Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 5 (2023)

Analyzing Dementia Prediction Models with Deep Neural Networks on OASIS Data

[1]Hemanth Kumar H S, [2]Tanuja R, [3]S H Manjula, [4]Venugopal K R

[1] Research Scholar, Bangalore University, Bengaluru
[2][3][4] Department of Computer Science and Engineering
University Visvesvaraya College of Engineering, Bengaluru, India.

Abstract— This study presents a comprehensive investigation into the early prediction of dementia using longitudinal data, focusing on the evaluation of various machine learning classifiers. In the methodology, we meticulously prepare and preprocess the data, including data visualization, imputation, transformation, and feature selection. Subsequently, we apply an array of hyper-parametric classifiers, including logistic regression, linear discriminant analysis, k-nearest neighbors, decision trees, naive Bayes, support vector machines, ensemble methods like random forest and XGBoost, among others. The performance of these classifiers is rigorously assessed, with random forest and XGBoost emerging as the top performers, achieving accuracy rates.

Index Terms — Dementia Prediction; Longitudinal Data Analysis; Machine Learning Classifiers; OASIS Dataset; Early Detection

1. INTRODUCTION

Dementia is a debilitating condition characterized by cognitive decline and memory loss, severely impacting an individual's daily life and functioning. Alzheimer's disease, one of the most common forms of dementia, accounts for a significant portion of dementia cases. It is a progressive neurodegenerative disorder with no known cure, making early detection and intervention crucial for better patient outcomes. The causes of dementia are multifaceted, with age being a significant risk factor. Other contributing factors include genetics, lifestyle, and underlying medical conditions. AD, specifically, is associated with the accumulation of abnormal protein deposits in the brain, leading to the death of brain cells and cognitive impairment. Its effects are profound, affecting memory, reasoning, language, and the ability to perform daily tasks independently [1]. Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) play pivotal roles in dementia research and diagnosis[2] .In the context of dementia, MRI is used to examine the brain's structural changes, such as atrophy or the presence of lesions, which are common indicators of various dementia subtypes, including Alzheimer's disease. For instance, the OASIS dataset employs MRI to study brain changes in dementia patients. On the other hand, PET scans are employed to assess brain function and metabolic activity, aiding in the early detection of dementia-related abnormalities, such as reduced glucose metabolism, often seen in Alzheimer's disease.

Motivation: The motivation behind this research is driven by the critical need for early dementia detection. Timely identification can enable healthcare professionals to initiate interventions and support for affected individuals, potentially slowing the progression of the disease and improving their quality of life. Additionally, early detection aids in patient care planning and resource allocation within the healthcare system. The paper's exploration of advanced algorithms and datasets aims to contribute to the ongoing efforts in dementia prediction, bringing us closer to effective early diagnosis and intervention strategies.

2. LITERATURE REVIEW

Kuo et al.,[6] focused on predicting dementia using MRI data, employing a range of ensemble methods including Random Forest, AdaBoost, and Gradient Boosting. Among these algorithms, Random Forest emerged as the most successful, achieving an impressive accuracy rate of 92.67\% in dementia prediction based on MRI data. This outcome underscores the effectiveness of ensemble techniques in harnessing the power of multiple

models to make highly accurate predictions in the field of dementia diagnosis, showcasing their potential for enhancing early detection and intervention strategies. Wang et al., [7] explained the primary objective used to predict cognitive decline through the utilization of longitudinal latent variable models and Cox proportional hazards models. The study resulted in the development of a robust predictive model specifically designed to anticipate cognitive decline. This model exhibited the capability to forecast changes in cognitive status over time, providing valuable insights into the complex task of modelling cognitive trajectories for the purpose of dementia prediction. This work contributes to the growing body of knowledge in understanding the progression of cognitive disorders and lays the foundation for more effective early intervention strategies.

Merkin et al.,[8] aimed to assess the performance of multiple machine learning algorithms in dementia prediction. They conducted extensive analyses using ten different machine learning algorithms on two distinct datasets: the Sydney Memory and Age Study (MAS) and the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. Their research yielded notable results, with an accuracy of 0.82 achieved for the MAS dataset and an impressive accuracy of 0.93 for the ADNI dataset. These findings underscore the significance of comparative evaluations, demonstrating the varying performance of algorithms in dementia prediction and providing valuable insights for selecting the most effective approaches for different data sources and research contexts.

