International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Prediction of Alzheimer's Disease using Machine Learning Technique

Vindhya G B[1], Mahera Alam[2], Medha Mansi[3], Muskan Kedia[4], Anjal D[5]

[1]*Assistant Professor, Dept. of Computer Science Engineering, Sapthagiri College of Engineering, Bangalore, India*
[2-5]*Students, Dept. of Computer Science Engineering, Sapthagiri College of Engineering, Bangalore, India*

Abstract-*Alzheimer's disease (AD) is a neurodegenerative disease and the most common cause of dementia in older adults. The part of brain that gets affected in this disease is hippocampus degeneration. Detection of Alzheimer's disease at preliminary stage is very important as it can prevent serious damage to the patient's brain. It becomes dangerous and sometimes fatal in case of people of 65 years of age or above. The main objective of this paper is to use machine learning algorithms that is CNN and feature extraction and selection to predict the Alzheimer's disease and build a useful model. The dataset is taken in the form of MRI images. The proposed approach detects the various stages of Alzheimer's Disease such as moderate-demented and non-demented using CNN algorithm.*

***Key Words*: Alzheimer's disease, CNN, hippocampus**.

1. INTRODUCTION

Alzheimer's Disease (AD) is the most common cause of Dementia in people of the age 65 years and above. It is a progressive and irreversible neurological disease which follows a distinct pattern of brain damage as the disease progresses. Alzheimer's disease is a very common type of dementia. Dementia is an umbrella term describing a variety of diseases and conditions that develop when nerve cells in the brain (called neurons) die or no longer function in a normal way. The death or malfunction of neurons causes abnormalities in one's memory, behavior, and ability to think in a clear way. In Alzheimer's disease, these brain conditions eventually impair an individual's ability to perform even basic functions such as walking, speaking, and swallowing. Development of AD can be classified into three stages. First, is the asymptomatic stage, changes in the brain, blood, or cerebrospinal fluid (CSF) may begin to occur without the patient showing any particular symptoms. After the first stage comes the second stage, that is mild cognitive impairment (MCI) stage, memory complaints and other cognitive behaviour may start to be noticeable for the patients themselves and for close family or friends, which affects day to day activities but the symptoms are mild. In the final stage of the disease, or the dementia stage, memory, thinking, and behavioral symptoms are evident and significant, and it is noticeable. The neurons of brain starts degenerating and the synapses are slowly dissolved. The patient loses the ability to respond to the environment. The hippocampus abnormally shrinks in its size 2.2 to 5.9 percent annually. AD is the 6th leading cause of death in the United States. According to a report from the Alzheimer's Association, AD and other dementias are predicted to cost the nation $1.1 trillion in 2050. Currently, 5.8 million Americans are living with AD and by 2050 this number is expected to rise to closely 14 million. It is difficult to manually diagnose AD, or any other types of dementia at an early stage before most of its symptoms are noticeable. Therefore, it is important to use computer analysis to analyse as much patient's data as possible for a better evaluation and more accurate diagnosis. The methods detect AD at a very late stage when all of the symptoms appear. However, this research will focus on developing an evolving framework to effectively diagnosis and predict AD at a very early stage using the data collected for AD patients[2]. The framework will continuously use large sets of related data to AD patient collected from multiple sources like medical sources, lifestyle and demography. The datasets are taken in the form of MRI images. MRI, which reflects the spontaneous blood-oxygen-level dependent signal fluctuations when a subject is not performing an explicit task. The detection of Alzheimer's disease using conventional technique is time consuming, so we apply machine learning technique CNN to predict the Alzheimer's disease.

