Grenze International Journal of Engineering and Technology, June Issue

Importance of Biomarkers and Demographic Data for Alzheimer's Disease Detection in the Early Stage

B A Sujathakumari[1], Usha Rani C M[2], Shalini M S[3] and Sharayu R Siddhanti[4]
[1-4]Department of Electronics and Communication Engineering, Sri Jayachamarajendra College of Engineering, JSS Science and Technology University, Mysuru, Karnataka, India
Email: sujathakumari@sjce.ac.in, usharani@jssstuniv.in, shalinims@jssstuniv.in, sharayusiddhanti@gmail.com

***Abstract*— Type of Dementia is a Neuro-Degenerative disorder that affects memory and causes difficulty in problem solving and thinking is. Instead of recognizing AD in clinical trials, research in MCI is likely to move the threshold of recognition to a stage earlier in the disease process to allow intervention at an earlier point. At this point, treatments and drugs can either halt or reduce the development of AD. As the outcome, it's critical to diagnose AD at the Mild Cognitive Impairment stage as soon as possible. This research suggests a unique method for early AD detection that combines biomarkers, clinical data, and MRI imaging. The ADNI provided the data set for the Cognitive Normal cases and Mild Cognitive Impairment cases. The clinical dataset includes information on every participant, including demographics, physical exam results, and cognitive test results. To improve the accuracy of AD detection, a range of clinical and genetic data, including CSF biomarkers and amyl-beta biomarkers, are taken into account. In addition to the features derived from the MRI data, clinical data is incorporated for improved classification outcomes. After including each of the chosen biomarkers, a 91% accuracy rate is attained.**

***Index Terms*— AD, Biomarkers, Deep learning, CN, MCI.**

I. INTRODUCTION

Alzheimer's disease (AD) is the one among the cause resulting for dementia [1], which is defined by a progressive deterioration in mental, behavioural, and social skills as well as challenges with an individual's ability to think and act for themselves. Globally, 35.6 million people are estimated to have dementia [2]. Dementia is characterized as "a substantial loss of global cognitive ability in a formerly normal person" [3], and AD is a common cause for it. Forgetting earlier encounters or occurrences is one of the early indicators of the condition. A person with advanced AD will have severe memory loss and lose their capacity to carry out basic duties. Patients with AD experience difficulties speaking, reading, and writing in addition to confusion and memory loss. In the end, AD results in death by destroying the area of the brain in charge of regulating breathing and heart rate [4].

Typically, early diagnosis of AD is characterized as a challenge for classification, such as the separation of sMCI cases from pMCIs. However, the cut-off follow-up period employed to distinguish between pMCI and sMCI affects the performance. In addition, regardless of the chosen criterion, the cohort's participants are frequently heterogeneous.

Grenze ID: 01.GIJET.12.2.549
© Grenze Scientific Society, 2026

Predicting it at an early stage is critical because if we predict it, we will have easier access to different treatments [5], emotional support for patients, social benefits for their parents, and the cost of the treatment will be reduced. Biomarkers can help with diagnosis, promote target engagement, support disease change, and provide safety monitoring. Key biological occurrences are being recognized, and our understanding of AD is quickly advancing. A biomarker is a trait that may be scientifically investigated and assessed as a signal for biological processes that are healthy or unhealthy, as well as for physiologically healthy reactions to therapeutic interventions[6]. Biomarkers have the ability to identify diseases, show target engagement, help modify diseases, and check for safety. The amyloid (A), CSF, and neuro-degeneration Research Framework can be used in clinical trials and pharmaceutical research, and it highlights the importance of brain imaging and CSF data for disease diagnosis and stage [7].