3. Aim Of Research

The primary aim of this research is to create an effective predictive model for dementia based on longitudinal data sourced from the OASIS dataset [12] The study adopts several key strategies to achieve this goal:

1. **Dataset Analysis and Feature Selection**: The research commences by thoroughly analyzing the OASIS dataset, scrutinizing its various features to identify the most significant ones for predicting dementia.
2. **Explanatory Data Analysis**: In-depth exploratory data analysis is performed on the OASIS dataset to gain insights into its characteristics, distribution, and relationships among variables.
3. **Feature Selection Techniques:** Various feature selection techniques are employed to pinpoint the most relevant variables that contribute to dementia prediction. This step streamlines the model by focusing on the most informative features.
4. **Efficiency Enhancement:** The efficiency of the proposed predictive model is improved through the application of feature selection and cross-validation techniques. This optimization process aims to enhance the model's accuracy and computational efficiency.
5. **Identifying the Best Model:** The research aims to propose the most effective predictive model that achieves the highest accuracy in distinguishing between elderly/healthy individuals and those with dementia. This involves a meticulous evaluation of various machine learning algorithms.

4. Problem Definition

The objective of this predictive model is to assess the likelihood of patients diagnosed with Mild Cognitive Impairment (MCI) transitioning to Dementia within two distinct timeframes: 0-1 years and 1-5 years. The model classifies patients into two categories: Class 1, indicating those at risk of developing Dementia during the specified time intervals, and Class 0, representing those not expected to progress to Dementia within these timeframes [13]. Leveraging patient-specific data, including demographics, medical history, and cognitive assessments, the model employs binary classification techniques to make these predictions. Successful implementation can offer valuable insights for healthcare providers, enabling early intervention and tailored care plans for patients at higher risk of cognitive decline.

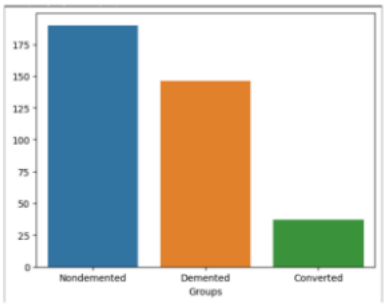

Fig.1. Subjects in the datasets into groups

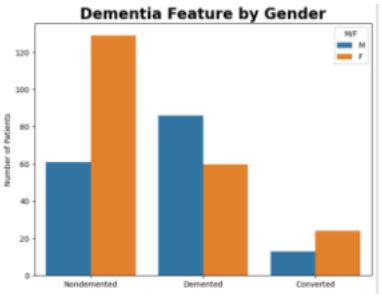

Fig.2. Dementia Feature with Gender

5. Dataset Descriptions

The dataset utilized in this research originates from the Open Access Series of Imaging Studies (OASIS) project, which aims to facilitate open access to MRI (Magnetic Resonance Imaging) brain datasets for scientific exploration. Specifically, the dataset under investigation focuses on Longitudinal MRI data in older adults, encompassing both nondemented and demented individuals. This dataset holds significant relevance for dementia research, providing insights into the progression of cognitive health over time [14]. The dataset comprises a cohort of 150 subjects, ranging in age from 60 to 96 years. These individuals participated in multiple MRI scans conducted across two or more separate visits. These visits were scheduled with a minimum interval of one year, resulting in a total of 373 imaging sessions. This longitudinal aspect of the dataset offers a unique opportunity to investigate cognitive health changes over time in the elderly population. The dataset encompasses gender diversity, with both male and female subjects included. It is noteworthy that all individuals featured in the dataset are right-handed, which may be of significance in studies related to brain lateralization.

6. Methods And Methodologies

In this study, the methodology is structured around the development of a robust dementia prediction model using the OASIS dataset. Initially, data collection is emphasized as a pivotal phase, wherein MRI data from the OASIS project is gathered, serving as the foundational element. Following this, comprehensive data preparation and preprocessing are carried out, involving essential tasks such as data cleaning, imputation of missing values, label encoding, data transformation, feature selection, and feature scaling. These measures are aimed at enhancing data quality and relevance for subsequent analysis.

Subsequently, the methodology includes an exploratory data analysis (EDA) phase, where the dataset is scrutinized to unveil concealed relationships and patterns. EDA leverages data visualization techniques, including charts and histograms, to gain valuable insights into the dataset's characteristics. Feature selection techniques, including diverse variations, are employed to identify the most pertinent attributes contributing significantly to dementia prediction. The dataset is then thoughtfully partitioned into training and testing subsets in a 3:1 ratio, allocating 70\% for training purposes and 30\% for testing. Following data splitting, inputs crucial for predicting dementia outcomes are meticulously chosen and fed into the machine learning classifiers. Finally, a qualified model is formed, which is applied to the testing dataset to categorize individuals into 'demented' and

'non-demented' categories. The study adopts supervised classifiers to facilitate this prediction process, ensuring accuracy and reliability in dementia diagnosis and prediction.

