Cognitive Algorithms: Machine Learning's Role in Alzheimer's Early Detection

Lakshmi Sathvika Kurmala
Department of CSE,
SRM University-AP,
sathvika_kurmala@srmap.edu.in

Srujitha Devineni
Department of CSE,
SRM University-AP,
srujitha_devineni@srmap.edu.in

Tanya Kavuru
Department of CSE
SRM University-AP,
tanya_kavuru@srmap.edu.in

Vijaya Vyshnavi Muvvala
Department of CSE,
SRM University-AP,
vyshnavi_vijaya@srmap.edu.in

Srilatha Tokala
ACT Lab, Department of CSE,
SRM University-AP,
srilatha_tokala@srmap.edu.in

Murali Krishna Enduri
ACT lab, Department of CSE,
SRM University-AP,
muralikrishna.e@srmap.edu.in

***Abstract*—Alzheimer's disease is a neurogenerative disorder that produces a particular global healthcare challenge. Early diagnosis is required for effective treatment. This research explores the potential of machine learning and deep learning techniques for predicting Alzheimer's disease. The datasets encompass both numerical data and structural MRI scans, including cognitive test scores and genetic markers from individuals both with and without Alzheimer's. A dataset containing various MRI scans, neuroimaging, and features was used to train and contemplate machine learning models. Numerous engineering and selection techniques were applied to enhance the model's performance. Various classification algorithms were used in the implementation that predict Alzheimer's disease. These models were strictly evaluated using different measures. The results indicate that ML models can effectively predict the disease based on a combination of neuroimaging features. This demonstrates the potential of ML in aiding early Alzheimer's disease diagnosis, which is important for personalized treatment. Future work may involve refining and validating these models and exploring the integration of multi-modal data sources for even more robust predictions.**

Keywords-*Machine learning algorithms; Convolutional Neural Networks; Magnetic Resonance Imaging; Exploratory data analysis*

I. Introduction

Alzheimer's disease is a terrible neurological ailment that has arisen as a global health concern as the world's population continues to age [1]. It is distinguished by a steady decline in cognitive function, which results in memory loss as well as a variety of behavioral and psychological symptoms. The impact of this condition extends beyond individual patients to their families, carers, and healthcare systems [2]. Due to the absence of a cure and the limited effectiveness of available treatments, early diagnosis, intervention, and detection have become more important.

The beginning stages of Alzheimer's disease are frequently obscure, making correct diagnosis even harder. Cognitive tests and medical imaging play crucial roles in diagnosis but may lack the specificity required for early detection. Machine learning techniques are capable of handling complex and multidimensional data, they offer the potential to augment and, in some cases, transfigure Alzheimer's disease diagnosis [3]. The sub-sections of this research will dive into the methodology, dataset, feature engineering, and various machine learning algorithms used in this study. Results and discussions will be present in the performance of these models, along with insights into the importance of specific features [4].

The main purpose of this project is to elaborate on the details of the project we had done, it has several potential uses and applications, which benefit various stakeholders, including healthcare professionals, researchers, and the public. This report precisely focuses on Early detection and diagnosis, Support, Research and Development, Public Awareness and Education, and also to improve caregiver support.

Early detection of this condition is difficult, but it is critical for reducing the disorder's advancement and improving the quality of life for people impacted [5]. ML-based prediction models offer to detect the disease in its early stages. We gathered a large number of datasets which consist of medical records, MRI preprocessed data, MRI scans, genetic information, and other information of the relevant patient. These datasets serve for training predictive models. ML models require careful selection of features that are relevant to Alzheimer's disease prediction [6].

The fundamental purpose of this research is predict the risk of Alzheimer's disease or assess its presence in each individual. This model will use many data sources, which include, medical records, MRI scans, survey data, neuroimaging data, genetic information, and required data types. The reason for early detection is it leads to more effective interventions, improved patient care, and a better healthcare system. The development of reliable predictive models can impact the lives of individuals who are at risk of or currently living with Alzheimer’s disease.

