A Machine Learning Model for Alzheimer's Disease Prediction

Pooja Rani[1], Rohit Lamba[2*], Ravi Kumar Sachdeva[3], Karan Kumar[4*], Celestine Iwendi[5*]
[1]MMICTBM, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, India
[2,4]Electronics and Communication Engineering Department, Maharishi Markandeshwar Engineering College, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, India-133207
[3]Department of Computer Science & Engineering, Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
[5]School of Creative Technologies, University of Bolton, Bolton, BL3 5AB, UK
[1]poojasachdeva1886@gmail.com, [2]rohitlamba14@gmail.com, [3]ravisachdeva1983@gmail.com, [4]karan.170987@gmail.com, [5]celestine.iwendi@ieee.org
*Corresponding Author

Abstract
Alzheimer's disease (AD) is a neurodegenerative disorder that mostly affects old aged people. Its symptoms are initially mild, but they get worse over time. Although this health disease has no cure, its early diagnosis can help to reduce its impacts. In this paper, a methodology SMOTE-RF is proposed for AD prediction. Alzheimer's is predicted using machine learning (ML) algorithms. Performance of three algorithms decision tree (DT), extreme gradient boosting (XGB), and random forest (RF) are evaluated in prediction. Open Access Series of Imaging Studies (OASIS) longitudinal dataset available on Kaggle is used for experiments. Dataset is balanced using synthetic minority oversampling technique (SMOTE). Experiments are done on both imbalanced and balanced datasets. DT obtained 73.38% accuracy, XGB obtained 83.88% accuracy and RF obtained a maximum 87.84% accuracy on the imbalanced dataset. DT obtained 83.15% accuracy, XGB obtained 91.05% accuracy and RF obtained maximum 95.03% accuracy on the balanced dataset. Maximum accuracy of 95.03% is achieved with SMOTE-RF.
Keywords: Extreme Gradient Boosting, Alzheimer's Disease, Decision Tree, Random Forest.

1. Introduction

The most prevalent neurodegenerative disorder is Alzheimer's disease. Its symptoms are firstly mild, but with time symptoms increase. Ten to twenty years before symptoms appear, the brain begins to change in the early stages of this disease. It gradually impairs thinking skills and damages memories. A group of symptoms linked to cognitive impairment make up dementia. Memory, thinking, reasoning, and the capacity to carry out daily duties are all impacted by dementia. The most typical cause of dementia is Alzheimer's disease. Over 70% of dementia patients come from low-income nations. Dementia patients face difficulty in managing their emotions. Mostly, old age persons are affected by this disease. The person suffering from this illness could have anxiety or memory problems, such as forgetting familiar names and places. The person's close friends and family have noticed that they have trouble remembering their names. A doctor can identify a patient's memory and attention issues by performing a thorough medical interview[1].

Alzheimer's disease symptoms persist and worsen with time. This development impairs a person's capacity for efficient communication, environment adaptation, and finally even the execution of simple movements. It gets harder for them to verbally express their pain or suffering. They frequently need significant support for daily tasks due to the continuing decrease in memory and cognitive abilities. At this point in the disease, Alzheimer's patients could encounter the following difficulties:

1. Everyday activities and personal care require round-the-clock assistance.
2. They lose awareness of their surroundings and recent events.
3. With time, their physical capabilities, such as sitting, walking, and swallowing, may change.
4. Interacting with others is getting harder and harder.
5. An increase in the prevalence of infections, particularly pneumonia.

Although this disease has no treatment, its symptoms can be reduced with timely detection[2]. Therapeutic options that can slow disease development using an early and precise diagnosis. The impact of ML techniques on the healthcare industry has grown significantly recently[3]. The pressure to create new methods for early identification and intervention of AD has increased due to the lack of a permanent treatment.

Scenario

Alzheimer's disease develops gradually, starting with mild symptoms that get worse over time. A prompt and correct diagnosis is essential for enabling actions and support that may enhance the quality of life. This study aims to improve the accuracy and responsiveness of Alzheimer's disease prediction. Early detection and intervention are crucial to improve patient outcomes and lower the societal and financial burden due to the aging population and lack of a recognized treatment for the illness. Machine learning presents a promising approach for locating subtle patterns that can help in early diagnosis due to its ability to analyze complex data[4]. By giving medical practitioners early detection tools, this research has the potential to change clinical practice. We can significantly contribute to the ethical and social imperative of improving the lives of millions of people impacted by this disease by utilizing machine learning for prediction.

Motivations

This research is motivated by the pressing global health challenge posed by Alzheimer's disease. There are two main motivations. First, the need for rapid diagnosis and treatment of Alzheimer's is made more urgent by the aging world population. Second, there is an urgent need for automated technologies that can support medical professionals in their diagnostic work in the current diagnostic landscape.

Proposed Solution

In this research, methodology for Alzheimer's disease prediction called SMOTE-RF is proposed. Authors have carefully analyzed the performance of three well-known ML algorithms: DT, XGB, and RF in their effort to create a reliable predictive model. The well-known OASIS dataset, which is available on Kaggle, is utilized to carry out experiments. Predictive model performance can be greatly impacted by class imbalance, a typical problem in medical data. SMOTE, a method that rebalances the dataset by creating synthetic samples for the minority class is used to overcome this issue. By ensuring that models are not biased towards the majority class, this stage enables more precise predictions.

Main Contributions

The Main Contributions of this research are:

1. Machine learning algorithms namely: decision tree, extreme gradient boosting, and random forest are used for model building to predict Alzheimer’s disease.
2. Experiments are performed in two ways, first on original dataset and then on class balanced dataset.
3. As the dataset is highly imbalanced, so the class imbalance problem is overcome by SMOTE technique.

Various existing systems of predicting Alzheimer's are discussed in section 2. Methodology used in this paper is explained in section 3. Results of experiments performed are demonstrated in section 4. Work is concluded in section 5.

2. Related Works

Antor et al. [2] used SVM, DT, RF, and LR algorithms for early detection of Alzheimer's. Eight attributes of OASIS dataset was used for prediction. Missing values in all features were deleted except socioeconomic status (SES). Missing values in this feature was filled using median. Ratio of 80:20 was used for training and testing data. RF suffered from the problem of overfitting. SVM provided the best accuracy of 92%. Kavitha et al. [3] performed prediction using DT, SVM, RF, voting and gradient-boosting classifiers. Missing data of SES was filled using median method. RF obtained the highest accuracy of 86.92%.

