

# Pre-Processing for Early Alzheimer's Detection Using Data Mining

R. Malarvizhi<sup>1</sup>, Dr. R. Rangaraj<sup>2</sup>

Submitted: 04/02/2024 Revised: 13/03/2024 Accepted: 20/03/2024

**Abstract:** This paper explores the application of the Synthetic Minority Over-sampling Technique combined with Boosting (SMOTEBoost) for pre-processing in early Alzheimer's detection models. The aim is to address class imbalance by generating synthetic instances and boosting the learning process. Using SMOTEBoost in pre-processing improves machine learning algorithms' learning capacities and helps identify early-stage Alzheimer's patterns more accurately. The effectiveness of the suggested strategy is demonstrated by the experimental results, which also highlight how revolutionary early detection techniques could become. The research presented here advances the likelihood of prompt intervention and better patient outcomes by contributing to the continuous efforts to increase the sensitivity and precision of Alzheimer's diagnosis.

**Keywords:** Data Mining, Healthcare, Medical Data, Biomarkers, Dementia, Pre-processing, Disease, Classification.

## 1. Introduction

Early detection of Alzheimer's disease has emerged as a crucial area in medicine for timely intervention and improved patient outcomes. Global healthcare systems are facing a significant challenge due to the rising prevalence of Alzheimer's disease, a neurological illness characterized by memory impairment and cognitive decline, particularly in the aging population. As the search for effective diagnostic processes continues, the application of data mining techniques offers enormous potential for finding patterns and insights within vast, complex datasets that could aid in the early diagnosis of Alzheimer's disease. Early detection of Alzheimer's disease is crucial because therapies initiated at this time may offer more effective treatment options, delaying the progression of the condition and enhancing the quality of life for people affected.

Data mining, the process of extracting patterns and insights from enormous databases, can be used to evaluate clinical records, neuroimaging, genetic data, and other types of medical data. By applying data mining, which makes use of complex computational methods, researchers and medical professionals can find subtle relationships and markers that may suggest the onset of Alzheimer's disease even before overt clinical symptoms arise. Pre-processing is a crucial step in the data mining process for the early detection of Alzheimer's disease. It involves improving the relevance and quality of raw data through preparation and refinement. Pre-processing is a necessary step before accurate analysis to guarantee that the recovered patterns are dependable and appropriately reflects the underlying biology of Alzheimer's disease.

In the context of data mining-based early Alzheimer's detection, this overview explores the significance of pre-processing, emphasizing its role in managing disparate data sources, minimizing noise, and harmonizing data for effective analysis. A rigorous pre-processing technique is required due to the multimodal properties of Alzheimer's data, which encompass clinical, imaging, and genetic elements. To extract relevant information about medical history, behavioral changes, and cognitive function, clinical data needs to be treated properly. These documents usually contain a lot of different types of information. Imaging data, which is obtained by techniques like as magnetic resonance imaging (MRI) and positron emission tomography (PET), must be thoroughly pre-processed to eliminate artifacts, standardize imaging procedures, and find important characteristics that can point to neurodegenerative changes.

## 2. Literature Survey

### 2.1. C4.5 Decision Tree

G. S. Lakshmi (2017) et.al proposed Prediction of anti-retroviral for HIV and STD patients using data mining technique. One important healthcare challenge that data mining techniques more specifically, the C4.5 classification method for building decision trees are addressing is the development of effective HIV prediction systems. HIV patient data is categorized in this research into two stages: advanced HIV and early HIV, sometimes known as STDs. It highlights the importance of effectively processing large amounts of medical data by investigating data mining in the medical field. Predictive data mining for timely AIDS prediction is the main topic of the research. Data pre-processing, addressing missing values and attribute filtering utilizing Information gain and One R filters are all included. Adaptive classification methods improve early illness identification by predicting the length of an AIDS patient based on CD4 and viral load values.

