

# A Review on the Method of Diagnosing Alzheimer's Disease using Data Mining

Bhagya Shree S R\*

Assistant Prof. Department of E & C  
ATME college of Engineering  
Mysore, India

Dr. H. S. Sheshadri

Prof. Department of E & C  
PES College of Engineering  
Mandya, India

Dr. Sandhya Joshi

Associate Prof. Department of MCA  
Regional Center, VTU, Gulbarga,  
India

**Abstract**— All over the world a large number of people are suffering from brain related diseases. Studying and finding the solutions for those diseases is the requirement of the day. Dementia is one such disease of the brain. This causes loss of cognitive functions such as reasoning, memory and other mental abilities which may be due to trauma or normal ageing. Alzheimer's disease is one of the types of the dementia which accounts to 60-80% of mental disorders [1]. Diagnosis of this disease at an early stage will help the patients to lead a quality life for the remaining tenure of their life. The goal of the paper is to have a review on different neuro psychological tests, the various algorithms used for the purpose of diagnosis, and the tool that may be used for the analysis.

**Keywords**— Alzheimer's disease, Machine Learning, Weka

## I. INTRODUCTION

According to the report generated in 2013, by NCBI funded by US Government, an estimated 5.2 million Americans have AD. Approximately 2,00,000 are below 65 and 5 million are above 65. In America today someone develops AD once in 68 sec. By 2050 one new case of AD is expected to develop every 33secs. Between 2000 and 2010, the proportion of deaths resulting from heart disease, stroke, and prostate cancer has decreased to 16%, 23%, and 8%, respectively, whereas the proportion resulting from AD is increased 68% [2]. These facts and figures reveal us the need for early diagnosis of AD.

Diagnosis can be done at three different stages, namely consulting the G.P, Conducting Neuro Psychological tests and taking MRI or PET scans [3].

In the Neuro psychological tests, there are different batteries like MMSE, BIMC, ADAS and SKT etc. The reliability, practicality and validity of these batteries are discussed by authors in their paper on An Approach in the Diagnosis of Alzheimer's Disease - A Survey [4]. These tests are well suited for people belonging to a particular community. A neuro psychological test is needed which can suit the person from any background. The 10/66 research group has suggested 10/66 CoG battery which may be used to conduct the neuro psychological test for subjects from any background, irrespective of their culture, religion and education.

## II. LITERATURE SURVEY

Dementia is the disease of the brain, causing loss of cognitive functions like reasoning, memory and other mental abilities due to trauma or normal ageing.

Dementia is classified into Alzheimer's disease, Dementia with lewy bodies, Parkinson's disease, Creutzfeldt-Jakob disease, Normal Pressure Hydrocephalus, Vascular Dementia, Front temporal Dementia [4].

Out of all the above mentioned diseases, 2/3 of the demented patients suffer from Alzheimer's disease. Alzheimer's disease is officially listed as the sixth-leading cause of death in the United States. It is the fifth-leading cause of death for those aged 65 and above. However, it may cause even more deaths than official sources recognize [5].

An estimated 5.2 million Americans have AD. Approximately 200,000 people younger than 65 years with AD comprise the younger onset AD population; 5 million comprise the older onset AD population. A projected 450,000 older Americans with AD will die in 2013, and a large proportion will die as a result of complications of AD [2].

Though the mortality is high for the people aged above 65, the number of people affected by the disease is more in the age group of 40-65. Thus early diagnosis is needed to have quality life.

There are various risk factors which contribute to the development of the disease namely age, genetics, smoking, consuming alcohol, cholesterol, Down syndrome [5].

The symptoms of Alzheimer's diseases are decision making, poor judgment, misplacing things, impairment of movements, verbal communication, abnormal moods, complete loss of memory.

From the literature survey, it is very much clear that there is a requirement for early diagnosis of the disease to ensure quality life to the diseased.

## III. PROPOSED METHODOLOGY

Ferri et al. [1] estimated a worldwide increase of 4.6 million new dementia cases every year. Without changes in mortality and new effective prevention strategies or curative treatments, the numbers of affected people will double every 20 years to 81.1 million by 2040[6].

To do the neuro psychological tests MMSE, BDIMC, COG, BOMC, MOCA, AD8 and GP CoG are used. The disadvantage of MMSE is its insensitivity to the early

changes of dementia. As the test relies on the verbal response, it's difficult to use this for the patients with language problems like dysarthria, aphasia etc. [7]. In these screening tests the questionnaire is meant for a set of people. There is a requirement for a screening test which may be used to the subjects irrespective of gender, religion, culture and education.