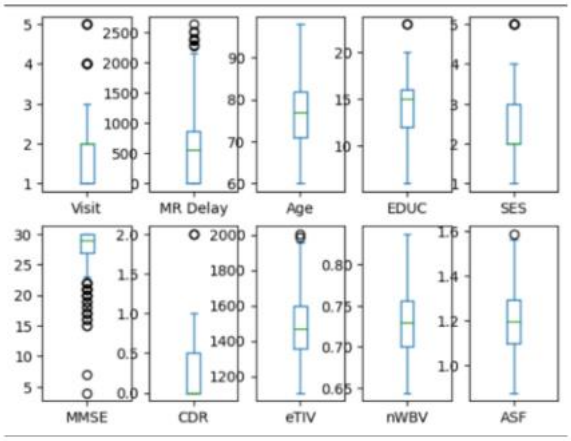

Fig. 3. Box plot for different attributes

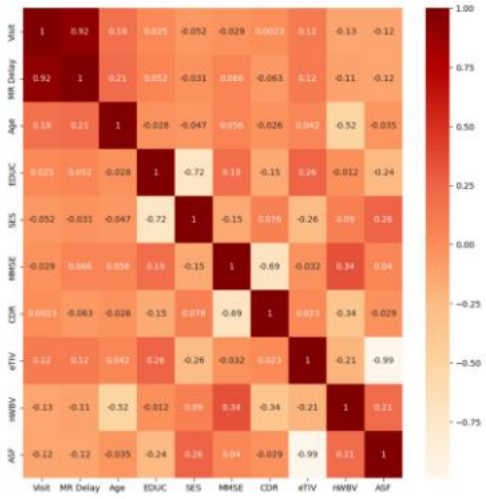

Fig. 4. Heatmap for the entire dataset

To construct an early prediction model for dementia based on longitudinal data, the initial step involves training the dataset to acquire a concise representation and encode the evolving dynamics of longitudinal measurements for each subject. This work encompasses a diverse set of hyper-parametric classifiers including Support Vector Machine (SVM), XGBoost, Logistic Regression. These classifiers collectively form the arsenal of techniques employed to analyze and predict the onset of dementia in individuals. Leveraging a wide array of machine learning algorithms, this predictive model aims to discern early signs of dementia, thereby facilitating timely intervention and support for affected individuals.

7. Experimental Results And Discussion

Develop an early prediction model for dementia based on longitudinal data. The journey began with meticulous data pre-processing, which involved cleaning, transforming, and optimizing the dataset to ensure its suitability for machine learning. This phase also included crucial tasks such as handling missing values, encoding categorical variables, and selecting relevant features while maintaining data consistency. With the data primed for analysis, it was then segregated into training and testing sets, with a 70:30 ratio allocation, allowing for the assessment of the model's generalization capabilities. Recognizing the need for robust validation, the research adopted a 10-fold cross-validation technique, further enhancing the model's reliability. This method

divided the data into ten distinct subsets or "folds," ensuring that each data point served as both training and testing data during the iterative validation process.

The combination of data pre-processing, effective data segregation, and comprehensive cross-validation techniques collectively contributed to a robust foundation for building a predictive model for early dementia detection. These steps collectively enhanced the model's ability to generalize and provided a thorough evaluation of its performance on unseen data.

