.37	65.18	61.68	64.71	57.21	66.09	60.04
MICIF	59.01	56.88	58.44	64.87	61.04	74.02	71.12	76.41	62.56	73.94	68.98
PCA	61.96	63.66	67.42	70.90	65.11	81.65	78.86	83.71	67.05	89.49	67.20
KNN	62.02	61.93	55.96	68.24	61.83	70.30	67.03	70.30	62.37	73.42	65.10
DTC	61.28	66.79	59.46	67.77	61.77	76.89	69.71	76.94	61.84	79.29	66.25
ETC	60.75	65.51	60.68	67.19	60.90	76.10	70.86	76.77	65.01	80.63	66.11
RFC	57.89	61.34	60.59	65.97	60.54	75.11	70.79	76.17	63.67	79.64	66.29
XGB	60.93	62.01	58.53	63.73	62.57	74.26	72.44	76.44	65.15	77.77	66.61

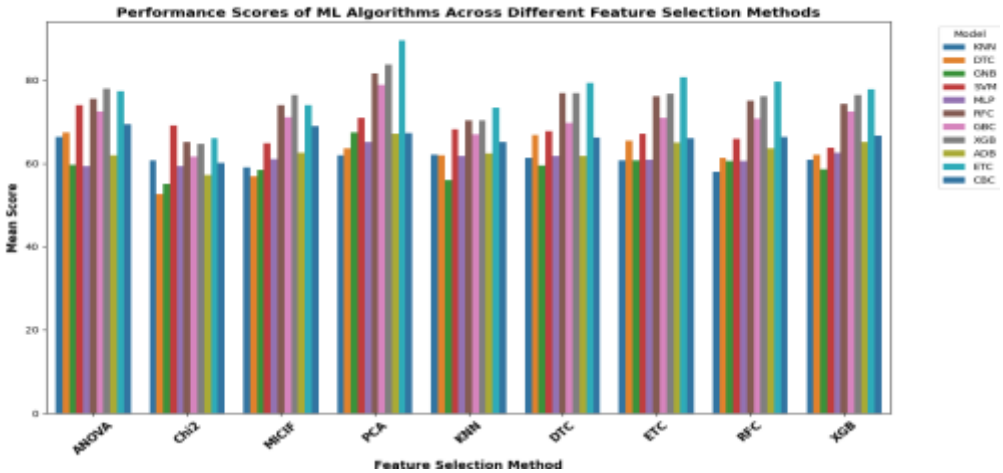

Figure 2. *ROC-AUC score of ML algorithm across different feature selection method.*

- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Basheera, S., & Ram, M. S. S. (2019). Convolution neural network–based Alzheimer's disease classification using hybrid enhanced independent component analysis based segmented gray matter of T2 weighted magnetic resonance imaging with clinical valuation. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *5*, 974-986.

[2] Angkoso, C. V., Tjahyaningtijas, H. P. A., Adrianto, Y., Sensusiati, A. D., Purnama, I. K. E., & Purnomo, M. H. (2022). Multi-features fusion in multi-plane MRI images for Alzheimer's disease classification. *Int. J. Intell. Eng. Syst*, *15*(4), 182-197.

[3] Ammal, S. M., & Manoharan, P. S. (2023). Multi-Headed Deep Learning Models to Detect Abnormality of Alzheimer's Patients. *Computer Systems Science & Engineering*, 44(1).

[4] AV, A., KUMAR, D. A. S., & LATIP, D. R. (2023). CNN-MOBILENETV2-DEEP LEARNING-BASED ALZHEIMER'S DISEASE PREDICTION AND CLASSIFICATION. *Journal of Theoretical and Applied Information Technology*, 101(9).

[5] Lee, J., Meijer, E., Langa, K. M., Ganguli, M., Varghese, M., Banerjee, J., & Dey, A. B. (2023). Prevalence of dementia in India: National and state estimates from a nationwide study. *Alzheimer's & Dementia*, *19*(7), 2898-2912.

[6] AlSaeed, D., & Omar, S. F. (2022). Brain MRI analysis for Alzheimer's disease diagnosis using CNN-based feature extraction and machine learning. *Sensors*, 22(8), 2911.

[7] Alshamlan H, Alwassel A, Banafa A, Alsaleem L. (2024). Improving Alzheimer's Disease Prediction with Different Machine Learning Approaches and Feature Selection Techniques. *Diagnostics*.; 14(19):2237. https://doi.org/10.3390/diagnostics14192237.

[8] Arjaria, S. K., Rathore, A. S., Bisen, D., & Bhattacharyya, S. (2024). Performances of Machine Learning Models for Diagnosis of Alzheimer's Disease. *Annals of Data Science*, 11(1), 307-335.

[9] Uddin, K.M.M., Alam, M.J., Jannat-E-Anawar. (2023). A Novel Approach Utilizing Machine Learning for the Early Diagnosis of Alzheimer's Disease. *Biomedical Materials & Devices* 1, 882–898. https://doi.org/10.1007/s44174-023-00078-9.

[10] Gu, F., Ma, S., Wang, X., Zhao, J., Yu, Y., & Song, X. (2022). Evaluation of feature selection for Alzheimer's disease diagnosis. *Frontiers in aging neuroscience*, 14, 924113.