II. Related Works

This section discusses a few studies that are pertinent to this subject. To create a prediction model of the progression of AD dementia, researchers developed a model based on RNNs. Through the integration of baseline hippocampal MRI data with longitudinal cognitive assessments of individual subjects, the model learns about the temporal dynamics and informative representation of these regions. The deep learning model was able to learn relevant metrics from the data to describe the progression in participants to AD, according to experimental results on a sizable cohort of MCI individuals. Furthermore, the early onset of AD was correctly predicted by the prognostic model. [1]. Two objectives were accomplished by developing a new DL methodology that combined CNN and was applied to the analysis of T1WI, the most common anatomical MRI of the brain: it was able to recognize the intricate change patterns associated with AD and categorize it as MCIc, HC, and MCInc. The suggested technique uses a two-stage EL system to improve robustness and generalization [8]. It is demonstrated that while classification approaches for AD identification based on the characteristics extracted from neuro-images have attained excellent accuracy, their performance is significantly worse for AD detection in its early stages. The classification accuracy has increased by about 5-7% because of the use of statistical approaches in classifiers. Lack of defined measurements that describe how much the brain shrinks as AD progresses and a dearth of data to convert patterns of diminished brain activity taken from functional neuro-images into diagnostic data are two issues that hinder early AD detection [3].

As part of a modelling technique of supervised feature extraction is the most effective method for accounting dimensional imaging modalities with multiple less dimensional indicators.. The goal concentrates to identify imaging biomarkers and combine these with existing ones to predict the prognosis of people with a high risk of acquiring AD. This methodology evaluated the relative value of various imaging techniques for foretelling conversion. Using only data from baseline, it is evaluated that suggested methodology on ADNI database to forecast the progression of people with MCI to AD [4]. The classifier results showed that, in terms of AUC, the random forest classifier performed better than the logistic regression method. Thus concluded by incorporating biomarkers and demographic information with MRI data, the prediction algorithm's precision and accuracy can be increased [9].

Deep learning has received considerable attention for its potential application in the detection of AD. Biomarker has important role in AD detection. It has been noted that contemporary research makes use of deep learning models to improve experiment prediction accuracy. There are several ways to contribute more data to the system and get the classification results. Researchers from across the world are putting a lot of effort into finding better ways to identify people and objects over time, extract features from them, and classify them. There are already a few biomarkers for AD, including biochemical, genetic, and neuroimaging indicators. The addition of biomarkers helps in detecting the disease at an early stage more accurately.

III. Methodology

The model includes various steps and Fig. 1 depicts these as shown,

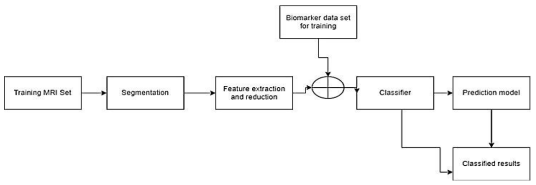

Fig. 1. Block diagram depicting the steps of a system

A. Data acquisition

The initial step involves gathering information from the standard ADNI website. The particular source of the data is ADNI-1, which has patient records dating back four years. It includes data and samples from cognitive test results, blood biomarkers, CSF biomarkers[10], PET scans, and MRI scans. The development of ADNI aimed to yield biomarkers that could serve as indicators of the results of clinical trials. ADNI aims to monitor the disease's progression and discover early detection techniques. With the use of state-of-the-art diagnostic instruments, it seeks to promote developments in AD detection, prevention, and intervention. The ADNI provides data, which include information on people with MCI, CN and AD. A total of 1200 MRI image samples were acquired.

B. Data pre-processing and segmentation

The categorization of AD using the ADNI dataset requires proper feature segmenting approaches to find the existent and important finer smaller brain area features, hence eliminating a laborious, time-consuming, and labour-intensive voxel-based morphometry technique. a deep learning-based segmentation method that uses SegNet to accurately categorize AD and dementia conditions by identifying structural magnetic resonance imaging (sMRI) properties of important brain regions for AD. Brain MRI segmentations taken at different times are also used to measure structural alterations in the brain[11].Much more data is required for more accurate diagnoses. Picture segmentation is the process of dividing a picture into many pixel-rich regions, each represented by a tagged image or mask[12]. By segmenting an image, you can process only the important parts of it instead of the entire thing. CN and MCI segmented scans are displayed in Figs. 2. We use the SPM toolbox in MATLAB to compare each MRI scan to a standard template to extract the best parts of the skull from the scans.

Fig. 2. Segmented NC scans: a) Grey Matter, b) White matter, c) CSF

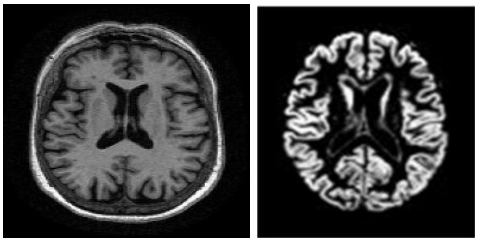

Fig. 3. Original Image Fig. 4. Segmented Image

As observed in Fig.4, the data is visible in axial, coronal, and sagittal views.Data cleansing needs to be done after the data is acquired. For the aforementioned process, a technology known as SPM (Statistical Parameter Mapping) is utilized. This is accomplished using MATLAB 2016a, and the resulting segmented image and original image are displayed in Fig. 5 and 6, respectively. Using the graphical user interface, the modulated normalized MRI is selected from the Batch Editor. But for this project, only the grey matter is selected. Additionally, you must register for a basic spot. Each person's brain's grey matter is where additional processing happens. A tool called MRIcro analyses brain segmented MRI images[13]. The brain images that were not able to be segmented correctly are removed.

C. Feature extraction

The process of dimensionality reduction, which divides an initial raw data collection's size and complexity into manageable portions, includes feature extraction. This will make processing simpler[14]. The importance of these massive data sets is the variety of unique variables they include. It requires a lot of computing power to process them. By selecting and merging these into features, it helps to extract the best feature from the data in order to effectively reduce the number. Despite being easy to use, these features are able to uniquely and accurately characterize the genuine data set.

D. Biomarkers and clinical data

These aid in disease modification, aid in target engagement, aid in diagnostics, and check for safety. Biomarker trial results are more important for establishing if a treatment is safe and effective physiologically, but they are also more limited and less comprehensive than trials required to show a clinical benefit. Companion diagnostics, which may be developed in AD medication development projects, are necessary for the safe and efficient administration of therapies[15]. If an objectively measured and evaluated trait can reveal biological responses to a therapeutic intervention, pathological processes, or normal biological processes, it is called a biomarker. Using clinical evaluations, cognitive testing, positron emission tomography, other biological markers, and serial MRI, ADNI's primary goal has been to keep track of MCI and early AD. For our investigation, we have extracted the raw data and biomarkers from the ADNI dataset.Some of the clinical and demographic data[16] are considered along with the biomarkers to improve the efficiency. The project's implementation flow, which includes the use of biomarkers and MRI scans to improve prediction accuracy. Fig 5 shows the flow chart for implementation with biomarkers.

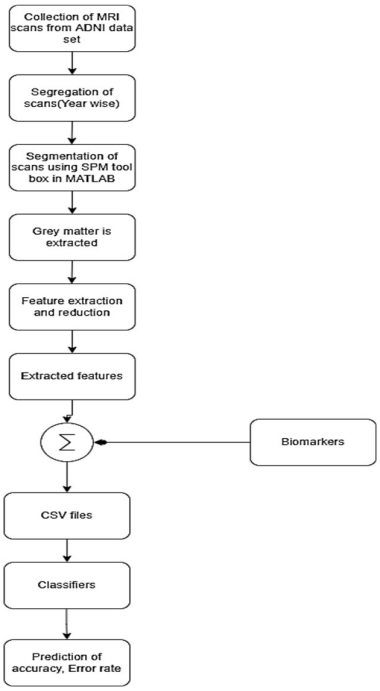

Fig.5. Flow chart for Implementation with Biomarkers

E. Classification

In reference to this work, the division into stages is classifying the data based on some considerations. In this paper, the accuracy as well as other metrics are computed for CN vs MCI individuals.

Neural Networks – DL

An ANN has several sensing units, usually known as neurons. Numerous sensing units, commonly referred to as neurons, are found in an artificial neural network (ANN). Because all inputs are processed forward, Artificial Neural Network (ANN) is frequently referred to as a feed-forward neural network. One of the simplest types of neural networks to comprehend is this one. Data is sent in a single direction and passes via a few input nodes before arriving at the output node.

The network may or may not have hidden node levels, which contributes to the explanation of its operation. Here in this study 4-layer architecture is used, has an enormous number of neurons, that perform all functions. The basic unit of information contained in neurons is the weighted connection of neurons.The later layers are merely retrained in transfer learning; only the early and center layers are used. It makes use of the task's labelled data, which acted as its training set. 91 image slices are acquired for a single patient; the 45th slice yields the best image, as seen in Fig 6.

Fig. 6. 45th slice of MRI image

IV. Results and Discussions

The results of neural networks are discussed here. Some of the important biomarkers, clinical and demographic data are added to ANN to improve the results. Heat map is generated to analyse the results and F1 score, Accuracy are calculated.

TABLE I: ACCURACY AND F1-SCORE OF ANN WITHOUT BIOMARKERS

Neural Network	Accuracy	F1-score
ANN	81%	0.79

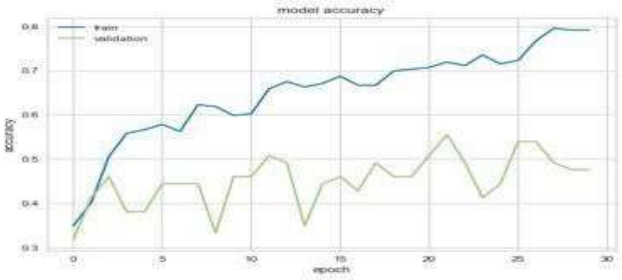

Fig 7. Shows the Model accuracy curve of ANN.

Without adding biomarkers or clinical data the accuracy given by ANN is 81%. To improve the accuracy and to make it more efficient biomarkers are added. Table I shows the result of ANN.

TABLE II: ACCURACY AND F1-SCORE OF ANN WITH BIOMARKERS

ANN WITH BIOMARKERS	Accuracy	F1- score
ANN+BM(C3+FH)	90%	0.90
ANN+BM(Alpha_syn+Hemoglobin)	93%	0.92
ANN+BM(Abeta42/40)	89%	0.88
ANN+BM(CDR)	88%	0.88
ANN+BM(All)	91%	0.91

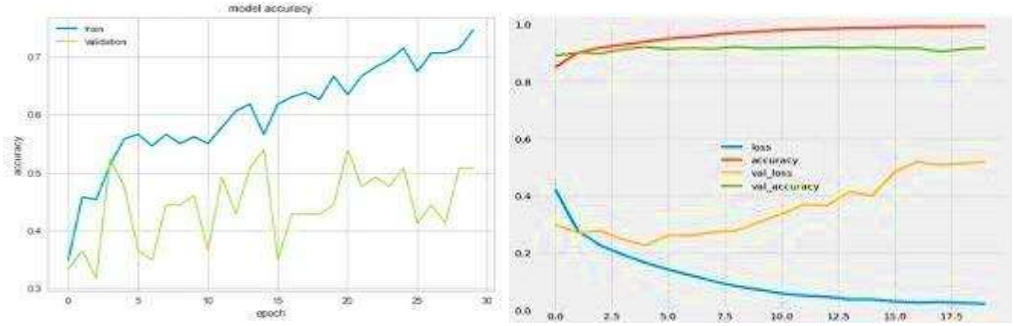

Fig. 8. Model accuracy curves of ANN with Biomarkers

Biomarkers are added to inorder to increase the accuracy of the systems. Here in this case c3 and fh protein are together added to improve the efficiency. The obtained accuracy of the ANN classifier along with biomarkers c3 and fh protein is 90%, F1- score is obtained. i.e 90%, The obtained accuracy of the ANN classifier along with biomarkers Alpha syn and Hemoglobin is 93%, F1-score is obtained. i.e 92%, The obtained accuracy of the ANN classifier along with biomarker Abeta (42/40) is 89%, F1-score is calculated, and it is obtained. i.e 88%, The obtained accuracy of the ANN classifier along with biomarker CDR is 88%, F1-score is calculated, and it is obtained. i.e 88%, The obtained accuracy of the ANN classifier along with all biomarkers together is 91%, F1-score is calculated, and it is obtained. i.e 91%, which is tabulated in Table II. Fig 10 shows the Model accuracy curves of ANN with different Biomarkers

TABLE III: ACCURACY AND F1-SCORE OF ANN WITH CLINICAL AND DEMOGRAPHIC DATA

ANN WITH CLINICAL & DEMOGRAPHIC DATA	Accuracy	F1-score
APOE4+PSYCHATRIC	84%	0.83
NEURALOGIC+MMSC	82%	0.83
ALL(DEMOGRAPHIC)	82%	0.81

Some of the clinical and demographic data are added to improve the accuracy. Here Apoe4 and Medical history of psychiatric patients who can develop AD are analyzed and accuracy obtained is 84%, MMSC and Medical history of neurological patients apart from AD, who can develop AD and accuracy obtained is 82%, and all clinical and demographic data are together added, who can develop AD and accuracy obtained is 82%, the accuracy is tabulated in the Table III.

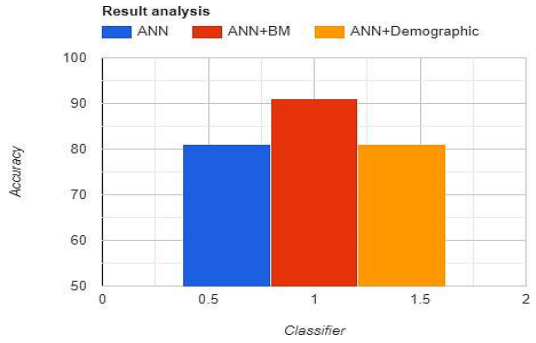

Fig. 9. Shows the analysis of the result. It is clearly visible that after adding all the biomarkers to the MRI image the accuracy improved for about 10% more than that of only MRI image.

V. CONCLUSION

This paper's main goal is to determine the best method for detection of AD in early stages. When the provided strategy is put into practice, it is shown that AD can be predicted early using deep learning neural networks, and that accuracy will increase when biomarkers and clinical data are added. The suggested approach is put into practice, and the anticipated outcomes show that AD can be easily identified by utilizing a few key indicators that are discussed, allowing for early detection of the disease. The accuracy we obtained is 91% after adding all the chosen biomarkers, and 93% after adding the haemoglobin and alpha-syn biomarkers to the ANN, demonstrating its effectiveness. The future scope includes considering of more classification of data from ADNI dataset and also to work on different classifiers to obtain more accurate results to provide detection of AD in early stage for this progressing disease.

REFERENCES

[1] J. Islam and Y. Zhang, "Early Diagnosis of Alzheimer's disease: A Neuroimaging Study with Deep Learning Architectures," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018

[2] J. Fritsch, S. Wankerl and E. Nöth, "Automatic Diagnosis of Alzheimer's Disease Using Neural Network Language Models," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019

[3] Li H, Fan Y; "Alzheimer's Disease Neuroimaging Initiative. Early Prediction Of Alzheimer's disease Dementia Based on Baseline Hippocampal MRI And 1-Year Follow-up Cognitive Measures Using Deep Recurrent Neural Networks". Proc IEEE Int Symp Biomed Imaging. 2019

[4] M. Amin-Naji, H. Mahdavinataj and A. Aghagolzadeh, "Alzheimer's disease diagnosis from structural MRI using Siamese convolutional neural network," 2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA), 2019.
[5] P. Bhatkoti and M. Paul, "Early diagnosis of Alzheimer's disease: A multi-class deep learning framework with modified k-sparse autoencoder classification," 2016 International Conference on Image and Vision Computing New Zealand (IVCNZ), 2016
[6] Capecci, Elisa & Doborjeh, Zohreh & Mammone, Nadia & Foresta, Fabio & Morabito, Francesco & Kasabov, Nikola. "Longitudinal Study of Alzheimer's Disease Degeneration through EEG Data Analysis with a NeuCube Spiking Neural Network Model".(2016)
[7] Dan Pan1, An Zeng1,2*, Longfei Jia1, Yin Huang1, Tory Frizzell3 and Xiaowei Song3, " Early Detection of Alzheimer's Disease Using Magnetic Resonance Imaging: A Novel Approach Combining Convolutional Neural Networks and Ensemble Learning" Frontiers in Neuroscience, Volume 14, 2020.
[8] Khan, Protima & Kader, Md Fazlul & Islam, S. M. Riazul & Rahman, Aisha B & Kamal, Md & Toha, Masbah & Kwak, Kyung. "Machine Learning and Deep Learning Approaches for Brain Disease Diagnosis: Principles and Recent Advances". IEEE Access, (2021).
[9] Peng Cao, Xiaoli Liu, Jinzhu Yang, Dazhe Zhao, Osmar Zaiane, "Sparse Multi-kernel Based Multi-task Learning for Joint Prediction of Clinical Scores and Biomarker Identification in Alzheimer's Disease", Springer International Publishing(2017).
[10] L. Lu, S. Elbeleidy, L. Z. Baker, H. Wang and F. Nie, "Predicting Cognitive Declines Using Longitudinally Enriched Representations for Imaging Biomarkers," in IEEE Transactions on Medical Imaging, vol. 40, no. 3, pp. 891-904, March 2021
[11] Song M, Jung H, Lee S, Kim D, Ahn M. Diagnostic Classification and Biomarker Identification of Alzheimer's Disease with Random Forest Algorithm. Brain Sci. 2021 Apr 2
[12] Singh, N., Wang, A.Y., Sankaranarayanan, P., Fletcher, P.T., Joshi, S. (2012). Genetic, Structural and Functional Imaging Biomarkers for Early Detection of Conversion from MCI to AD. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012. MICCAI 2012. Lecture Notes in Computer Science, vol 7510. Springer, Berlin, Heidelberg.
[13] K. Skolariki, T. Exarchos and P. Vlamos, "Contributing factors to Alzheimer's Disease and biomarker identification techniques," 2020 5th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), 2020
[14] Yan, J. et al. (2012). Multimodal Neuroimaging Predictors for Cognitive Performance Using Structured Sparse Learning. In: Yap, PT., Liu, T., Shen, D., Westin, CF., Shen, L. (eds) Multimodal Brain Image Analysis. MBIA 2012. Lecture Notes in Computer Science, vol 7509. Springer, Berlin, Heidelberg.
[15] Sujathakumari, B. A., M. Charitha Shetty, H. M. Lakshitha, P. Jain Mehulkumar, and S. Suma. "Predictive Analysis for Early Detection of Alzheimer's Disease." In Data Intelligence and Cognitive Informatics: Proceedings of ICDICI 2020, pp. 709-723. Springer Singapore, 2021.
[16] Yan J, Huang H, Kim S, Moore J, Saykin A, Shen L. Joint Identification Of Imaging and Proteomics Biomarkers Of Alzheimer's Disease Using Network-Guided Sparse Learning. Proc IEEE Int Symp Biomed Imaging. 2014 May