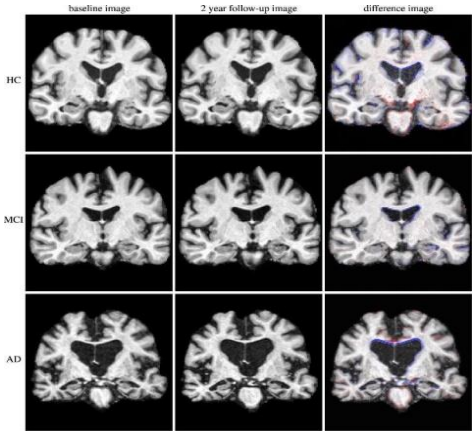

Fig-1: Three examples of MR images (brain-extracted) from the ADNI database showing different AD stages.

2. METHODOLOGY

Detection of Alzheimer's disease at prodromal stage is very important as it can prevent serious damage to the patient's brain. The detection of the Alzheimer's disease is

done the medical professionals through certain steps such as CT scan. The detection of the Alzheimer's disease is time consuming using CT scan. Hence we are applying the machine learning technology to predict the Alzheimer's disease. The proposed approach detects the various stages of Alzheimer's Disease such as moderate-demented and non-demented using CNN algorithm. It reduces the time required to predict the output and can be used for real time predictions. We describe the datasets used in this study and how the data was pre-processed before the machine learning task. Feature extraction using principal component analysis and feature selection techniques were also employed. After the data preprocessing is done, the efficient machine learning algorithm that is CNN is applied to predict the disease and categorized it into moderate demented and non demented.

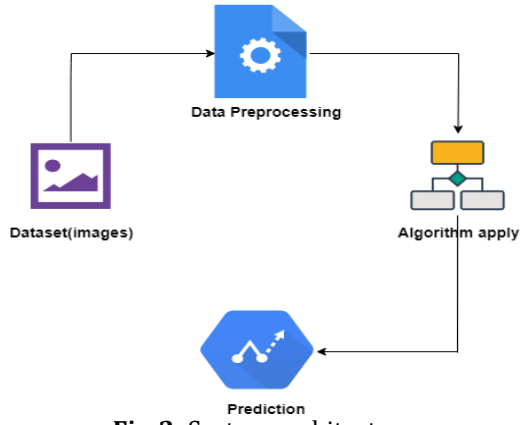

Fig-2: System architecture

2.1 Data preprocessing

Data preprocessing refers to all transformations on the raw data before it is fed to the machine learning or deep learning algorithm. For instance, training a convolutional neural network on raw images will probably lead to bad classification performances. The preprocessing is also important to speed up training such as clustering and scaling technique. Real- world data is often incomplete, inconsistent, and lacking in certain behaviours or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues.

2.2 Feature selection and extraction

Feature selection is an effective strategy to optimize the predictive performance of machine learning algorithms. In this study, we have tested a number of feature selection techniques including SelectKBest. Sequential Forward Selection Sequential Backward Selection, and Recursive Feature Elimination implemented in the sklearn package. The models with the best performance resulted from the SelectKbest, which is a univariate correlation feature selection method. We explored values for different number of features (k) with the highest correlation to the dependent variable (AD/control)[1].

2.3 Convolution Neural Network

Artificial Neural Networks are used in various classification task like image, audio, words. Different types of Neural Networks are used for different purposes, for example for predicting the sequence of words we use Recurrent Neural Networks more precisely an LSTM, similarly for image classification we use Convolution Neural Network. In the next section, we are going to build basic building block for CNN.

In a regular Neural Network there are three types layers.

Input Layer: It's the layer in which we give input to our model. The number of neurons in this layer is equal to total number of features in our data (number of pixels in case of an image).

Hidden Layer: The input from Input layer is then feed into the hidden layer. There can be many hidden layers depending upon our model and data size. Each hidden layers can have different numbers of neurons which are generally greater than the number of features. The output from each layer is computed by matrix multiplication of output of the previous layer with learnable weights of that layer and then by addition of learnable biases followed by activation function which makes the network nonlinear.

