IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-3, Mar 2019 E-ISSN: 2347-2693

Prevention of Alzheimer's Disease using Decision tree and Association rule mining Algorithms.

Yash Shetty[1*], Linda John[2], Vikrant Patil[3]

[1,2] Dept. of Information Technology, S.T John college of engineering and management., Mumbai University, Mumbai, India

[*]*Corresponding Author: yashshetty381@gmail.com ,Tel.: 02228960737*

DOI: https://doi.org/10.26438/ijcse/v7i3.10591064 | Available online at: www.ijcseonline.org

Accepted: 12/Mar/2019, Published: 31/Mar/2019

***Abstract*—** Early diagnosis of Alzheimer's Disease is important for the progress of more predominant treatments. Machine learning (ML), a branch of artificial intelligence, provides a variety of probabilistic and upsurge techniques that permits PCs to gain from vast and complex datasets. As a result, researchers concentrate on using machine learning often for diagnosis of early stages of Alzheimer's Disease. This paper represents a review, analysis and critical evaluation of the recent work done for the early detection of Alzheimer's Disease using Machine Learning techniques. Several methods achieved promising prediction accuracies, however they were calculated on different pathologically unproven data sets from different imaging modalities making it difficult to compare among them. Moreover, many other factors such as pre-processing, the number of important attributes for feature selection, class imbalance distinctively affect the computation of the prediction accuracy. To overcome these flaws, a model is proposed which comprise of initial pre-processing step followed by imperative attributes selection and classification is achieved using association rule mining. Furthermore, this proposed model-based approach gives the right path for research in early diagnosis of AD and has the potential to distinguish AD from healthy controls.

***Keywords*——** *Alzheimer's, Mental Disorder, Association Rule Mining.*

I. INTRODUCTION

Alzheimer's disease is a progressive disorder that causes brain cells to degrade and results in loss of life. Alzheimer's disease is the most common cause of dementia — a continuous decline in thinking, behavioural and social skills that disturb a person's ability to function independently. The initial signs of the disease may be forgetting recent events or conversations. As the disease progresses, a person with Alzheimer's disease will develop severe memory ruination and lose the capability to carry out everyday tasks. Current Alzheimer's disease medications may temporarily improve symptoms or slow the rate of failure. These treatments can sometimes help people with Alzheimer's disease to maximize function and maintain autonomy for a time. Different programs and services can help support people with Alzheimer's disease and their custodian. There is very less or no treatment that cures Alzheimer's disease. In progressive stages of the disease, complications from severe loss of brain function — such as dehydration, malnutrition or infection — can result in death[1].

A. Early Stage:

In the initial stage of Alzheimer's, a person may function independently. He or she could drive, work and be part of social activities. Despite this, the person may feel as if he or she is having memory breach, such as forgetting recognizable words or the location of everyday objects.

Friends, family or others on the point of the individual begin to note difficulties. During a close medical interview, doctors could also be able to discover issues in memory or concentration. Common difficulties include:

Problems coming up with spelling or writing right word or name. Problem in memorizing names when introduced to new people. Challenges in performing tasks in social or work placed. Not able to remember material that one has just read. Losing or misplacing a valuable object. Increasing trouble with planning or organizing the tasks or activities.

B. Middle Stage:

Moderate Alzheimer's is often the longest stage and might last for several years. As the sickness progresses, the person with Alzheimer's would force a bigger level of care.

During the moderate stage of Alzheimer's, the insanity symptoms square measure additional pronounced. A person might have bigger problem playing tasks, like paying bills, however they will still bear in mind vital details regarding

their life. At this point, symptoms will be identifiable to others and may include Forgetfulness of familiar things and own things. Feeling moody or stressed, especially in socially or mentally challenging situations. Not able to remember their own address or telephone number or the high school or college from which they graduated. Confusion about the location they are at and what day it is the need for help choosing proper outfit for the season they are experiencing or the occasion they are attending. Changes in sleep habits, such as sleeping during the day and becoming restless at night. An increased risk of wandering and becoming lost or loosing concentration in the work they are doing

C. Late Stage:

In the ending of this malady, insanity symptoms area unit severe. Individuals lose the power to retort to their atmosphere, to hold on a spoken language and, eventually, to manage movement. They may still say words or phrases; however, communication pain becomes tough. As memory and psychological feature skills still worsen, important temperament changes could present itself and people want intensive facilitate with daily activities. At this stage, individuals may: Need 24*7 assistance with daily activities and personal care. Lose attention of recent experiences as well as of their surroundings. Experience changes in physical fitness, including the ability to walk, sit and, eventually, swallow become sensitive to infections, especially pneumonia.