The 10/66 Dementia Research Group (10/66) founded in 1998 is a network of over 100 researchers from mainly developing countries. 10/66 is committed to encourage more good quality research in those regions, where an estimated two-thirds of all those with dementia live. It represents a collaboration of academics, clinicians, and an international non-governmental organization, Alzheimer's disease International (ADI). The 10/66 research group has suggested a battery which fulfills the above requirements [8].

In this paper the author focuses on the new battery, suggested by the 10/66 research group. This battery is preferred compared to the most popular MMSE battery as it is applicable to anyone irrespective of gender, religion, culture and education [9].

In this battery a predefined questions will be asked to the subject. Each answer will be evaluated. Depending on the score, the subjects' will be classified as AD patient or not.

Analysis of data and decision making is a crucial step. Many a times the analysis and decision making depends on the mood of the Psychologist. In addition to that, the humane error cannot be avoided.

Powerful and versatile tools are really needed to automatically uncover valuable information from the tremendous amount of data and transform such data into organized knowledge. This necessity has led birth to data mining. Many people treat data mining as the synonym for Knowledge Discovery from Data (KDD).

The knowledge Discovery process is a procedure that comprises of Data Cleaning, Data integration, Data selection, Data Transformation, Data mining, Pattern evaluation, Knowledge presentations[10].

Various techniques are used for discovering the knowledge, namely Association, Sequential pattern, Classifiers, Decision trees, Neural networks, Visualization, Clustering, Collaborative filtering, Data transformation and cleaning, Deviation and fraud detection, Estimation and forecasting, Bayesian and dependency networks, OLAP and dimensional analysis, Statistical analysis, Text analysis, Web mining etc.

Of the all the above mentioned techniques, the most commonly used techniques are association, classifiers, visualization and clustering [11].

Data mining finds its applications in almost all the fields which include cattle farming, molecular biology, drug discovery, process based industries, Pharmacy, Astronomy, Medicine, geophysics, Fraud detection, Intrusion detection and many more.

Data mining is very useful in diagnosing many of the life threatening diseases at early stage.

Jyothi Sony has used supervised machine learning namely Naïve Bayes, K-NN, Decision List algorithm to analyze the datasets of heart disease patients [12].

Ruijuan Hu has suggested data mining algorithm in the diagnosis of breast cancer [13].

Bichen Zheng and team have used data mining approach in the diagnosis of breast cancer [14].

Amir Fallahi et.al have used Bayesian network for the detection of breast cancer [15].

Mahjabeen Mirza Beg et.al has used artificial neural networks for the diagnosis of breast cancer [16].

Shu-Ting Luo & Bor-Wen Cheng in their paper have discussed about using decision tree, Support vector Machine sequential minimal optimization in diagnosing breast cancer [17].

Breetha S and Kavinila R have discussed about using hierarchical clustering in the diagnosis of cancer and classification of cancer [18].

Rashedur M. Rahman and Farhana Afroz have tested the various classification techniques using various tools like WEKA, Mat lab, Tanagra for the data sets of diabetes patients [19].

Abhishek Taneja in his paper has discussed about using data mining for the prediction of heart disease [20].

Duarte Ferreira et.al has used decision trees, neural networks in the diagnosis of neonatal Jaundice [21].

Abhishek et al have discussed two different types of neural networks called Back Propagation Algorithm, Radial Basis Function and one non-linear classifier Support Vector Machine and comparison is made. Weka is used as a tool for diagnosis purpose [22].

P.Rajeswari in her paper has used Naïve Bayes in the analysis of liver disorder and WEKA tool is used for analysis [23].

Tarigoppula V.S Sriram et.al has used classification algorithms to detect Parkinson's disease [24].

From the above references it is evident that data mining may be used for analyzing medical data of different diseases.

Data mining also finds applications in the diagnosis of Alzheimer's disease in particular.

Sandhya Joshi et.al have used the various machine learning methods such as neural networks, multilayer Perceptron, Bagging, Decision tree, CANFIS and Genetic algorithms for the classification of different stages of Alzheimer's disease[25].

Devi Parikh et.al in their paper on early diagnosis of Alzheimer's disease, have discussed about classifiers base data fusion approach to data from two different sources, containing complementary information [26].

Claudia Plant et.al in their paper on Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease have used combination of Support Vector Machine, Bayes statistics and voting feature Interval to derive a quantitative index of pattern matching for the prediction of conversion from MCI to AD[27].

Stefan Kloppel et.al in their paper on Automatic classification of MR scans in Alzheimer's disease have used linear support Vector achiness to classify the grey matter segment of TI weighed MR scans obtained from two different centers and two different equipment [28].

Ali Hamouet.al in their paper on cluster analysis of MR Imaging in Alzheimer's disease using decision tree refinement have used clustering algorithm to analyze the data

and they have used decision tree algorithm to model the level of importance of variants influencing the decision [29].

J. Shane Kippenhan et.al and group in their paper on Neural-Network Classification of Normal and

Alzheimer's disease Subjects Using High-Resolution and Low-Resolution PET Cameras have trained neural networks to distinguish between normal and abnormal subjects [30].

Obi J.C1.and Imainvan A.A in their paper discussed about analyzing the traditional procedure employed by the physician in diagnosis of AD [31].

Sandhya Joshi et.al in their paper have used data mining approach to classify various neuro degenerative disorders like Alzheimer's disease, vascular disease and Parkinson's disease by considering the common risk factors. In this paper they have used machine learning and neural networks. Under machine learning Decision tree, Bagging, BF tree, Random Forest tree, and RBF Networks are being used [32].

Javier Escudero et.al in their paper have discussed about detection of Alzheimer's disease using machine learning [33].

Roman Filipovych et.al in their paper have discussed about using supervised classification approach for images using SVM technique [34].

From the references 24 to 32 it is evident that the different techniques of data mining are used for the diagnosis of Alzheimer's disease.

In data mining there are various algorithms which may be used for extracting information. The award winners of IEEE International conference on data mining, Xindong Wu and team have identified the ten most popular algorithms used in data mining. They said C4.5, K-means, SVM, Apriori, EM, Page Rank, Ada Boost, KNN, Naïve Bayes and CART are the top ten, most popular data mining algorithms used in research. In this survey paper the authors have discussed about the description of the algorithm and the impact of the algorithm [35].

In data mining a number of software like WEKA, See5 and Wiz Why are used for the purpose of analysis of data.

#### IV. ABOUT WEKA

. Weka (Waikato Environment for Knowledge Analysis) is a popular suite of machine learning software written in Java, developed at the University Of Waikato, New Zealand. The Weka (pronounced Weh-Kuh) work bench contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality. Weka is a collection of machine learning algorithms for data mining tasks. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. Weka supports several standard data mining tasks more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection.

Advantages of Weka over other tools are free availability under the GNU General Public License, Portability, a comprehensive collection of data preprocessing and modeling techniques, ease of use due to its graphical user interfaces.

All of Weka's techniques are predicated on the assumption that the data is available as a single flat file or relation, where each data point is described by a fixed number of attributes.

Weka's main user interface is the *Explorer*. There is also the *Experimenter*, which allows the systematic comparison of the predictive performance of Weka's machine learning algorithms on a collection of datasets.

The *Explorer* interface features several panels providing access to the main components of the workbench.

The *Preprocess* panel has facilities for importing data from a database, a CSV file, etc., and for preprocessing this data using a so-called *filtering* algorithm. These filters can be used to transform the data (e.g., turning numeric attributes into discrete ones) and make it possible to delete instances and attributes according to specific criteria.

The *Classify* panel enables the user to apply classification and regression algorithms (indiscriminately called *classifiers* in Weka) to the resulting dataset, to estimate the accuracy of the resulting predictive model, and to visualize erroneous predictions, ROC curves, etc., or the model itself (if the model is amenable to visualization like, e.g., a decision tree).

The *Associate* panel provides access to association rule learners that attempt to identify all important interrelationships between attributes in the data.

The *Cluster* panel gives access to the clustering techniques in Weka, e.g., the simple k-means algorithm. There is also an implementation of the expectation maximization algorithm for learning a mixture of normal distributions.

The *Select attributes* panel provides algorithms for identifying the most predictive attributes in a dataset.

The *Visualize* panel shows a scatter plot matrix, where individual scatter plots can be selected and enlarged, and analyzed further using various selection operators [36].