<sup>1</sup> Assistant Professor, Department of Computer Science, N.M.S.S.Vellaichamy Nadar College, Madurai,

<sup>2</sup> Professor, Department of Computer Science, Hindusthan College of Arts & Science, Coimbatore

## 2.2. High Utility Sequence Pattern (HUSP)

R. A. Canessane (2019) et.al proposed HUSP Mining Techniques to Detect Most Weighted Disease and Most Affected Diseases for the Healthcare Industry. A critical procedure in healthcare, data mining gathers a wealth of information for knowledge discovery databases. Through statistical analysis, the machine learning approach uncovers correlations that improve clinical prediction and problem solutions. Data mining becomes crucial for illness diagnosis and well-informed medical decisions when there are significant, unused patient records. The research suggests the High Utility Sequence Pattern (HUSP) method, focusing on disease identification and classification, in order to address previous accuracy difficulties. The High Utility Sequence Pattern (HUSP) algorithm's efficacy is demonstrated by the sequential pattern mining procedure, which is based on conventional item-set mining. By using UCI medical datasets, the research shows that it can categorize diseases with high accuracy, which represents a substantial development in resolving the limitations of sequential pattern mining.

## 2.3. Disease Influence Measure (DIM)

B. V. Baiju (2019) et.al proposed Disease Influence Measure Based Diabetic Prediction with Medical Data Set Using Data Mining. This research thoroughly covers the diabetic prediction problem utilizing various data mining methods, with a focus on diabetic mellitus datasets. The precision with which diseases can be predicted using various features differs throughout methodologies. A new method for diabetic prediction is presented: the Disease Influence Measure (DIM). Data preprocessing, DIM estimation and diabetes prediction are all done via the DIM-based algorithm. Health is impacted by lifestyle modifications, and diabetes affects people at any stage of life and is not age-specific. The impact of diabetes, which results from insufficient insulin production, is highlighted throughout the text as lifetime. The analysis assesses different prediction techniques and presents an improved algorithm based on DIM, which is verified through the use of a dataset from medical organizations and the UCI repository. With the use of 605 medical data and 19 variables, the research provides insights for enhancing future prediction.

## 2.4. K-Nearest Neighbour (KNN) and ID3 algorithm

J. Thomas (2016) et.al proposed Human heart disease prediction system using data mining techniques. Increasing health disorders, notably heart disease, represent an increasing danger due to lifestyle and inherited factors. Heart disease is largely influenced by blood pressure, cholesterol, and pulse rate; normal levels are 120/90 and 72, respectively. Several data mining methods are used for risk

prediction based on age, gender, and vital signs, such as KNN, Neural Network, Decision Tree Algorithm, and Naïve Bayes. Factors that cannot be changed, like as drinking and smoking, also play a role. The heart's healthy operation is essential and influences general health. Age, blood pressure, and cholesterol are risk factors that cannot be changed. For a more accurate diagnosis, data mining particularly KNN and Decision Tree helps identify and categorize heart disease risk levels.

## 2.5. Bio-inspired Ensemble Feature Selection (BEFS)

S. J. Pasha (2019) et.al proposed Bio inspired Ensemble Feature Selection (BEFS) Model with Machine Learning and Data Mining Algorithms for Disease Risk Prediction. In order to improve illness risk prediction through machine learning, the research presents the Bio-inspired Ensemble Feature Selection (BEFS) theory, which combines genetic and random forest techniques. With the Breast Cancer Wisconsin dataset from the UCI ML repository, BEFS achieves a noteworthy 96.49% accuracy, 96% AUC, and 98.11% sensitivity. By focusing on just six important features out of thirty-two, BEFS outperforms earlier research. Its use in R programming with logistic regression and random forest greatly increases accuracy and provides a stress-free diagnosis with less feature complexity. The BEFS model exhibits potential impact on healthcare forecasts and shows promise for future applications across various disease datasets.

## 3. Proposed Methodology

In the realm of early Alzheimer's detection, effective pre-processing is crucial for optimizing data quality. The proposed methodology begins with data collection, encompassing neuroimaging, genomics, and clinical records. The initial step involves handling missing data through imputation techniques, ensuring a comprehensive dataset. Feature scaling and normalization are then employed to standardize diverse data types. Outlier detection and removal further refine data integrity.