The urgent need to address the issues associated with this neurogenerative ailment necessitates the development of a predictive model of Alzheimer's disease prediction using machine learning [7]. This research study is important because as millions of individuals are affected by this disorder, this disease is expected to rise, placing an increase of burden in healthcare systems and in society as well. This disorder currently has no cure expect of few medications.

Early prediction can help the disorder to slow down its progression, and also improves the quality of life for affected individuals, and probably we can delay the onset of severity of this disorder [8]. With the present health records, neuroimaging technologies, and genetic testing, there is huge amount of data tat is available for research. ML offers a way to evaluate this data for predictive modeling.

Predictive models can help in the development of personalized treatments for individuals who are victims of Alzheimer’s disease [9]. This type of care can lead to more effective interventions and best outcomes. By identifying and predicting the risk outcomes as early as possible can reduce the long-term cost of medical facilities. As the field of ML and AI is evolving day by day and is becoming advanced, every researcher is motivated by the opportunity to apply cutting- edge technology to a crucial medical problem, which leads to breakthroughs in diagnosis and treatment. Precisely the main motivation is to improve the quality of individuals life which is at risk of or affected by Alzheimer’s disease. Early detection and prediction can lead to a greater level of independence and a better standard of living for affected patients and their families [10].

Several common knowledge patterns and observations have emerged from research and clinical practice in Alzheimer's illness diagnosis using machine learning techniques. These patterns provide insights into the disease and guide the development of predictive models. Some of the common knowledge patternsobserved include:

- Progressive Cognitive Decline
- Age as a Significant Risk Factor
- Neuroimaging Abnormalities
- Longitudinal Data
- Clinical and Medical History
- Ethical Considerations

II. Related Work

Our research adventure started with an intensive evaluation of academic papers, aiming to pinpoint an most useful research consciousness for our Alzheimer’s prediction study. Our primary hobby lay in the application of numerous machine mastering algorithms and picture processing techniques, seeking a challenge that could be intellectually stimulating and nearly possible. After thorough attention, we decided to middle our research on predicting Alzheimer’s chance using those techniques.

The research papers we examined provided precious insights that enriched our understanding of the diverse strategies hired in Alzheimer’s prediction through system gaining knowledge of and photograph processing techniques. These readings illuminated the strengths and weaknesses of different approaches, presenting important perspectives at the hurdles faced whilst utilizing these methods for Alzheimer’s threat evaluation. They additionally guided us in formulating our studies problem: the improvement of a new, extra accurate, and efficient approach for Alzheimer’s hazard prediction, surpassing the capabilities of current methods.

Kavitha *et al.* used various ML techniques like Random Forest, SVM, XG Boost, Max Voting, and Decision tree classifiers for predicting AD vs normal subjects [11]. For this, they used a dataset consisting of 150 patients’ MRI data. They used various data mining techniques to pre-process data for adjusting missing values and extracting features. After analysis of the data they finally took parameters like EDUC, MMSE, SES, CDR, Etiv, nWBv, and ASF into consideration. They measured the performance using accuracy, precision, recall, and F1 score. Out of all the used ML techniques Random Forest and XG boost received higher accuracies and precision. XG Boost and voting received higher recall values and F1 scores. Masurkar *et al.* tried to find Alzheimer’s disease at an early stage. This stage is known as the MCT stage which results in memory loss [12]. By predicting the disease at this stage it helps to decrease the risk to life. For this, they used Machine learning algorithms like SVM and Decision Trees. For predicting the disease they took parameters like MMSE score, no. of visits, and education of patients into consideration out of all other parameters available in the dataset. By using the SVM algorithm they attained an accuracy of 85percent and by using Decision trees they attained an accuracy of 83%. A novel method employing 3D deep convolutional neural networks to effectively differentiate mild Alzheimer’s from cognitively normal and cognitively impaired people is created by Neelavani *et al.* [13].