Leong and Abdullah[5] performed feature selection to select significant attributes from OASIS-I cross-sectional dataset. There were 5.63% missing values in the dataset and these values were removed from the dataset. RF provided the best accuracy of 94.39%. Alickovic and Subasi [6] used histogram for extraction of features and RF for classification. Histogram transformed images of brain into feature vector. After that classification was done attaining 85.77% accuracy on ADNI dataset. Neelaveni [7] presented an Alzheimer's disease prediction methodology by using psychological parameters. Two machine learning classifiers SVM and decision tree were applied for classification and best 85% accuracy was obtained by SVM classifier. Shahbaj et al. [8] used KNN, DT, NB, rule induction and linear model for predicting Alzheimer's. Algorithms were applied on ADNI dataset. Highest accuracy of 88.24% was achieved with linear model.

Velazquez et al.[9] considered nine features from ADNI dataset for prediction of Alzheimer's. Imbalanced classes were balanced using oversampling. After oversampling, prediction was done using RF attaining 93.6% accuracy. Bashir et al. [10]developed a deep neural network (DNN) based system for predicting Alzheimer's. OASIS dataset was used for experiments and 92.39% accuracy was obtained. Vashishtha et al. [11] performed feature selection using wrapper methods. Experiments were performed with and without feature selection methods. RF, DT, SVM, XgB classifiers were used. Classifiers provided increased accuracy with feature selection.

Martinez-Murcia et al. [12] have presented the deep learning-based auto encoders for AD prediction. The features from MRI pictures had been extracted to describe a person's cognitive symptoms. The distribution of the collected features is then examined using classification analysis, and the effect of each coordinate of the auto encoder manifold on the brain is estimated.

According to studies [13], 30% of instances of AD can be delayed by early detection and cure of modifiable risk factors[14]. The Lifestyle for Brain Health (LIBRA) index is one method suggested by the Innovative Midlife Intervention for Alzheimer's Deterrence (In-MINDD) project [15], [16]. The three primary types of dementia intervention[17], [18], are cognitive training, increased physical activityandhypertension treatment. AD is the most prevalent variety of the disease (AD). Vascular Alzheimer's (VaD), the another most prevalent form, is followed by Alzheimer's with Lewy bodies. Other forms of AD are linked to alcohol misuse, infections, and brain trauma.

In their work, Tatiq and Barber [19] hypothesised that Alzheimer's disease could be avoided by focusing on vascular risk factors because these two types frequently coexist in the brain and share some vascular risk variables. Williams et al. [20] used four alternative models-Decision Tree, SVM, NN, and Naïve Bayes (NB)to produce predictions of cognitive

performance based on demographic and neuropsychological data. The accuracy of NB was the highest in this situation because average values were used to fill in the gaps left by the missing values.

Menagadevi et al. [21] presented an automated Alzheimer's detection model using MRI datasets. Two datasets were used for experiments: one is ADNI and the other one is from kaggle repository. Image preprocessing was done by curvelet thresholding then image enhancement was done by octagon histogram equalization. The classification was performed by SVM, KNN and extreme learning machine. The results were validated by k-fold validation method with values of k 3, 5 and 10. The SVM classifier provided the best accuracy of 98.21% on ADNI dataset and 99.77% on kaggle dataset.

Prasath et al. [22] proposed an Alzheimer's disease detection model based on deep learning architectures. The MRI images dataset was taken from Kaggle repository. During the preprocessing stage, the images were resized and then image enhancement was done by fusion methods. the LTP features were extracted from the processed fused images and then fed into classifiers. The best 99.5% accuracy was obtained by the proposed Alzheimer's detection system

Shukla et al. [23] presented an Alzheimer's disease detection and diagnosis methodology using machine learning algorithms. The dataset set was gathered from ADNI website and then images were converted from 4D to 2D format. Then clipping, histogram equalization and grayscale conversion were done during the preprocessing stage. The feature extraction and selection were performed by principal component analysis method. The dataset was split into 7:3 for train and test and classification was performed by CNN, XgBoost and RF classifier. The best accuracy of 97.57% was achieved by the proposed method.

3. Materials and Methods

3.1 Dataset

OASIS dataset available on Kaggle is used in experiments [24]. This dataset is a popular collection of clinical and neuroimaging data that is largely targeted towards Alzheimer's disease research. Here's a detailed explanation of each feature:

1. **MRI ID:** This is an identification code assigned to each MRI session. It uniquely identifies each imaging session or scan.
2. **Subject ID:** A unique identification code assigned to each subject in the dataset. This code allows researchers to associate multiple MRI sessions with the same individual.
3. **Age:** The age of the subject at the time of the MRI session. Age can be an important factor in Alzheimer's disease research, as the risk of developing the disease often increases with age.
4. **M/F:** This feature indicates the sex of the subject. It typically uses 'M' for male and 'F' for female.
5. **Hand:** This feature represents the subject's significant hand of use, which can be either right-handed (R), left-handed (L), or ambidextrous (both hands) (B).
6. **EDUC:** The number of years of education received by the subject. Education level can be relevant when studying cognitive function and Alzheimer's disease.
7. **SES (Socio-Economic Status):** SES is a 5-level categorization of the subject's socio-economic status, which reflects their societal class:
    - Level 1: Lower class
    - Level 2: Lower-middle class
    - Level 3: Middle class
    - Level 4: Middle-upper class

- Level 5: Upper class

8. **MMSE (Mini-Mental State Examination) Score:** MMSE is a cognitive assessment score that measures various cognitive abilities. Scores range from 0 to 30, where lower scores indicate a higher likelihood of dementia, and higher scores suggest better cognitive health.
9. **CDR (Clinical Diagnosis Rating):** CDR is a clinical rating that categorizes the subject's cognitive status based on assessments, including MMSE and MRI scans. It has four levels:
    - CDR = 0: Cognitively normal
    - CDR = 0.5: Very mild dementia
    - CDR = 1: Mild dementia
    - CDR = 2: Moderate dementia
10. **eTIV (Estimated Total Intracranial Volume):** This represents the estimated total volume of the brain, often measured in unspecified units.
11. **nWBV (Normalized Whole Brain Volume):** nWBV is a measure of the normalized whole brain volume. It represents the proportion of the brain's total volume relative to the estimated total intracranial volume.
12. **ASF (Atlas Scaling Factor):** ASF is the determinant of an affine transformation matrix applied to brain MRI data points.
13. **Delay:** The interval, measured in days, between the previous and the current MRI session. This can be relevant for tracking changes over time.
14. **Visit:** This feature indicates the ordinal number of the visit to the testing facility for the MRI session. It helps track the sequence of visits for each subject.
15. **Group:** The dementia group to which the subject belongs. It is categorized into three levels:
    - Dementia: Indicates significant dementia.
    - Converted: Denotes subjects who transitioned to a significant dementia state after the initial assessment.
    - Non-Demented: Represents subjects who do not have dementia.

There are three output classes: demented, non-demented, and converted represented by 0, 1, and 2. The distribution of output classes is shown in Figure 1.

