ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 12, December 2022

Early detection of Alzheimer's using blood plasma proteins with Recurrent Neural Networks

[1] Monisha M, [2]Harshitha K M, [3]Dhanalakshmi N H, [4]Kokatam Sai Prakash Reddy, [5]Nagarathna C R, [6]Dr. Kusuma M

[1][2][3][4] B.E Students, [5]Assistant Professor, BNM Institute of Technology, 6Professor, Department of Information and Science,Dayananda Sagar Academy of Technology and Management, Bangalore – 560082
[1] monishasmekala@gmail.com, [2]harshithakmurali@gmail.com, [3]dhanalakshminh.work@gmail.com, [4]saiprakash2610@gmail.com

Abstract*— Alzheimer's disease (AD) which is a disease that belongs to the group of neurodegenerative diseases and is considered one of the most destructive and severe diseases of the human nervous system. Presently there is no quick cost-effective method for routinely screening of persons with Alzheimer's disease. The problem is how to diagnose it at the earliest possible stage before specific symptoms begin to appear. The main idea is to build an intelligent system that will be able to answer, based on certain biomarkers from the subject, whether the disease is present or not. This paper presents how machine learning concepts are used that have upgraded the detection of Alzheimer's disease in the early stage. In addition, the proposed does hierarchal classification into stages: CN, EMCI, LMCI and AD. Experimental results show that the proposed method achieves classification accuracy of 92-95 % for AD demonstrating the promising performance for RNN analysis.

Index Terms*— Alzheimer's disease, Cognitive Normal, predictive testing, Positron emission tomography, Support vector machine, Machine learning.

I. INTRODUCTION

Alzheimer's disease is the neurological disease that makes brain to shrink and destroys the brain cells. Alzheimer's is a type of common dementia that impacts on the memory, thinking skills and patients lose the ability to implement simple everyday activities or function independently.

Alzheimer's disease generally starts slowly and gradually worsens overtime. As most of the people's common initial symptom is trouble in recalling recent activities. As the condition advances the person will develop critical memory impairment and will lose the ability to perform everyday activities. Alzheimer's disease has no treatment to cure it. Even after the diagnosis a person can only live up to 4 to 8 years on average. Alzheimer's disease generally progresses slowly in three stages: CN stage, MCI stage and AD stage .i.e., early, middle and late stages. CN stage is the initial stage, where a person will have symptom to forget minor things like he or she has kept a book on a table, but the person is facing difficulty in recalling it after a few minutes or the person facing difficulty because of forgetting that was just read. MCI stage is the longest stage and can last for many years. The symptoms in this stage is, where the person may confuse words, mood swings such as getting frustrated or angry and may act in unexpected ways. Destruction of brain cells can also make it difficult for the person to express their thoughts and implement routine tasks. The final stage of the disease is the AD stage which has severe symptoms. The patient with MCI will progress to Alzheimer's disease at a rate of 15% per year. The symptoms of AD stage are individuals lose the tendency to respond to their environment, to convey or maintain a conversation and eventually to control movement.

Hopeful results were seen in the classification of Alzheimer's disease (AD) using Machine Learning methods, but combination of high accuracy and short processing time is required for successful application in clinical settings. Our Objective is to Build and analyse models for early prediction of Alzheimer's disease (AD). In this study, Alzheimer's disease will be analysed. The problem is how to diagnose it at the earliest possible stage before specific symptoms begin to appear. The main idea is to build an intelligent system that will be able to answer, based on certain biomarkers from the subject, whether the disease is present or not.

II. MODALITIES

PET Scan: Positron emission tomography (PET)[13] along with other imaging techniques are used to let out several details of dementia physiology, such as tau and amyloid accumulation in brain, neuron-inflammation and information about irregular metabolism in dementia patients. One of the methods is imaging with PET by using radioliglands that bind to amyloid beta (Aβ), in particular [11C] PiB compound i.e., Pittsburgh B compound are used to determine the amount brain amyloid burden and hence the pre-symptomatic stages of the Alzheimer's disease.

