Penerbit Akademia Baru

Journal of Advanced Research Design

Journal homepage:
https://akademiabaru.com/submit/index.php/ard
ISSN: 2289-7984

Alzheimer's Disease Prediction Using ANOVA with t-SNE Feature Selection Techniques and Ensemble Learning

Ferdib-Al-Islam[1], Mostofa Shariar Sanim[1], Kah Ong Michael Goh[2,*], S M Hasan Mahmud[3,*], Dip Nandi[3]

[1] Department of Computer Science and Engineering, Northern University of Business and Technology Khulna, Khulna, 9100, Bangladesh
[2] Faculty of Information Science & Technology (FIST), Multimedia University, Jalan Ayer Keroh Lama, 75450, Melaka, Malaysia
[3] Department of Computer Science, American International University-Bangladesh (AIUB), 408/1, Kuratoli, Khilkhet, Dhaka 1229, Bangladesh

ARTICLE INFO

Article history:
Received 4 August 2025
Received in revised form 8 September 2025
Accepted 9 September 2025
Available online 23 September 2025

Keywords:

Alzheimer's Disease; Neurodegenerative Disorder; t-SNE; ANOVA; Machine Learning; SHAP

ABSTRACT

Alzheimer's disease stands as one of the most common neurodegenerative disorders, and currently, there is no cure for it. Early identification is pivotal for delaying disease continuation. The current approaches to Alzheimer's disease early detection rely on handwriting activities, which provide a significant quantity of data. Because of its great dimensionality, the final data obscures the significance of pertinent features. The challenge of dimensionality in data arises when there are too many features but not enough data samples, making it difficult for a model to discover a pattern in the data, affecting the many approaches used to diagnose or classify Alzheimer's disease. In this study, a way has been provided to overcome the curse of dimensionality by applying t-SNE and improve the efficacy of early Alzheimer's diagnosis by selecting key features using ANOVA; apart from that, seven machine learning algorithms have been used as base classifiers. These base classifiers were then used to create voting classifier results. The results of the studies indicate that the voting ensemble technique (approximately 94.28%) had the highest classification testing accuracy. Our approach has demonstrated its effectiveness by surpassing the latest benchmarks with our proposed technique. To comprehend how different features influence the model's outcome, we utilized Explainable AI (XAI) techniques, specifically SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME). Our proposed method has the potential to significantly improve the accuracy of early Alzheimer's disease diagnosis, laying the foundation for timely interventions and better patient outcomes.

1. Introduction

Alzheimer's Disease (AD), known as a progressive neurodegenerative [34] disorder, is an escalating worrisome in our population ages [1]. For improved treatment and possible intervention throughout the disease's course, early detection is essential. Analyzing handwriting is one such new technique. Given the complexity and uniqueness of handwriting, it may be possible to get important

* *Corresponding author.*
E-mail address: michael.goh@mmu.edu.my

https://doi.org/10.37934/ard.144.1.123147

insights regarding the neurological and cognitive abnormalities linked to Alzheimer's disease. This technique provides a noninvasive, affordable, and easily accessible way to identify minute changes in cognitive function, language proficiency, and motor control. Additionally, someone in the world is diagnosed with Alzheimer's dementia every three seconds. In 2020, there were approximately 55 million people living with Alzheimer's dementia globally. This number is expected to rise to 152 million by 2050, nearly doubling every 20 years to reach 78 million by 2030 [1]. As people grow older, the prevalence of Alzheimer's disease increases significantly. It affects 5.0% of adults aged 65 to 74, 13.1% of those aged 75 to 84, and 33.3% of individuals aged 85 and older. Although there is no known cure for the illness, people can control their deterioration over time with appropriate care to prolong their ability to live independently [2]. However, early detection of the condition allows for initiating therapy at the earliest stage, yielding the greatest outcomes. This makes it imperative that, as recommended by national and international health organizations, the population at risk of contracting the disease be screened regularly.

Even with the tremendous advancements in medicine, a complete treatment for AD is still unattainable. Remarkably, the diagnosis of Alzheimer's is still difficult, with a failure rate of more than 20%, despite the warning signs, which include personality changes, language difficulty, and memory loss [3]. The necessity of early diagnosis is highlighted by this high failure rate, which presents significant challenges to the development of effective therapeutics [4]. Early detection provides the best opportunity for therapeutic intervention; estimates indicate that detection can take place up to 9 years before symptoms of dementia that can be diagnosed appear [5]. Novel diagnostic approaches have been stimulated by AD's recognized impact on motor and cognitive skills, which are strongly correlated with handwriting motions [6, 7]. The majority of methods use biomarkers, including clinical, imaging, biochemical, and genetic ones, although it has been questioned whether these methods can accurately predict outcomes [8]. In addition, these methods are costly, invasive, time-consuming, and need supplies and equipment that are mostly found in hospitals and specialized laboratories. Since handwriting demands precise and well-coordinated body control [10], which relies on cognitive and motor functions that can be impaired by the progression of diseases, handwriting dynamics analysis may offer a low-cost and noninvasive method for evaluating the progression of the disease [9]. Subsequently, affordable and widely available graphic tablets were introduced to capture the kinematic and dynamic aspects of movement and to conduct various handwriting and drawing tests. Research on the neural processes related to motor learning and execution, as well as the decline of motor skills in writing and drawing among patients, has identified key characteristic features of their movements [11].

Based on the ensemble concept of link analysis, a technique created for feature ranking and dimensionality reduction and investigated the new approach's application potential in feature ranking, visualization, and other areas. The experimental findings demonstrate MRMD3.0's excellent dimensionality reduction capabilities. Although MRMD3.0 can support a wide range of feature rank methods, it still has issues with processing huge datasets slowly [12]. An innovative Convolutional Neural Network (CNN) was presented as an inexpensive, quick, and precise fix. The trial results demonstrated that the proposed model achieved satisfactory accuracy, surpassing the state-of-the-art baselines, which included 17 commonly used classifiers. [13]. The performance of three one-class classifiers: the Negative Selection Algorithm, the Isolation Forest, and the One-Class Support Vector Machine, was also evaluated [14]. The approaches attain state-of-the-art performance when compared to the state-of-the-art, suggesting that they might be a good substitute for the prevalent strategy. A methodology that avoids Curse of Dimensionality and enhances early Alzheimer's detection performance was published [15] and that method surpassed state-of-the-art benchmarks, showcasing the technology's effectiveness. According to the data that have been published, this

strategy can greatly enhance the accuracy of early AD detection and lay the groundwork for timely treatments and improved patient outcomes. In recent years, both the scientific and medical fields have experienced a rise in data complexity. High dimensionality, high noise, and data diversity are typically the result of many techniques and experimental setups. To investigate the essential information in the data, which is crucial for resolving several biological and medical issues, a quick, efficient, and potent machine learning technique is required. A growing number of researchers are currently acknowledging the significance of data dimensionality reduction for machine learning models. High-dimensional data will result in a larger search space and longer computation times. For the purpose of training and visualizing data preprocessing models, it is crucial to extract relevant information from data and eliminate superfluous and redundant information through the use of dimensionality reduction and feature selection.

This paper suggests a new approach that lowers dimensionality while keeping only the most significant features. Our method is intended to improve the ML models computational complexity, mitigate overfitting, learn feature maps, and improve performance. Our suggested method seeks to improve the effectiveness of Machine Learning models for Alzheimer's disease identification by producing a condensed feature subset, allowing neurologists to concentrate on particular features while diagnosing patients. This strategy may enable early diagnosis and treatment, which would eventually improve patient outcomes.

The rest of the research paper is structured as follows: The "Related Works" section offers a thorough literature review of relevant experiments. The "Proposed Method" section outlines the methodology of this research, which includes the data acquisition, the proposed feature selection technique and dimensionality reduction technique, followed by the performance of the machine learning algorithms and evaluation metrics has been described. The statistical information about selected features and dataset's composition has been provided. Afterwards, in the "Results and Discussion" section presents the experimental results and comparisons using various metrics. Finally, the "Conclusions" section summarizes the findings and discusses their implications for Alzheimer's diagnosis using machine learning. It also addresses the study's limitations and suggests potential future research directions for the scientific community.

2. Literature Review

Alzheimer's and Parkinson's diseases are among the most common causes of cognitive deficits today, impacting millions of people worldwide. It is anticipated that their occurrence would rise over the next several decades due to the global average lifetime augmentation. Handwriting is one of the first daily tasks that is impacted by cognitive disorders. For such reasons, scientists have also started looking at the examination of changes in handwriting as diagnostic indicators for these kinds of illnesses.

Shida *et al.* [12] introduced MRMD3.0, a dimensionality reduction tool that utilizes an ensemble approach grounded in link analysis. Their research focused on assessing the utility of that novel technique in various domains such as feature ranking and visualization. The process can be summarized in two key stages: initially, the ensemble method integrates various feature ranking algorithms to assess feature importance. Then, the forward feature search strategy, combined with cross-validation, is used to identify the optimal feature combinations. In MRMD3.0, additional link-based ensemble algorithms like PageRank, HITS, LeaderRank, and TrustRank have been introduced compared to the previous version. Additionally, more feature ranking algorithms have been included, as a result both the performance and computation time had been increased. Moreover, the approach was mainly developed to present the method for categorization jobs. Further work will focus on

creating techniques that can handle regression problems so that this method might perhaps be used to unsupervised learning settings.

Erdogmus et al. [13] introduced an innovative Convolutional Neural Network (CNN) that offers an affordable, fast, and highly accurate solution. They conducted the study using the novel DARWIN dataset, which includes data from 174 participants, 89 with Alzheimer's disease and 85 healthy individuals. This dataset initially featured 1D attributes derived from handwriting analysis, which were subsequently transformed into 2D attributes. Their proposed model was trained and assessed using this modified dataset. Based on the experimental results, the new model demonstrated an impressive accuracy rate of 90.4%. To evaluate its efficiency, the model was compared against 17 cutting-edge traditional machine learning algorithms and deep neural networks (DNNs), which served as benchmarks. The proposed novel model's inference time was obtained as an average of 2 ms, which demonstrated that the proposed novel model was almost 3.5 times faster than MobileNetV2, a widely-used CNN that is well-known to have a lightweight architecture [29]. This experimental result proves the promise of the proposed model to be employed on a lightweight or real-time system.

Njimbouom *et al.* [15] employed three dimensionality reduction methods together with six ML classifiers and identified a subset of the most important features for accurate diagnosis support. Their findings indicated that utilizing the most crucial features resulted in comparable or superior performance when contrasted with state-of-the-art models. Furthermore, this approach markedly enhanced the accuracy of Alzheimer's disease (AD) patient detection, surpassing current benchmarks [2]. Evaluating the importance of each feature improved the performance of the machine learning (ML) model while also reducing computational complexity. In summary, this proposed method represents a substantial advancement in AD detection and lays the groundwork for the development of more precise and efficient diagnostic support systems. Despite the promising results achieved with this approach, opportunities for further improvement exist. For example, using an effective hyperparameter tuning method could enhance the performance of the various machine learning models used. Furthermore, exploring and comparing new approaches might result in even higher prediction accuracy than what was achieved in the current study.