For this, they used structural MRI and data sets were from ADNI and NACC. Based on the sizes and thickness of formerly identified brain regions that are known to be involved in the course of disease, we have created a reference model for comparison [14]. When classifying individuals with moderate Alzheimer’s dementia and mild cognitive impair- ment from those with cognitively normal subjects, the deep learning model produced an AUC of 85.12. This indicates that the model is accurate. It obtains an AUC of 62.45 in the more difficult task of identifying MCI. In comparison to the reference model built that is volume/thickness model, which requires the volumes and thickness to be retrieved beforehand, it is also much faster. The paper discussed the software of system mastering fashions to expect dementia in patients using longitudinal Magnetic Resonance Imaging (MRI) statistics from the OASIS venture [15]. Various device mastering fashions, such as SVM, Logistic Regression, Decision Tree, and Random Forest, were employed to make predictions. Before great tuning, those fashions performed exclusive ranges of accuracy, with SVM displaying the highest trying out accuracy of 74 percent. After satisfactory tuning, SVM outperformed other fashions, accomplishing 92 percent accuracy. However, some overfitting troubles had been mentioned inside the Decision Tree and Random Forest models.

They also supplied insights into facts preprocessing, consisting of filling missing values and function scaling, and using assessment metrics like accuracy, remember, AUC, and the confusion matrix for version overall performance evaluation. It highlights that SVM is the satisfactory appearing model across all metrics and recommends using a bigger dataset and additional machine studying fashions, inclusive of AdaBoost, KNN, Majority Voting, and Bagging, for further development [16]. The gadget goals to offer the public with a tool for assessing the chance of dementia in person patients via inputting MRI information.

In precise, the paper centered on the improvement and assessment of gadget studying fashions to expect dementia the use of MRI records. It demonstrates that SVM is the most accurate model, with 92 percent accuracy after first-rate- tuning, and emphasizes the ability of device studying in early dementia detection and public use.

III. Methodology

A major global health concern is Alzheimer’s disease, for which early detection is essential to successful treatment [17]. The objective of this research is to enhance the accuracy and robustness of the identification by applying a dual-modality strategy that makes use of both numerical and MRI image datasets. Several classification techniques were used on the numerical dataset to find patterns indicating of the disease. Cross-validation was then used to carefully evaluate the models’ reliability. The study also included an MRI image dataset, from which relevant characteristics were extracted using image processing techniques [18]. This pre-processed MRI dataset was then subjected to deep learning algorithms, opening up a new method for Alzheimer’s diagnosis. In the upcoming sections, we’ll explain in detail how we selected our data, the algorithms we used, and the steps we took to ensure the study was conducted ethically and fairly [19]. The overall methodology that is used in our analysis and the procedure that we have followed is shown in Fig.1.

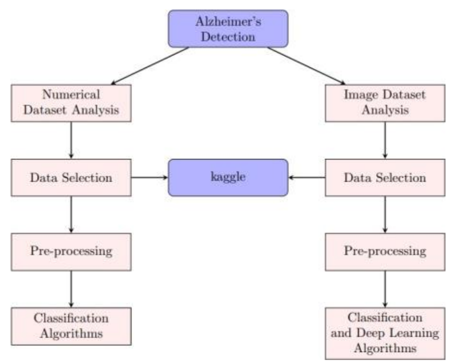

Fig. 1. Visualizing the steps behind the model.

A. For Numerical Dataset

We used the "oasis longitudinal.csv" file from Kaggle as the source of the numerical dataset, which focused on labels related to Alzheimer's disease and important characteristics like demographics, cognitive scores, and brain attributes. During the preprocessing stage, "Converted" and "Demented" were combined to standardize the groups, and median imputing was used to address missing values in columns like "SES" and "MMSE." Insights into gender-specific dementia patterns and the correlation matrix, which explains correlations between numerical variables, were provided via exploratory data analysis (EDA) as shown in Fig.2. Predictive accuracy ranged from 56% to 84% when different classification algorithms and were implemented.