Although few studies have proved good sensitivity in predicting AD at pre-symptomatic stages using PET scan, an absolute threshold for shaping the test positivity is not defined. Moreover, PET scan is an expensive method so

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 12, December 2022

standardising the process is important. However, PET scan as limited advantages because it is costly and it is peculiar for centres [2].

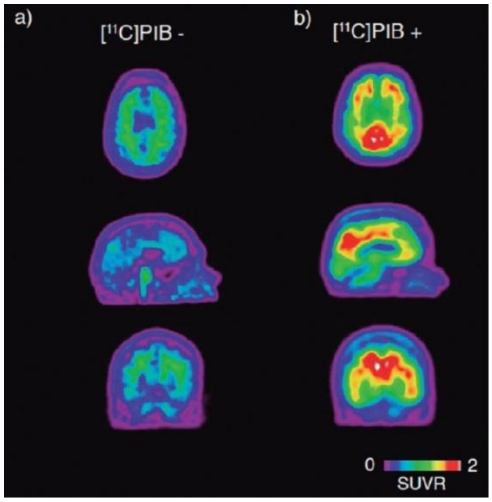

Fig 1: [11C] PiB PET scan image: The white matter is the growth of: [11C]
PiB compound in a CBS patient (a; age 75) and immense cortical in a AD patient. [13]

MRI Scan: Deep learning, a concept in machine learning has won a lot of advantages over traditional machine learning. Especially when using a high-dimensional and complex data for training the model. The same feature of deep learning has been used for early detection of Alzheimer's disease and gained attention. Deep learning papers on Alzheimer's disease were found and out of 16 papers, combination of deep learning and usual machine learning concepts were used in 4 papers and 12 were about using deep learning concepts alone. After studying the above 16 papers, it could be concluded that 83.7% accuracy was achieved in prediction when the combination was used and about 84.2% accuracy was achieved in prediction of MCI conversion [14].

Deep learning which is a part of machine learning uses raw neuro-imaging data [14]. Neuro-imaging data means a data that is both complex and high-dimensional (very large).

But detecting Alzheimer's disease through MRI scan images is possible only the amyloid is already accumulated around brain cells in the AD stage. And at this stage it's too late to know about the disease.

Blood Plasma: So far, only amyloid biomarkers were used to detect Alzheimer's disease. But by using non-amyloid biomarkers Alzheimer's disease can be detected in early stages. Using amyloid biomarkers will only give bounded information about progress and are inefficient in detecting the disease before remarkable accumulation of amyloid around the brain cells. By using blood plasma Alzheimer's disease can be detected in early stages with the help of few non-amyloid proteins present in plasma, thrifty and convenient as well [15].

III. EXISTING METHODS

While modern clinical, cerebrospinal fluid and neuroimaging studies are highly accurate in diagnosing Alzheimer's disease, these methods are prohibitively expensive for extensive screening. Furthermore, these technology and specialized services are not voluntarily available to everyone, such as rural seniors and ethnic minorities, limiting their use as AD screeners. A blood-based test, on the other hand, would give a speedy and thrifty way of screening for Alzheimer's disease at the high-population level, therefore expanding worldwide access to care [6].

As a part of multi-stage approach, this sort of blood test would be an absolute initial screening tool, with advanced neuroimaging, clinical,and/or cerebrospinal fluid investigations available for screen-positive patients. On top of this, a precise and simple screen would result in a thrifty method of screening for clinical trials. Molecular biomarkers have traditionally focused on single molecules, but as proteomic, genomic, and metabolomic technology advances, it is becoming more viable to construct classifiers based on a variety of complicated disease signatures [6].

List of available datasets for Alzheimer's disease prediction is shown below:

1	Kaggle [20]	Alzheimer's dataset. The data is manually collected and consists of MRI images.
2	ADNI [21]	*A. Alzheimer's Disease Neuroimaging Initiative began in 2004. The main goal of ADNI is to detect AD at the earliest possible stage. The data is available in many types like clinical, genetic, MRI image, PET image and biospecimen.*
3	Alzheimer's Disease and Healthy Aging Data [22]	Published by Centers for Disease Control and Prevention, USA on November 20, 2020. The data is available in csv, JSON, RDF, XML formats.