Cilia et al. [16] describe an experimental protocol designed to analyse the handwriting dynamics of patients with cognitive impairments. The main goal of this protocol is to create a comprehensive database that facilitates the effective training of various classifier systems. Additionally, it outlines the prevalent and efficient features employed in prior research to portray the handwriting dynamics of individuals with cognitive impairments. In the feature extraction process, two main categories are considered: function features and parameter features. Function features describe handwriting movements using temporal functions, while parameter features are obtained by transforming these function features. Common function features include (x, y) coordinates, pressure, azimuth, altitude, displacement, velocity, and acceleration. Some of these features are directly captured by the data acquisition device, while others are calculated numerically.

Cilia et al. [17] assessed the performance of nine top-performing and widely-used classification models. The study implemented twenty multi classifier systems, utilizing various classification models for the basic classifiers and different strategies for constructing the classifier pool. Their findings showed that a feature vector consisting of 18 features extracted from 25 tasks (450 features in total) generally outperformed feature vectors based on 18 features from individual tasks. Statistical analysis confirmed that certain classifier pools exhibited significantly better performance than others, and for each task, there was a classifier that outperformed the rest. These findings supported the idea that combining data from multiple tasks provided better characterization of AD patients' handwriting compared to any single task.

Subha *et al.* [18] proposed a hybrid ML model for predicting AD has been created using PSO-based feature selection, six different classifier models. Notably, when considering a minimum feature subset size of 20, the model featuring the Random Forest (RF) classifier demonstrated superior performance. To determine the best-performing feature dimension from an initial feature size of 450, a Particle Swarm Optimization (PSO) method was employed. A total of 20 particles were initialized and updated, optimizing an objective function geared toward minimizing prediction error. The machine learning model was trained on the training data for feature selection and classification across 100 iterations. Its performance was then evaluated using the testing data, which had reduced feature dimensions. After each iteration, the solution set was updated to reflect the results, incorporating the error computed from the classifier model for the selected feature subset. Throughout this process, specific parameters for the PSO algorithm were inputted.

3. Methodology

This section presents our proposed approach for detecting Alzheimer's disease (AD) in patients using handwriting task-driven features. Fig. 1. outlines five key stages of the approach: data description, data preprocessing stage, Feature Selection stage, ML model implementation stage and evaluation stage. At first, relevant data has been collected. After that, the label encoding and feature scaling methods are implemented in the data preprocessing phase. Next, the ANOVA feature selection approach and dimensionality reduction techniques, such as t-SNE, are employed.. In the fourth stage, the machine learning algorithms namely Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Extra Tree (ET), Ada Boost (AB), Gradient Boosting (GB), and Voting ensemble methods implemented to create an ensemble of classifiers that determines class labels. Finally, in the evaluation stage involves assessing the performance of the proposed model in accurately identifying AD. Moreover, the performance of the suggested system has been assessed using the confusion matrix, precision, recall, f1-score, and accuracy.

3.1 Dataset Description

To ensure an unbiased assessment, the DARWIN dataset has been utilized [17], which is the largest publicly available dataset specifically designed for detecting Alzheimer's disease (AD). This dataset comprises handwriting samples produced by 174 individuals, with 89 of them being AD patients and the remaining 85 serving as healthy controls. The recruitment of subjects was carefully conducted to match patients and the control group, considering factors such as age, education level, occupation type, and gender. Each participant completed 25 different motor tasks, which were divided into four groups based on increasing difficulty: dictation, graphic activities, copy and reverse copy tasks, and memory tasks. A Bamboo Wacom tablet was used to record the handwriting samples, enabling the replication of a pen-and-paper setting while also digitizing the handwritten results. Eighteen parameters, including execution time, mean speed, acceleration, jerk (both on paper and in the air), mean pressure, and the discernible tremor in the trace, were used to define each handwriting sample.

The system architecture, shown in Fig. 1, was designed with the primary objective of accurately diagnosing Alzheimer's disease using handwriting features. The dataset's dimensions are (174, 451), where the last column indicates the type of individuals, including two primary binary classes: 'P' and 'H', represents patients and healthy individuals. As the distribution of the constructed dataset, it was a balanced dataset, consisting of 85 normal and 89 samples with AD. The dataset was divided into 80% for the training set and 20% for the test set.

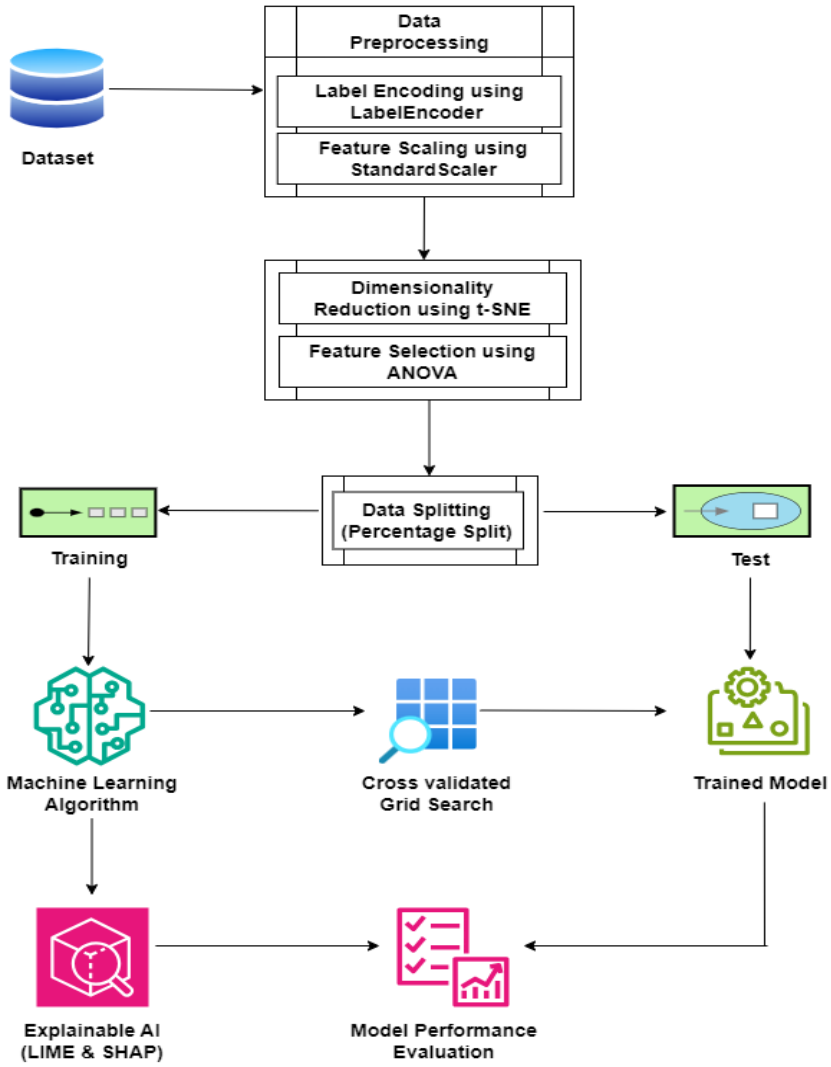


Fig. 1. Proposed System Architecture

3.2 Data Pre-processing

First, the initial phase in the data preprocessing procedure is label encoding. Label encoding is a fundamental machine learning approach used for handling categorical data, which represents categories or groups rather than numerical values. Since many machine learning algorithms and models require input data in numerical form, label encoding plays a crucial role in preparing data for analysis. Label encoding is a simple and widely used method to convert categorical data into numerical values. Label encoding that makes it possible to convert a numerical representation for a categorical variable. Each category or label in the dataset is given a unique integer as part of how it operates. Since categorical variables are hard for machines to manage, most machine learning (ML) techniques only accept numerical values as input. To input categorical data into machine learning classifiers, they need to be transformed into a numerical representation. In this experiment, the target feature, which included categorical variables, was transformed into numeric inputs using the label encoding approach. The "StandardScaler" method for scaling features has been applied. First, the mean and standard deviation for each feature in the dataset are calculated using the StandardScaler. To do this, each feature's mean value is subtracted from the data points, and the

result is divided by the standard deviation. This feature scaling technique standardizes the scale of all variables, ensuring they have comparable sizes, effectively transforming all attributes into a consistent range. This process normalizes the values to lie within the range of 0 to 1. Equation (1) expresses the following mathematical interpretation of this normalization process:

$$z = \frac{x-\mu}{\sigma} \tag{1}$$

This expression, (z) illustrates the scaled value, (x) the original data point, (μ) the feature mean, and (σ) the feature standard deviation. While StandardScaler () will separately normalize each column of (x), if with_std = False, then all samples' standard deviations (σ) are equal to 1, and if with_mean = False, then the training samples' means (μ) are equal to 0.

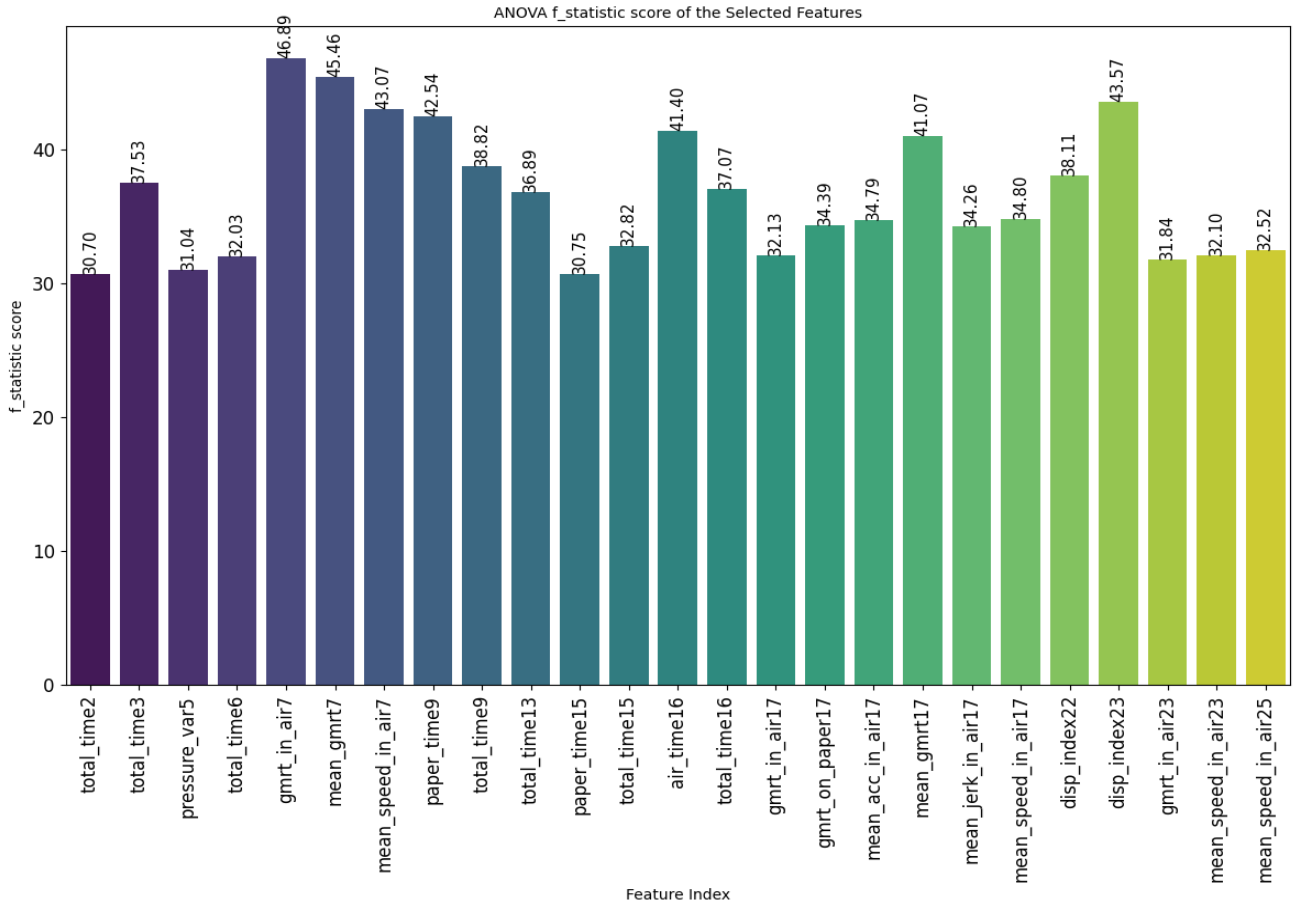


Fig. 2. f-statistic value of the ANOVA selected features

3.3 Feature Selection Using t-SNE and ANOVA

Dimensionality reduction isn't solely for data visualization; it also helps address the 'curse of dimensionality' by identifying key structures in high-dimensional space and preserving them in a lower-dimensional representation. Firstly, t-SNE has been applied for dimensionality reduction on the scaled feature to obtain a two-dimensional representation [24]. The experiment has been divided into three individual tasks, for the first task, t-SNE has been combined for dimensionality reduction technique and ANOVA as feature selection technique and then applied eight machine learning classifiers and finally, used the Explainable AI to interpret the machine learning models. t-SNE has