Machine learning is used to read and analyse the data. Furthermore, it can analyse patterns and model data. It grants decisions to be made that couldn't be made generally utilizing routine systems while saving time and struggle[6].

II. RELATED WORK

According to research there are many Data Mining and Machine Learning techniques have been refined. Among these techniques, Naïve Bayes, Decision Tree and Clustering have been widely used due to its modesty and certainty. Decision Tree refers to algorithm that performs decision based on the data set provided to the system and based on that data set an appropriate decision is generated.

It is found that Alzheimer's disease (AD) is the most popular dementia in elderly people worldwide. Various classification methods were used for the early detection of AD. These methods include Random Forests, which is based on a group of decision trees and Support Vector Machines which is used for classification and pattern recognition in a different application for its ability for police investigation patterns in experimental databases. Lastly, K Nearest Neighbour was used. KNN may be a process} formula with an outsized vary of applications within the image processing domain. data mining and machine learning.

The algorithms used were Naïve Bayes and KNN (KNN can be used for both classification and regression predictive problems.

Classification is a popular data mining technique. The use of information mining ways as well as classification to help in diagnosing and prediction of medical and health conditions has received substantial attention from researchers in recent years. Decision trees represent a simple, yet powerful classification method. Decision trees are produced by algorithms that identify various ways of splitting a data set into branch-like fragments. These fragments form an inverted decision tree whose root node is at the top of the tree. In our sample data there are five important attributes that contribute to the likelihood of getting sick with Alzheimer disease. These attributes are gender (G), age (A), genetic causes (GC), brain injury (BI), and vascular disease (VD). These attributes represent the risk factors that have impact on developing Alzheimer's disease.

III. DATA MINING

Data mining is that the method of analyzing hidden patterns data in line with completely different views for categorization into helpful information, that is collected and assembled in common areas, such as data warehouses, for economical analysis, data -processing algorithms, facilitating business deciding and different data needs to ultimately cut prices and increase revenue. Data mining is additionally referred to the knowledge discovery and data discovery.

Data mining is that the method of analyzing hidden patterns data in line with completely different views for categorization into helpful information, that is collected and assembled in common areas, such as data warehouses, for economical analysis, data -processing algorithms, facilitating business deciding and different data needs to ultimately cut prices and increase revenue. Data mining is additionally referred to the knowledge discovery and data discovery[10].

The second step in data processing is choosing an acceptable algorithmic rule - a mechanism manufacturing a knowledge mining model. The general operating of the algorithmic rule involves distinctive trends during a set of knowledge and victimisation the output for parameter definition. The most standard algorithms used for data processing square measure classification algorithms and regression algorithms, that square measure want to establish relationships among information parts. Major information vendors like Oracle and SQL incorporate data processing algorithms, like agglomeration and regression hairdo, to satisfy the demand for data processing.

Machine learning is associate degree application of computer science (AI) that gives systems the flexibility to mechanically learn and improve from expertise while not being expressly programmed. Machine learning focuses on the event of pc programs which will access knowledge and use it learn for themselves. The primary aim is to permit the computers learn mechanically while not human intervention or help and modify actions consequently.

Machine learning (ML) may be a class of algorithmic rule that enables computer code applications to become additional correct in predicting outcomes while not being expressly programmed. The basic premise of machine learning is to create algorithms which will receive knowledge input file computer file Associate in Nursing use applied mathematics analysis to predict an output whereas change outputs as new data becomes on the market[3],[9].
The processes concerned in machine learning are just like that of knowledge mining and prophetic modeling. Both need exploring through information to appear for patterns and adjusting program actions consequently. Many people are conversant in machine learning from looking on the net and being served ads associated with their purchase. This happens as a result of recommendation engines use machine learning to individualize on-line ad delivery in nearly real time. Beyond personalised promoting, different common machine learning use cases embrace fraud detection, spam filtering, network security threat detection, prophetic maintenance and building news feeds[4],[5].

Decision Tree:
A decision tree could be a flowchart-like structure within which every internal node represents a "test" on associate attribute (e.g. whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). The ways from root to leaf represent classification rules. Decision trees square measure ordinarily employed in research and operations management. If, in apply, selections ought to be taken on-line with no recall below incomplete information, a choice tree ought to be paralleled by a likelihood model as a best option model or online selection model algorithm. Another use of call trees is as a descriptive suggests that for scheming conditional possibilities[11].

Association Rule Mining:
It is supposed to spot robust rules discovered in databases mistreatment some measures of interest. This rule-based approach conjointly generates new rules because it analyzes a lot of knowledge.

