CLASSIFICATION METHODS FOR DETECTION OF DEMENTIA USING MACHINE LEARNING

[1]R. Mohan Reddy, [2]S. Anvitha, [3]Syed Fardeen Ahmed ,[4]J. Ashwitha

[1,2,3,4]Final Year – Department of Computer Science and Engineering(AI&ML)

[5]Dr. G. Srilatha , [5]Associate Professor CSE Dept., Jyothishmathi Institute of Technology and Science Karimnagar, Telangana.

ABSTRACT

According to the World Health Organization (WHO), dementia is "an umbrella term for several diseases affecting memory, and behavior that interferes with a per- son's ability to maintain their daily living activities. It is not a normal part of aging". Dementia is caused by physical changes in the brain. Drugs are available to alleviate some of the symptoms, but they do not cure them. In the existing system, it is differ- cult to identify if a person is suffering from Dementia. It can be only done with the help of clinical history and by knowing if the person has some genetic disorder. Many machine learning algorithms like SVM, KNN were used to overcome the" quantify ability of the stages in dementia" but there was a decline in the accuracy. To overcome the limitations of the existing system, the proposed system was structured in a way to use classification based on the features like Years of Education, Socio-Economic Status, Mini-Mental State Examination, Clinical Dementia Rating, Estimated Total Intracranial Volume, Whole Brain Volume, Atlas Scaling Factor. The purpose of this study was to provide a new clinical tool based on ensemble learning techniques like Random Forest, Adobos and LightGBM which can increase the accuracy of the final outcome.

KEYWORDS: Dementia, SVM, KNN, Random Forest, AdaBoost, Light GBM, CDR.

1. INTRODUCTION

Dementia is the failure of brain function, understanding, recognizing, thinking, and behavioral skills to such a level that an individual faces problem in everyday life and behavior. From the mildest stage, dementia varies in severity andprogressively damages the brain cells.

This detection and diagnosis of the patient can be done faster using machine learning meth- odds like classification based on the information of the patient to be diagnosed. The classification methods like Random Forest, Ada Boost, LightGBM and Boost are used which can increase the accuracy of the final outcome. In addition, the project will investigate theimportance of different features in the detection of dementia. Ultimately, the goal is to develop a robust and accurate classification model for the early detection of dementia, which can assist healthcare professionals in providing early intervention and management for individuals with dementia, ultimately improving their quality of life.

2. LITERATURE SURVEY

The purpose of this paper is to summarize the relation between prodromal and manifest AD and language functions and language domains. Based on the research, they claimed that AD can be more sensitively detected with the help of a linguistic analysis than with other cognitive examinations. The temporal characteristics of spontaneous speech, such as speech tempo, number of pauses in speech, and their length are sensitive detectors of the early stage of the disease, which enables an early simple linguistic screening for AD. However, knowledge about the unique features of the language problems associated with different dementia variants still has to be improved and refined.[1] Alzheimer's is a progressive, irreversible, neurodegenerative brain disease. Even with prominent symptoms, it takes years to notice, decode, and reveal Alzheimer's. However, advancements in technologies, such as imaging techniques, help in early diagnosis. Still, some-times the results are inaccurate, which delays the treatment. Thus, this research focused on identifying the molecular biomarkers that differentiate the genotype and phenotype Chirac-touristic. However, the gene expression dataset's

generated features are huge, 1,000 or even more than 10,000. To overcome such a curse of dimensionality, feature selection techniques are introduced. They designed a gene selection pipeline combining a filter, wrapper, and unsung-previsedmethod to select the relevant genes. They used a Bayesian Optimization technique to tune the hyper parameters in the Improved Deep Belief Network. The tabulated results showed that the proposed pipeline shows promising results.[2] In this study, a multimodal approach to Alzheimer's dementia detection based on the patient's spontaneous speech is presented. This approach was tested on a standard, publicly available Alzheimer's speech dataset for comparability. The data comprise voice samples from156 participants (1:1 ratio of Alzheimer's to control), matched by age and gender. The proposed combination of ADR audio and textual features is capable of successfully modelling temporal aspects of the data. The machine learning approach toward dementia detection achieves best performance when ADR features are combined with strong semantic bag-of-n-grams features. This combination leads to state-of-the-art performance on the AD classification task.[3]

3. PROPOSED METHOD

The detection and diagnosis of the patient can be done faster using machine learning methods like classification, based on the information of the patient to be diagnosed. In this system, the ensemble learning techniques like Random Forest, Adobos and LightGBM are used which can increase the accuracy of a single classifier by combining the classification results from different trained classifiers. It has been demonstrated to increase the performance when processing the imbalance problem.