been used on the scaled feature of the dataset to obtain a lower-dimensional representation of the present data. Fig. 3, shows the distribution of the constructed dataset after applying t-SNE embedding. ANOVA was then used for feature selection to reduce the number of features and identify those most significant in driving these relationships.

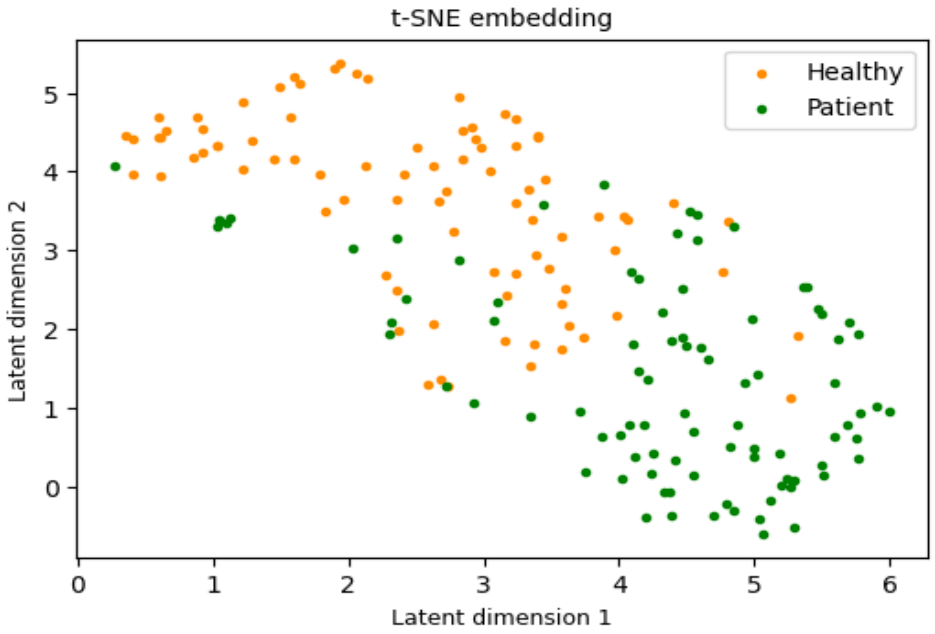


Fig. 3. The distribution of constructed dataset after using t-SNE embedding

t-distributed stochastic neighbor embedding (t-SNE) is a statistical technique used to visualize high-dimensional data by assigning each data point a location on a two- or three-dimensional map [24]. Unlike traditional methods, the Stochastic Neighbor Embedding variation of t-SNE reduces the tendency to cluster points at the center of the map, resulting in noticeably better visualizations and easier optimization. The t-SNE algorithm operates in two main stages. First, it constructs a probability distribution over pairs of high-dimensional objects, assigning higher probabilities to similar objects and lower probabilities to dissimilar ones. Second, it creates a similar probability distribution for points in the low-dimensional map and minimizes the Kullback–Leibler (KL) divergence between the two distributions based on the points' locations in the map. t-SNE excels at generating a single map that reveals structures at various scales, making it particularly useful for high-dimensional data that reside on multiple, related low-dimensional manifolds, such as images of objects from different classes viewed from multiple angles [25]. The visualizations produced by t-SNE are superior to those generated by other techniques across almost all datasets. For a dataset with n elements, t-SNE operates in $O(n^2)$ time and requires $O(n^2)$ space [26]. t-SNE is a powerful tool for visualizing high-dimensional data in a lower-dimensional space, facilitating exploration and understanding of the structure and relationships between data points. It creates visualizations that highlight clusters or patterns in the data that might not be evident in the high-dimensional space, making it invaluable for exploratory data analysis and gaining deeper insights into the dataset. After using t-SNE the reduced-dimensional visualizations are more interpretable and useful for communicating results to non-technical stakeholders. Our approach can make the data more manageable designed to enhance the performance, mitigating overfitting, and decreasing the computational complexity of subsequent machine learning models.

Analysis of Variance (ANOVA), a term introduced by Ronald Fisher in 1918, is also referred to as Fisher's Analysis of Variance among statisticians [22]. ANOVA examines samples from different data groups to assess the impact of differences among them. It is a statistical technique that separates systematic factors from random ones to explain the total observed variability within a dataset. The dataset is statistically influenced by systematic factors rather than random ones. In general, ANOVA statistic component is denoted as F, and the ANOVA statistical component is computed by using the formula given below:

$$F = \frac{MST}{MSE} \tag{2}$$

Here, (F) represents the ANOVA coefficient, (MST) stands for the Mean Sum of Squares due to treatment, and (MSE) denotes the Mean Sum of Squares due to errors. To find significant differences across groups within a dataset, ANOVA is commonly employed. After using t-SNE, feature selection using ANOVA has been performed on the original scaled features to select the most informative features. It helps identify which features (independent variables) have a statistically significant impact on the dependent variable. In the context of feature selection, ANOVA can help to rank and select the most relevant features.

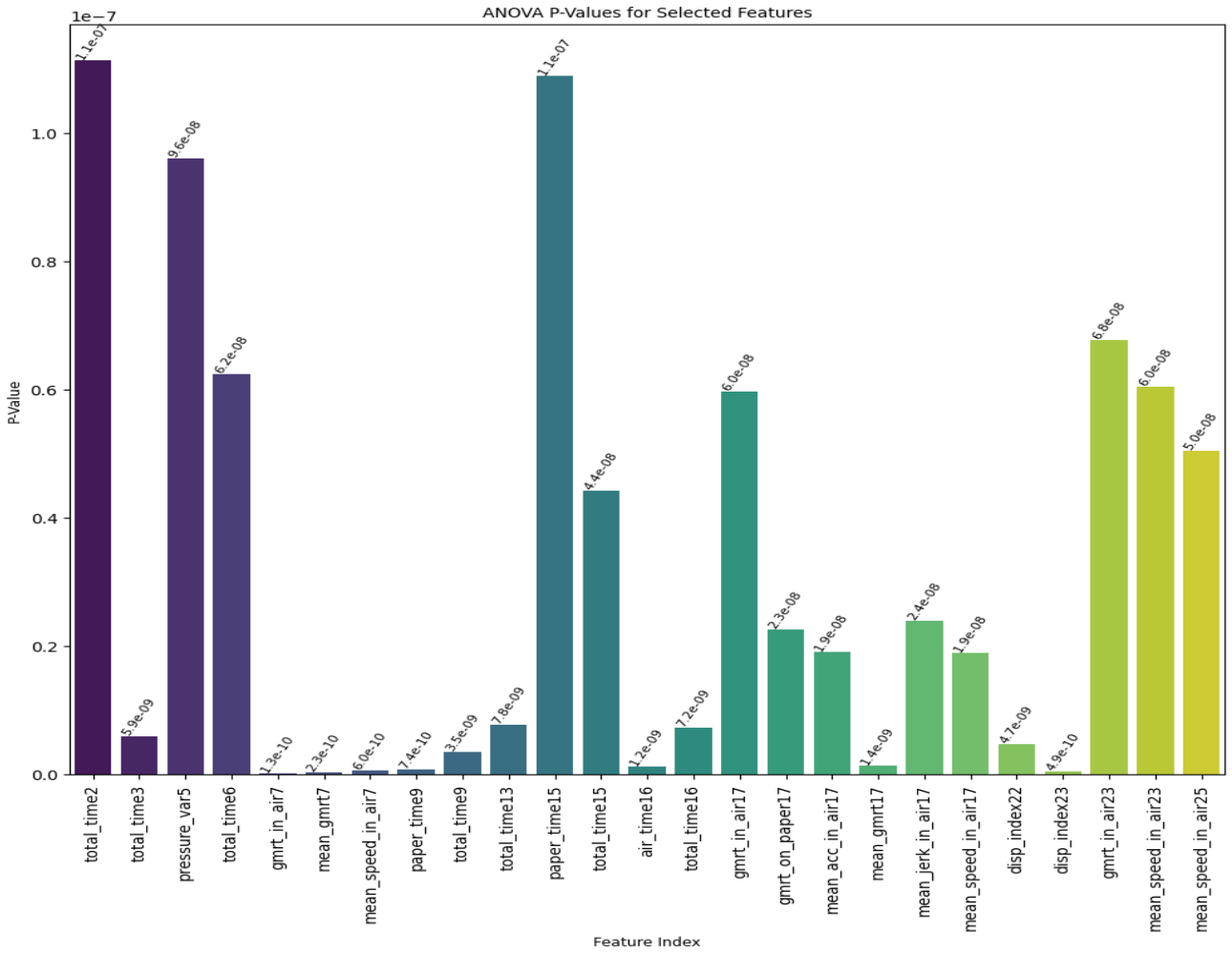


Fig. 4. p-value of the ANOVA selected features

The ANOVA test can compare more than two groups simultaneously to determine if there is a correlation between them. The result of the ANOVA formula is the F statistic, or F-ratio, which helps examine multiple data sets to assess variability within and across samples. So, by using this F-statistic score, each feature of the data can be ranked accordingly, and the features with higher ranks can be considered as the optimal set of features. Moreover, the F statistic value of the selected features by using ANOVA is calculated in Fig. 4. The p-value is also calculated in this study. The p-value in ANOVA is a key statistical measure used to assess the significance of differences between group means. ANOVA is a robust statistical method that compares the means of three or more groups to determine if there are statistically significant differences among them. In ANOVA, the null hypothesis posits that there is no significant difference between the group means, while the alternative hypothesis suggests that at least one group mean differs from the others. The p-value helps us decide whether to reject the null hypothesis.

