Diagnosis of Alzheimer's Disease using Machine Learning

Priyanka Lodha
Department of EnTC
Vishwakarma Institute of Technology
Pune, India
priyanka.lodha94@gmail.com

Ajay Talele
Department of EnTC
Vishwakarma Institute of Technology
Pune, India
ajay.talele@vit.edu

Kishori Degaonkar
Department of EnTC
Vishwakarma Institute of Technology
Pune, India
kishori.degaonkar@vit.edu

Abstract— **Machine learning is being widely used in various medical fields. Advances in medical technologies have given access to better data for identifying symptoms of various diseases in early stages. Alzheimer's disease is chronic condition that leads to degeneration of brain cells leading at memory enervation. Patients with cognitive mental problems such as confusion and forgetfulness, also other symptoms including behavioral and psychological problems are further suggested having CT, MRI, PET, EEG, and other neuroimaging techniques. The aim of this paper is making use of machine learning algorithms to process this data obtained by neuroimaging technologies for detection of Alzheimer's in its primitive stage**.

Keywords— Linear Regression, Machine Learning algorithm, Alzheimer's disease symptoms, Random Forest, Gradient Boosting Algorithm, Neural Network.

I. INTRODUCTION

Alzheimer's disease (AD), a regular dementia type, a distant syndrome which develops retrogression in the intellectual capabilities and regular operating [1]. AD is a syndrome which leads to continuous degeneration of memory and other important mental functions. At first, someone with Alzheimer's disease might get the symptoms of mild confusion and have difficulty remembering things. Alzheimer's is a leading cause of dementia in today's world. Alzheimer's disease damages the brain, resulting in death of brain cells and decline of memory, and thinking, and affects behavior. The risk of acquiring Alzheimer's disease increases with increasing age. While there is no proven cure or method to prevent Alzheimer's, there are a few efficient treatments that can stop or limit its progression.

Dementia is a growing and major health issue in public, of all the cases of Alzheimer's disease about 70% of cases suffer from dementia. Though this condition is more occurring than some cancers, it has received significantly less investment in research and health initiatives for public. Urgent research is needed to investigate this problem for getting better and efficient solutions for Alzheimer's disease.

Machine Learning is an application of data analysis. These algorithms try to build a model that predicts desired output for available data through statistical and predictive features. These models can identify patterns and correlations in data provided and eventually try to make decisions without or with minimal intervention from human.

We determine to design a Machine Learning model to accurately predict Alzheimer's disease of a person from given parameters which consist of different cognitive and medical factors. RAVLT tests, MOCA and FDG score etc are used for prediction. Alzheimer's disease Neuroimaging Initiative (ADNI) has provided a base where it provides research data for a significant number of patients and controls that have been through the diagnosis processes (ADNI1, ADNI2, and ADNI-GO). We have chosen ADNI2 for the training.

II. METHODOLOGY

There has been considerable research in field of Alzheimer's disease. A comprehensive algorithm has been implemented to predict Alzheimer's from available data like whole brain volume which is extracted from MRI scans and other cognitive and biological features.

Clustering algorithms have also been implemented along with Fuzzy interference system [2]. The algorithms Logistic Regression, Support Vector Machine, Gradient boosting, Neural Network, K-Nearest Neighbor, Random Forest are implemented. Deterioration is defined as "the scientific problem of the cognitive energy is symbolized by the serious global decrease in mental work is especially not because of adjustment in the carefulness [3].

III. DATA ANALYSIS

Using descriptive statistics, we first examine the questionnaire data using. By implementing a framed coding structure, we examine on the transcripts of direct content in language of design session, which was evaluated by integrating concepts. We approved and concentrated in the coding structure before examining the transcripts.

Female sex was associated with a larger caudate nucleus volume. Comparable to the results of the present study, gender differences in sub cortical volumes with large volumes in females have been reported in other studies, 34 although such results have been inconsistent.17, 34 Furthermore," men have been found to experience greater age-related atrophy of both cortical and sub cortical volumes than women [4]. Analysis of males and females is done separately. We calculated ratio between.