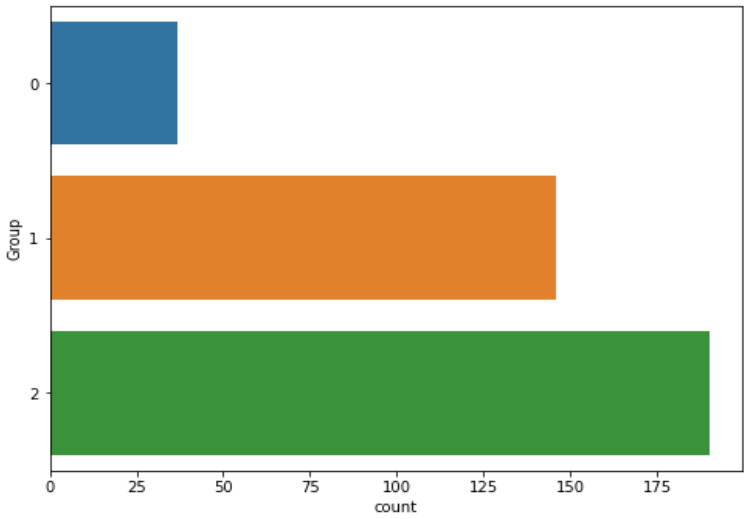

Figure 1: Distribution of Output Classes

Data preprocessing has been done by removing unnecessary attributes: Subject ID, MRI ID, and visit. The correlation between features is shown in Figure 2 and histogram of features is

shown in Figure 3. Analysis of Figure 1 indicates that distribution of output classes is not even. Uneven distribution of classes can adversely affect the performance of the models when trained. Therefore, SMOTE was applied to make the even distribution of classes.

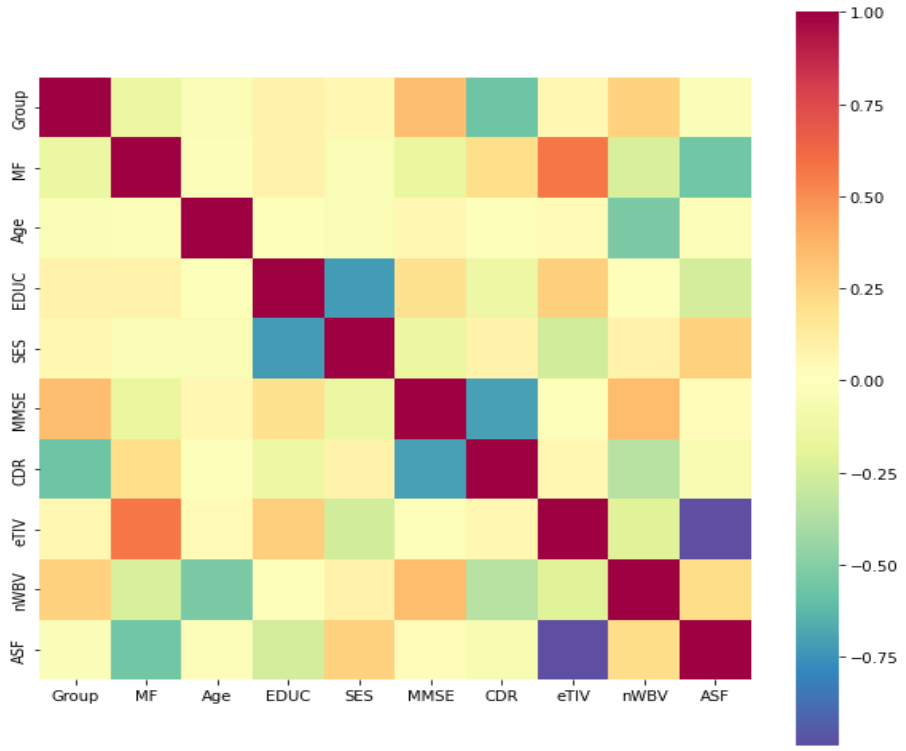

Figure2: Correlation between features

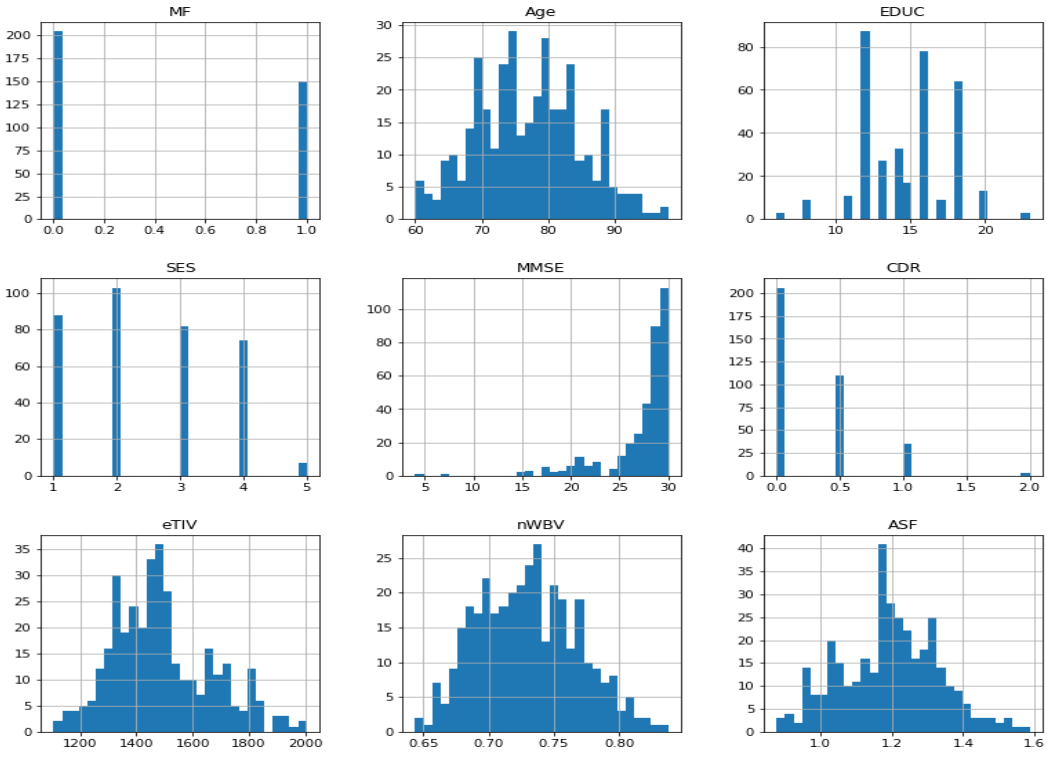

Figure 3: Histogram of features

3.2 Methodology

The proposed methodology for predicting Alzheimer's is shown in Figure 4. The collection of the dataset pertinent to Alzheimer's disease is the first step in the methodology. Data preprocessing is then carried out to make sure the dataset is appropriate for additional analysis. As a part of data preprocessing unnecessary attributes are removed and the balancing of dataset is done. Subject ID, MRI ID, and visit attributes are removed from the dataset. The number of output classes wasn't uniform. Therefore, balancing is done using SMOTE. SMOTE creates synthetic data points for minority classes, resulting in a more accurate representation of all output classes.