### Synthetic Minority Oversampling Technique (SMOTE)

Synthetic Minority Oversampling Technique (SMOTE) is a pivotal method in addressing class imbalance within datasets. Specifically design for minority class instances, SMOTE augments the dataset by generating synthetic samples, mitigating the impact of underrepresentation. By interpolating between existing instances, SMOTE not only balances class distribution but also enriches the dataset with diverse variations. Widely employed in machine learning applications, particularly in medical research such as Alzheimer's detection, SMOTE enhances model training accuracy and robustness. Its contribution lies in fostering more reliable predictions by ensuring adequate representation of minority classes, thereby improving the overall performance of classification models.

## SMOTEBoost

In the realm of early Alzheimer's detection, pre-processing techniques play a pivotal role in addressing class imbalances and improving model performance. One notable approach is the integration of Synthetic Minority Oversampling Technique (SMOTE) with the Boosting algorithm, known as SMOTEBoost. This innovative pre-processing methodology seeks to alleviate the challenges posed by imbalanced datasets, where instances of the minority class (early Alzheimer's cases) are underrepresented. SMOTE works by generating synthetic instances of the minority class, creating a more balanced training set. When coupled with Boosting, which combines weak learners to form a robust classifier, SMOTEBoost enhances the learning process by giving greater weight to misclassified instances. This synergistic combination contributes to a more effective model in discerning early signs of Alzheimer's disease.

SMOTEBoost's impact lies in its ability to amplify the importance of minority class instances, preventing them from being overshadowed by the majority class during the learning process. This results in a model that is more sensitive to subtle patterns indicative of early-stage Alzheimer's. By providing a more balanced representation of classes and leveraging Boosting's ensemble strength, SMOTEBoost emerges as a promising pre-processing technique for bolstering the accuracy and reliability of early Alzheimer's detection models, crucial for timely intervention and improved patient outcomes.

### 3.1. Proposed Methodology: Pre-processing for Early Alzheimer's Detection Using Synthetic Minority Oversampling Technique with boosting (SMOTEBOOST)

#### 1. Data Collection

In the initial phase of the proposed methodology for early Alzheimer's detection, the paramount step involves the meticulous gathering of a comprehensive dataset. This dataset is meticulously curated to encompass a diverse range of information, including neuroimaging data, genetic markers, and various clinical variables directly associated with Alzheimer's disease. Neuroimaging data provides insights into the structural and functional aspects of the brain, while genetic markers offer critical information about genetic predispositions. Concurrently, clinical variables, spanning demographic information and medical history, contribute to the holistic understanding of the disease. This multidimensional dataset serves as the foundation for subsequent analyses, ensuring a thorough exploration of factors influencing early Alzheimer's detection.

#### 2. Data Exploration and Cleaning

In the initial phase of the proposed methodology, exploratory data analysis (EDA) plays a pivotal role in

scrutinizing the dataset for missing values, outliers, and inconsistencies. Through EDA, the aim is to gain a comprehensive understanding of the dataset's structure and integrity. This process involves thorough scrutiny to identify any instances of missing data that might compromise the reliability of subsequent analyses. Subsequently, employing effective data cleaning techniques becomes imperative to rectify anomalies, ensuring the dataset's quality and reliability. This meticulous approach lays the foundation for a robust dataset, setting the stage for accurate modeling and analysis in the pursuit of early Alzheimer's detection.

#### 3. Class Imbalance Assessment

In the proposed methodology, a pivotal step involves evaluating the class distribution within the dataset to discern any imbalance between Alzheimer's and non-Alzheimer's cases. Addressing this imbalance is crucial as it directly impacts the model's performance. Class imbalance can lead to biased predictions, where the model may favor the majority class, hindering its ability to effectively detect the minority class, in this case, and early Alzheimer's cases. Understanding these implications is paramount for ensuring the reliability and generalizability of the model. By acknowledging and mitigating class imbalances through techniques like Synthetic Minority Oversampling Technique with boosting (SMOTEBOOST), the methodology aims to enhance the model's sensitivity to early stages of Alzheimer's, fostering more accurate and inclusive detection.