IV. Materials and Methods

Data utilized in these tasks is taken by Alzheimer's Disease Neuroimaging Initiative (ADNI) index (http://adni.loni.usc.edu). A National Institute of Bio-medicinal Image and Bio-engineering (NIBIB), a National principle on developing and, Food and Drug had launched ADNI project in year 2003. To calculate the development for Medical Council of India and advance Alzheimer's disease in the ADNI goal is to investigate whether Positron Emission Tomography (PET), sequential Magnetic Resonance Image (MRI), and the biotic markers, neuropsychological and the objective evaluation are being connected. Total details of information is taken from http://www.loni.usc.edu/ADNI/. "Very" premature stage of an Alzheimer disease. Diagnosis of the distinct causes an Alzheimer disease could be of great aid to the clinicians and researcher for development of new and effective treatments and monitoring mechanisms. The ages between 55 and 90 were applied in the ADNI study.

A. Support Vector Machine (SVM)

Support Vector Machine Algorithm administered researched technology with associated method that analyses data used for Classification and Regression. SVM is used for both Classification and Retrogression challenges. SVM is popular algorithm. SVM is a fast and dependable Classification algorithm that performs very well with a limited amount of data. Microbe's al. separated the whole-intellect within primary sectors, restricted among the one-dimensional SVM [5]. Fung. Correlated change connecting the SVM algorithm and fisher limited discriminate [6]. The technique, pending an inadequacy of an intellect MRI, enhances an achievement of innovation [7]. There are various types of kernels in SVM. A kernel is a similarity function. Machine learning algorithm provides kernel as domain expert. Equal type of two inputs is present for spiting. Example of kernel are polynomial, Gaussian, Sigmoid, ANOVAs Radial Basis, Gaussian Radial basis Function (RBF), Bessel Function, Laplace RBF, Hyperbolic tangent.

$$K\left(\overline{x}\right) = \begin{cases} 1 & \text{if } \|\overline{x}\| \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Window Function

B. Gradient Boosting Algorithm

Gradient boosting is a supervised learning technique. Gradient boosting is effective techniques for building predictive models. Regression, Ranking and Classification perform different task to perform state of art accuracy. Gradient boosting model is effective in the machine learning algorithm for Regression and Distribution issues, which build group of poor prevision models, mostly decision trees. The module is built in a correlative technique similar to other methods.

To generate an accurate model multiple single faint module are collectively used. Gradient boosting is a prolongation of boosting the development of additively developing faint models is methodize as a gradient descent algorithm over a function. To create a relation with the numerical structure, the gradient-descent planted the regulation of a boosting technique have being copied [8]

Moreover, the Gradient Boosting has shown reasonable achievement along with practical operations, as well as in the different machine-teaching and fact-drilling interrogation [9].

$$F(x) = \sum_{i=1}^{M} \gamma_i h_i(x)$$

Approximation Function

C. Neural Network

The neural network is wide matrix of neurons and interconnected group of nodes. Artificial neurons are a collection based on units or nodes. There is an exchange of signals from one to another between each artificial neuron. In ANN the problem is to be solved in a way like an intelligent brain. In due time ANN have been used in multiple tasks like medical diagnosis, machine translation, and computer vision. The purpose was to identify the indications of an ANN present in drug, eminently in the Alzheimer disease [10]. Earlier investigation indicates ANN remain superior in the anticipating the conclusion than academic numerical analysis [11]. According to function expert straight feed forward (FF) neural system created ANN which is satisfactory to observe the coordination of an input picture [12].

D. K-Nearest Neighbor (KNN)

The K-Nearest Neighbor is not an invariable technique. KNN is reason-based learning. KNN is machine learning algorithm to find matching ratings and average ratings of top KNN. For Classification of different objects KNN method is used. KNN is a dominant tool, simple algorithm to understand and to execute. A peculiarity of the KNN algorithm is that it is sensitivity to local structure of the data.