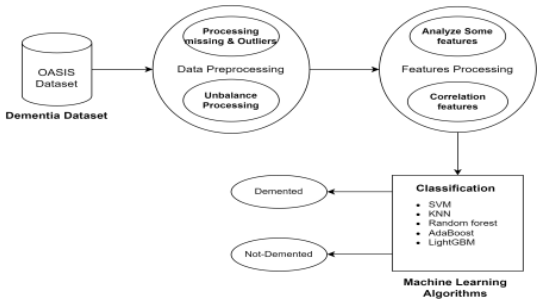

Figure 1: System Architecture

4. MODULES

4.1 RANDOM FOREST

Random Forest is an ensemble learning method that can be used for both classification and regression tasks. It is a type of decision tree algorithm that creates multiple decision trees, also known as" forests", and combines their predictions to make a final decision. The basic idea behind Random Forest is to generate multiple decision trees, each built on a different subset of the data, and then average the predictions of all the trees to make a final prediction. This helps to reduce the overfitting that can occur with single decision trees, and also improves the overall accuracy of the model.

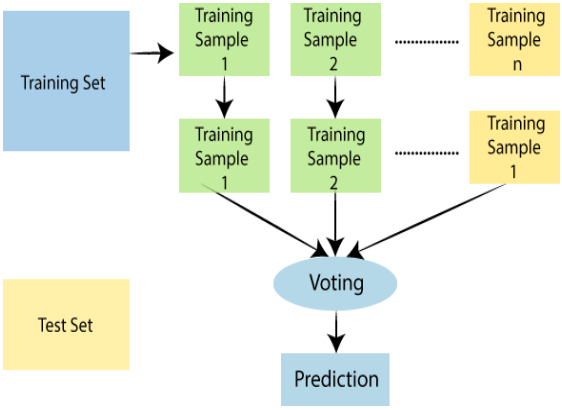

Figure 4.1: Random Forest Algorithm

4.2 ADABOOST ALGORITHM

Adaboost (Adaptive Boosting) is an ensemble learning algorithm that can be used for both classification and regression tasks. It is a type of boosting algorithm that works by com- binning multiple weak learners to create a strong ensemble model. The basic idea behind Adobos is to iteratively train a series of weak classifiers on differ-Ent subsets of the data, and then combine their predictions to make a final decision. In each iteration, the algorithm assigns a weight to each sample in the dataset, with higher weights being assigned to samples that are misclassified by the previous weak classifiers. This helps to focus the training of the next weak classifier onthe samples that are most difficult to classify.

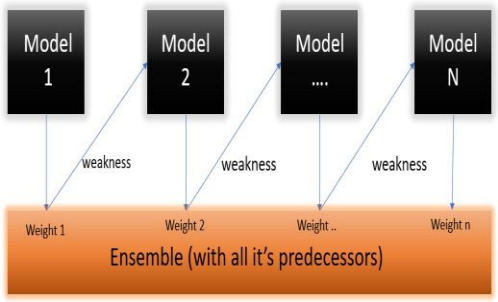

Figure 4.2: Adabost Algorithm

4.3 LIGHTGBM ALGORITHM

LightGBM is a gradient boosting framework that uses tree-based learning algorithms. It is designed to be efficient and scalable, and it is particularly well-suited for large datasets. It grows tree leaf-wise while other algorithms grow level-wise. It will choose the leaf with max delta loss to grow. When growing the same leaf, leaf-wise algorithm can reduce more loss than a level-wise algorithm. If we grow the full tree, best-first (leaf-wise) and depth-first (level-wise) will result in the same tree. The difference is in the order in which the tree is expanded. Since we don't normally grow trees to their full depth, order matters.

Applications of early stopping criteria and pruning methods can result in very different trees. Because leaf-wise chooses splits based on their contribution to the global loss and not just the loss along a particular branch, it often will learn lower-error trees" faster" than level- wise. For a small number of nodes, leaf-wise will probably out-perform level-wise. As we add more nodes, without stopping or pruning they will converge to the same performance because they will literally build the same tree eventually.

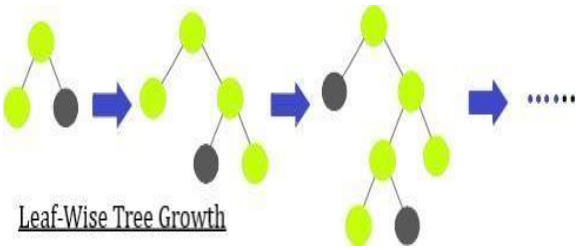

Figure 4.3: LightGBM Algorithm (Leaf-Wise tree growth)

4.4 XGBOOST ALGORITHM

Boost (eXtreme Gradient Boosting) is an open-source library for gradient boosting. It is designed to be efficient and scalable, and it has been widely used in machine learning competitions and industry applications.

XGBoost (Extreme Gradient Boosting) is preferred in various scenarios where its unique characteristics offer significant advantages. It excels in tasks requiring high performance, such as machine learning competitions, where its state-of-the-art results are highly valued. XGBoost's efficiency with large datasets makes it a top choice when computational speed is crucial, as it can handle substantial data volumes efficiently.

Boost uses a technique called gradient boosting to improve the performance of decision trees. It builds multiple decision trees and combines their predictions to make a final prediction. In this algorithm, decision trees are created in sequential form. Weights play an important role in Boost. Weights are assigned to all the independent variables which are then fed into the decision tree which predicts results. Weight of variables predicted wrong by the tree is increased andthese the variables are then fed to the second decision tree. These individual classifiers/predictors then assemble to give a strong and more precise model.

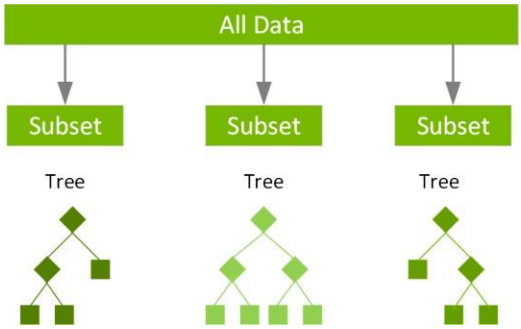

Figure 4.4: Boost Algorithm (Level-Wise tree growth)

RESULT

The results of Detection of Dementia with machine learning Techniques can vary depending on many factors, including different attributes such as age, gender, MMSE, CDR, etc., Overall, Ensemble

models have shown good results Detection of Dementia. Early Diagnose can be done accurately by using all those ensemble methods.

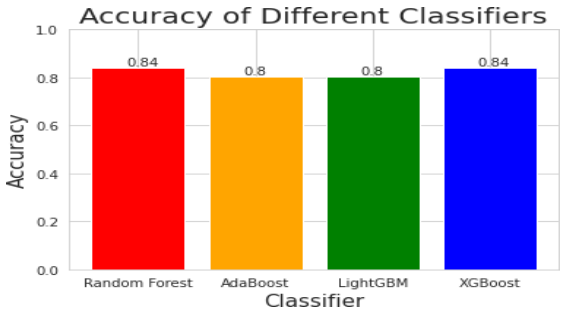

Figure 2: Accuracy Comparison of the Classification Models

CONCLUSION

Dementia cannot be cured as a syndrome that affects memory, but its symptoms can be mitigated. In order to act upon this syndrome, dementia needs to be identified effectively, this is done using Random Forest, Adobos, LightGBM, Boost in the proposed system. Multi-plea classification methods help us to obtain the best accuracy, result and gives a clear idea on which classification method is beneficial with inadequate data and model. Additionally, early detection and treatment of dementia can potentially improve the quality of life for individualswith the disease, allowing them to maintain their independence and autonomy for longer peril-odds of time. It is also important to consider the ethical implications of using these classification methods, particularly in terms of privacy and discrimination. The use of sensitive personal in- formation and the potential for biased algorithms must be carefully evaluated and addressed. Furthermore, it is crucial to ensure that the results of these classification methods are properly interpreted and communicated to patients, caregivers, and healthcare professionals.

Ultimately, the development of a reliable method for detecting dementia has the potential to improve the lives of individuals with the disease and their caregivers, as well as have a positive impact on society as a whole. While there is still much work to be done to fully understand and effectively treat dementia, the project represents an important step forward in the fight against this debilitating condition.

REFERENCES

[1] Julian Appal, Andrew Kerecz, and Michael Cisman. A study of language functioning in Alzheimer patients. *Brain and language*, 17(1):73–91, 1982.

[2] Nivedhitha Mahendran, PM Duran Raj Vincent, Kathiravan Srinivasan, and Chan-Yu Chang. Improving the classification of Alzheimer's disease using hybrid gene selection pipeline and deep learning. *Frontiers in Genetics*, 12, 2021.

[3] Mate Martin, Fish Haier, Sonja Pollack, and Saturnine Luz. Temporal integration of text transcripts and acoustic features for Alzheimer's diagnosis based on spontaneous speech. *Frontiers in Aging Neuroscience*, page 299, 2021.

[4] Robe Sidechain, J David Schaffer, and Stephen a Zaharias. Towards an automatic speech-based diagnostic test for Alzheimer's disease. *Frontiers in Computer Science*, 3:624594, 2021.

[5] www.who.int/news room/factsheets/detail/dementia.

[6] Mate Martin, Fish Haier,Temporal integration of text transcripts and acoustic features for Alzheimer's diagnosis based on spontaneous speech.

[7] Westman, E., Simmons, A., Zhang, Y., Muehlboeck, J. S., Tunnard, C., Liu, Y., ... & Evans, A. (2011). Multivariate analysis of MRI data for Alzheimer's disease, mild cognitive impairment, and healthy controls. Neuroimage.

[8] Liu, S., Liu, S., Cai, W., Che, H., Pujol, S., Kikinis, R., ... & Feng, D. (2014). Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease. IEEE transactions on bio-medical engineering.

[9] Eskildsen, S. F., Coupé, P., García-Lorenzo, D., Fonov, V., Pruessner, J. C., Collins, D. L.,

& Alzheimer's Disease Neuroimaging Initiative. (2013). Prediction of Alzheimer's disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning. Neuroimage.

[10] Khedher, L., Ramírez, J., Górriz, J. M., & Brahim, A. (2015). Alzheimer's disease neuroimaging initiative. Automated diagnosis of Alzheimer's disease based on three- dimensional MRI features using probabilistic neural network. BioMed research international, 2015.