3.4 ML Classifiers for Classification

In this work, Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Extra Tree (ET), Ada Boost (AB), Gradient Boosting (GB) were implemented as the base classifiers. Afterwards, the Voting classifiers were then created by utilizing these base classifiers. Also, the "GridSearchCV" approach has been used to identify the classifiers' optimal parameters [19]. "GridSearchCV" combines each of the provided hyperparameters and their values distinctively with regards to that it computes the performance of each combination and then chooses the hyperparameters with the best value. Cross-validated grid search is employed to fine-tune the estimator's parameters over a range of values, optimizing the model's performance. With so many hyperparameters involved, processing becomes time-consuming and costly. Table 1. demonstrates the values of the selected optimal parameters. Subsequently, the ensemble method improves machine learning performance by integrating all the base classifiers [20].

Table 1
Optimal parameter's value for the classifiers

No. Classifier	Parameter	Value
1. Random Forest (RF)	i. bootstrap	i. True
	ii. criterion	ii. 'gini'
	iii. max_depth	iii. 5
	iv. max_leaf_nodes	iv. 5
	v. n_estimators	v. 100
	vi. random_state	vi. 42
2. Logistic Regression (LR)	i. C	i. 1
	ii. class_weight	ii. 'balanced'
	iii. penalty	iii. 'l1'
	iv. solver	iv. 'saga'
	v. multi_class	v. 'auto'
3. Gradient Boost (GB)	i.loss	i. 'exponential'
	ii. max_features	ii. 'sqrt'
	iii. n_estimators	iii. 100
	iv. subsample	iv. 0.8
4. Ada Boost (AB)	i. algorithm	i. 'SAMME.R'
	ii. learning_rate	ii. 0.1
	iii. n_estimators	iii. 200
5. Decision Tree (DT)	i. criterion	i. 'entropy'
	ii. min_samples_leaf	ii. 6
	iii. max_depth	iii. None

	iv. splitter	iv. 'best'
	v. min_samples_split	v. 2
6. Support Vector Machine (SVM)	i. cache_size	i. 10
	ii. C	ii. 1
	iii. degree	iii. 1
	iv. gamma	iv. 0.1
	v. probability	v. True
	vi. shrinking	vi. True
7. Extra Tree (ET)	i. bootstrap	i. True
	ii. n_estimators	ii. 200
	iii. max_samples	iii. 1.0
	iv. criterion	iv. 'gini'
	v. max_features	v. 'sqrt'
	vi. max_depth	vi. 10

In this context, seven distinct classifiers were employed, and their outcomes were integrated using majority voting to yield the final model's prediction. This ensemble approach led to a noticeable improvement in performance when contrasted with individual classifiers. Unlike relying on a single model, machine learning ensemble methods [35] utilize a variety of models, leading to significant improvements, more accurate predictions, and better overall performance. Ensemble algorithms are especially effective for both regression and classification tasks because they reduce bias and variance, thereby enhancing accuracy [21]. To enhance the efficiency of the proposed system, the "GridSearchCV" algorithm is applied, incorporating both parallel and sequential ensemble techniques. Combining "GridSearchCV" with both parallel and sequential ensemble techniques allow us to optimize the model's hyperparameters and take advantage of various ensemble methods to enhance the efficiency and performance our machine learning system.

This approach frequently results in improved model accuracy and expedited execution, delivering significant benefits in numerous scenarios. After testing various machine learning algorithms, our experiments showed that accuracy levels ranged from 86.81% to 94.28%, with AdaBoost achieving 84.69%. The highest classification accuracy, around 94.28%, was achieved by the Voting ensemble method. This technique uses a Voting Classifier, which is trained on multiple models and predicts the output class based on the highest probability among the selected options [27].

In our study, we utilized the hard voting technique for our ensemble model. Each base classifier within the ensemble makes its own prediction, and the final prediction is determined by a majority vote. Essentially, the class that receives the most votes from the individual models is selected as the ensemble's final prediction for that input. This method counts the number of times each class is predicted and chooses the one with the highest count. Ensemble methods, such as voting, play a vital role in addressing overfitting issues as they diminish the influence of noise present in the training data. Furthermore, when individual classifiers exhibit biases or shortcomings, combining them in a voting ensemble helps to balance and mitigate these issues [28]. The amalgamation of predictions from multiple base classifiers frequently results in enhanced overall accuracy compared to using individual classifiers.

3.5 Explainable AI for Model Explanation

In this study, LIME (Local Interpretable Model-Agnostic Explanations) has been employed to analyse system architecture models. LIME is an AI-driven technique that enhances interpretability by creating a local, easily understandable model around predictions. It achieves this by generating a synthetic dataset from a single sample and then permuting it. Subsequently, LIME computes

similarity metrics between the permuted data and the original observations, providing accurate and comprehensible explanations for classifier predictions.

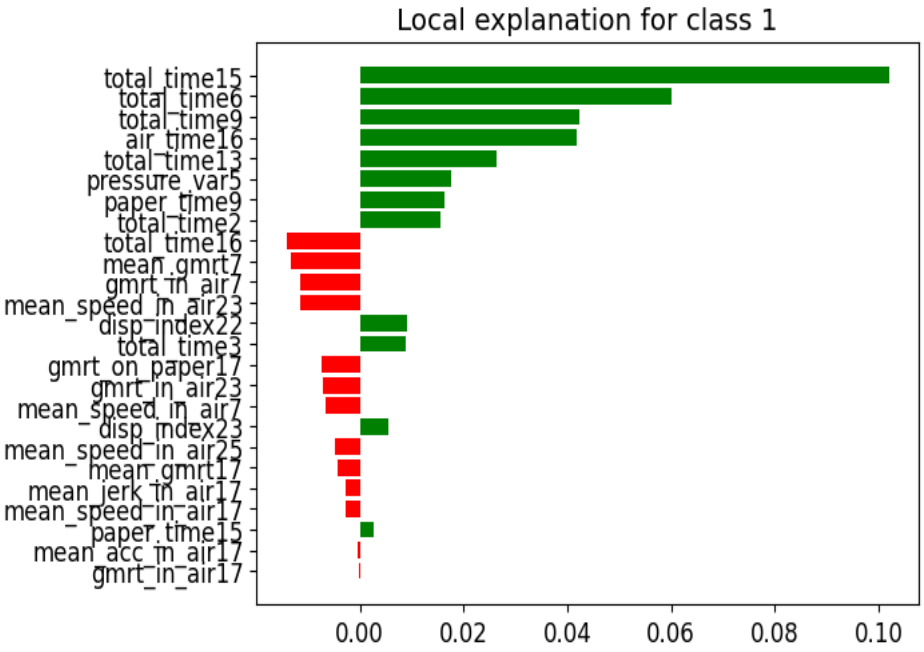


Fig. 5. Feature's Contribution Using LIME

In this context, the comparability between falsified and accurate data when subjected to permutations has been observed. Leveraging the LIME functionality, various similar metrics have been explored. Our approach involves utilizing a previously sophisticated model to predict outcomes using newly permuted fraudulent data. Subsequently, features that best represent the complex model's performance on permuted false data have been identified. By adjusting the number of selected features via the Lime library, we proceed to fit a basic model (such as a random forest or decision tree) using the permuted false information, chosen attributes, and previously computed similarity scores. This process allows to create a straightforward model for deployment, facilitated by the LIME library.

Here, test data is repeatedly used as input, with different machine learning algorithms applied each time to generate new data. Specifically, LIME was applied to a stacking model, resulting in the selection of 13 features out of the initial 25 attributes (excluding the target attribute) for the 0 (0 to -0.02) prediction probability category. In contrast, 12 features from the same set of 25 traits fell into the 1 (0 to 0.10) prediction probability group. Notably, the attributes associated with the 0 category include total_time16, mean_gmrt7, gmrt_in_air7, mean_speed_in_air23, gmrt_on_paper17, gmrt_in_air23, mean_speed_in_air7, mean_speed_in_air25, mean_gmrt17, mean_jerk_in_air17, mean_speed_in_air17, mean_acc_in_air17, and gmrt_in_air17.

On the other hand, the attributes corresponding to the 1 category are total_time15, total_time6, total_time9, air_time16, total_time13, pressure_var5, paper_time9, total_time2, disp_index22, total_time3, disp_index23, and paper_time15. Notably, total_time15 and gmrt_in_air17 exhibit the highest and lowest prediction scores among the 25 selected features (excluding the target feature). For a detailed illustration of LIME feature contribution and sample predictions, refer to Fig. 5.

SHAP (SHapley Additive exPlanations) is a game-theoretic method used to explain the outputs of machine learning models [31]. It links optimal credit allocation with local explanations through Shapley values from game theory and their extensions [32]. SHAP assigns an importance value to

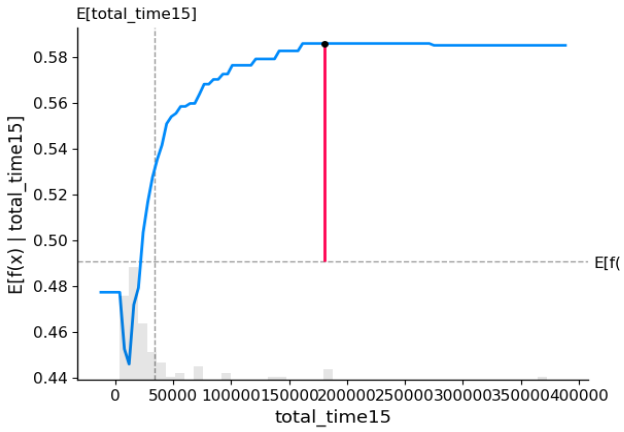
each feature for a specific prediction. Its innovative aspects include the introduction of a new class of additive feature importance measures and theoretical results demonstrating that there is a unique solution within this class that possesses desirable properties.

The SHAP framework identifies a class of additive feature importance methods, which includes six previously established methods, and demonstrates that there is a unique solution within this class that has desirable properties [33]. SHAP builds on the concept of Shapley values, originally used to fairly distribute the rewards in a cooperative game among its players. In machine learning, the 'players' are the features of a dataset, and the 'reward' is the difference between the model's prediction and a baseline prediction.

To determine each feature's contribution to a prediction, SHAP requires a baseline, which can be the mean prediction of the model on the training dataset or a reference point chosen based on domain knowledge. SHAP calculates Shapley values for each feature by evaluating its marginal contribution to the difference between the model's prediction for a specific instance and the baseline. This involves considering all possible combinations of features and measuring how the prediction changes when a feature is included or excluded. The Shapley value for each feature is the average of these marginal contributions across all possible feature combinations.

SHAP offers various algorithms for efficiently calculating Shapley values based on the model and dataset characteristics. For instance, TreeSHAP is designed for tree-based models such as decision trees and random forests, while KernelSHAP uses kernel methods to estimate Shapley values for models with continuous features. DeepSHAP extends SHAP to deep learning models. Once Shapley values are computed, they indicate feature importance. Positive Shapley values suggest that the presence of a feature increases the model's prediction, whereas negative values indicate a reduction in the prediction. The magnitude of the Shapley value reflects the strength of a feature's influence on the prediction.

SHAP provides several visualization tools to help interpret these values, including summary plots, individual feature importance plots, and dependence plots. These tools aid users in understanding the relationships between features and predictions. Ultimately, SHAP uses Shapley values to explain individual model predictions, offering insights into why a model made a specific decision and enhancing transparency and trust in the model's outputs. The most influential features of the model were analysed using standard partial dependence plots and scatter plots from the SHAP library, as shown in Fig. 6 and 7. Fig. 6. illustrates a Standard Partial Dependence Plot, which depicts the relationship between the target response and a specific feature while keeping other variables constant. This plot highlights the feature with the highest contribution to the model's performance.