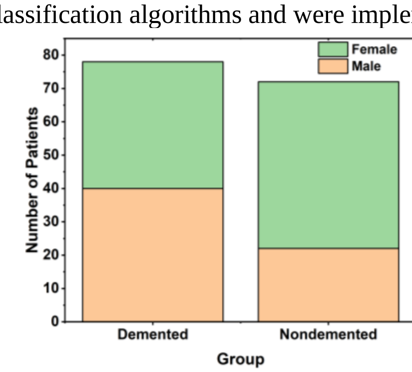

Fig.2: Gender vs Demented Graph

B. For MRI Image Dataset

The preprocessed images for the MRI image dataset were carefully chosen from Kaggle and classified according to the phases of Alzheimer's disease. Pixel values were normalized, photos were scaled for consistency, and data augmentation techniques were used to vary the dataset throughout the preprocessing stages. The VGG16 model's feature extraction improved the dataset's depth. When advanced classification methods were used, accuracies ranging from 66.56% to 98.91% were attained. This methodical approach, which combines imaging and numerical data with a range of algorithms, adds to our comprehensive understanding of Alzheimer's disease and opens the door to new discoveries and possible developments in diagnosis and research.

C. Classification Of Algorithms

For the analysis at this level, we used a variety of machine learning and deep learning methods. The following is a simple explanation.

When logistic regression was used on a single numerical dataset, its accuracy was 64%; however, when applied to the MRI dataset, it was 95.16%. It works well in both cases due to its adaptability when managing linear relationships, but with different levels of accuracy. Gaussian Naive Bayes showed its flexibility to a variety of numeric properties in the numerical dataset, achieving an accuracy of 73.33%. Its performance was unrated in the MRI dataset. By utilising closeness in the feature space, KNN was able to attain an accuracy of 73.33% for the numerical dataset. KNN performed exceptionally well in the MRI dataset, achieving an astounding accuracy of 98.91%, highlighting its effectiveness in image-based classifications. Random Forest attained an accuracy of 78.67% in the numerical dataset, demonstrating its capacity to grasp complicated relationships [20].

Using a linear kernel, SVM achieved an accuracy of 78.67% for the numerical dataset and 96.25% for the MRI dataset. Its adaptability to a variety of traits is visible in both cases. AdaBoost obtained an accuracy of 80% in the numerical dataset, indicating its capacity to repair errors and increase overall predictive performance. XGBoost achieved an amazing accuracy of 84% on the numerical dataset. It demonstrated solid performance in the MRI dataset, with an accuracy of 87.5%, emphasizing its versatility to both feature types. When applied to the MRI dataset, the CNN attained an accuracy of 87.5%, demonstrating its ability to detect complicated patterns in visual data. The ANN attained an accuracy of 80.86% in the numerical dataset, demonstrating its capacity to recognize intricate patterns in numerical features.

These algorithmic evaluations shed light on how each technique performs when applied to particular numerical and MRI datasets, demonstrating their strengths and adaptability to different feature types.

IV. Result Analysis

A. Dataset Statistics

We acquired both the datasets from Kaggle. The details of both the datasets are explained below. The numerical dataset consists of the “oasis_longitudinal.csv” file, we noticed that some important columns namely, “SES” (Socioeconomic Status) and “MMSE” (Mini-Mental State Examination) that had missing values. We chose a calculated approach to deal with these missing values in order to improve the robustness of our models i.e., to replace the missing values in these columns with the corresponding medians. This method of imputation helps in keeping the integrity of the dataset by decreasing the effect of missing data on future analyses. We observed that the MRI Image dataset contains the following figures: “Moderate_Demented”, “Mild_ Demented”, “Very_ Mild_ Demented”, and “Non_Demented”. Images that are exclusive to their respective categories are gathered in these directories to guarantee a fair representation. ‘Moderate_Demented’ comprises 64 images, ‘Mild_Demented’ 896 images, ‘Very_Mild_Demented’ 2240 images, and ‘Non_Demented’ 3200 images. Our deep learning and classification algorithms are built on this careful selection, which guarantees that they are trained on a wide range of visual data.