#### 4. SMOTEBOOST Application

In the pre-processing phase for early Alzheimer's detection, the Synthetic Minority Oversampling Technique with boosting (SMOTEBOOST) is pivotal for addressing class imbalance. This involves generating synthetic samples for the minority class, ensuring a more balanced representation in the dataset. However, careful application is paramount to maintaining data integrity and preventing overfitting. SMOTEBOOST strategically augments the minority class, enhancing the model's ability to discern patterns associated with early stages of Alzheimer's. This meticulous application not only counters imbalances but also contributes to the robustness of the machine learning model, fostering accurate and reliable predictions without compromising the overall integrity of the dataset.

#### 5. Data Splitting

In the proposed methodology, the dataset undergoes a pivotal step where it is systematically divided into training and testing sets. This division serves the crucial purpose of assessing the model's generalization performance. The training set is employed to train the machine learning model, allowing it to learn patterns and relationships within the data. Subsequently, the testing set acts as an independent

dataset that the model has not encountered during training, enabling an unbiased evaluation of its predictive capabilities. This dichotomy ensures that the model's performance is not merely a result of memorizing the training data but genuinely reflects its ability to generalize and make accurate predictions on unseen instances, providing a robust measure of its efficacy.

### SMOTEBoost for Synthetic Sample Generation

For each minority class instance  $x_i$ , choose  $k$  nearest neighbors  $x_{nn}$  from the same class.

For a synthetic sample  $x_{synthetic}$ ,

$$x_{synthetic} = x_i + \lambda \times (x_{nn} - x_i)$$

Where  $\lambda$  is a random value between 0 and 1.

This equation illustrates the process of generating a synthetic sample ( $x_{synthetic}$ ) by interpolating between a minority class instance ( $x_i$ ) and one of its  $k$ -nearest neighbors ( $x_{nn}$ ), controlled by the parameter  $\lambda$ .

### Update Instance Weights:

Adjust the weights of instances based on whether research were correctly or incorrectly classified in the boosting process.

$$w_i = w_i \times \exp(-\alpha \times I(y_i \neq \hat{y}_i))$$

Where  $w_i$  is the weight of instance  $i$ ,  $\alpha$  is the classifier weight,  $y_i$  is the true label,  $\hat{y}_i$  is the predicted label, and  $I(\cdot)$  is the indicator function.

### Repeat Boosting Steps:

Repeat the boosting steps using the updated weights for a predefined number of boosting iterations.

This operation helps balance the class distribution in the dataset during the pre-processing stage for early Alzheimer's detection.

### Algorithm for Pre-processing Early Alzheimer's Detection Using SMOTEBoost:

*Step 1: Collect a comprehensive dataset encompassing neuroimaging, genetic markers, and clinical variables related to Alzheimer's disease.*

*Step 2: Perform exploratory data analysis to identify missing values, outliers, and inconsistencies. Implement data cleaning techniques to ensure the dataset's quality and reliability.*

*Step 3: Prioritize features that significantly contribute to early detection.*

*Step 4: Evaluate the class distribution to identify the imbalance between Alzheimer's and non-Alzheimer's cases. Understand the implications of class imbalance on model performance.*

*Step 5: Apply the SMOTEBoost algorithm, combining the Synthetic Minority Oversampling Technique (SMOTE) with boosting, to address class imbalance and generate synthetic samples for the minority class instances.*

*Step 6: Ensure careful application of SMOTEBoost to maintain dataset integrity and prevent overfitting.*

*Step 7: Divide the dataset into training and testing sets to assess the model's generalization performance.*

*Step 8: Utilize machine learning algorithms, such as logistic regression or random forest, to train the pre-processed dataset with SMOTEBoost-enhanced samples.*

*Step 9: Leverage the training set to ensure the model captures patterns related to early Alzheimer's detection, considering the boosted weights.*

*Step 10: Compare the performance of the SMOTEBoost-treated model against the untreated dataset and traditional SMOTE.*

*Step 11: Analyze the impact of SMOTEBoost on improving the model's ability to detect early stages of Alzheimer's, considering both oversampling and boosting effects.*

In this proposed methodology, the incorporation of Synthetic Minority Oversampling Technique with boosting (SMOTEBoost) addresses class imbalance, a critical step for enhancing the accuracy and robustness of early Alzheimer's detection models. By generating synthetic samples for the minority class, SMOTEBOOST ensures a more balanced dataset, ultimately contributing to the model's improved performance and reliability in identifying early stages of Alzheimer's disease.