Plamena Andreeva and group have tested the parameters of data sets of three different diseases namely breast cancer, Diabetes Pima and IRIS and published a scholarly article in Google scholar. They have analyzed the data using various types of classification by using different tools namely See5, Wiz Why and Weka. From the results the authors suggest that WEKA is better in terms of usage, consistency etc. The authors also say that, of the all, WEKA predicts the majority of the data [37].

#### V. CONCLUSION AND FUTURE WORK

The current study focuses on a new emerging approach in the diagnosis of AD. There are various neuro psychological tests that may be conducted for the diagnosis of AD. Though MMSE is a popular test of all, it too has a disadvantage. The disadvantage is that it cannot be used for the people who are having problem with verbal communication. The 10/66 battery designed by the research group of Alzheimer's association will overcome this disadvantage.

Though there are various ways to analyze the data, the application of data mining in the field of medicine makes,

data mining as a more appropriate method of discovering the knowledge.

There are various tools like Weka, C4.5, WEKA etc. The WEKA tool has an added feature to predict majority of the data and hence Weka tool may be used for the purpose of analysis.

The future work is to design a methodology to diagnose Alzheimer's disease using 10/66 CoG battery by applying data mining techniques and to analyze the data using Weka tool.

## ACKNOWLEDGMENT

The authors are thankful to Dr. Murali Krishna, Earlier Scientist Research Fellow, Wellcome DBT Allianz, CSI Holdsworth Memorial Mission Hospital, Mysore, Dr. L Basavaraj, Principal, ATME, Mysore and to the research colleagues who supported with the data in respect of the Alzheimer's disease.