E. Random Forest

Random Forest is Supervised Classification algorithm. Random Forest is adjustable, effortless to use machine learning algorithm. Random Forest is one of the incredible and most effective machine learning algorithms. Random Forest uses both Classification and Regression. Classification and Regression are ensemble learning methods that manage by constructing a multitude at training time and outputting the classification mode of prediction mean. We can maintain the accuracy even if the data is not present. For identification of the important features from the training dataset we can used Random Forest Algorithm.
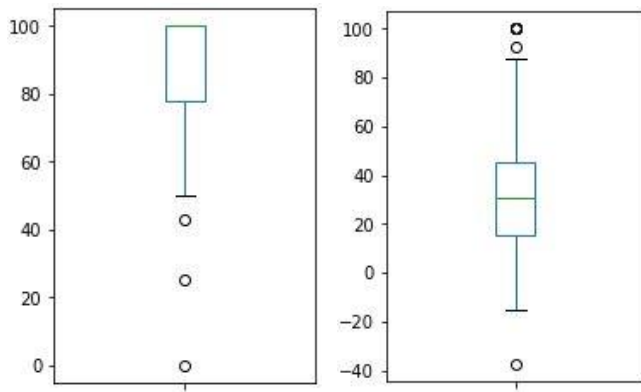
F. Graphs Analysis

Fig1. RAVLT % forgetting score for subjects with AD and normal controls respectively.

As evident from the scatter plot figure 1, those who have Alzheimer disease their RAVLT percentage forgetting score is very high and those who have don't have Alzheimer disease their RAVLT percentage forgetting score is not high. Evaluation of oral learning and memory is done by Rey Auditory Verbal Learning Test (RAVLT). RAVLT is an efficient neuropsychological instrument for learning strategies, rate of learning, short term auditory-verbal, retroactive, proactive interference, retention of information, confusion in memory processes, and difference between learning and retrieval.

G. Scatter Plots

Similar to line graphs, scatter plots use vertical and horizontal axes. Data points are plotted on this graph. It depicts how one variable affects other. This relationship is termed as correlation between the variables. The points on the Scatter (XY) Plot depict the relationship between two different sets of data. Scatter plot summarized a set of two variables (bivariate data). It is usually drawn before fitting parameters for a regression curve (or line) or figuring out a linear correlation coefficient. Along with aid in interpretation of the data in consideration and giving their regression model or correlation coefficient, it gives a visual representation of the bivariate data's relationship.
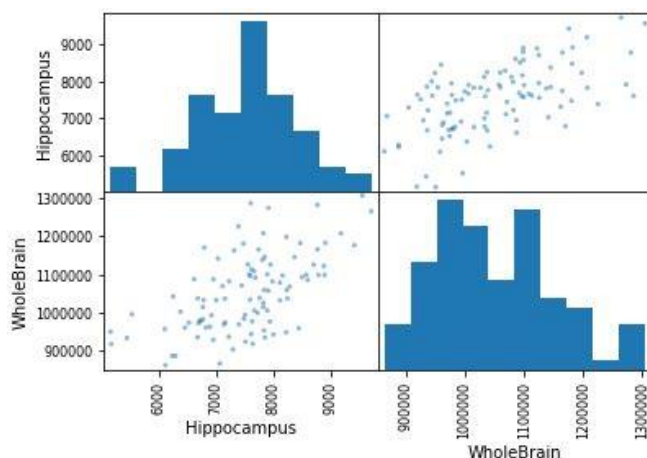

Fig2. Scatter plot for whole brain and hippocampus volume.

As evident from the scatter plot figure 2, the hippocampus volume is strongly correlated with whole brain volume. The more the whole brain volume the more is the hippocampus volume.
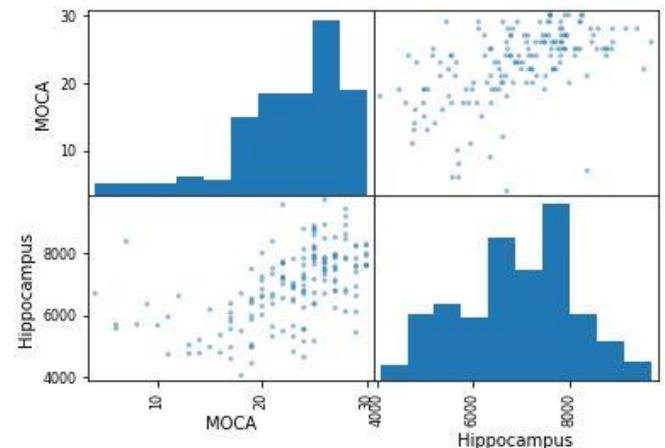

Fig3. Scatter plot for MOCA and hippocampus volume.

As evident from the scatter plot figure 3, the more the hippocampus volume higher the MOCA score.