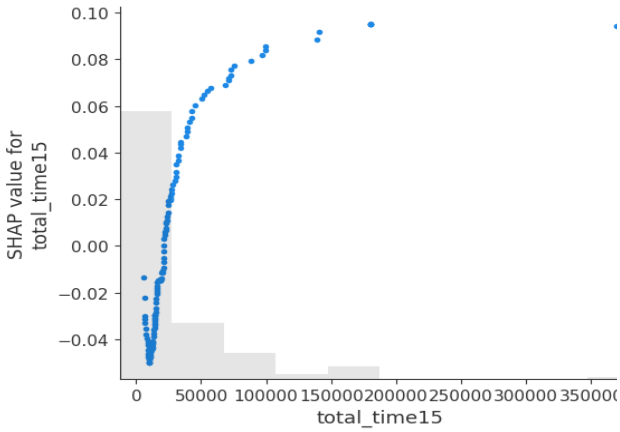


Fig. 6. Standard partial dependence plot with highest contributing feature's SHAP value

Fig. 7. Scatter plot of highest contributing feature's SHAP value

Fig. 8. illustrates the waterfall plot for feature importance using SHapley Additive exPlanations (SHAP), which highlights the contributions of various features to the model's output. Each bar in the plot represents the impact of a specific feature on the predicted output, with positive impacts in pink (indicating an increase in prediction score) and negative impacts in blue (indicating a decrease in prediction score). Starting from the base value of $E[f(X)] = 0.491$, each feature incrementally adjusts the prediction towards the final output of $f(x) = 0.767$. Key features such as "total_time15," "total_time6," and "total_time2" contribute significantly with positive SHAP values, while others like "total_time13" and "total_time9" show negative contributions. This plot enables a clearer understanding of feature importance and their respective directional influences on the prediction, aiding in model interpretability and decision-making processes. Notably, `total_time15` shows the highest positive impact, increasing the prediction by 0.1, followed by `total_time6` and `total_time2` with contributions of 0.07 and 0.06, respectively. On the other hand, `total_time13` and `total_time9` exhibit negative contributions, slightly reducing the prediction by 0.03 and 0.02, respectively. The cumulative effect of these features results in a final predicted value of 0.767. This detailed breakdown enhances the interpretability of the model, enabling a deeper understanding of how different features influence the prediction outcomes.

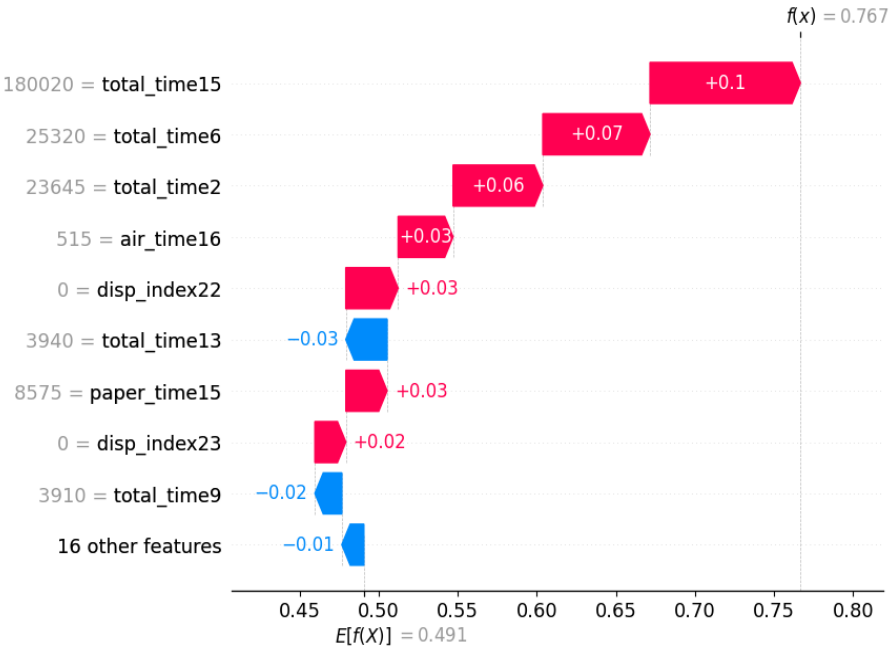


Fig. 8. Feature's Waterfall Plot Using SHAP

Additional SHAP interpretability methods were applied to better understand the relationships between the most impactful features and their contributions to the model. Fig. 9. representing the contribution of individual features to the model's predictions. Features are ranked based on their SHAP value, with higher values indicating a more significant impact on the model's output. The most influential feature, "total_time15," exhibits the highest SHAP value at 0.56, suggesting it has the strongest effect on the prediction outcome. Other prominent features, such as "disp_index22," "disp_index23," and "total_time13," also show substantial contributions, each with SHAP values above 0.5. The presence of multiple "total_time" and "disp_index" metrics among the top features

implies that these temporal and displacement indices play a critical role in model prediction. Lesser contributions are observed from variables like "mean_speed_in_air25" and "mean_jerk_in_air17," which display SHAP values of 0.07, indicating a relatively minor influence.

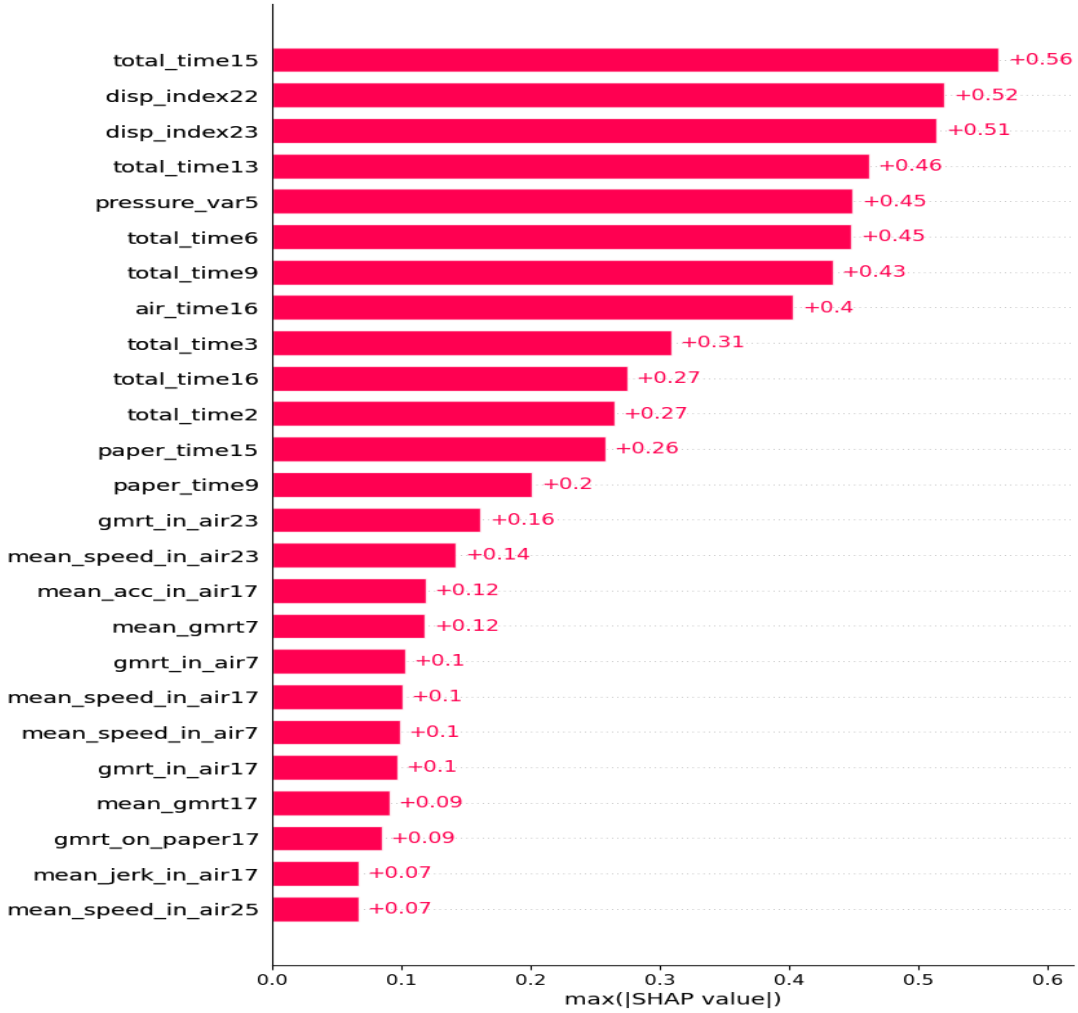


Fig. 9. Feature's Contribution Max SHAP Value

Fig. 10. presents the mean absolute SHAP values for various features, indicating the contribution of each feature to the model's predictions. The features with the highest contributions include `total_time15`, `total_time6`, `total_time9`, `air_time16`, `total_time13`, `pressure_var5`, `paper_time9`, `total_time2`, `disp_index22`, `total_time3`, and `disp_index23`. `total_time13` stands out with the highest SHAP value, suggesting it has the most significant impact on the model's predictions. Fig. 10 and 11 further confirm this, with `total_time13` showing the highest SHAP values of +0.05 and +6.57, respectively. In contrast, `mean_speed_in_air25` is among the least contributing features.

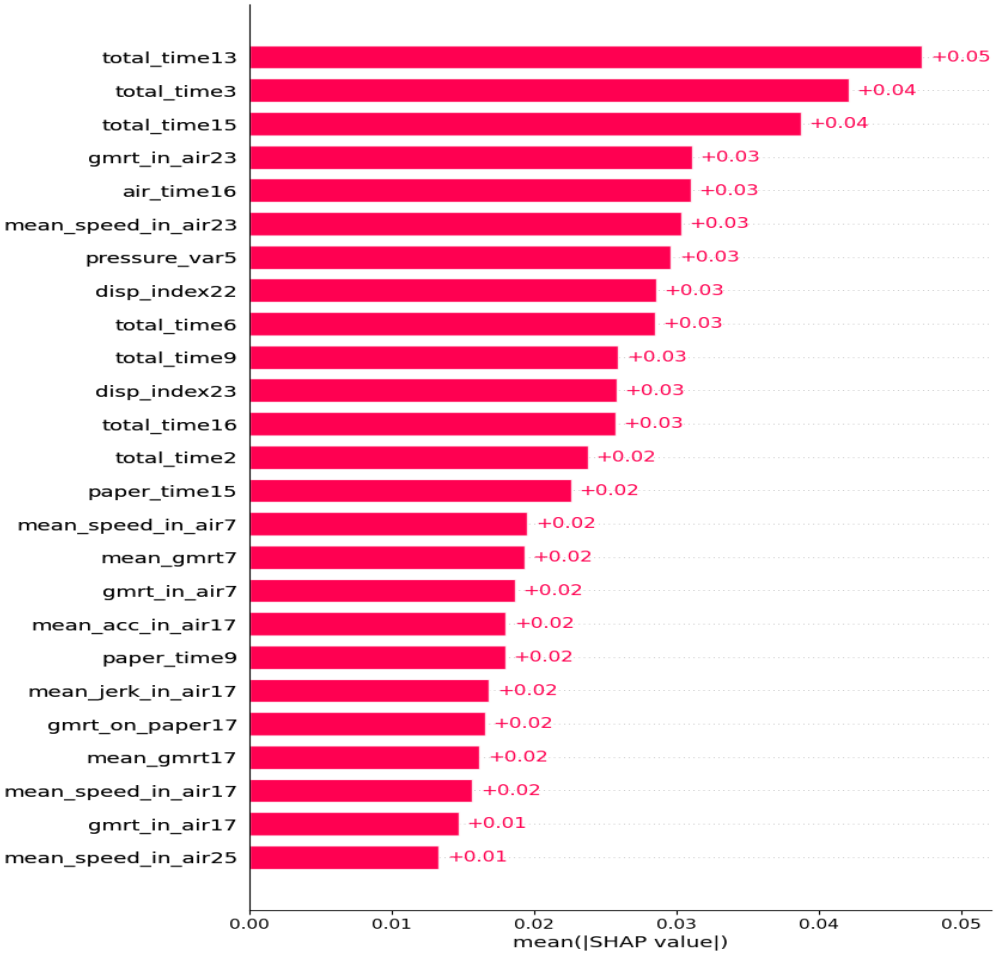


Fig. 10. Feature's Contribution Using Mean SHAP Value

Fig. 11 displays the sum of SHAP values for different features, reinforcing that `total_time13` has the most significant impact on the model. Other notable features include `total_time3`, `total_time15`, `gmrt_in_air23`, `air_time16`, `mean_speed_in_air23`, `pressure_var5`, `disp_index22`, `total_time6`, `total_time9`, `disp_index23`, `total_time16`, `total_time2`, `paper_time15`, and `mean_speed_in_air7`.