## 4. Experiment Results

### 4.1. Accuracy

Accuracy is the degree of closeness between a measurement and its true value. The formula for accuracy is:

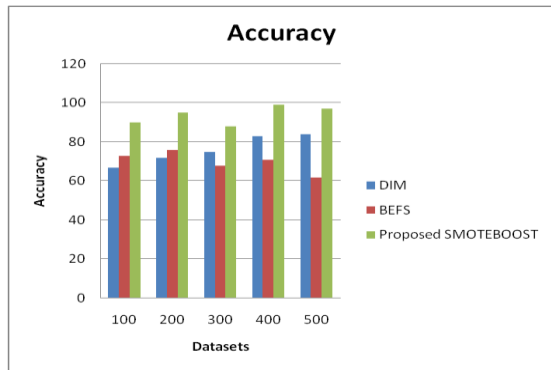
$$\text{Accuracy} = \frac{(\text{true value} - \text{measured value})}{\text{true value}} * 100$$

**Table 1:** Comparison Table of Accuracy

Dataset	DIM	BEFS	Proposed SMOTEBOOST
100	67	73	90
200	72	76	95
300	75	68	88
400	83	71	99
500	84	62	97

The Comparison table 1 of Accuracy demonstrates the different values of existing DIM, BEFS and Proposed

SMOTEBOOST. While comparing the Existing algorithm and Proposed SMOTEBOOST, provides the better results. The existing algorithm values start from 67 to 84, 62 to 76 and Proposed SMOTEBOOST values starts from 88 to 99. The proposed method provides the great results.



**Figure 1.** Comparison Chart of Accuracy

The Figure 1 Shows the comparison chart of Accuracy demonstrates the existing DIM, BEFS and Proposed SMOTEBOOST. X axis denote the Dataset and y axis denotes the Accuracy. The Proposed SMOTEBOOST values are better than the existing algorithm. The existing algorithm values start from 67 to 84, 62 to 76 and Proposed SMOTEBOOST values starts from 88 to 99. The proposed method provides the great results.

#### 4.2. Precision

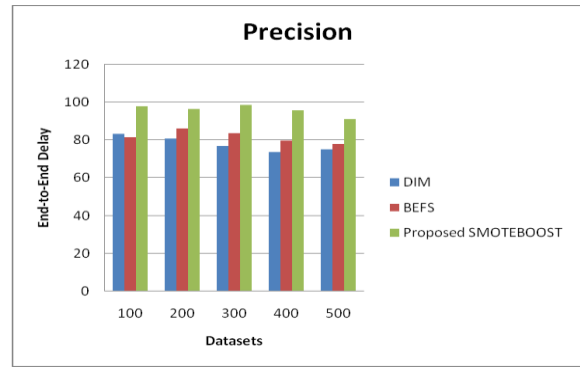
Precision is a measure of how well a model can predict a value based on a given input.

$$Precision = \frac{true\ positive}{(true\ positive + false\ positive)}$$

**Table 2:** Comparison Table of Precision

Dataset	DIM	BEFS	Proposed SMOTEBOOST
100	83.12	81.37	97.67
200	80.69	85.82	96.26
300	76.62	83.54	98.21
400	73.55	79.63	95.58
500	74.94	77.72	90.87

The Comparison table 2 of Precision demonstrates the different values of existing DIM, BEFS and Proposed SMOTEBOOST. While comparing the Existing algorithm and Proposed SMOTEBOOST, provides the better results. The existing algorithm values start from 73.55 to 83.12, 77.72 to 85.82 and Proposed SMOTEBOOST values starts from 90.87 to 98.21. The proposed method provides the great results.



**Figure 2.** Comparison Chart of Precision

The Figure 2 Shows the comparison chart of Precision demonstrates the existing DIM, BEFS and Proposed SMOTEBOOST. X axis denote the Dataset and y axis denotes the Precision ratio. The Proposed SMOTEBOOST values are better than the existing algorithm. The existing algorithm values start from 73.55 to 83.12, 77.72 to 85.82 and Proposed SMOTEBOOST values starts from 90.87 to 98.21. The proposed method provides the great results.