[11] Yue, L., Chen, W. G., Liu, S. C., Chen, S. B., & Xiao, S. F. (2023). An explainable machine learning based prediction model for Alzheimer's disease in China longitudinal aging study. *Frontiers in Aging Neuroscience*, 15, 1267020.

[12] Reddy, G. N., & Nagireddy, K. (2022). A Robust Machine Learning Approach for Multiclass Alzheimer's Disease Detection using 3D Brain Magnetic Resonance Images. *Journal of Engineering Research* (2307-1877), 10.

[13] AlSaeed, D., & Omar, S. F. (2022). Brain MRI analysis for Alzheimer's disease diagnosis using CNN-based feature extraction and machine learning. *Sensors*, 22(8), 2911.

[14] Ahmadi M, Javaheri D, Khajavi M, Danesh K, Hur J (2024) A deeply supervised adaptable neural network for diagnosis and classification of Alzheimer's severity using multitask feature extraction. *PLoS ONE* 19(3): e0297996. https://doi.org/10.1371/journal.pone.0297996.

[15] Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., & Weiner, M. W. (2010). Alzheimer's disease Neuroimaging Initiative (ADNI) clinical characterization. *Neurology*, *74*(3), 201-209.

[16] Altaf, T., Anwar, S. M., Gul, N., Majeed, M. N., & Majid, M. (2018). Multi-class Alzheimer's disease classification using image and clinical features. *Biomedical Signal Processing and Control*, *43*, 64-74.

[17] SonicStrain. (2023). Skull stripping using DeepBrain. GitHub repository. https://github.com/SonicStrain/skull-stripping-using-deepbrain/tree/main, accessed on 3/03/2023.

[18] Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, & Smith SM. (2011) Fsl. *NeuroImage*. 2012;62:782–790. doi: 10.1016/j.neuroimage.09.015.

[19] Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*,SMC-3(6), 610-621.

[20] Soh, L. K., & Tsatsoulis, C. (1999). Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on geoscience and remote sensing*, *37*(2), 780-795.

[21] Gonzalez, R. C., & Woods, R. E. (2002). *Digital Image Processing* (2nd ed.). Prentice-Hall, Inc.

[22] Amadasun, M., & King, R. (1989). Textural features corresponding to textural properties. *IEEE*

Transactions on systems, man, and Cybernetics, *19*(5), 1264-1274.

[23] Montgomery, D. C. (2012). *Design and Analysis of Experiments* (8th ed.). Hoboken, NJ: John Wiley & Sons.

[24] Agresti, A. (2007). *An Introduction to Categorical Data Analysis* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

[25] Brown, G., Pocock, A., Zhao, M.J., & Luján, M. (2012). Conditional likelihood maximisation: A unifying framework for information theoretic feature selection. *Journal of Machine Learning Research*, 13, 27-66.

[26] Rupapara, V., Rustam, F., Ishaq, A., Lee, E., & Ashraf, I.(2023). Chi-square and PCA based feature selection for diabetes detection with ensemble classifier. *Intelligent Automation & Soft Computing*, 36(2), 1931-1949.

[27] Altman, N.S. (1992). An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *The American Statistician,* 46*(3)*, 175-185.

[28] Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning,*63(1), 3-42.

[29] Breiman, L.(2001). Random Forests. *Machine Learning*, *45*(1), 5-32.

[30] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

[31] Panigrahi, R., Borah, S., Day, N., Babo, R., & Ashour, A. S. (2018). Classification and analysis of facebook metrics dataset using supervised classifiers. *Social Network Analytics: Computational Research Methods and Techniques*, *1*.

[32] Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*(1), 119-139.

[33] Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. *arXiv* preprint arXiv:1810.11363.

[34] Hand, D. J., & Yu, K. (2001). Idiot's Bayes—not so stupid after all?. *International Statistical Review*, 69(3), 385-398.

[35] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273-297.

[36] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536.

[37]Johnsymol Joy, & Mercy Paul Selvan. (2025). An efficient hybrid Deep Learning-Machine Learning method for diagnosing neurodegenerative disorders. *International Journal of Computational and Experimental Science and Engineering*, 11(1). https://doi.org/10.22399/ijcesen.701

[38]P., A. M., & R. GUNASUNDARI. (2024). An Interpretable PyCaret Approach for Alzheimer's Disease Prediction. *International Journal of Computational and Experimental Science and Engineering*, 10(4). https://doi.org/10.22399/ijcesen.655

[39]S. Amuthan, & N.C. Senthil Kumar. (2025). Emerging Trends in Deep Learning for Early Alzheimer's Disease Diagnosis and Classification: A Comprehensive Review. *International Journal of Computational and Experimental Science and Engineering*, 11(1). https://doi.org/10.22399/ijcesen.739

[40]Rama Lakshmi BOYAPATI, & Radhika YALAVARTHI. (2024). RESNET-53 for Extraction of Alzheimer's Features Using Enhanced Learning Models. *International Journal of Computational and Experimental Science and Engineering*, 10(4). https://doi.org/10.22399/ijcesen.519