BRAIN DISEASE PREDICTION BY MACHINE LEARNING OVER BIGDATA FROM HEALTHCARE

Vijayan A,
Assistant Professor,
Department of Information Technology,
Velammal Engineering College,
Chennai,Tamilnadu,India.

Athithiyan G,
Student,
Department of Information Technology,
Velammal Engineering College,
Chennai,Tamilnadu,India.

Anandaraj K,
Student,
Department of Information Technology,
Velammal Engineering College,
Chennai,Tamilnadu,India.

Abstract: Due to the rapid growth of computerized techniques in healthcare, it gives raise to the rapid development of computer technology in healthcare for early diagnosis. Especially, recent progress of machine learning in various fields such as marketing, banking, insurance, weather etc. The machine learning will help in predicting the future happenings based on past history. The early diagnosis of any types of disease is not possible at all cases. In order to overcome from this disadvantage, we are introducing the datamining using machine learning algorithms for early detection of disease to prevent loss of life. In this paper, we are discussing about the introduction of machine learning in brain diseases prediction. The decision tree algorithm is used in predicting which type of disease. The types of disease considered are Parkinson's disease, stress, amnesia, bipolar disorder. The symptoms and physical behavior of the user has been considered. The result has been discussed through the graphical representation.

Keywords: Machine Learning, brain diseases, decision tree algorithm.

I.INTRODUCTION

The brain diseases we consider for prediction are Parkinson's disease, stress, amnesia, bipolar disorder. The psychological behavior of a person varies due to the change in environment, working style, genetic factors etc. The brain diseases prediction system includes the various parameters such as head ache, confusion, stiff neck, nausea, insomnia, etc. The person is diagnosed by using CT images, physical verbal input. In this paper, the input is collected from the user and gets saved in database. The datamining process is done with existing dataset to train the machine with algorithms. The datamining is a process of extracting knowledge from the huge data. It is also known as Knowledge Discovery in Databases. It has various algorithms such as Classification, Clustering, prediction. The datamining follows the collection of process such as Data Collection, Data Exploration, Data Analysis and Data Visualization. The brain diseases prediction is able to predict the result of the given input using Decision Tree Algorithm. The data might be available in relational databases, file format, etc. The result is viewed through the graphical representation

II.RELATED WORK

Ritika Agarwal et.al done a survey on the information retrieval from lung images. The paper (1) is based on the support vector machine algorithm (SVM) Kernel functions. This designed a framework called Content Based Retrieval System. This framework involves the following steps such as Pre-Processing, Segmentation process and feature extraction. Based on the result from feature extraction or feature selection, the conclusion arrived stating the image is normal image or abnormal image. Feature selection is based

on three types of kernel function partial kernel, linear kernel, and polynomial kernel.

Saranya P et.al does a survey (2) on the available data mining algorithms to classify the data and extract the knowledge from it. It discusses about the difficulties in classification, segmentation, extraction and selection. It compares the different algorithms like Support Vector Machine, Naïve Bayesian classification, Rough set theory, Decision Tree.

Dmitry Ignatov et.al (3) does the study on decision trees and their behavior in arriving to the conclusion. Tree node splitting based on relevant feature selection is a key step of decision tree learning, at the same time being their major shortcoming: the recursive nodes partitioning leads to geometric reduction of data quantity in the leaf nodes, which causes an excessive model complexity and data over fitting. In this paper, the author presented a novel architecture called a Decision Stream.

Kelvin KF Tsoi et.al does the survey (4) on the smoking behavior of the user. The e-cigarette has a small electrical resistance coiled wire in 1.5 ohms which is connected to the positive and negative poles of the device. When the button of e-cigarette is pressed, the resistance coil can be connected with electrical supply under the immersion of some "E-liquid", the coil heats up and transform the E-liquid to vapor, which can be inhaled by the smokers. It monitors the smoking behavior of the user in order to prevent the patient from cancer.

International Journal of Contemporary Research in Computer Science and Technology (IJCRCST)   ISSN: 2395-5325

Volume 4, Special Issue 3 (March '2018)

Proceeding of - National Conference on Recent Trends in Information Technology and Communication,
Conference held at Velammal Engineering College,Chennai,India.