The methodology's next step involves applying classification algorithms to data that has been preprocessed in order to produce predictions about Alzheimer's disease. Three classifiers DT, XGB, and RF are used for classification. DT is known for its ability to capture complex decision boundaries in the data. XGBoost is an ensemble learning method having ability to handle complex datasets. RF is good in lowering overfitting and improving the model's predictive capability. In order to produce predictions about the output class, which in this case refers to the likelihood of Alzheimer's disease, trained classifiers are applied to the preprocessed data.

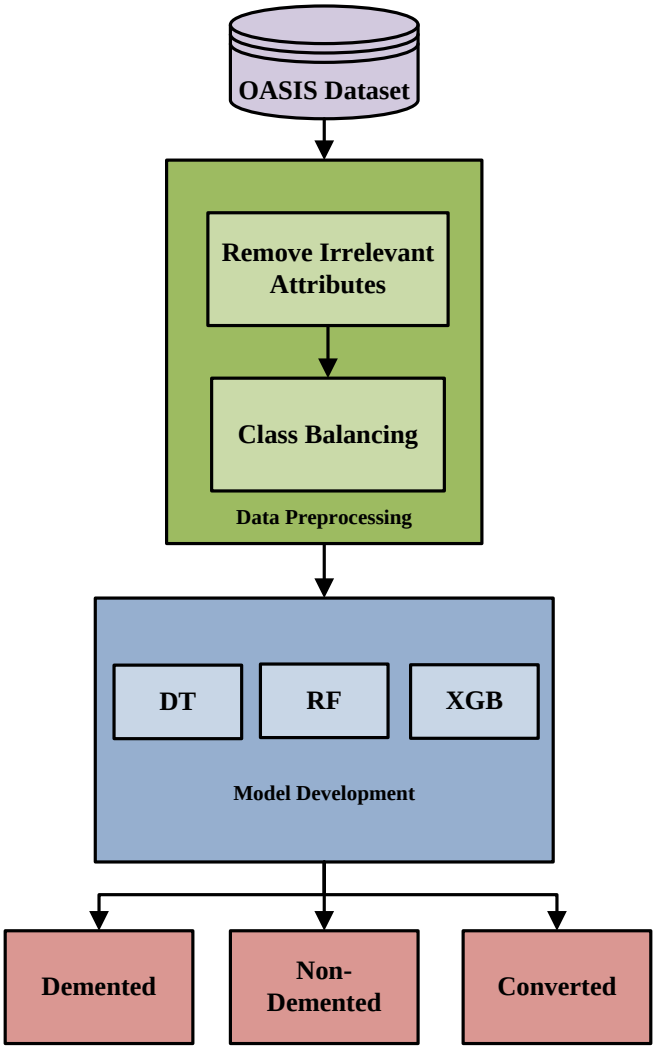

Figure 4: Proposed methodology for predicting Alzheimer'sDisease.

The system's early prediction capability, which is essential for prompt intervention and treatment, is one of its strong points. It lessens the chance of human mistakes in diagnosis and prognosis. This might result in more trustworthy outcomes. The model has some limitations, including the possibility of false positives (forecasting Alzheimer's when it isn't there) and false negatives (failing to detect Alzheimer's when it is present). Patients may experience unneeded stress in case of false positives, and diagnoses may be missed in case of false negatives. The knowledge of healthcare practitioners is complemented, not replaced, by proposed model. To understand the model's predictions and decide on patient care, doctors and specialists are required.

Performance of classifiers is calculated in terms of accuracy, sensitivity, specificity, precision and F-measure.

Accuracy: Calculates the percentage of predictions made by the system that are accurate.

$$\text{Accuracy} = \left(\frac{\text{TP} + \text{FP}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}\right) * 100$$

Sensitivity: Measures the system's accuracy rate for positive predictions.

$$\text{Sensitivity} = \left(\frac{\text{TP}}{\text{FN} + \text{TP}}\right) * 100$$

Specificity: Measures the system's accuracy rate for negative predictions.

$$\text{Specificity} = \left(\frac{\text{TN}}{\text{FP} + \text{TN}}\right) * 100$$

Precision: Calculates the percentage of relevant findings the system produced.

$$\text{Precision} = \left(\frac{\text{TP}}{\text{FP} + \text{TP}}\right) * 100$$

F-Measure: To calculate F-Measure, the harmonic mean of sensitivity and precision is used.

$$\text{F} - \text{Measure} = 2 * \frac{\text{Sensitivity} * \text{Precision}}{\text{Sensitivity} + \text{Precision}}$$

The number of correctly detected negatively classified cases is denoted by the acronym TN, which stands for true negatives. The term "true positives," or "TP," refers to the quantity of correctly identified positive cases. False negatives, or FN, are positive cases that were inadvertently classified as negative. False positives (FP), denoted by the initials FP, are the number of negative cases that were incorrectly classified as positive [25], [26].

3.3 Classifiers

Three Classifiers used in proposed methodology are described in this section. DT has a root node, internal nodes, and leaf nodes. Nodes are connected through branches. At every internal node attribute's value is tested and the test's result is on the branch. Leaf nodes contain the class labels that represent the outcome. A root node is the parent of all other nodes. In DT, features are represented by internal nodes, and rules are represented by branches. Data is arranged in a tree manner, processing a particular outcome at each leaf. It is very easy to gather the data and come up with some insightful insights because decision trees imitate the thinking of humans[27], [28]. The structure of DT is demonstrated in Figure 5.

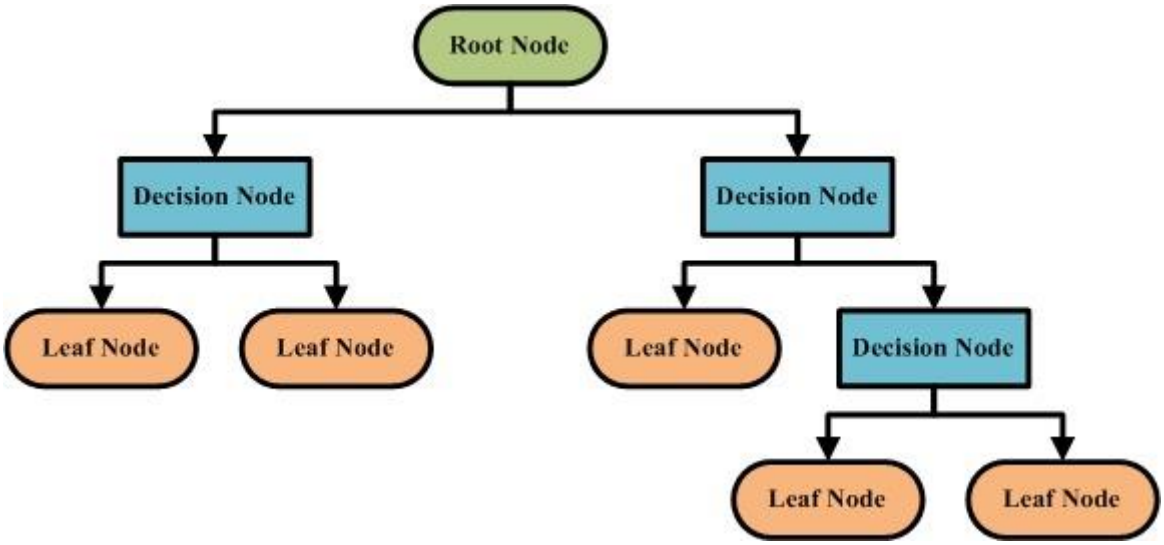

Figure 5: Structure of Decision Tree.