Following the training phase, an in-depth evaluation was carried out to determine how well the developed Alzheimer’s disease classification model was performed using data that had never been seen before. To ensure the model’s generalizability, real-world scenarios were simulated using a different test set from the training data. A variety of assessment metrics were used to give a thorough grasp of the model’s advantages and disadvantages.

1) ***Accuracy:*** In the context of Alzheimer's detection, accuracy represents the total correctness of the model's predictions. High accuracy suggests that the model can generate accurate predictions at various stages of the disease.
2) ***Precision:*** Precision evaluates how many of the projected positive cases are actually true positive cases in Alzheimer's detection. The model's high precision implies that when it predicts Alzheimer's, it is often correct, minimizing false positives.
3) ***Recall:*** Recall evaluates how well the model recognizes real positive cases among all actual positive instances in Alzheimer's detection. The model's high recall shows that it effectively detects

people with Alzheimer's disease while minimizing false negatives.

4) ***F1-Score:*** The F1 score, which is the harmonic mean of precision and recall, provides a balanced evaluation of the model's accuracy in Alzheimer's disease detection, taking into account both false positives and false negatives. A high F1 score suggests a model that performs well in terms of precision and recall while avoiding false positives and false negatives.

5) ***Confusion Matrix:*** It provides a detailed summary of the model's predictions, emphasising true positives, false positives, and false negatives. This matrix aids in determining the precise locations where the model flourishes or need improvement.

6) ***Receiver Operating Characteristic (ROC) Curve:*** In binary or multiclass classification settings, the ROC curve represents the trade-off between a true positive rate and a false positive rate. It aids in visualizing the model's discriminatory capacity across various thresholds.

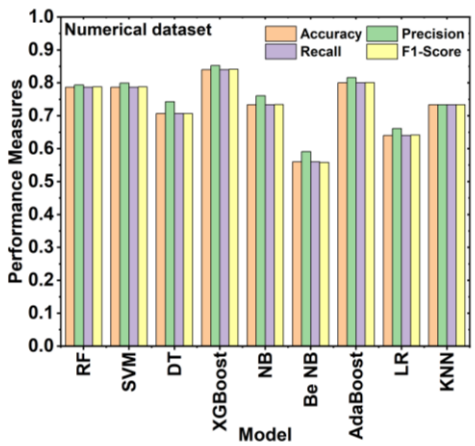

Fig.3: Performance measures of different models performed on the numerical dataset.

Fig.3 provides a brief analysis of different models with their performance measures on the numerical dataset. It is observed in the dataset that for the accuracy measure, it is seen that the XGBoost model gives better than all the different models, the lowest accuracy is provided by the Bernoulli Naive Bayes method. For the precision measure, it is seen that a better score is produced by the XGBoost model, and the lowest score is produced by the Bernoulli Naïve Bayes method. For the recall measure, it is seen that the XGBoost model and the lowest score is by the Bernoulli Naïve Bayes method. And finally by the F1-Score measure, better performance is shown by the XGBoost model and the least performance is shown by the Bernoulli Naive Bayes model. From the overall analysis of the numerical dataset, it is observed that XGBoost model is give better performance and the Bernoulli Naive Bayes model gives the least performance than the other models.

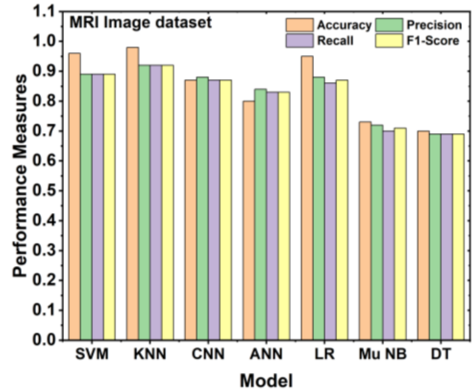

Fig.4: Performance measures of different models performed on the MRI image dataset.