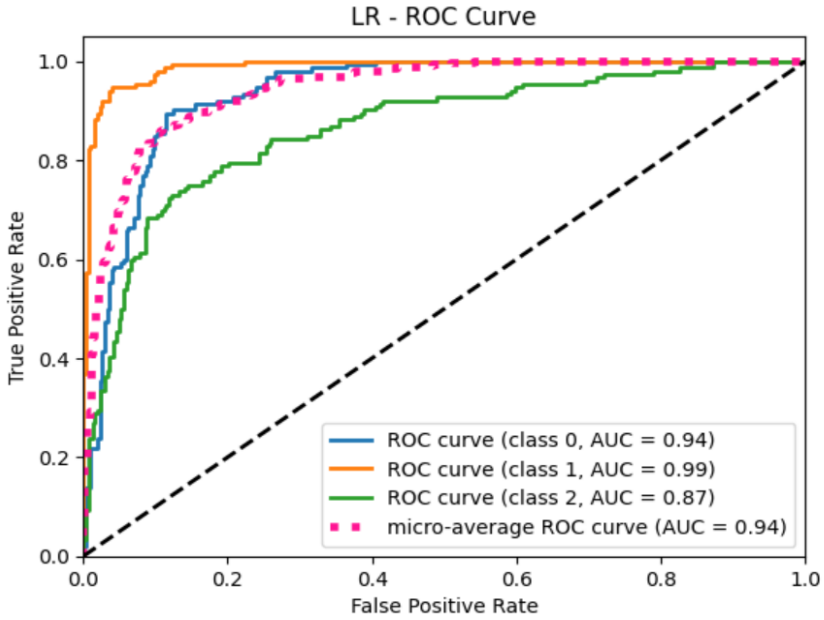

Fig. 5. ROC AUC value for LR Classifier

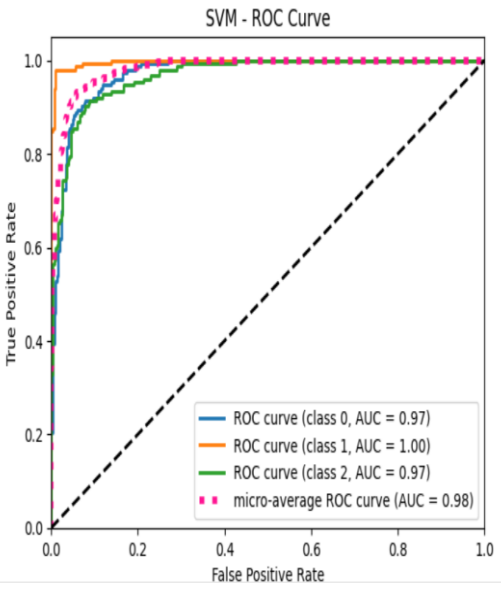

Fig. 6. ROC AUC value for SVM Classifier

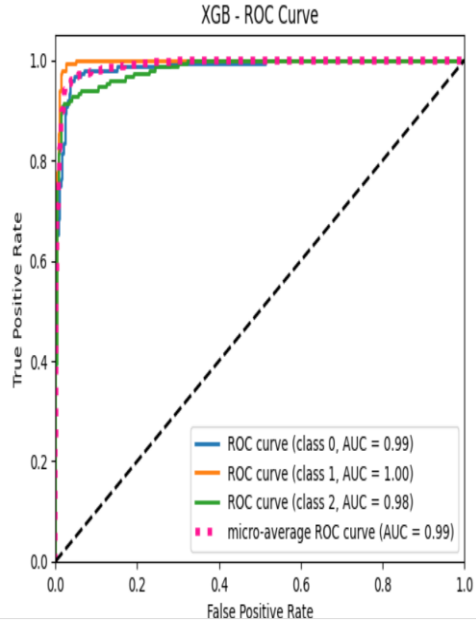

Fig. 7. ROC AUC value for XGB Classifier

Logistic Regression is a linear classification algorithm that models the probability of an instance belonging to a particular class. It achieved an accuracy of 82.24\%, making it a competitive choice. The ROC AUC value of 94.06\% indicates its ability to differentiate between dementia and non-dementia cases with good precision and recall. Support Vector Machine is a powerful algorithm that finds an optimal hyperplane to separate classes. It achieved an accuracy of 89.04\%, indicating strong performance. The high ROC AUC value

of 97.89\% demonstrates its excellent class separation as shown in Figure-6. XGBoost is a gradient boosting algorithm known for its speed and performance. It achieved an accuracy of 93.20\%, showing its effectiveness. The ROC AUC value of 99.09\% indicates strong class separation.

8. CONCLUSIONS

This research represents a significant step forward in the pursuit of early dementia prediction based on longitudinal data analysis. Leveraging a diverse set of machines learning classifiers, including logistic regression, support vector machine and XGBoost, we conducted a comprehensive evaluation of their performance. Among the classifiers, random forest and XGBoost emerged as the standout performers, achieving the highest accuracy rates of 94.30\% and 93.20\%, respectively, along with remarkably high ROC AUC values of 99.10\% and 99.09\%. These ensemble methods demonstrated their prowess in effectively distinguishing between dementia and non-dementia cases based on longitudinal data.