One of the most effective ways to create gradient-boosted decision trees is XGBoost. It is specifically developed to optimize memory utilization. Building successive sub-trees from an original tree to reduce the errors of the preceding one is the core concept behind boosting. The new sub-trees will so update the older ones. In this manner, the new sub-trees will update the earlier residuals to lessen the cost function's error. RF is a set of trees trained using samples obtained from a random resampling of the training set [29], [30]. Bootstrap samples are those produced by randomly resampling the training set. A collection of bootstraps that exclude records from the original dataset is used as the test set once the tree has been formed. Each tree casts a vote for one class, and the forest calculates which class will receive the majority of votes to categorize the input data [31], [32]. Random forest is shown in Figure 6.

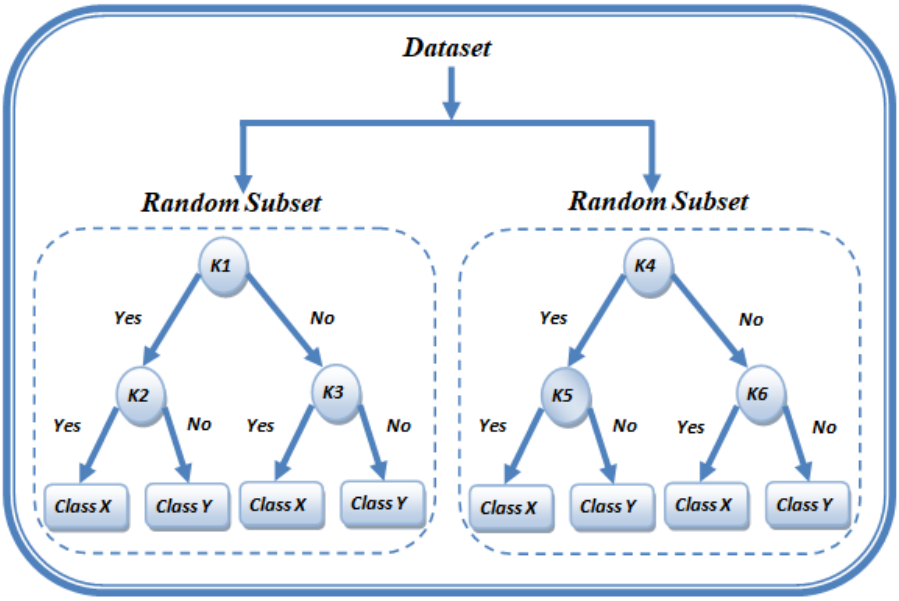

Figure 6: Random Forest

4. Results and Discussion

There are two sets of experiments. Experiments are done using three classifiers on the original dataset and balanced dataset. The performance of classifiers on the original imbalanced dataset is shown in Table 1. DT obtained 73.38% accuracy, XGB obtained 83.88% accuracy and RF

obtained a maximum 87.84% accuracy. DT obtained 84.67% precision, XGB obtained 91.33% precision and RF obtained 95.17% precision.DT obtained 84.67% specificity, XGB obtained 87.21% specificity and RF obtained 87.68% specificity. Very low sensitivity was obtained by all classifiers in imbalanced dataset. DT obtained 19.04% sensitivity whereas XGB and RF obtained only 15% sensitivity. Results in Table 1 indicate that sensitivity is quite low in all classifiers. This was due to imbalanced dataset. Therefore, balancing is further done to improve performance. Performance of classifiers after balancing is shown in Table 2.

Table1: Performance of classifiers on imbalanced dataset

Classifier	Accuracy	Precision	Sensitivity	Specificity	F-Measure
DT	73.38	84.67	19.04	84.67	85.36
RF	87.84	95.17	15.00	87.68	91.32
XGB	83.88	91.33	15.00	87.21	89.23

Table 2: Performance of models on balanced dataset

Classifier	Accuracy	Precision	Sensitivity	Specificity	F-Measure
DT	83.15	85.63	91.71	85.63	88.46
RF	94.03	95.26	93.95	94.27	94.76
XGB	91.05	93.68	93.85	94.17	93.93

Balancing increased performance of all parameters. There was an effective improvement in sensitivity. DT obtained 91.71% sensitivity, XGB obtained 93.85% sensitivity and RF obtained maximum of 95.26% sensitivity. In addition to sensitivity, other performance parameters of classifiers also improved. DT obtained 83.15% accuracy, XGB obtained 91.05% accuracy and RF obtained maximum of 94.03% accuracy. DT obtained 85.63% precision, XGB obtained 93.68% precision and RF obtained 95.26% precision. DT obtained 85.63% specificity, XGB obtained 94.17% specificity and RF obtained 94.27% specificity. DT obtained 88.46% F-measure, XGB obtained 93.93% F-measure and RF obtained 94.76% F-measure. Classifiers' performance comparison before and after balancing is shown in Figures 7 to 11. Analysis of these figures indicates that balancing has significantly improved the performance of all the classifiers and RF has provided the highest value of all the performance parameters.

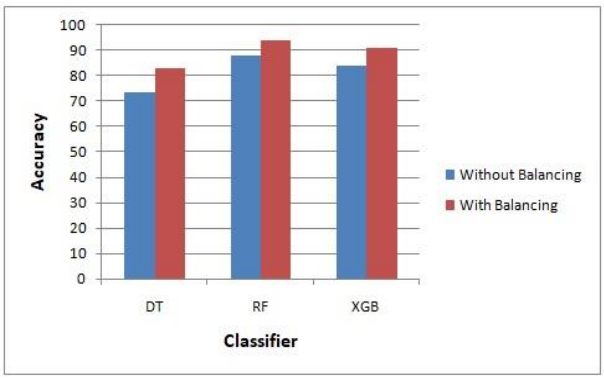

Figure 7: Classifiers' accuracy comparison

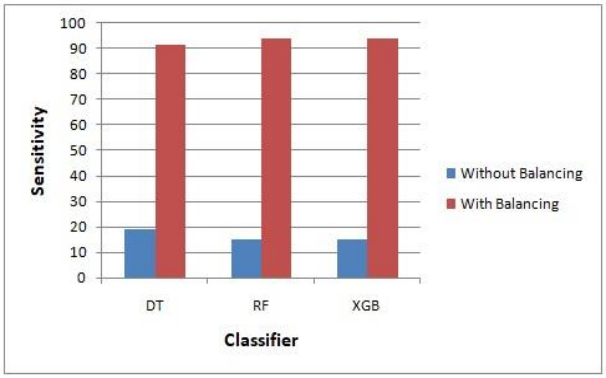

Figure 8: Classifiers' sensitivity comparison

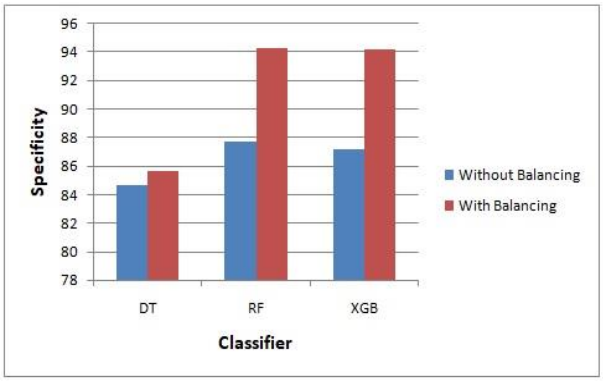

Figure 9: Classifiers' specificity comparison

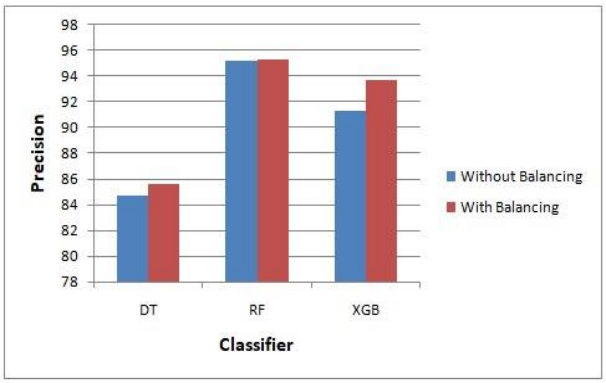

Figure 10: Classifiers' precision comparison

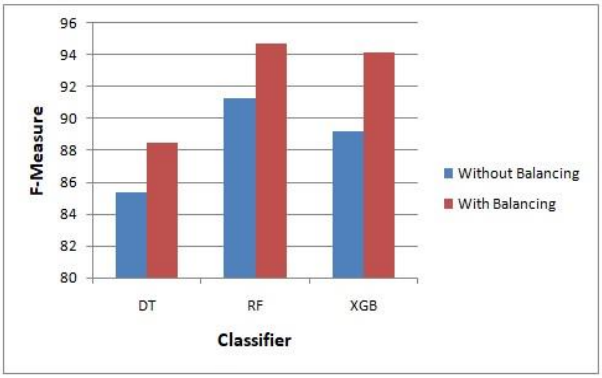

Figure 11: Classifiers' F-measure comparison

Classifier's confusion matrix is also evaluated to determine performance. Confusion matrices show counts of expected and observed values. The anticipated and actual classification done

by classifier is displayed in a confusion matrix of size n x n, where n is the total number of classes. By examining the diagonal values for determining the number of accurate classifications, one might assess the model's accuracy by visualizing the confusion matrix.Classifier's confusion matrix before and after balancing is shown in Figures 12 to 17. There was increase in correct predictions after balancing.