Fig.4 describes the performance measures of different models that are used for analyzing the MRI image dataset. It is observed in the dataset that, compared to all the models for the accuracy measure SVM, KNN, and logistic regression models gives better performance. The lowest performance for the accuracy measure is given by the decision tree. For the precision measure, better performance is given by the KNN method, and the lowest score is given by the multinomial Naive Bayes model. In the recall measure, a better score is given by the KNN model, and the lowest score is given by the decision tree model. For the F1-score model, it is seen that the KNN model gives better performance, and the lowest score is given by the multinomial Naïve Bayes model. From the overall analysis of the MRI image dataset, the KNN model gives a better performance than all the other models for the dataset.

V. CONCLUSION AND FUTURE WORK

In concluding our study on Alzheimer's disease identification, our innovative approach merges MRI and numerical datasets with diverse machine learning and deep learning algorithms. Our in-depth analysis of the numerical dataset highlighted the effectiveness of XG Boost, achieving an outstanding 84% accuracy, particularly notable when incorporating cross-validation. The MRI image dataset exhibited remarkable accuracies, notably with CNN (87%), KNN (98%), and SVM (96.25%), showcasing the capability to accurately diagnose Alzheimer's using MRI images. Significantly, our model distinguishes itself by presenting a comprehensive diagnostic methodology, seamlessly integrating numerical and image data. The meticulous application of cross-validation ensures the reliability and consistency of our findings. In essence, our study not only propels advancements in Alzheimer's detection techniques but positions our model as a promising breakthrough that surpasses certain existing research models. It provides a nuanced perspective for swift and precise diagnosis, emphasizing the potential of integrated datasets and sophisticated algorithms for enhancing healthcareoutcomes.

In light of the study, future research could focus on improving the numerical dataset pre-processing procedures even more. A deeper understanding might be gained by investigating sophisticated feature engineering methods unique to Alzheimer's symptoms. Further research into the effects of unbalanced datasets on model performance may open the door to more focused enhancements. Practically speaking, working with healthcare organizations could make it easier to integrate patient data from the real world, overcoming issues with differentiating data quality and demographics. It might also be investigated how federated learning techniques might be used to enable models to learn from decentralized data sources while maintaining privacy. To further improve the precision, understanding, and relevance of Alzheimer's dis- ease detection, a comprehensive strategy comprising domain- specific improvements, cutting-edge technology integration, and ethical considerations is outlined in the future trajectory.

VI. REFERENCES

[1]. Kapasi, A., & Schneider, J. A. (2016). Vascular contributions to cognitive impairment, clinical Alzheimer's disease, and dementia in older persons. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1862*(5), 878-886.

[2]. Porsteinsson, A. P., Isaacson, R. S., Knox, S., Sabbagh, M. N., & Rubino, I. (2021). Diagnosis of early Alzheimer's disease: clinical practice in 2021. *The journal of prevention of Alzheimer's disease*, *8*, 371-386.

[3]. Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C. E., ... & van der Flier, W. M. (2021). Alzheimer's disease. *The Lancet*, *397*(10284), 1577-1590.

[4]. Blennow, K., de Leon, M. J., & Zetterberg, H. (2006). Alzheimer's disease. *The Lancet*, *368*(9533), 387-403.

[5]. Battineni, G., Chintalapudi, N., Amenta, F., & Traini, E. (2020). A comprehensive machine-learning model applied to magnetic resonance imaging (mri) to predict alzheimer's disease (ad) in older subjects. *Journal of Clinical Medicine*, *9*(7), 2146.

[6]. Ballard, C., Gauthier, S., Corbett, A., Brayne, C., Aarsland, D., & Jones, E. (2011). Alzheimer's disease. *the Lancet*, *377*(9770), 1019-1031.