Association rules are if-then statements that facilitate to indicate the chance of relationships between knowledge things among giant knowledge sets in numerous forms of databases. Association rule mining contains a variety of applications and is wide wont to facilitate discover sales correlations in transactional knowledge or in medical knowledge sets.

Popular algorithms that use association rules embrace AIS, SETM, Apriori and variations of the latter.

With the AIS algorithm, itemsets are generated and counted as it scans the data.

The SETM formula conjointly generates candidate itemsets because it scans a info, but this algorithm accounts for the itemsets at the end of its scan. New candidate itemsets are generated the same way as with the AIS algorithm, but the transaction ID of the generating transaction is saved with the candidate itemset in a sequential structure. At the top of the pass, the support count of candidate itemsets is created by aggregating the sequential structure. The downside of both the AIS and SETM algorithms is that each one can generate and count many small candidate itemset.

With the Apriori algorithm, candidate itemsets are generated using only the large itemset of the previous pass. The large itemset of the previous pass is joined with itself to generate all itemsets with a size that's larger by one. Each generated itemset with a subset that is not large is then deleted. The remaining itemsets are the candidates. The Apriori formula considers any set of a frequent itemset to even be a frequent itemset. With this approach, the formula reduces the quantity of candidates being thought of by solely exploring the itemsets whose support count is bigger than the minimum support count[1],[7].

Visualization:
Data image is a very important talent in applied statistics and machine learning. Statistics will so target quantitative descriptions and estimations of information. Data image provides a very important suite of tools for gaining a qualitative understanding. This can be useful once exploring and attending to understand a dataset and may facilitate with distinguishing patterns, corrupt data, outliers, and much more. With a bit domain information, knowledge visualizations will be wont to categorical and demonstrate key relationships in plots and charts that area unit a lot of visceral to yourself and stakeholders than measures of association or significance. Data image and preliminary knowledge analysis area unit whole fields themselves and that can advocate a deeper dive into some the books mentioned at the top. In this tutorial, let's check out basic charts and plots you'll use to raised perceive your knowledge[2],[8].

IV. PROPOSED SYSTEM

PHASE 1 – A list of questions are prepared by a medical practitioner that can ensure that it becomes easy for the doctor to diagnose the patient. The list of questions can be ranged from personal details to logical and mathematical questions. These questions can be easily divided into levels, which can help the doctor to detect the extent i.e., the stage of the disease. Each stage of the disease has symptoms that are particular to that stage. Extensive research of the disease can help determine the various questions that can easily help to diagnose the stage. These questions are to be answered by both the patient as well as the patient's relative. This is done to ensure that answers are correctly answered by the patient. Both the answer sets are compared and the stage one answers are extracted. The other stage answers given by the patient are entered into the database. A resultant dataset is created which is then used later on for disease detection and stage prediction.

PHASE 2- Data Mining algorithms like Decision Trees and Navies Bayes are applied to the dataset and the output of these help the doctor to understand the presence or absence of the disease on the patient.

PHASE 3- The results of the above phase also provides the patients with the impact of the disease. This phase helps the patient to know which stage of Alzheimer's Disease he/she is suffering from. Recommendation is provided to the patients by the system to help reduce the future impact of the disease.

PHASE 4- An e-report is generated which provides the patient with diagnosis and the do's-and-don'ts to reduce the disease impact.

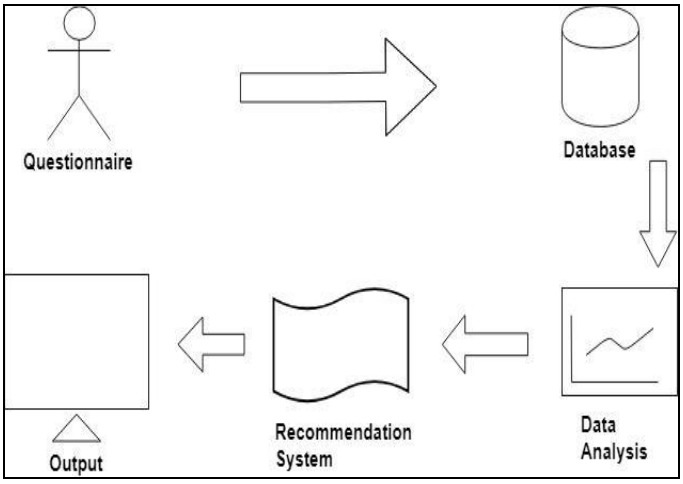

Fig. 1

V. RESULTS

1	236	yes
2	503	no
3	435	no
4	175	yes
5	153	yes
6	329	no
7	404	yes
8	171	yes
9	25	yes
10	436	no
11	54	yes
12	433	yes
13	195	yes
14	40	yes
15	538	yes
16	402	yes
17	420	yes
18	3	no
19	390	no
20	221	yes
21	528	yes
22	521	no
23	464	yes
24	157	no
25	513	yes
26	149	yes
27	477	no
28	427	no