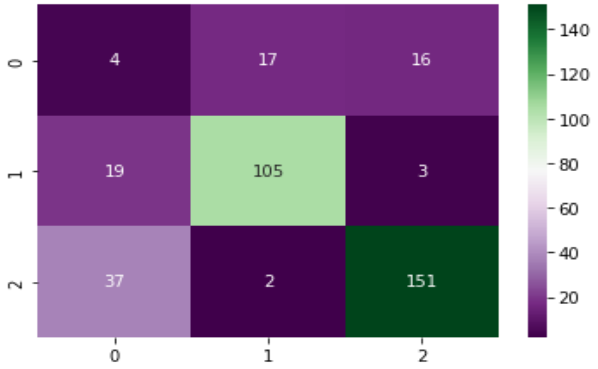

Figure 12: DT confusion matrix without balancing

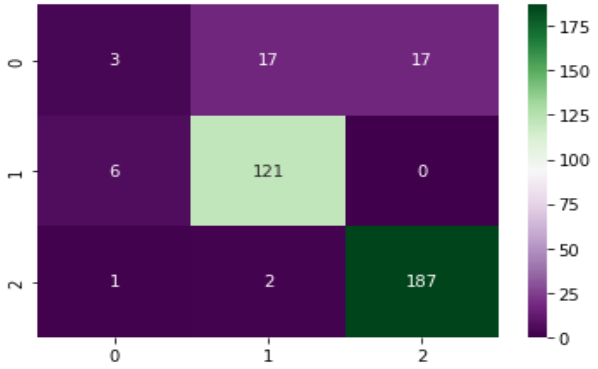

Figure 13: RF confusion matrix without balancing

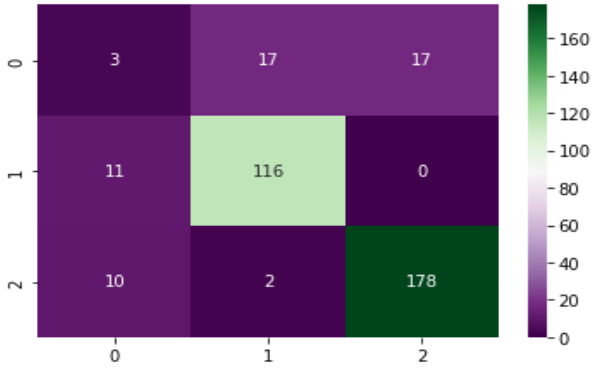

Figure 14: XGB confusion matrix without balancing

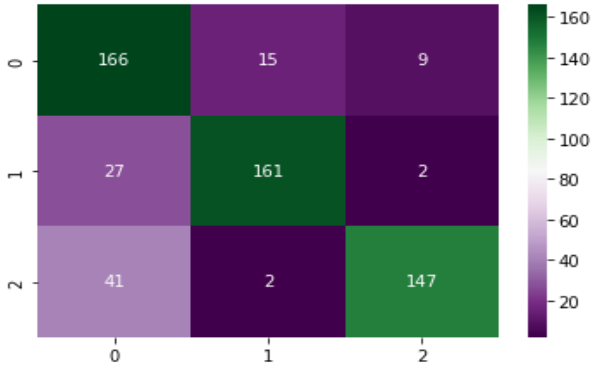

Figure 15: DT confusion matrix with balancing

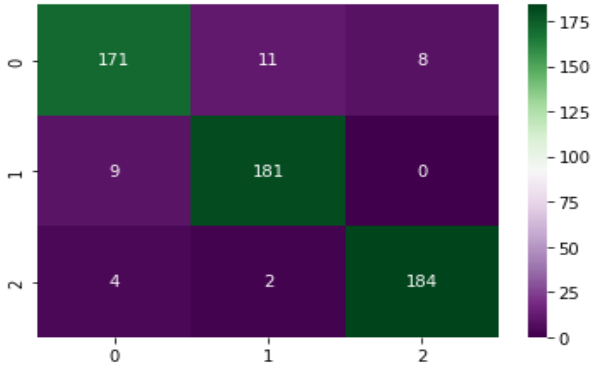

Figure 16: RF confusion matrix with balancing

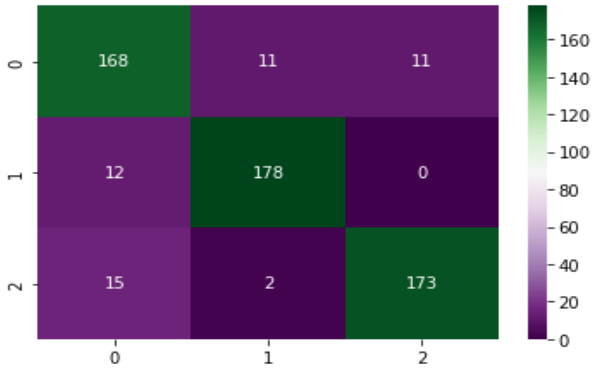

Figure 17: XGB confusion matrix with balancing

Experimental analysis indicated that the RF along with SMOTE provided the best performance. Comparison of work in this paper is paper is also done to recent existing methodologies in Table 3. Existing research used different machine learning methods to categorize Alzheimer's disease using the OASIS dataset. SMOTE-RF provided better results than recent work in literature. The maximum accuracy was reached by the SMOTE-RF, which merged SMOTE and RF, with a 94.03% accuracy, demonstrating its better performance. Improvement in performance with SMOTE-RF as compared to existing methods is shown in Figure 18.

Table 3: Comparison with existing methodologies

Authors	Year	Dataset	Methodology	Accuracy
Antor et al. [1]	2021	OASIS	SVM	92%
Basheer et al. [9]	2021	OASIS	DNN	92.39%
Kavitha et al. [2]	2022	OASIS	RF	86.92%
Proposed methodology	-	OASIS	SMOTE-RF	94.03%

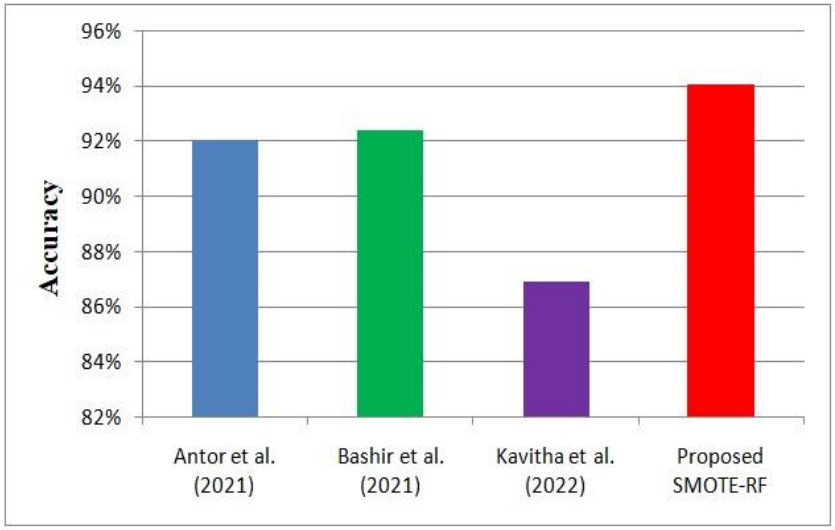

Figure 18: Improvement in performance with SMOTE-RF

This study's limitation is its reliance on pre-existing neuroimaging datasets like OASIS, which by nature have little demographic variety, potential biases, and difficulties generalizing to real-world clinical scenarios. The model's capacity to adjust to varied demographics and data sources may be hampered by its reliance on a single dataset for both training and evaluation. Despite the model's promise in research environments, integrating it seamlessly into clinical practice is a difficult and time-consuming procedure that requires regulatory permissions and validation in clinical settings, which could delay its practical deployment.

5. Conclusion & Future Work

Early diagnosis of Alzheimer's can help in slowing the disease's development. In this paper, DT, RF, and XGB machine learning algorithms are used to predict Alzheimer's. Experiments are performed without balancing and with balancing. Balancing was done using SMOTE. OASIS dataset available on kaggle is used for experiments. Without balancing DT, XGB, and RF obtained 73.38%, 83.88%, and 87.84% accuracy respectively. With balancing DT provided 83.15% accuracy, XGB provided 91.05% accuracy and RF provided 94.03% accuracy.

In future, authors will develop a system with improved accuracy using ensemble methods. Work can also be extended to perform prediction using brain images.

References

[1] H. A. Helaly, M. Badawy, and A. Y. Haikal, “Deep Learning Approach for Early Detection of Alzheimer’s Disease,” *Cognitive Computation*, 2021, doi:

10.1007/s12559-021-09946-2.

[2] M. Bari Antor *et al.*, "A Comparative Analysis of Machine Learning Algorithms to Predict Alzheimer's Disease," *Journal of Healthcare Engineering*, vol. 2021, 2021, doi: 10.1155/2021/9917919.

[3] C. Kavitha, V. Mani, S. R. Srividhya, O. I. Khalaf, and C. A. Tavera Romero, "Early-Stage Alzheimer's Disease Prediction Using Machine Learning Models," *Frontiers in Public Health*, vol. 10, 2022, doi: 10.3389/fpubh.2022.853294.

[4] P. Rani, R. Kumar, and A. Jain, "Coronary artery disease diagnosis using extra tree-support vector machine: ET-SVMRBF," *International Journal of Computer Applications in Technology*, vol. 66, no. 2, 2021, doi: 10.1504/IJCAT.2021.119772.

[5] L. K. Leong and A. A. Abdullah, "Prediction of Alzheimer's disease (AD) Using Machine Learning Techniques with Boruta Algorithm as Feature Selection Method," in *Journal of Physics: Conference Series*, 2019, vol. 1372, no. 1. doi: 10.1088/1742-6596/1372/1/012065.

[6] E. Alickovic and A. Subasi, "Automatic Detection of Alzheimer Disease Based on Histogram and Random Forest," in *CMBEBIH 2019*, 2020, pp. 91–96.