[7]. Masters, C. L., Bateman, R., Blennow, K., Rowe, C. C., Sperling, R. A., & Cummings, J. L. (2015). Alzheimer's isease. *Nature reviews disease primers*, *1*(1), 1-18.

[8]. Paulsen, J. S., Nance, M., Kim, J. I., Carlozzi, N. E., Panegyres, P. K., Erwin, C., ... & Williams, J. K. (2013). A review of quality of life after predictive testing for and earlier identification of neurodegenerative diseases. *Progress in neurobiology*, *110*, 2-28.

[9]. Mittelman, M. S., Ferris, S. H., Steinberg, G., Shulman, E., Mackell, J. A., Ambinder, A., & Cohen, J. (1993). An intervention that delays institutionalization of Alzheimer's disease patients: treatment of spouse-caregivers. *The Gerontologist*, *33*(6), 730-740.

[10]. Ajagbe, S. A., Amuda, K. A., Oladipupo, M. A., Oluwaseyi, F. A., & Okesola, K. I. (2021). Multi-classification of Alzheimer disease on magnetic resonance images (MRI) using deep convolutional neural network (DCNN) approaches. *International Journal of Advanced Computer Research*, *11*(53), 51.

[11]. Kavitha, C., Mani, V., Srividhya, S. R., Khalaf, O. I., & Tavera Romero, C. A. (2022). Early-stage Alzheimer's disease prediction using machine learning models. Frontiers in public health, 10, 853294.

[12]. Liu, S., Masurkar, A. V., Rusinek, H., Chen, J., Zhang, B., Zhu, W., . . . Razavian, N. (2022). Generalizable deep learning model for early Alzheimer's disease detection from structural MRIs. Scientific reports, 12(1), 17106.

[13]. Neelaveni, J., & Devasana, M. G. (2020, March). Alzheimer disease prediction using machine learning algorithms. In *2020 6th international conference on advanced computing and communication systems (ICACCS)* (pp. 101-104). IEEE.

[14]. Korolev, I. O., Symonds, L. L., Bozoki, A. C., & Alzheimer's Disease Neuroimaging Initiative. (2016). Predicting progression from mild cognitive impairment to Alzheimer's dementia using clinical, MRI, and plasma biomarkers via probabilistic pattern classification. *PloS one*, *11*(2), e0138866.

[15]. Bari Antor, M., Jamil, A. H. M., Mamtaz, M., Monirujjaman Khan, M., Aljahdali, S., Kaur, M., ... & Masud, M. (2021). A comparative analysis of machine learning algorithms to predict alzheimer's disease. *Journal of Healthcare Engineering*, *2021*.

[16]. Kato, S., Homma, A., Sakuma, T., & Nakamura, M. (2015, August). Detection of mild Alzheimer's disease and mild cognitive impairment from elderly speech: Binary discrimination using logistic regression. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 5569-5572). IEEE.

[17]. Wang, J., Huang, S., Wang, Z., Huang, D., Qin, J., Wang, H., ... & Liang, Y. (2023). A calibrated SVM based on weighted smooth GL1/2 for Alzheimer's disease prediction. *Computers in Biology and Medicine*, *158*, 106752.

[18]. Porsteinsson, A. P., Isaacson, R. S., Knox, S., Sabbagh, M. N., & Rubino, I. (2021). Diagnosis of early Alzheimer's disease: clinical practice in 2021. *The journal of prevention of Alzheimer's disease*, *8*, 371-386.

[19]. Seixas, F. L., Zadrozny, B., Laks, J., Conci, A., & Saade, D. C. M. (2014). A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer′ s disease and mild cognitive impairment. *Computers in biology and medicine*, *51*, 140-158.

[20]. Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C. R., Jagust, W., ... & Beckett, L. (2005). Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimer's & Dementia*, *1*(1), 55-66.