REFERENCES

[1] Inoue, Y., Shue, F., Bu, G. & Kanekiyo, T. Pathophysiology and probable etiology of cerebral small vessel disease in vascular dementia and Alzheimer's disease. *Molecular Neurodegeneration*. 18, 1-22 (2023)

[2] Yan, S., Liu, M., Qi, Z. & Lu, J. Research Applications of Positron Emission Tomography/Magnetic Resonance *(PET/MR) Imaging in Alzheimer's Disease (AD). PET/MR: Functional And Molecular Imaging Of Neurological Diseases And Neurosciences*. pp. 161-186 (2023).

[3] Shah, J., Rahman Siddiquee, M., Krell-Roesch, J., Syrjanen, J., Kremers, W., Vassilaki, M., Forzani, E., Wu, T. & Geda, Y. Neuropsychiatric Symptoms and Commonly Used Biomarkers of Alzheimer's Disease: A Literature Review from a Machine Learning Perspective. *Journal Of Alzheimer's Disease.,* 1-16 (2023)

[4] Junaid, M., Ali, S., Eid, F., El-Sappagh, S. & Abuhmed, T. Explainable machine learning models based on multimodal time-series data for the early detection of Parkinson's disease. *Computer Methods And Programs In Biomedicine*. 234 pp. 107495 (2023).

[5] Veena, K., Priya, M. & Sumathi, D. Predictive Diagnostic Analysis for Early Detection of Alzheimer's disease Using Machine Learning. *Journal Of Algebraic Statistics*. 13, 586-592 (2022).

[6] Kuo, P., Huang, C. & Yao, T. Optimized Transfer Learning Based Dementia Prediction System for Rehabilitation Therapy Planning. *IEEE Transactions On Neural Systems And Rehabilitation Engineering*. (2023) .

[7] Wang, M., Greenberg, M., Forkert, N., Chekouo, T., Afriyie, G., Ismail, Z., Smith, E. & Sajobi, T. Dementia risk prediction in individuals with mild cognitive impairment: a comparison of Cox regression and machine learning models. *BMC Medical Research Methodology*. 22, 284 (2022).

[8] Merkin, A., Krishnamurthi, R. & Medvedev, O. Machine learning, artificial intelligence and the prediction of dementia. *Current Opinion In Psychiatry*. 35, 123-129 (2022).

[9] Kavitha, C., Mani, V., Srividhya, S., Khalaf, O. & Tavera Romero, C. Early-stage Alzheimer's disease prediction using machine learning models. *Frontiers In Public Health*. 10 pp. 853294 (2022).

[10] Liu, K., Li, Q., Yao, L. & Guo, X. The Coupled Representation of Hierarchical Features for Mild Cognitive Impairment and Alzheimer's Disease Classification. *Frontiers In Neuroscience*. 16 pp. 902528 (2022).

[11] Saha Bharati, S., Podder, P., Thanh, D. & Prasath, V. Dementia classification using MR imaging and clinical data with voting based machine learning models. *Multimedia Tools And Applications*. 81, 25971-25992 (2022).

[12] Kanikar, P., Sankhe, M. & Patkar, D. A Comparative Analysis of Classification Algorithms for Dementia Prediction. *International Conference On Innovations In Bio-Inspired Computing And Applications*. pp. 658-668 (2022).

[13] Erickson, C. Promoting Brain Health Through Identifying and Communicating Dementia Risk to Cognitively Unimpaired Older Adults. (The University of WisconsinMadison,2022).

[14] Huang, Y., Shan, Y., Qin, W. & Zhao, G. Apolipoprotein E 4 accelerates the longitudinal cerebral atrophy in open access series of imaging studies-3 elders without dementia at enrollment. *Frontiers In Aging Neuroscience*. 15 pp. 1158579 (2023).
[15] Carcagnı, P., Leo, M., Del Coco, M., Distante, C. & De Salve, A. Convolution Neural Networks and Self-Attention Learners for Alzheimer Dementia Diagnosis from Brain MRI. *Sensors*. 23, 1694 (2023).
[16] Khan, A., Zubair, S. & Khan, S. A systematic analysis of assorted machine learning classifiers to assess their potential in accurate prediction of dementia. *Arab Gulf Journal Of Scientific Research*. 40, 2-24 (2022)
[17] Chen, T., Su, P., Shen, Y., Chen, L., Mahmud, M., Zhao, Y. & Antoniou, G. A dominant set-informed interpretable fuzzy system for automated diagnosis of dementia. *Frontiers In Neuroscience*. 16 pp. 867664 (2022).