Gonzalez-Alonso P et.al (5) has discussed about the methodology, challenges and design complications in meeting technological development for health care. Since health care deals with sensitive and critical data. The data is dependent on the environment and human behavior. It changes according to region, climatic conditions etc. The authors designed a conceptual framework. It comprises of metadata layer such as cognitive layer, data layer, source and storage layer

   Lakshmi et.al(6) has done a survey of various datamining techniques such as clustering technique, classification technique, prediction algorithm. In the paper, they discussed about the datamining steps or process, types of input data, format of the data and finally data security. They discuss about the architecture of the data mining.

   Arpita Rautet.al(7) has proposed a system to detect the neurological alzheimer' disease. Mostly, these kinds of disease are been affected at the age of 65. They have categorized into training phase and testing or prediction phase. It takes MRI images as a input to the system. It extract the feature from the hippocampus region. The extracted data is used for training using ANN classification.

III.ARCHITECTURE
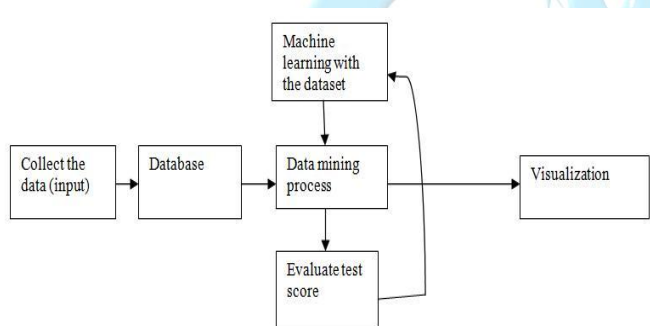
IV.SYSTEM DESIGN
A.   Existing System
   The traditional methodology for finding any kind of disease is through Computerized Technology Images.
      Disadvantages:
      • Not earlier prediction
      • Helps in to take earlier steps to prevent

B.   Proposed System
      The disadvantages of the existing system were overcome by the proposed work. The machine learning algorithms has been introduced in healthcare. This will predict whether the person is having the disability.
         The following are the modules of the proposed work:

      • User Interface design
      • Data mining for individual
      • Finding vulnerability
      • Admin

Table 1: List of Attributes

S.No	Attributes	Description	Values
1	Age	Current age of a person	Numerical values years
2	Gender	Describes the state of being male or female	Alphabetical values M-Male F-Female
3	Head Ache	Pain in the region of head and neck	Yes, No
4	Nausea	Vomiting	Yes, No
5	Slurred Speech	Weak Speech	Yes, No
6	Blurred Vision	Not cleared vision	Yes, No
7	Loss of Sensory Touch	Can't able to sense	Yes, No
8	Memory Loss	Forgot the happenings	Yes, No
9	Confusion	Not clear idea with the work	Yes, No
10	Inability to Recognize	Not able to recognize people, object or place	Yes, No
11	Numbness	Damage, irritation or compression of the nerves	Yes, No
12	Dizziness	Losing one's balance	Yes, No
13	Unconscious	Part of mind is inaccessible	Yes, No
14	Insomnia	Difficulty falling asleep	Yes, No
15	Impaired Voice	Communication disorder or voice impairment	Yes, No
16	Muscle Stiffness	Muscle feel tight and difficult to	Yes, No

International Journal of Contemporary Research in Computer Science and Technology (IJCRCST)   ISSN: 2395-5325
Volume 4, Special Issue 3 (March '2018)

Proceeding of - National Conference on Recent Trends in Information Technology and Communication,
Conference held at Velammal Engineering College,Chennai,India.

		move	
17	Mood Swings	Unstable emotional	Yes, No
18	Weight Loss	Change in weight	Yes, No
19	Difficulty in Swallowing	Not able to swallow food, water.	Yes, No
20	Slow Movement	Due to muscle stiffness, the body cannot function properly	Yes, No
21	Smoking	Habit of having smoking	Yes, No
22	Head Injury	Any injury or gets hurt accidently	Yes, No
23	Alcohol Consuming	Habit of having liquor drinking	Yes, No
24	Work Environment	Discussed about the working nature	Includes all sorts of work
25	Depression	Stressed	Yes, No

Advantages:
- Earlier prediction
- Easy to find solution or remedy

Disadvantages:
- Does not provide all time accurate values

IV. IMPLEMENTATION

A. User Interface Design
The user interface has been designed to interact with the user. The following various attributes are collected from the interface.