While the hunt for blood basednon-amyloid biomarkers of Alzheimer's disease has been mainly futile for years, substantial progress has lately been made. Ray et al. looked at a variety of non-amyloid plasma-based proteins and came up with a model of algorithm that correctly categorized AD and predicted the disease progress between mild cognitive impairment and AD. Very recently, Booij et al. and Rye et al. completed few studies using arrays of gene expression and identified that overall diagnosis accuracy ranged from 74% to

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 12, December 2022

92%. A serum-based algorithm was developed based on the Texas Alzheimer's Research Consortium's longitudinal cohort that had good diagnostic accuracy, properly categorizing 95 percent of AD patients and controls [6].

There is a conspicuous requirement for precise and powerful analytic and prognostic biomarkers for Alzheimer's sickness, and there has been a monstrous ascent in endeavours to track down such markers as of late. It has recently been suggested that the ideal biomarkers would be gotten from fringe blood because of impressive advantages. Fringe blood might be drawn at any centre (or during an in-home visit), however, most facilities can't do lumbar cuts. Progressed neuroimaging strategies are likewise regularly just accessible in large clinical offices in thickly populated regions. While practically all patients are prepared to persevere through venepuncture, less more seasoned people agree to lumbar cut, and many can't go through neuroimaging for an assortment of reasons [4].

Biomarkers can recognize endophenotypes inside AD populaces connected with specific sickness pathways, as well as give more open, fast, cost-and time-compelling methods for assessment. Designated medicines custom fitted to endophenotype status may be investigated once the endophenotype status is not set in stone [4].

Plasma cholesterol, for instance, is an important biomarker in the treatment of coronary conduit illness since it can recognize a subgroup of people whose atherosclerosis is pathogenically associated with hypercholesterolemia. Plasma cholesterol levels can be utilized to survey the viability of reductase inhibitor treatment. It would be a huge advance forward in this area in case this calculated structure could be meant AD. The revelation of a favourable to incendiary endophenotype of Alzheimer's infection could prompt designated therapeutics for a subgroup of patients, with those with an over-articulation of the supportive of provocative biomarker profile profiting from mitigating drugs and those with an under-articulation profiting from calming drugs [4]. Enormous scope proteomic profiling of the blood has been broadly taken on to concentrate on cardiovascular sicknesses and maturing, bringing about the ID of novel biomarkers and giving organic comments to illness stages, on account of on-going advances in ultrasensitive andhigh- throughput protein estimation innovations. Thus, the vicinity augmentation measure procedure was utilized to inspect the protein profiles of AD plasma in this examination. After evaluating 1160 plasma proteins,it was identified that 429 plasma proteins were deregulated in patients with AD in a Hong Kong Chinese AD accomplice comprising of 106 AD patients and 74 solid controls for whom all the segment information, cerebrum locale volumes, intellectual measures, and plasma biomarker levels were cordial. Likewise,a 19-protein based biomarker board was found that addresses the plasma proteomic mark that belong to Alzheimer's sickness and affirmed its brilliant exactness for distinguishing AD and related endophenotypes in a different example [5].

Besides, it was observed that specific proteins in plasma are deregulated at various periods of Alzheimer's sickness. Subsequently, fostering of a total profile of the AD plasma proteome and a superior presentation plasma biomarker board for AD was done, laying the preparation for the advancement of a blood-based test for AD screening and arranging [5].

Although there is no remedy for Alzheimer's infection, analysts are striving to find novel treatment techniques that may help slow or stop the sickness. Such medicines are aimed at patients in the beginning phases of the illness before patients have experienced serious cell harm when treatment is bound to be helpful. The utilization of perceived biomarkers, for example, those dependent on Aβ in the cerebrospinal fluid and sub-atomic imaging of cerebrum amyloidal affidavit utilizing positron outflow tomography, is encouraged to help early determination [15].

Regardless of progressions in the production of amyloidal biomarkers and tests for early Alzheimer's infection identification, developers face two key difficulties. Amyloid-based biomarkers just give a restricted measure of data in regards to disease etiology and pathways. Besides, testing dependent on these biomarkers can't identify those in danger of Alzheimer's infection before there is a crucial amyloid-beta collection in the mind. There is a requirement for the biomarkers that are capable of distinguishing natural cycles that happen before amyloid-beta development in the mind during infection movement. Such biomarkers may assist specialists with bettering comprehending the condition, just as distinguish individuals in the beginning phases of the infection and make novel medicines [15].

A study was also done with hybrid model, a combination of VGG19 and other additional layers along with Convolutional Neural Network for detecting and classifying various stages of AD [16].

High throughput tests have as of late become more practical on account of the presentation of board-based proteomics. SOMA scan, for instance, accommodates concurrent estimation of 41000 proteins and has effectively been utilized in Alzheimer's exploration. This technique still can't seem to be utilized to find blood-based biomarkers for AD-related qualities in symptomless individuals. Although most blood protein-based biomarker studies have used gatherings of individuals, a single exploration of APOA1 blood DNA methylation and protein level utilized 24 twin sets that were dissonant in verbose memory work. Dissonant verbose memory has been connected to two APOA1 CpG districts. Nonetheless, in light of the little example size accessible, the APOA1protein level was not shown to vary across conflicting twins [12].

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 12, December 2022

The summarisation of the referred papers is shown below:

Sl No	AUTHOR	YEAR	DATASET	FEATURES	METHOD	ACCURACY
[1]	Elisabeth H. Thijssen*et al.*	2020	ARTFL	Plasma phosphorylated tau181	-	-
[2]	Alessandro Rabbito et al.	2020	-	Biomarkers	-	-
[3]	LinLiu	2020	Dem@Care project	The spectrogram features from speech data	Logistic regression	-
[4]	Sid E. O'Bryant*et al.*	2011	TARC	A Multiplex Serum Protein-based marker	Random Forest	90%
[5]	Yuanbing Jiang *et al.*	2021	Hong Kong Chinese AD Cohort	Plasma Biomarkers	Linear Regression	-
[6]	Sid E. O'Bryant*et al.*	2011	TARC	30 Serum protein markers	Linear Regression	89%
[7]	Sid E. O'Bryant*et al.*	2011	TARC, ADNI	Protein Biomarkers	Random Forest	70%
[8]	Jinny Claire Lee *et al.*	2019	-	Fluid Biomarkers	-	-
[9]	Henrik Zetterberg *et al.*	2019	-	Blood-based Molecular Biomarkers	-	-
[10]	Nicholas J. Ashton *et al.*	2019	AIBL KARVIAH	Plasma Proteins	Support Vector Machine (SVM)	81%
[11]	A. Hye*et al.*	2006	2- DGE	Proteome-based plasma biomarkers	Support Vector Machine (SVM)	56%

IV. DATASET AND PROPOSED METHODOLOGY

A study on Alzheimer's Disease Neuroimaging Initiative (ADNI) data set gave the information on how blood plasma proteins are influential over causing Alzheimer's diseases and that was utilized in this examination. ADNI, a private-public association began in 2004 by Dr. Michael W. Weiner, the prime investigator. In each phase of the study new participants were recruited. ADNI–1, ADNI–GO, ADNI-2, ADNI-3 are the different phases of studies made by ADNI. The primary goal of ADNI is to detect the dementia at the earliest possible stage and to improve the data access for scientists worldwide.

The participants were diagnosed at their first visitation and the participants blood samples were collected for CN, MCI and AD diagnosis according to guidelines in [23]. After twelve months of the first diagnosis, plasma samples from the subset of these participants were collected. This examination has taken around 9,500 samples of blood plasma.

The following is the proposed methodology:

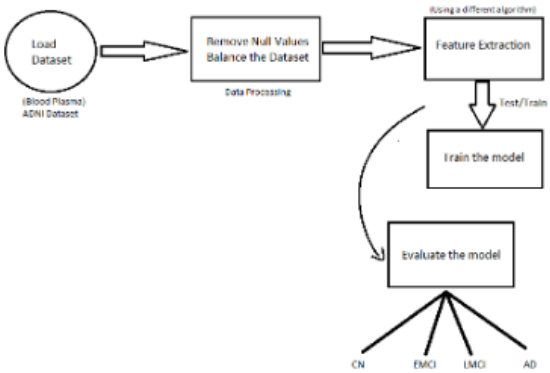

Fig 2: Proposed model

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 12, December 2022

Step 1: Load the data set

Step 2: Standardize the values.

Step 3: Among all the features in the dataset, consider only those features which will play important role in determining the stages of Alzheimer's disease.

Step 4: With a part of records in the dataset, the algorithm is trained.

Step 5: After training, using rest of the records in dataset the to evaluate the model i.e., to determine the stage of Alzheimer's disease, if the person is in CN, EMCI, LMCI, AD stage.

V. DATA PREPROCESSING

Before feeding the data to the model, the data is first prepared to make it acceptable and befitting for the model. The raw data is often known to have lot of noise, missing values and probably in a unusable format for a machine learning model. Thus, the available data set needs to be cleaned to make it acceptable by model and this might also increase the accuracy.

Normalization of data is done to bring all the values between certain range i.e., between 0 and 1. The following formula is used for normalization of the dataset:

$$\left[\frac{Considered\ point - min\,range}{max\,range - min\,range} * (new\,min - new\,max)\right] + new\,min$$

Handling the missing values is the next step in data pre-processing. There are ways to handle the missing data: by deleting that particular row, by filling with zero or by calculating the mean. As the dataset contains numeric values, the strategy of calculating the mean of that column which contains missing or null values and filling the missing values with this mean is appropriate.

After this, dividing of dataset for training and testing is done. It is one of the crucial steps in pre-processing of data and enhances the performance of the model as well. Testing the model on a different dataset other than what was used for training might bring down its performance so the same dataset is divided for training and testing. 80% of dataset is used for training and 20% is used for testing.

VI. MODAL

Recurrent Neural Network (RNN) is a type of Neural Network. In RNN, the current state takes the output of previous step as input. Traditional neural networks had all the inputs and outputs independent of each other, but in cases where we have to predict next word, the previous word plays an important role in prediction of next consecutive word. There is a need to remember previous state or previous word to predict the next word and hence there is a need to remember the previous words. This issue was thus solved by using RNN which has the concept of hidden layers. This important feature will thus reduce the complexity of huge number of parameters by remembering the previous states and giving them as input to the next hidden layer.

First, a single input is fed to the network. With the previous state and current input, the current state is calculated. Then the ht of current state becomes ht-1 of the next step. This can be repeated as many times as required according to the problem statement. The output is taken from the final current state once all the time steps are done. The error is generated by comparing the final output with the target output or actual output. Finally, back-propagation is done i.e., the error is fed back to the network for renewal of weights and that's how RNN is trained.

This examination uses bidirectional RNN where we duplicate the chain of RNN so that the inputs are refined in both forward and backward directions. At the both ends the input is fed and the refining of network is done.

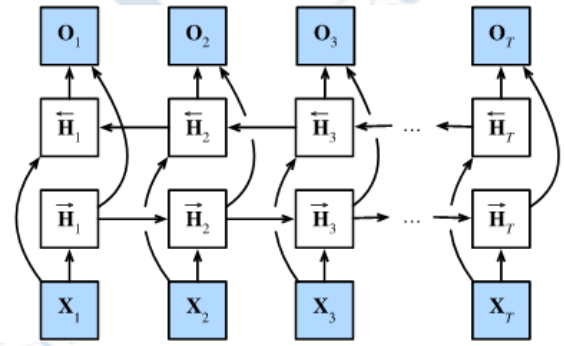

Fig 3: Architecture of a bidirectional RNN [24]

VII. RESULT

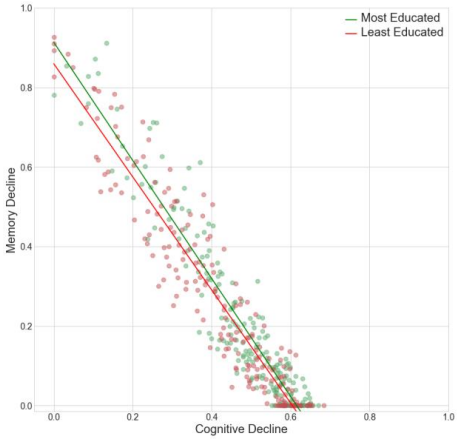

Fig 4: Memory decline vs Cognitive decline

The above figure shows how cognitive decline affects memory decline. As the cognitive decline increases, the patient looses memory.

After data pre-processing, as per the studies made, RNN would be the best modal to detect the Alzheimer's disease in the early, keeping in mind the complex data. To get an

accuracy around 92-95% is the aim of this examination Furthermore, using bidirectional recurrent neural network would help in getting more accurate results.

REFERENCES

[1] Quitterer, U., &AbdAlla, S. (2020). Improvements of symptoms of Alzheimers disease by inhibition of the angiotensin system. Pharmacological research, 154, 104230.

[2] Rabbito, A., Dulewicz, M., Kulczyńska-Przybik, A., &Mroczko, B. (2020). Biochemical markers in Alzheimer's disease. International journal of molecular sciences, 21(6), 1989.

[3] Liu, L., Zhao, S., Chen, H., & Wang, A. (2020). A new machine learning method for identifying Alzheimer's disease. Simulation Modelling Practice and Theory, 99, 102023.

[4] O'Bryant, S. E., Xiao, G., Barber, R., Reisch, J., Doody, R., Fairchild, T., ...& Texas Alzheimer's Research Consortium. (2010). A serum protein–based algorithm for the detection of Alzheimer disease. Archives of neurology, 67(9), 1077-1081.

[5] Jiang, Y., Zhou, X., Ip, F. C., Chan, P., Chen, Y., Lai, N. C., ...&Ip, N. Y. (2021). Large-scale plasma proteomic profiling identifies a high-performance biomarker panel for Alzheimer's disease screening and staging. Alzheimer's & Dementia.

[6] O'Bryant, S. E., Xiao, G., Barber, R., Reisch, J., Hall, J., Cullum, C. M., ...& Diaz-Arrastia, R. (2011). A blood-based algorithm for the detection of Alzheimer's disease. Dementia and geriatric cognitive disorders, 32(1), 55-62.

[7] O'Bryant, S. E., Xiao, G., Barber, R., Huebinger, R., Wilhelmsen, K., Edwards, M., ...& Alzheimer's Disease Neuroimaging Initiative. (2011). A blood-based screening tool for Alzheimer's disease that spans serum and plasma: findings from TARC and ADNI. PloS one, 6(12), e28092.

[8] Lee, J. C., Kim, S. J., Hong, S., & Kim, Y. (2019). Diagnosis of Alzheimer's disease utilizing amyloid and tau as fluid biomarkers. Experimental & molecular medicine, 51(5), 1-10.

[9] Zetterberg, H., & Burnham, S. C. (2019). Blood-based molecular biomarkers for Alzheimer's disease. Molecular brain, 12(1), 1-7.

[10] Ashton, N. J., Nevado-Holgado, A. J., Barber, I. S., Lynham, S., Gupta, V., Chatterjee, P., ...&Hye, A. (2019). A plasma protein classifier for predicting amyloid burden for preclinical Alzheimer's disease. Science advances, 5(2), eaau7220.

[11] Hye, A., Lynham, S., Thambisetty, M., Causevic, M., Campbell, J., Byers, H. L., ...&Lovestone, S. (2006). Proteome-based plasma biomarkers for Alzheimer's disease. Brain, 129(11), 3042-3050.