Fig2:Results for the classification

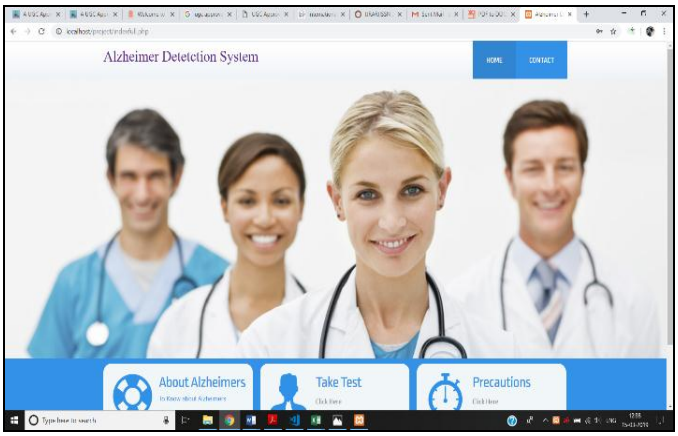

Fig 3:Results for GUI

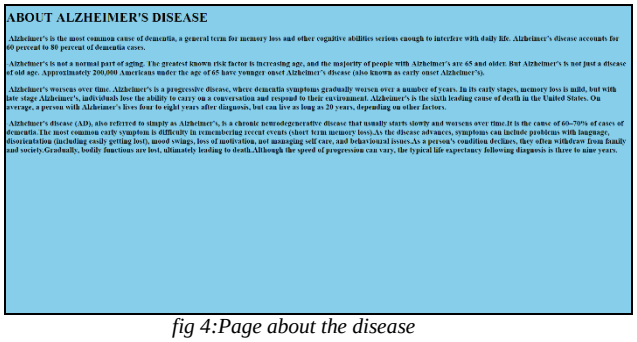

fig 4:Page about the disease

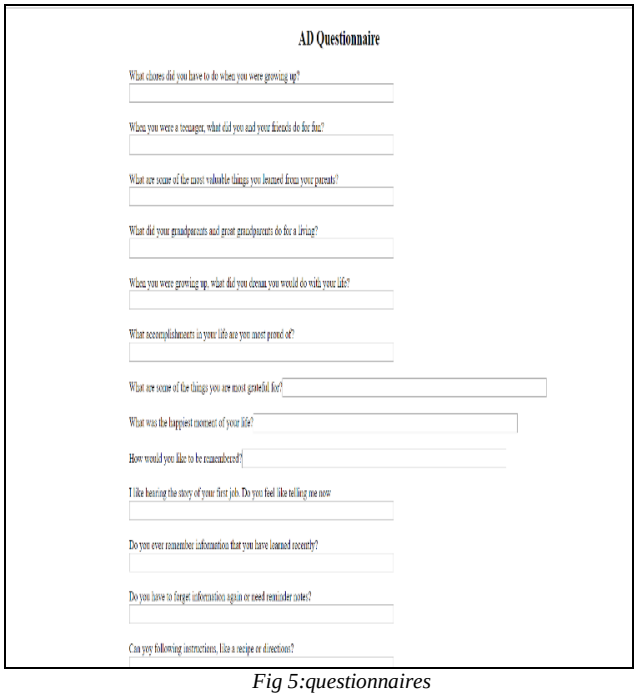

Fig 5:questionnaires

In the above result, the decision tree algorithm used provides the classification of the patients, if he /she is demented or non-demented

VI . CONCLUSION

According to the 2018 reports by Alzheimer's Association, sharp increase has been witnessed in Alzheimer's prevalence, deaths and the costs of care. In America alone, around 5.7 million people are suffering from this disease which has led to researchers and doctors trying to find ways to reduced the diseases' presence. Machine Learning is one of the ways in which the data can be successfully analyzed, patterns can be detected and appropriate actions can be taken. In this paper, the two main Machine Learning algorithms selected are Association rule mining which includes AIS,SETM and Apriori with that also the visualization technique is used. These algorithms were taken into consideration due to their accuracy and efficiency. These algorithms were convinient to use with the dataset available with us. The sensitivity provided by these algorithms helped in the successful classification of the data. In review, other algorithms which provide high level of efficiency as well as accuracy can also be used. Also, the Virtualization Techniques helps the doctor and the patient to understand the analysis more precisely.