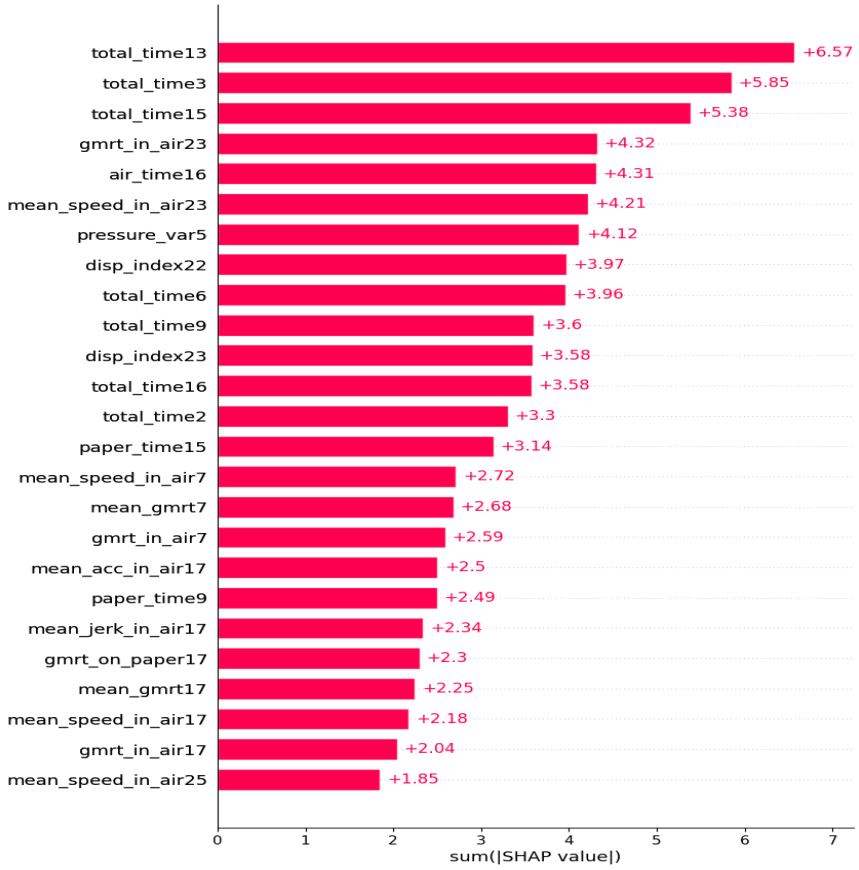


Fig. 11. Feature's Contribution Using Sum of the SHAP Value

Fig. 12 shows that a clustering cut-off of 1.8 was used to identify correlated features through SHAP clustering. Fig. 13 presents a SHAP heatmap that visualizes the contribution of various features to the model. The heatmap plots SHAP values on the x-axis and feature values on the y-axis, with colors indicating the impact of each feature on the model output. The heatmap suggests that features like `total_time13`, `total_time3`, `total_time15`, `gmrt_in_air23`, `air_time16`, and `mean_speed_in_air23` have a high impact on the model.

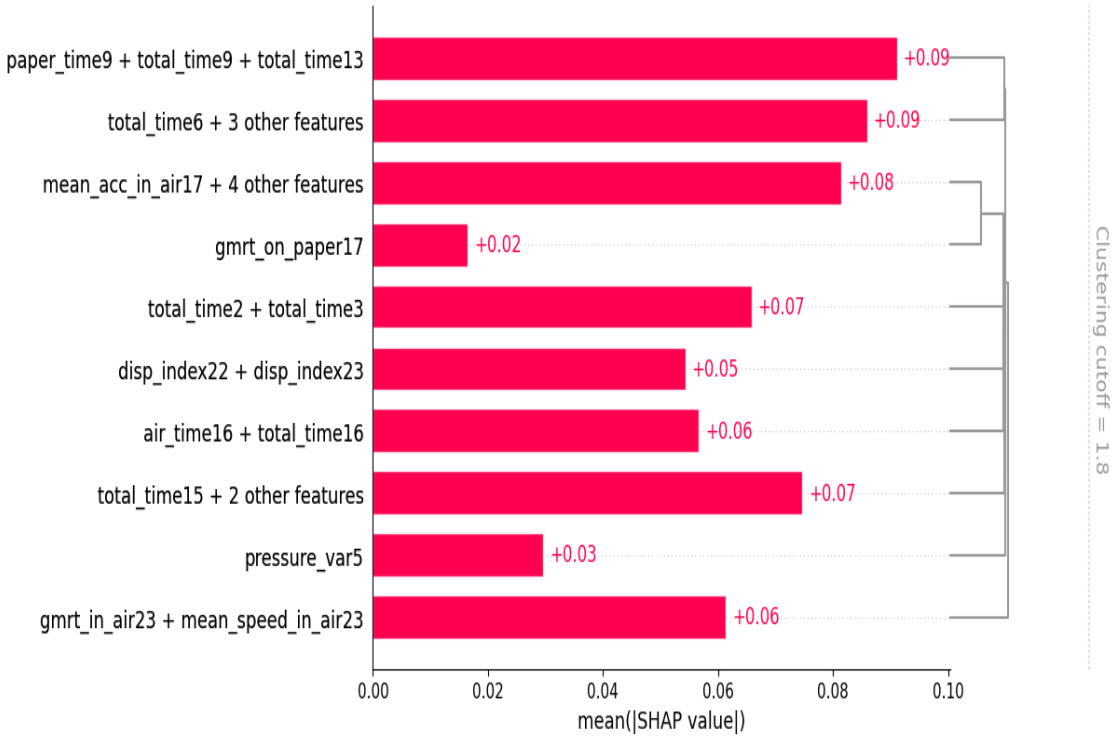


Fig. 12. Correlated Features of the Model Using SHAP Clustering

Overall, SHAP summary plots provide a comprehensive view of feature importance, helping to understand how each feature influences the model's predictions. It displays the impact of each feature on model predictions, providing insights into how variables contribute to outcomes. By showcasing both the magnitude and direction of influence, it aids in understanding model behaviour and identifying influential factors. This visual representation helps analysts and stakeholders grasp complex model interactions and make informed decisions. With its ability to highlight key drivers behind predictions, the SHAP summary plot enhances transparency and trust in machine learning models, fostering better interpretation and utilization in various domains. Also, the Fig. 13 and Fig. 14 represent the Heatmap and Summary plot of the highest contributing feature's where in both figures the showed that the highest contributing feature is total_time13 with highest impactful SHAP value and mean_speed_in_air25 is less contributing feature among other 25 features.