B.Datamining for individual
This module involves the training of the system with the dataset using Machine Learning algorithm. This involves the following process: Data Collection, Data Cleaning, Data Modeling and Data Visualization. The analysis is done using decision tree algorithm. Decision tree is used for developing the Brain Diseases Prediction System. The attributes has the value either Yes or No. It is easy to cluster and classify the Patients into various brain diseases. The individual or person as the root node classifies into the various brain related diseases category. It is a recursive algorithm which repeated again to find root node as best classifier. This is sometimes called as Greedy algorithm. It covers both the classification and regression.

- Sensitivity or TPR= TP/P

- Specificity or SPC=TN/N

- Precision or Positive Predictive Value, PPV = TP/ (TP+FP)

- Accuracy, ACC = (TP+TN) / (TP+FP+FN+TN)

- F1 is the harmonic mean of precision and sensitivity.

$$F1 = 2TP / (2TP+FP+FN)$$

The efficient algorithm is chosen based on the above calculated results.

Method	AUC	CA	F1	Precision	Recall
Random Forest	0.375	0.500	0.385	0.385	0.385
SVM	0.600	0.500	0.579	0.440	0.846
Knn	0.400	0.438	0.273	0.333	0.231
Tree	0.350	0.344	0.231	0.231	0.231
Naïve Bayes	0.550	0.531	0.286	0.375	0.231
Logistic Regression	0.375	0.406	0.240	0.250	0.231

Figure 1: Represents the algorithms with accuracy

C. Finding Vulnerability
In this module, the huge amount of data has been fed into the application to find the probability of people will have high degree chance of having disease or not. It can be used in healthcare centre for predicting the number of patients will be affected based on the past history of data.
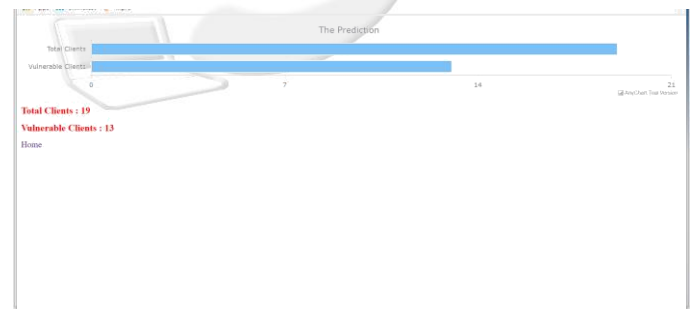

Figure 2: Number of vulnerable patients.

D. Admin
This module maintains the databases. It creates the record for each and every user. Since it deals with medical, sensitive and confidential data, the user has been authenticated and authorized to access.

V. CONCLUSION
Thus the development of brain diseases prediction system has been done. It helps in predicting the future in order to make necessary steps to prevent it from severe effects.In future, it can be used in many areas for predicting the future happenings. It can include other deadly diseases like dengue, cancer, etc to be predicted.

VI.REFERENCES

[1]. Reza Hassanzadeh, John H. Phan and May D. Wang, A Multi-Modal Graph-Based Semi-Supervised Pipeline for Predicting Cancer Survival", 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).

[2]. Saranya P and Satheeskumar B, "A Survey on Feature Selection of Cancer Disease Using Data Mining Techniques",
International Journal of Computer Science and Mobile Computing, Vol.5 Issue.5, May- 2016, pg. 713-719.

[3]. Dmitry Ignatov and Andrey Ignatov, "Decision Stream: Cultivating Deep Decision Trees", 3 Sep 2017 IEEE.

[4]. Kelvin KF Tsoi1, Yong-Hong Kuo and Helen M. Meng, "A Data Capturing Platform in the Cloud for Behavioral Analysis Among Smokers An Application Platform for Public Health Research", 2015 IEEE.

[5]. Gonzalez-Alonso P, Vilar R, Lupiañez-Villanueva F, "Meeting Technology and Methodology into Health Big Data Analytics Scenarios", 2017 IEEE.

[6]. B.N. Lakshmi, G.H. Raghunandhan "A Conceptual Overview of Data Mining", 2011 IEEE conference.pp.27-32.

[7]. "A Machine Learning Based Approach for Detection of Alzheimer's Disease Using Analysis of Hippocampus Region from MRI Scan", IEEE 2017.