V. RESULTS

A. Accuracy

Accuracy is defined as the number of correct predictions made [13]. It is a different label classification. It is a commonly used presentation measuring criterion in a vast range of implementation. For symmetric datasets it is the finest production measurement criterion i.e. datasets with similar number of fake positives and fake negatives.

It is given as:

Correct positive=A   Fake positive=X
Correct negative=B   Fake negative=Y

$$Accuracy = \frac{A + B}{A + B + X + Y}$$

Accuracy Table

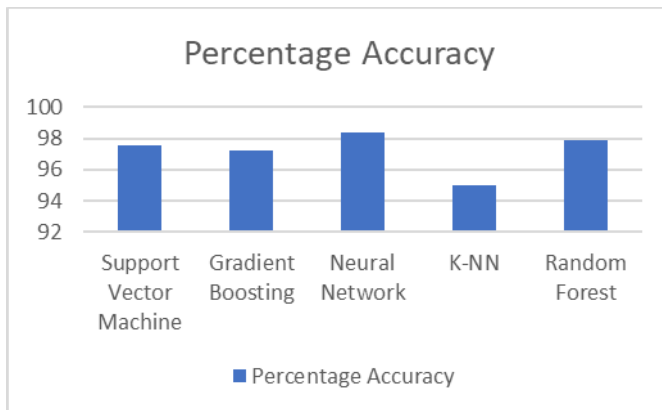| MODEL	ACCURACY
1. SUPPORT VECTOR MACHINE	97.56
2. GRADIENT BOOSTING	97.25
3. NEURAL NETWORK	98.36
4. K-NEAREST NEIGHBOUR	95.00
5. RANDOM FOREST	97.86

Fig4 Accuracy Graph

VI. CONCLUSION

In order to detect Alzheimer's Subjects and analyze images Alzheimer Disease related regions of brain, we use MRI images and process them to get numeric data which in turn is processed using machine learning algorithms. Neural Network and Random Forest have much better performs in accuracy, then other used methods. The implementation of this method will give instant and accurate results. Support Vector Machine and Gradient boosting are also powerful algorithms for classification problems and it works well with the problem at hand. At a premature stage Alzheimer disease can be detected and necessary treatment done at these early states it will minimize the possibility of creating further complications of Alzheimer disease patients.

REFERENCES

[1] Chakraborty A, de Wit NM, van der Flier WM, de Vries HE. 2016 The blood brain barrier in Alzheimer's disease. Vasc. Pharmacol. 89, 12–18.

[2] Breitner JC. Dementia—epidemiological considerations, nomenclature, and a tacit consensus definition. J Geriatr Psychiatry Neurol. 2006; 19(3):129–36.

[3] Kiraly A, Szabo N, Toth E, et al. Male brain ages faster: the age and gender dependence of subcortical volumes. Brain Imaging Behav. 2016; 10:901-910.

[4] Mesrob L, Magnin B, Colliot O et al (2008) Identification of Atrophy Patterns in Alzheimer's Disease Based on SVM Feature Selection and Anatomical Parcellation. Medical Imaging and Augmented Reality 5128:124–132.

[5] Glenn F, Jonathan S (2007) SVM feature selection for classification of SPECT images of Alzheimer's disease using spatial information. Knowl INF Syst 11(2):243–258.

[6] Zhu X, Lei Z, Zi H (2014) A sparse embedding and least variance encoding approach to hashing. IEEE Trans Image Process 23(9):3737.

[7] Torabi, M; Morad zadeh, h; Vaziri, R Ardekani, R.D; Fatemizadeh, E, (2007). "Multiple Sclerosis Diagnosis Based on Analysis of Subbands of 2-D wavelet Transform Applied on MR-images". Computer Systems and Applications. AICCSA '07. IEEE/ACS International Conference on, ON page(s): 717-721.

[8] Coppede, F., Grossi, E., Buscema, M., & Migliore, L. (2013). Application of Artificial Neural Networks to Investigate One-Carbon Metabolism in Alzheimer's disease and Healthy Matched Individuals.

[9] Badnjevic, A., Cifrek, M., Koruga, D., & Osmankovic, D. (2015). Neurofuzzy classification of asthma and chronic obstructive pulmonary disease. BMC Med Inform Decis Mak BMC Medical Informatics and Decision Making,15(SuppI3).