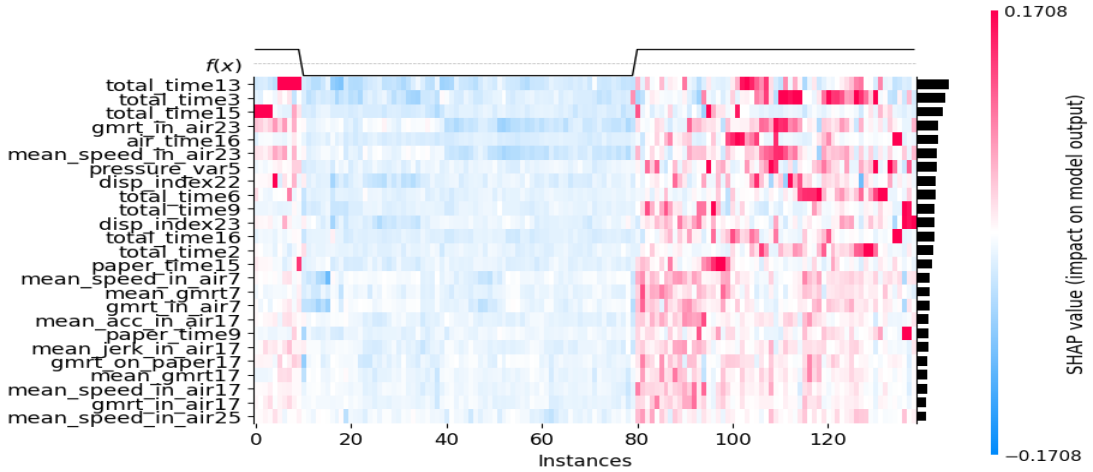


Fig. 13. Feature's Contribution Using SHAP Heatmap

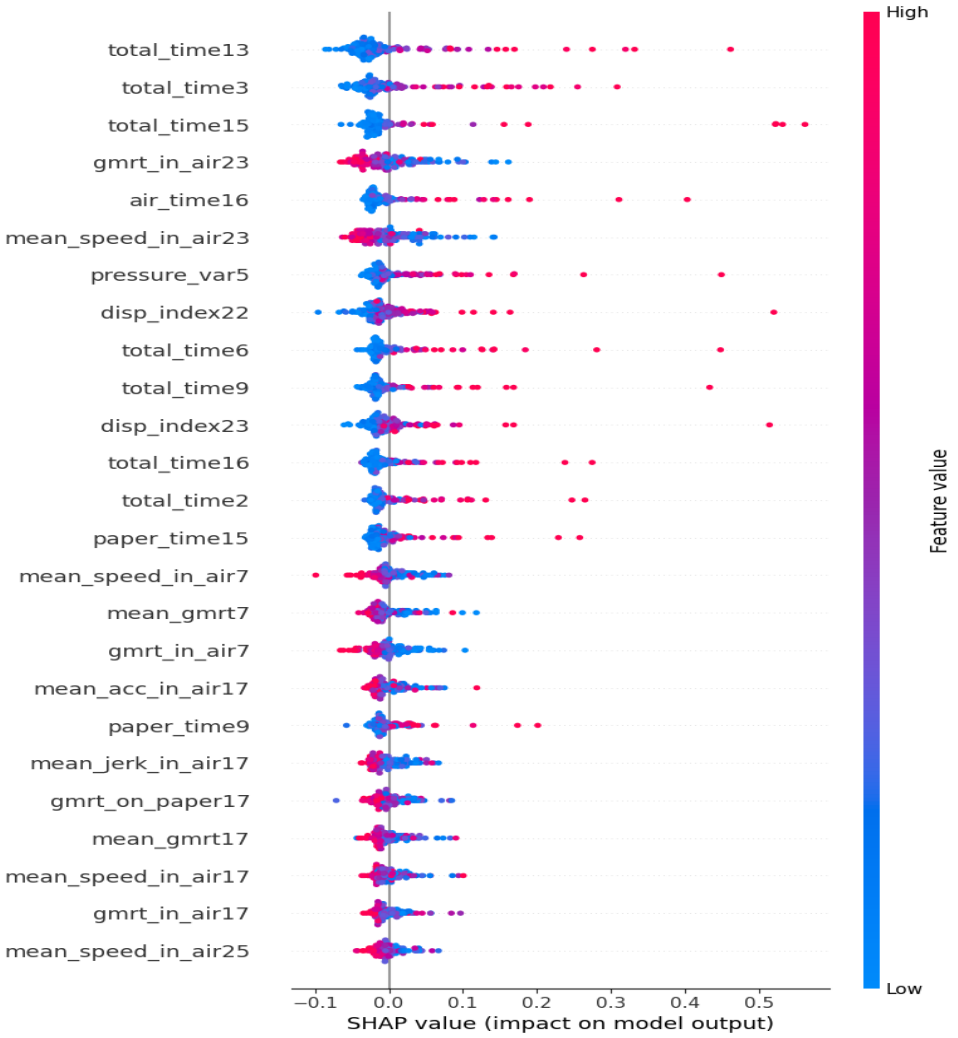


Fig. 14. Feature's Impact on Model Output Using SHAP Summary Plot

4. Results

To evaluate the proposed approach for predicting Alzheimer's Disease (AD), a set of widely recognized performance metrics have been applied. These metrics encompass accuracy, precision, f1-score, and recall. We used 10-fold cross-validation to assess the model's generalization ability and improve its performance. Additionally, we utilized the AUC-ROC (Area Under the Receiver Operating Characteristic Curve) to evaluate the effectiveness of different methods in accurately classifying patients with Alzheimer's Disease. The following formulas depict the measurement metrics utilized to evaluate the different approaches. The detailed classification report of each model has demonstrated in Table 2. Using t-SNE before feature selection with ANOVA for exploring and visualizing the data, high-dimensional data can be challenging to work with and visualize. Reducing the dimensionality to 2D or 3D can make it easier to explore and understand the data [30]. This method visualizes complex data in a lower dimensional space while maintaining its structure, which can help to gain insights into data, such as identifying clusters, outliers, or patterns.

The proposed method has been used to find the best performing feature dimension from a feature size of 450. Initially, the task has been divided into three distinct phases, in the first task t-SNE has been utilized for dimensionality reduction with the approach of ANOVA for feature selection. On the Other hand, dimensionality reduction and one feature selection method together have been

employed with eight machine learning classifiers including Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Extra Tree (ET), Ada Boost (AB), Gradient Boosting (GB), and Voting ensemble classifier and finally, used the Explainable AI to interpret the machine learning models.

Table 2
Classification report of the classifiers

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
LR	92.31	92.54	92.22	92.20	92
SVM	91.21	91.54	91.22	91.28	91
RF	93.83	93.81	93.69	93.71	94
DT	88.57	88.64	88.33	88.25	88
ET	90.89	90.87	90.63	90.65	91
GB	86.81	86.74	86.71	86.78	87
AB	84.69	84.30	84.13	84.12	84
Voting	94.28	94.53	94.48	94.51	94

The proposed approach has been evaluated after experimenting with dimensionality reduction technique namely t-SNE and feature selection technique ANOVA, and various machine learning (ML) models. Afterwards, the results in Table 3 indicate that dimensionality reduction namely t-SNE and the feature selection technique ANOVA selected most significant 25 features with Voting ensemble technique, yielded the best performance. This approach achieved an accuracy of 94.28%, precision of 94.53%, recall of 94.48%, and an F1 score of 94.51%. On the other hand, the Ada Boost Classifier had the lowest performance, with accuracy, precision, recall, and F1 scores of 84.69%, 84.30%, 84.13%, and 84.12%, respectively, compared to the other models. Others model's accuracy are lies in between approximately 87% to 91%. Fig. 15. illustrates an example of the confusion matrix for the all used algorithms.

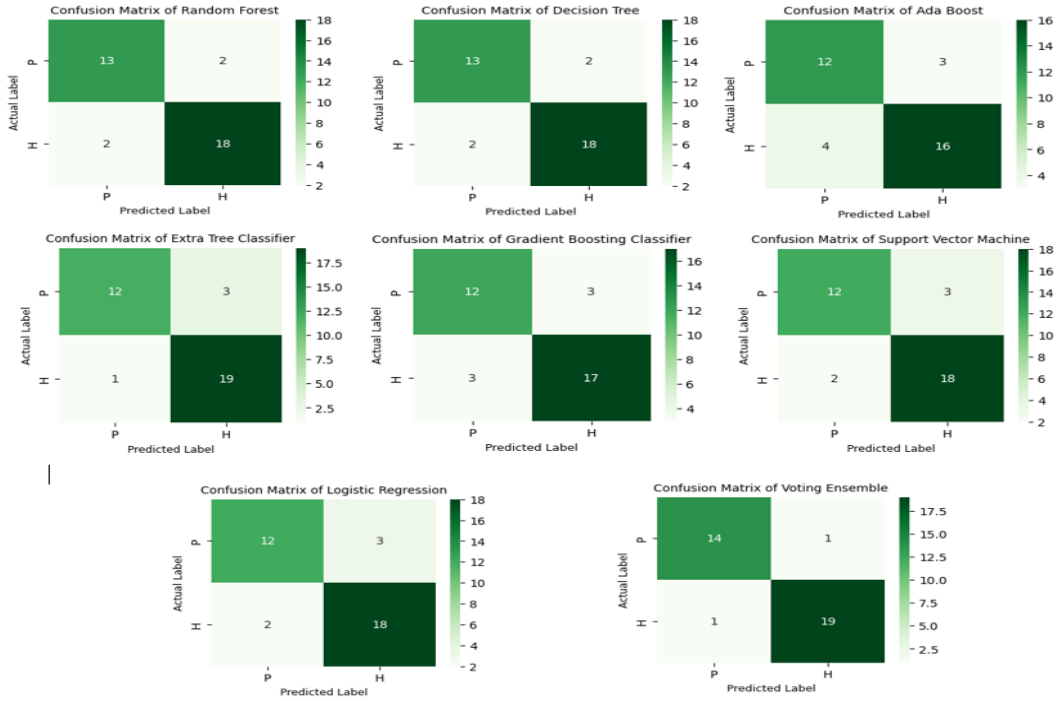


Fig. 15. Confusion Matrix for each of applied algorithms

Table 3
Classification Report of the different methods with distinct features

# of Features	Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC AUC (%)
10	(t-SNE+ ANOVA) + LR Classifier	84.45	84.47	84.45	84.41	84
25	(t-SNE+ ANOVA) + Voting Classifier	94.28	94.53	94.48	94.51	95
50	(t-SNE+ ANOVA) + Stacking Classifier	90.21	90.25	90.21	90.29	90
75	(t-SNE+ ANOVA) + RF Classifier	88.57	88.64	88.33	88.25	88
100	(t-SNE+ ANOVA) + Voting Classifier	87.88	87.75	87.56	87.59	88
125	(t-SNE+ ANOVA) + RF Classifier	87.76	87.72	87.65	87.63	88
150	(t-SNE+ ANOVA) + SVM Classifier	86.38	86.31	86.22	86.28	86
175	(t-SNE+ ANOVA) + ET Classifier	85.11	85.17	85.09	85.11	85
200	(t-SNE+ ANOVA) + GB Classifier	84.64	84.62	84.54	84.58	85
225	(t-SNE+ ANOVA) + SVM Classifier	84.78	84.62	84.68	84.69	84
250	(t-SNE+ ANOVA) + Stacking Classifier	86.59	86.47	86.42	86.44	86
275	(t-SNE+ ANOVA) + RF Classifier	85.93	85.88	85.88	85.91	86
300	(t-SNE+ ANOVA) + GB Classifier	83.49	83.41	83.46	83.47	83
325	(t-SNE+ ANOVA) + Voting Classifier	82.11	81.14	81.08	78.09	83
350	(t-SNE+ ANOVA) + GB Classifier	80.28	80.22	80.22	80.28	80
375	(t-SNE+ ANOVA) + Stacking Classifier	81.84	81.73	81.70	81.76	81
400	(t-SNE+ ANOVA) + ET Classifier	80.57	80.48	80.61	80.62	80
425	(t-SNE+ ANOVA) + RF Classifier	79.66	79.62	79.71	79.75	80
450	(t-SNE+ ANOVA) + Stacking Classifier	78.83	78.87	78.63	78.65	79

Firstly, the proposed model is check with selected 10 features with eight ML algorithms, with the results shown in Table 3. This table also displays the Accuracy, Precision, Recall, F1 score, and AUC score performance metrics for eight distinct ML classifiers with feature selection ranging from 25 to 450 in intervals of 25. With a minimal feature subset size of 25, it is discovered that the model with the voting ensemble classifier method (approximately 94.3%) performs best. In contrast, when the feature subset size is 450, it indicates that the model using the stacking ensemble classifier approach (about 79%) performs the worst. The results showed that using the most important features led to similar or better performance compared to state-of-the-art models.

Table 4
Performance comparison of the proposed system with existing works

Author	Dataset	Method Used	Accuracy (%)
Shida *et al.* [12]	DARWIN	Random Forest	88.7
Erdogmus *et al.* [13]	DARWIN	2D CNN	90.4
A. Parziale *et al.* [14]	DARWIN	Gaussian Naive Bayes	85.7
Njimbouom *et al.* [15]	DARWIN	Random Forest	87.7
Subha *et al.* [18]	DARWIN	Random Forest	90.5
Proposed Work	DARWIN	Voting Ensemble	94.3

A thorough performance comparison of the proposed work and existing works is shown in Table 4. The findings clearly demonstrate the recommended work's superiority over more recent advancements in the field. Numerous important performance indicators were considered, such as robustness, accuracy, and efficiency. The recommended work showed impressive efficiency gains in terms of computational resources and runtime. Additionally, it demonstrated a greater level of robustness, effectively handling data variations, noise, and unexpected inputs. First and foremost, accuracy is a crucial factor in evaluating the effectiveness of any work in this domain. In terms of accuracy, the recommended work regularly performed better than existing work. This increased accuracy is evidence of the advancements and innovations integrated into the suggested approach, resulting in more precise and reliable outcomes.

5. Conclusions

One of the most common neurodegenerative illnesses, AD affects a great number of people, and its progression might be slowed down by early detection. Alzheimer’s disease is becoming a growing public health issue. Detecting it early is crucial for providing timely care as well as improving outcomes for patients. Our work suggests a new method for identifying AD based on high-dimensional handwriting data from the noninvasive method presented already [19]. In order to create an ensemble of classifiers that determines class labels and identifies a subset of the most significant features for accurate diagnosis support, one feature selection and dimensionality reduction method in alongside eight machine learning classifiers: Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Extra Tree (ET), Ada Boost (AB), Gradient Boosting (GB), and Voting ensemble have been used. The Voting ensemble method (approximately 94.3%) turns out to be the better performing model with a minimal feature subset size of 25. According to our findings, employing the most essential features achieved such performance that is either identical to or superior to that of state-of-the-art models. In addition to that, our approach outperformed existing standards in terms of accuracy in recognizing patients with

AD [15]. Each feature's f-statistics value was examined, and the most important features were employed to improve the machine learning model's performance and lower computational complexity [31]. Overall, our suggested approach offers a substantial advancement in the detection of AD and provides a foundation for developing more accurate and efficient diagnosis support systems. Even though our method produced encouraging performance, there's still scope for enhancement. For example, choosing different features and combining various hyperparameter tuning methods for different machine learning algorithms and dimensionality reduction techniques can enhance the performance of ML models. To boost prediction accuracy, new methods could be developed and tested against the current study to achieve better results.