Output Layer: The output from the hidden layer then fed into a logistic function like sigmoid or softmax which converts the output of each class into probability score of each class. The data is then fed into the model and output from each layer is obtained this step is called feed forward, we then calculate the error using an error function, some common error functions are cross entropy, square loss error etc. After that, we back propagate into the model by calculating the derivatives. This step is called back propagation which basically is used to minimize the loss.

The implementation was carried out using Python programming language with Scikit-learn, Pandas, Numpy, and Matplotlib libraries using CNN algorithm.

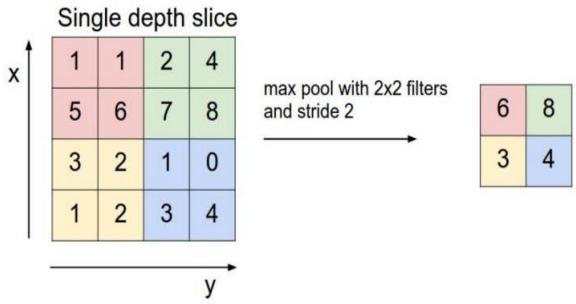

Fig-3 Filters of a matrix

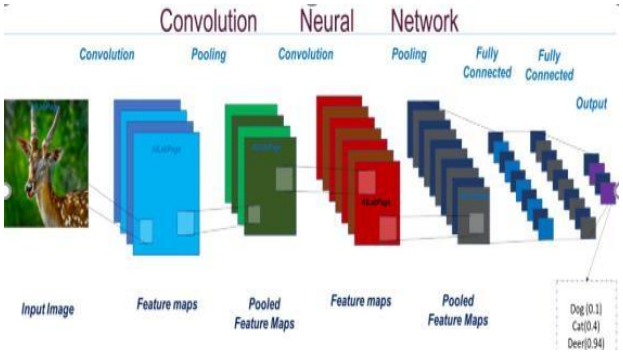

Fig-4: Convolution Neural Network

The main advantage is their accuracy in image recognition problems. The main advantage of CNN compared to its predecessors is that it automatically detects the important features without any human supervision. For example, given many pictures of cats and dogs it learns distinctive features for each class by itself. CNN is also computationally efficient.

3. CONCLUSION AND FUTURE WORKS

In this paper, we used machine learning algorithms along with data preprocessing and principal component analysis (PCA) and feature selection techniques to classify patients with mild AD and MCI from health volunteers using data obtained from MRI images taken from ADNI datasets.

The CNN algorithm has been employed to carry out he prediction. Compared with traditional methods, the proposed method has achieved about 20% of improvement on the classification accuracy, suggesting that neural network is a powerful tool for the diagnosis of neurological diseases. Based on our work, the same or similar methods can be used to diagnose other neurological diseases providing a intelligent healthcare systems. Potential future work includes evaluating our method on larger data sets and applying it to the diagnosis of other neurological diseases.

REFERENCES

1. J. Li et al., "Feature selection: A data perspective," ACM Comput. Surv. CSUR, vol. 50, no. 6, p. 94, 2018.
2. M. Fernández, A. L. Gobartt, and M. Balañá, "Behavioural symptoms in patients with Alzheimer's disease and their association with cognitive impairment.," BMC Neurol., vol. 10, no. 1, p. 87, Jan. 2010.
3. Mohamed Mahyoub, Martin Randles, Thar Baker, Po Yang,"Effective Use of Data Science Toward Early Prediction of Alzheimer's Disease" 2018 IEEE 20th International Conference on High Performance Computing and Communications.
4. Ibrahim Almubark1, Lin-Ching Chang1, Thanh Nguyen1, Raymond Scott Turner2, and Xiong Jiang3, "Early Detection of Alzheimer's Disease Using Patient Neuropsychological and Cognitive Data and Machine Learning Techniques",2019 IEEE International Conference.
5. Dana AL-Dlaeen, Abdallah Alashqur, "Using Decision Tree Classification to Assist in the Prediction of Alzheimer's Disease", 2014 6th International Conference on CSIT.