## REFERENCES

- [1] David P Salmon<sup>1</sup> and Mark W. Bondi<sup>2,3</sup> "Neuropsychological Assessment of Dementia" Access NIH public, PubMed central, US national library of medicine National Institutes of Health, May 2010.
- [2] Thies W, Bleiler L "2013 Alzheimer's facts and figures" Alzheimer's disease (Journal of Alzheimer's Association), published by Elsevier Inc. Mar-2013
- [3] James E. Galvin, MD, MPH, and Carl H. Sadowsky, MD "Practical Guidelines for the Recognition and Diagnosis of Dementia" Journal of the American Board of Family Medicine (JABFM)-Volume 25 No.3, pp 367-378, June 2012
- [4] Bhagya Shree S R and Dr.H.S.Sheshadri "An Approach in the Diagnosis of Alzheimer's Disease - A Survey" International Journal of Engineering Trends and Technology (IJETT) – Volume 7 No 1, PP 1-4, Jan 2014
- [5] L L Barclay, S Kheifets et.al "Risk Factors in Alzheimer's disease" Advances in behavioral biology, volume 29, pp. 141-146
- [6] Tobias Luck et.al "Incidence of Mild Cognitive Impairment: A Systematic Review", Journal of Dementia and Geriatric Cognitive Disorders; 29: 164–175; 2010
- [7] Robert M Herudon, Hand book of neurologic rating scales 2<sup>nd</sup> Ed. 2006
- [8] Prince M et.al "Alzheimer Disease International's 10/66 Dementia Research Group - one model for action research in developing countries" International Journal of Geriatric Psychiatry, 178-81, FEB 2004
- [9] Ana Luisa Sosa<sup>1</sup> et.al "Population normative data for the 10/66 Dementia Research Group cognitive test battery from Latin America, India and China: across-sectional survey" BMC Neurology, Vol9, pp 1-11, Aug 2009
- [10] Data Mining: Concepts and Techniques by Jiawei Han, Micheline Kamber, JianPei published by Elsevier, Third edition, 2012
- [11] K.P.Soman et.al "Insight into data mining Theory and concepts", PHI Learning Private Limited, 6<sup>th</sup> Ed, 2012
- [12] Jyoti Soni et.al "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction" International Journal of Computer Applications (0975 – 8887), Volume 17, No. 8, March 2011
- [13] Ruijuan Hu "Medical Data Mining Based on Decision Tree Algorithm" Published by Canadian Center of Science and education in volume 4, No5, pp 14-19, Sep.2011
- [14] Bichen Zheng, et.al "Breast cancer diagnosis based on feature extraction using a hybrid of k-means and support vector machine algorithms" Expert Systems with Applications, Volume 41, Issue 4, Part 1, pp 1476–1482 March 2014,
- [15] Amir Fallahi, Shahram Jafari "An Expert Detection of Breast Cancer Using Data Preprocessing and Bayesian Network" International Journal of Advanced Science and Technology Vol. 34, PP 65-70, September, 2011
- [16] Mahjabeen Mirza Beg, Monika Jain "An analysis of the methods employed for breast cancer diagnosis" International Journal of Research in Computer Science ISSN 2249-8265 Volume 2 Issue 3, PP 25-29, 2012
- [17] Shu-Ting Luo & Bor-Wen Cheng "Diagnosing Breast Masses in Digital Mammography Using Feature Selection and Ensemble Methods" Springer April 2010
- [18] Breetha S, Kavinila " Hierarchical clustering for cancer discovery using Range check and delta check" International Journal of Scientific and Research Publications, Volume 3, Issue 4, April 2013
- [19] Rashedur M. Rahman et.al "Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis" Journal of Software Engineering and Applications, , 6, PP 85-97, 2013
- [20] Abhishek Taneja "Heart Disease Prediction System Using Data Mining Techniques" Oriental Journal of Computer Science & technology, December, 2013, Vol. 6, No. 4, 457-466, Dec. 2013
- [21] Duarte Ferreira et.al "Applying data mining techniques to improve diagnosis in neonatal jaundice" BMC Medical Informatics and Decision Making, 0-5, 2012
- [22] Abhishek et al "Proposing Efficient Neural Network Training Model for Kidney Stone Diagnosis" International Journal of Computer Science and Information Technologies, Vol. 3 (3), 2012
- [23] P.Rajeswari, G.Sophia Reena "Analysis of Liver Disorder Using Data mining Algorithm" Global Journal of computer science & Technology Vol. 10 Issue 14 (Ver. 1.0) November 2010
- [24] Tarigoppula V.S Sriram et.al "Intelligent Parkinson Disease Prediction Using Machine Learning Algorithms" International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 1, PP 44-52, September 2010
- [25] Sandhya Joshi et.al "Classification and treatment of different stages of Alzheimer's disease using various machine learning methods" International Journal of Bioinformatics Research, Volume 2, Issue 1, pp-44-52, 2010
- [26] Devi Parikh, et.al "Ensemble Based Data Fusion for early Diagnosis of Alzheimer's Disease", Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, 1-4, September 2005
- [27] Claudia Plant et.al "Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease" Neuro Image 50 (2010), 162–174 published by ELSEVIER
- [28] Stefan Kloppel et.al "Automatic classification of MRI Scans in Alzheimer's disease" published in Brain (2008), 13, PP 681-689 2008
- [29] Ali Hamou et.al "cluster analysis of MR Imaging in Alzheimer's disease using decision tree refinement" International journal of Artificial intelligence, Vol6, PP 1-10, , spring 2011
- [30] J.Shane Kippenhan et.al "Neural-Network Classification of Normal and Alzheimer's disease Subjects Using High-Resolution and Low-Resolution PET Cameras Nuci Med; 35:7-15, 1994
- [31] Obi J.C1. and Imainvan A.A "Decision Support System for the Intelligent Identification of Alzheimer using Neuro Fuzzy Logic", International Journal on Soft Computing (IJSC), Vol.2, No.2, PP 25-38, May 2011
- [32] Sandhya Joshi et.al "Classification of neuro degenerative Disorders Based on Major Risk Factors Employing Machine Learning Techniques" IACSIT International Journal of Engineering and Technology, Vol.2, No.4, PP 350-355, August 2010
- [33] Javier Escudero "Machine Learning-Based Method for Personalized and Cost-Effective Detection of Alzheimer's Disease" IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 60, NO. 1, PP 164-168, JANUARY 2013
- [34] Roman Filipovych and Christos Davatzikos "semi supervised classification on medical images-Application to MCI neuro image Dec, 2012
- [35] Xindong Wuet.al "Top 10 algorithms in data mining", © Springer-Verlag London Limited, Volume 14, Issue 2, PP 1-37, 2008
- [36] <http://www.cs.waikato.ac.nz/ml/weka/>
- [37] Plamena Andreeva<sup>1</sup>, Maya Dimitrova<sup>1</sup>, Petia Radeva<sup>2</sup> "Data mining learning models and algorithms for medical applications" [http://scholar.google.co.in/scholar /data\\_mining](http://scholar.google.co.in/scholar/data_mining)