References

[1] "Adi - Dementia Statistics." Alzheimer's Disease International (ADI). Accessed November 6, 2024. https://www.alzint.org/about/dementia-facts-figures/dementia-statistics/

[2] Ferdib-Al-Islam, Mostofa Shariar Sanim, Md. Rahatul Islam, Shahid Rahman, Rafi Afzal, and Khan Mehedi Hasan. "Prediction of Dementia Using Smote Based Oversampling and Stacking Classifier." *Lecture Notes in Networks and Systems*, 2023, 441–52. https://doi.org/10.1007/978-3-031-27409-1_40

[3] Archer, Madison C., Patricia H. Hall, and John C. Morgan. "[P2–430]: Accuracy of Clinical Diagnosis of Alzheimer's Disease in Alzheimer's Disease Centers (ADCS)." *Alzheimer's & Dementia* 13, no. 7S_Part_16 (July 2017). https://doi.org/10.1016/j.jalz.2017.06.1086

[4] Lock, Margaret. *The alzheimer conundrum*, December 31, 2014. https://doi.org/10.1515/9781400848461.

[5] Swaddiwudhipong, Nol, David J. Whiteside, Frank H. Hezemans, Duncan Street, James B. Rowe, and Timothy Rittman. *Pre-diagnostic cognitive and functional impairment in multiple sporadic neurodegenerative diseases*, April 6, 2022. https://doi.org/10.1101/2022.04.05.22273468

[6] Myszczynska, Monika A., Poojitha N. Ojamies, Alix M. Lacoste, Daniel Neil, Amir Saffari, Richard Mead, Guillaume M. Hautbergue, Joanna D. Holbrook, and Laura Ferraiuolo. "Applications of Machine Learning to Diagnosis and Treatment of Neurodegenerative Diseases." *Nature Reviews Neurology* 16, no. 8 (July 15, 2020): 440–56. https://doi.org/10.1038/s41582-020-0377-8

[7] Albu, Adriana, Radu-Emil Precup, and Teodor-Adrian Teban. "Results and Challenges of Artificial Neural Networks Used for Decision-Making and Control in Medical Applications." *Facta Universitatis, Series: Mechanical Engineering* 17, no. 3 (November 29, 2019): 285. https://doi.org/10.22190/fume190327035a

[8] Li, Tianbai, and Weidong Le. "Biomarkers for Parkinson's Disease: How Good Are They?" *Neuroscience Bulletin* 36, no. 2 (October 23, 2019): 183–94. https://doi.org/10.1007/s12264-019-00433-1

[9] Impedovo, Donato, and Giuseppe Pirlo. "Online Handwriting Analysis for the Assessment of Alzheimer's Disease and Parkinson's Disease: Overview and Experimental Investigation." *Frontiers in Pattern Recognition and Artificial Intelligence*, June 2019, 113–28. https://doi.org/10.1142/9789811203527_0007

[10] Precup, Radu-Emil, Teodor-Adrian Teban, Adriana Albu, Alexandra-Bianca Borlea, Iuliu Alexandru Zamfirache, and Emil M. Petriu. "Evolving Fuzzy Models for Prosthetic Hand Myoelectric-Based Control." *IEEE Transactions on Instrumentation and Measurement* 69, no. 7 (July 2020): 4625–36. https://doi.org/10.1109/tim.2020.2983531

[11] Parziale, Antonio, Rosa Senatore, and Angelo Marcelli. "Exploring Speed–Accuracy Tradeoff in Reaching Movements: A Neurocomputational Model." *Neural Computing and Applications* 32, no. 17 (January 3, 2020): 13377–403. https://doi.org/10.1007/s00521-019-04690-z

[12] He, Shida, Xiucai Ye, Sakurai Tetsuya, and Quan Zou. "MRMD3.0: A Python Tool and Webserver for Dimensionality Reduction and Data Visualization via an Ensemble Strategy." *SSRN Electronic Journal*, 2022. https://doi.org/10.2139/ssrn.4258941

[13] Erdogmus, Pakize, and Abdullah Talha Kabakus. "The Promise of Convolutional Neural Networks for the Early Diagnosis of the Alzheimer's Disease." *Engineering Applications of Artificial Intelligence* 123 (August 2023): 106254. https://doi.org/10.1016/j.engappai.2023.106254

[14] Parziale, Antonio, Antonio Della Cioppa, and Angelo Marcelli. "Investigating One-Class Classifiers to Diagnose Alzheimer's Disease from Handwriting." *Lecture Notes in Computer Science*, 2022, 111–23. https://doi.org/10.1007/978-3-031-06427-2_10

[15] Ngnamsie Njimbouom, Soualihou, Gelany Aly Abdelkader, Candra Zonyfar, Hyun Lee, and Jeong-Dong Kim. "RD-Classifier: Reduced Dimensionality Classifier for Alzheimer's Diagnosis Support System." *Lecture Notes in Computer Science*, 2023, 3–17. https://doi.org/10.1007/978-3-031-39821-6_1

[16] Cilia, Nicole Dalia, Claudio De Stefano, Francesco Fontanella, and Alessandra Scotto Di Freca. "An Experimental Protocol to Support Cognitive Impairment Diagnosis by Using Handwriting Analysis." *Procedia Computer Science* 141 (2018): 466–71. https://doi.org/10.1016/j.procs.2018.10.141

[17] Cilia, Nicole D., Giuseppe De Gregorio, Claudio De Stefano, Francesco Fontanella, Angelo Marcelli, and Antonio Parziale. "Diagnosing Alzheimer's Disease from on-Line Handwriting: A Novel Dataset and Performance Benchmarking." *Engineering Applications of Artificial Intelligence* 111 (May 2022): 104822. https://doi.org/10.1016/j.engappai.2022.104822

[18] R, Subha, Nayana B R, and M Selvadass. "Hybrid Machine Learning Model Using Particle Swarm Optimization for Effectual Diagnosis of Alzheimer's Disease from Handwriting." *2022 4th International Conference on Circuits, Control, Communication and Computing (I4C)*, December 21, 2022, 491–95. https://doi.org/10.1109/i4c57141.2022.10057948

[19] S. O. Olatunji et al., "Machine learning based preemptive diagnosis of lung cancer using clinical data," *2022 7th International Conference on Data Science and Machine Learning Applications (CDMA)*, 2022. doi:10.1109/cdma54072.2022.00024

[20] Atif, Mohammad, Faisal Anwer, and Faisal Talib. "An Ensemble Learning Approach for Effective Prediction of Diabetes Mellitus Using Hard Voting Classifier." *Indian Journal Of Science And Technology* 15, no. 39 (October 21, 2022): 1978–86. https://doi.org/10.17485/ijst/v15i39.1520

[21] Dong, Xibin, Zhiwen Yu, Wenming Cao, Yifan Shi, and Qianli Ma. "A Survey on Ensemble Learning." *Frontiers of Computer Science* 14, no. 2 (August 30, 2019): 241–58. https://doi.org/10.1007/s11704-019-8208-z.

[22] Analysis of variance. Accessed November 5, 2024. https://www.mrc-cbu.cam.ac.uk/personal/rik.henson/personal/PennyHenson_SPM_06b.pdf

[23] Van der Maaten, Laurens, and Geoffrey Hinton. "Visualizing data using t-SNE." *Journal of machine learning research* 9, no. 11 (2008). http://jmlr.org/papers/v9/vandermaaten08a.html

[24] Van Der Maaten, Laurens. "Accelerating t-SNE using tree-based algorithms." *The journal of machine learning research* 15, no. 1 (2014): 3221-3245. https://doi.org/10.5555/2627435.2697068

[25] Van der Maaten, Laurens, and Geoffrey Hinton. "Visualizing non-metric similarities in multiple maps." *Machine learning* 87 (2012): 33-55. https://doi.org/10.1007/s10994-011-5273-4

[26] Pezzotti, Nicola, Boudewijn P. Lelieveldt, Laurens van Maaten, Thomas Hollt, Elmar Eisemann, and Anna Vilanova. "Approximated and User Steerable Tsne for Progressive Visual Analytics." *IEEE Transactions on Visualization and Computer Graphics* 23, no. 7 (July 1, 2017): 1739–52. https://doi.org/10.1109/tvcg.2016.2570755

[27] "Votingclassifier." scikit. Accessed November 6, 2024. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html

[28] Leon, Florin, Sabina-Adriana Floria, and Costin Badica. "Evaluating the Effect of Voting Methods on Ensemble-Based Classification." *2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, July 2017, 1–6. https://doi.org/10.1109/inista.2017.8001122

[29] Xu, William, Tingying Helen Zeng, and Mikhail Shalaginov. "Class Activation Mapping Enhanced AlexNet Convolutional Neural Networks for Early Diagnosis of Alzheimer's Disease." *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, December 6, 2022, 2550–55. https://doi.org/10.1109/bibm55620.2022.9994868

[30] Tong, Yinsheng, Zuoyong Li, Hui Huang, Libin Gao, Minghai Xu, and Zhongyi Hu. "Research of Spatial Context Convolutional Neural Networks for Early Diagnosis of Alzheimer's Disease." *The Journal of Supercomputing* 80, no. 4 (September 26, 2023): 5279–97. https://doi.org/10.1007/s11227-023-05655-9

[31] Cilia, Nicole Dalia, Claudio De Stefano, Francesco Fontanella, and Alessandra Scotto Freca. “Feature Selection as a Tool to Support the Diagnosis of Cognitive Impairments through Handwriting Analysis.” *IEEE Access* 9 (2021): 78226–40. https://doi.org/10.1109/access.2021.3083176

[32] “Welcome to the Shap Documentationℑ.” Welcome to the SHAP documentation - SHAP latest documentation. Accessed November 6, 2024. https://shap.readthedocs.io/en/latest/index.html

[33] Lu, Jieyu, and Yingkai Zhang. "Unified deep learning model for multitask reaction predictions with explanation." *Journal of chemical information and modeling* 62, no. 6 (2022): 1376-1387. https://doi.org/10.1021/acs.jcim.1c01467.s001

[34] Tan, Vincent Wei Sheng, Wei Xiang Ooi, Yi Fan Chan, Tee Connie, and Michael Kah Ong Goh. 2024. “Vision-Based Gait Analysis for Neurodegenerative Disorders Detection”. *Journal of Informatics and Web Engineering* 3 (1):136-54. https://doi.org/10.33093/jiwe.2024.3.1.9.

[35] Mushtaq, Shougfta, Tabassum Javed, and Mazliham Mohd Su’ud. 2024. “Ensemble Learning-Powered URL Phishing Detection: A Performance Driven Approach”. *Journal of Informatics and Web Engineering* 3 (2):134-45. https://doi.org/10.33093/jiwe.2024.3.2.10.