[7] J. Neelaveni and M. S. G. Devasana, "Alzheimer Disease Prediction using Machine Learning Algorithms," 2020. doi: 10.1109/ICACCS48705.2020.9074248.

[8] M. Shahbaz, S. Ali, A. Guergachi, A. Niazi, and A. Umer, "Classification of Alzheimer's disease using machine learning techniques," 2019. doi: 10.5220/0007949902960303.

[9] M. Velazquez and Y. Lee, "Random forest model for feature-based Alzheimer's disease conversion prediction from early mild cognitive impairment subjects," *PLoS ONE*, vol. 16, no. 4 April, 2021, doi: 10.1371/journal.pone.0244773.

[10] S. Basheer, S. Bhatia, and S. B. Sakri, "Computational Modeling of Dementia Prediction Using Deep Neural Network: Analysis on OASIS Dataset," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3066213.

[11] A. Vashishtha, A. K. Acharya, and S. Swain, "A Comparative Study on Various Machine Learning Approaches for the Detection of Alzheimer Disease," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 10, no. 3, pp. 294–304, 2022, [Online]. Available: https://ijisae.org/index.php/IJISAE/article/view/2168

[12] F. J. Martinez-Murcia, A. Ortiz, J. M. Gorriz, J. Ramirez, and D. Castillo-Barnes, "Studying the Manifold Structure of Alzheimer's Disease: A Deep Learning Approach Using Convolutional Autoencoders," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, 2020, doi: 10.1109/JBHI.2019.2914970.

[13] K. Yaffe, "Modifiable risk factors and prevention of dementia what is the latest evidence?," *JAMA Internal Medicine*, vol. 178, no. 2. 2018. doi: 10.1001/jamainternmed.2017.7299.

[14] G. Livingston *et al.*, "Dementia prevention, intervention, and care," *The Lancet*, vol. 390, no. 10113. 2017. doi: 10.1016/S0140-6736(17)31363-6.

[15] K. Deckers *et al.*, "Target risk factors for dementia prevention: A systematic review and Delphi consensus study on the evidence from observational studies," *International Journal of Geriatric Psychiatry*, vol. 30, no. 3. 2015. doi: 10.1002/gps.4245.

[16] K. Aggarwal, M. S. Bhamrah, and H. S. Ryait, *Texture Analysis of Ultrasound Images of Liver Cirrhosis Through New Indexes*, vol. 713. Springer, 2018. doi: 10.1007/978-981-10-4555-4_7.

[17] Y. F. Shea and S. T. DeKosky, "Preventing Cognitive Decline and Dementia," in *Cognitive Changes and the Aging Brain*, 2019. doi: 10.1017/9781108554350.019.

[18] K. Aggarwal, M. S. Bhamrah, and H. S. Ryait, "Detection of cirrhosis through ultrasound imaging by intensity difference technique," *Eurasip Journal on Image and Video Processing*, vol. 2019, no. 1, 2019, doi: 10.1186/s13640-019-0482-z.

[19] S. Tariq and P. A. Barber, "Dementia risk and prevention by targeting modifiable vascular risk factors," *Journal of Neurochemistry*, vol. 144, no. 5. 2018. doi: 10.1111/jnc.14132.

[20] J. A. Williams, A. Weakley, D. J. Cook, and M. Schmitter-Edgecombe, "Machine learning techniques for diagnostic differentiation of mild cognitive impairment and dementia," in *AAAI Workshop - Technical Report*, 2013, vol. WS-13-09.

[21] M. Menagadevi, S. Mangai, N. Madian, and D. Thiyagarajan, "Automated prediction system for Alzheimer detection based on deep residual autoencoder and support vector machine," *Optik*, vol. 272, 2023, doi: 10.1016/j.ijleo.2022.170212.

[22] T. Prasath and V. Sumathi, "Pipelined deep learning architecture for the detection of Alzheimer's disease," *Biomedical Signal Processing and Control*, vol. 87, p. 105442, Jan. 2024, doi: 10.1016/J.BSPC.2023.105442.

[23] G. P. Shukla, S. Kumar, S. K. Pandey, R. Agarwal, N. Varshney, and A. Kumar, "Diagnosis and Detection of Alzheimer's Disease Using Learning Algorithm," *Big Data Mining and Analytics*, vol. 6, no. 4, pp. 504–512, 2023, doi: 10.26599/BDMA.2022.9020049.

[24] "Alzheimers Dataset." https://www.kaggle.com/datasets/jboysen/mri-and-alzheimers?select=oasis_longitudinal.csv (accessed Aug. 10, 2022).

[25] R. Lamba, T. Gulati, and A. Jain, "An Intelligent System for Parkinson s Diagnosis Using Hybrid Feature Selection Approach," *International Journal of Software Innovation*, vol. 10, no. 1, 2022, doi: 10.4018/IJSI.292027.

[26] T. R. Ramesh, U. K. Lilhore, M. Poongodi, S. Simaiya, A. Kaur, and M. Hamdi, "PREDICTIVE ANALYSIS OF HEART DISEASES WITH MACHINE LEARNING APPROACHES," *Malaysian Journal of Computer Science*, vol. 2022, no. Special Issue 1, 2022, doi: 10.22452/mjcs.sp2022no1.10.

[27] R. Lamba, T. Gulati, A. Jain, and P. Rani, "A Speech-Based Hybrid Decision Support System for Early Detection of Parkinson's Disease," *Arabian Journal for Science and Engineering*, 2022, doi: 10.1007/s13369-022-07249-8.

[28] K. Verma *et al.*, "Latest tools for data mining and machine learning," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 9 Special Issue, 2019, doi: 10.35940/ijitee.I1003.0789S19.

[29] R. K. Sachdeva, P. Bathla, P. Rani, V. Kukreja, and R. Ahuja, "A Systematic Method for Breast Cancer Classification using RFE Feature Selection," in *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2022, pp. 1673–1676. doi: 10.1109/ICACITE53722.2022.9823464.

[30] P. Rani, R. Kumar, and A. Jain, “A Novel Hybrid Imputation Method to Predict Missing Values in Medical Datasets,” in *Lecture Notes in Networks and Systems*, 2022, vol. 339. doi: 10.1007/978-981-16-7018-3_16.

[31] R. Lamba, T. Gulati, and A. Jain, “A Hybrid Feature Selection Approach for Parkinson’s Detection Based on Mutual Information Gain and Recursive Feature Elimination,” *Arabian Journal for Science and Engineering*, vol. 47, no. 8, 2022, doi: 10.1007/s13369-021-06544-0.

[32] R. Lamba, T. Gulati, and A. Jain, “Comparative analysis of parkinson’s disease diagnosis system,” *Advances in Mathematics: Scientific Journal*, vol. 9, no. 6, 2020, doi: 10.37418/amsj.9.6.20.