VII. FUTURE SCOPE

In 2017, as per the Alzheimer's Association, 5.5 million people in America lives with Alzheimer's affectedness, and calculable 5.3 million are at 65 years previous or on top of. Moreover, according to the National Institute of Health, the global geriatric (RELATING TO OLD PEOPLE, ESPECIALLY WITH REGARD TO THEIR HEALTHCARE.) population is reportedly expected to reach 1.6 billion by 2050 from 617 million in 2016. Based on the numbers it is important to ensure that the disease is not taken for lower priority, but ways for successful detection and diagnosis are found rapidly rather than later. With the improvement in technology, various ways can be come up with for the treatment of this disease. It is important that the people have an perception about this disease and also better ways are found so that the disease can be diagnosed as early as possible. Also various techniques can be used in future as the technology progresses day by day. There could various advancement in the system such as, in the mere future the system can be upgraded with the record of the patient by various ways such as scanning and other tests.

REFERENCES

[1]https://searchbusinessanalytics.techtarget.com/definition/association-rules-in-data-mining, Accessed on 15/11/2018 at 2:28 P.M
[2]https://medium.com/@enjalot/machine-learning-for-visualization-927a9dff1cabaccessed on 14/01/2019 at 04:59 PM.
[3]https://en.wikipedia.org/wiki/Machine_learning, accessed on 12/09/2018 at 5:20 P.M.
[4]Early diagnosis of Alzheimer's disease using machine learning techniques: A review paper", IEEE, August 2016.
[5]Chenhui Hu , Ronghui Ju , Yusong Shen , Pan Zhou , Quanzheng Li, Chenhui Hu , Ronghui Ju Yusong Shen , Pan Zhou. & Quanzheng Li "Clinical decision support for Alzheimer's disease based on deep learning and brain network", IEEE explore, July 2016.
[6] Authors-Ranjan Duara; Malek Adjouadi, Chen Fang ; Chunfei Li ; Mercedes Cabrerizo ; Armando Barreto ; Jean Andrian ; David Loewenstein A Novel Gaussian Discriminant Analysis-based Computer Aided Diagnosis System for Screening Different Stages of Alzheimer's Disease, Published in: 2018 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE); Conference Location: Washington DC, USA.
[7] https://en.wikipedia.org/wiki/Association_rule_learningaccess ed on 16/01/2019 at 01:12 PM.
[8] https://en.wikipedia.org/wiki/Visualization, accessed on 17/01/2019 at 05:01 PM

[9] https://skymind.ai/wiki/machine-learning-algorithmsaccessed on 18/01/2019 at 03:55PM
[10] https://www.medgadget.com/2017/12/future-scope-of-alzheimers-disease-diagnostic-market-which-is-expected-to-grow-at-a-cagr-of-10-top-key-players-profile-forecast-to-2022.html, accessed on 09/01/2019 at 04:30 PM.
[11] https://alzheimerscareresourcecenter.com/2018-alzheimers-disease-facts-figures-report, accessed on 09/01/2019 at 04:43 PM.

Authors Profile

Mr. C T Lin pursed Bachelor of Science from University of Taiwan, Taiwan in 2006 and Master of Science from Osmania University in year 2009. He is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Computational Sciences, Department of Electronic and Communication, University of Taiwan, Taiwan since 2012. He is a member of IEEE & IEEE computer society since 2013, a life member of the ISROSET since 2013, ACM since 2011. He has published more than 20 research papers in reputed international journals including Thomson Reuters (SCI & Web of Science) and conferences including IEEE and it's also available online. His main research work focuses on Cryptography Algorithms, Network Security, Cloud Security and Privacy, Big Data Analytics, Data Mining, IoT and Computational Intelligence based education. He has 5 years of teaching experience and 4 years of Research Experience.

Mr C H Lin pursed Bachelor of Science and Master of Science from University of New York, USA in year 2009. He is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Telecommunication, University of New York, USA since 2012. He is a member of IEEE & IEEE computer society since 2013, a life member of the ISROSET since 2013 and ACM since 2011. He has published more than 20 research papers in reputed international journals including Thomson Reuters (SCI & Web of Science) and conferences including IEEE and it's also available online. His main research work focuses on Cryptography Algorithms, Network Security, Cloud Security and Privacy, Big Data Analytics, Data Mining, IoT and Computational Intelligence based education. He has 5 years of teaching experience and 4 years of Research Experience.