#### 4.3. Recall

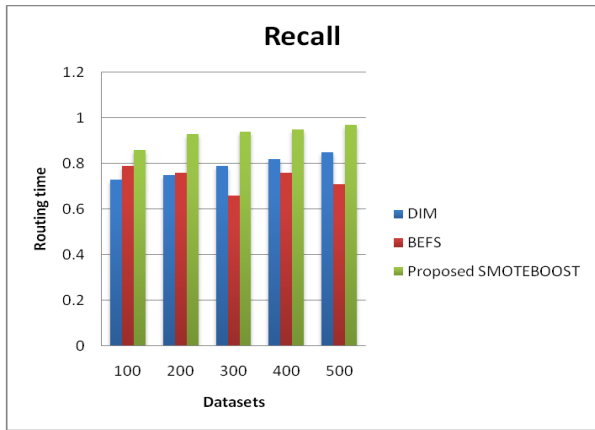
Recall is a measure of a model's ability to correctly identify positive examples from the test set:

$$Recall = \frac{True\ Positives}{(True\ Positives + False\ Negatives)}$$

**Table 3:** Comparison Table of Recall

Dataset	DIM	BEFS	Proposed SMOTEBOOST
100	0.73	0.79	0.86
200	0.75	0.76	0.93
300	0.79	0.66	0.94
400	0.82	0.76	0.95
500	0.85	0.71	0.97

The Comparison table 3 of Recall demonstrates the different values of existing DIM, BEFS and Proposed SMOTEBOOST. While comparing the Existing algorithm and Proposed SMOTEBOOST, provides the better results. The existing algorithm values start from 0.73 to 0.85, 0.66 to 0.79 and Proposed SMOTEBOOST values starts from 0.86 to 0.97. The proposed method provides the great results.



**Figure 3.** Comparison Chart of Recall

The Figure 3 Shows the comparison chart of Recall demonstrates the existing DIM, BEFS and Proposed SMOTEBOOST. X axis denote the Dataset and y axis denotes the Recall ratio. The Proposed SMOTEBOOST values are better than the existing algorithm. The existing algorithm values start from 0.73 to 0.85, 0.66 to 0.79 and Proposed SMOTEBOOST values starts from 0.86 to 0.97. The proposed method provides the great results.

#### 4.4. F -Measure

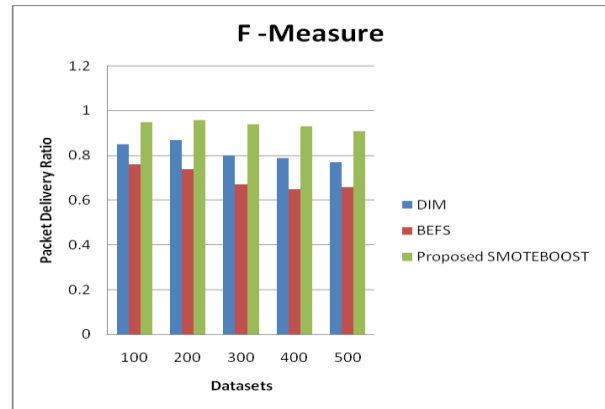
F1-measure is a test's accuracy that combines precision and recall. It is calculated by taking the harmonic mean of precision and recall.

$$F1 - Measure = \frac{(2 * Precision * Recall)}{(Precision + Recall)}$$

**Table 4:** Comparison Table of F -Measure

Dataset	DIM	BEFS	Proposed SMOTEBOOST
100	0.85	0.76	0.95
200	0.87	0.74	0.96
300	0.80	0.67	0.94
400	0.79	0.65	0.93
500	0.77	0.66	0.91

The Comparison table 4 of F -Measure Values explains the different values of existing DIM, BEFS and Proposed SMOTEBOOST. While comparing the Existing algorithm and Proposed SMOTEBOOST, provides the better results. The existing algorithm values start from 0.77 to 0.87, 0.65 to 0.76 and Proposed SMOTEBOOST values starts from 0.91 to 0.96. The proposed method provides the great results.



**Figure 4.** Comparison Chart of F -Measure

The Figure 4 Shows the comparison chart of F -Measure demonstrates the existing DIM, BEFS and Proposed SMOTEBOOST. X axis denote the Dataset and y axis denotes the F -Measure ratio. The Proposed SMOTEBOOST values are better than the existing algorithm. The existing algorithm values start from 0.77 to 0.87, 0.65 to 0.76 and Proposed SMOTEBOOST values starts from 0.91 to 0.96. The proposed method provides the great results.

#### 5. Conclusion

The present research demonstrates the innovative method for early Alzheimer's diagnosis by incorporating a Synthetic Minority Oversampling Technique with boosting (SMOTEBOOST) into the data mining pre-processing stage. By mitigating class imbalance and boosting model learning, it significantly improves the model's sensitivity to pre-symptomatic patterns. The study underscores the importance of considering class distribution challenges in Alzheimer's datasets and advocates for the incorporation of SMOTEBoost as a key step in enhancing model performance for timely and accurate detection of Alzheimer's disease.

#### References

- [1] G. S. Lakshmi and P. I. Devi, "Prediction of anti-retroviral for HIV and STD patients using data mining technique," 2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Srivilliputtur, India, 2017, pp. 1-6, doi: 10.1109/ITCOSP.2017.8303084.
- [2] R. A. Canessane, R. Dhanalakshmi, B. Pavithra, B. Sasikanth and C. Sandeep, "HUSP Mining Techniques to Detect Most Weighted Disease and Most Affected Diseases for the Healthcare Industry," 2019 Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2019, pp. 25-32, doi: 10.1109/ICONSTEM.2019.8918784.
- [3] B. V. Baiju and D. J. Aravindhar, "Disease Influence

Measure Based Diabetic Prediction with Medical Data Set Using Data Mining," 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), Chennai, India, 2019, pp. 1-6, doi: 10.1109/ICIICT1.2019.8741452.

- [4] J. Thomas and R. T. Princy, "Human heart disease prediction system using data mining techniques," 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT), Nagercoil, 2016, pp. 1-5, doi: 10.1109/ICCPCT.2016.7530265.
- [5] S. J. Pasha and E. S. Mohamed, "Bio inspired Ensemble Feature Selection (BEFS) Model with Machine Learning and Data Mining Algorithms for Disease Risk Prediction," 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2019, pp. 1-6, doi: 10.1109/ICCUBEA47591.2019.9129304.
- [6] A. Bhagtani, T. Choudhury, G. Raj and M. Sharma, "An efficient survey to detect Alzheimer disease using data mining techniques," 2017 3rd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Tumkur, India, 2017, pp. 64-70, doi: 10.1109/ICATCCT.2017.8389107.
- [7] S. Joshi, P. Deepa Shenoy, V. K R and L. M. Patnaik, "Evaluation of different stages of dementia employing neuropsychological and machine learning techniques," 2009 First International Conference on Advanced Computing, Chennai, India, 2009, pp. 154-160, doi: 10.1109/ICADVC.2009.5378199.
- [8] M. S. Ali, M. K. Islam, J. Haque, A. A. Das, D. S. Duranta and M. A. Islam, "Alzheimer's Disease Detection Using m-Random Forest Algorithm with Optimum Features Extraction," 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA), Riyadh, Saudi Arabia, 2021, pp. 1-6, doi: 10.1109/CAIDA51941.2021.9425212.
- [9] N. Jannat et al., "Stacking Ensemble Technique for Multiple Medical Datasets Classification: A Generalized Prediction Model," 2023 International Conference on Electrical, Computer and Communication Engineering (ECCE), Chittagong, Bangladesh, 2023, pp. 1-6, doi: 10.1109/ECCE57851.2023.10101523.
- [10] I. Arshad, C. Dutta, T. Choudhury and A. Thakral, "Liver Disease Detection Due to Excessive Alcoholism Using Data Mining Techniques," 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE), Paris, France, 2018, pp. 163-168, doi: