IACSIT International Journal of Engineering and Technology, Vol.2, No.4, August 2010
ISSN: 1793-8236

Classification of Neurodegenerative Disorders Based on Major Risk Factors Employing Machine Learning Techniques

Sandhya Joshi, *Member, IACSIT,* P. Deepa Shenoy, Vibhudendra Simha G.G., Venugopal K. R and L. M. Patnaik

***Abstract*—Medical data mining has great potential for exploring the hidden patterns in the data sets of the medical domain. These patterns can be utilized for the classification of various diseases. Data mining technology provides a user-oriented approach to novel and hidden patterns in the data. The present study consisted of records of 746 patients collected from ADRC, ISTAART, USA. Around eight hundred and ninety patients were recruited to ADRC and diagnosed for Alzheimer's disease (65%), vascular dementia (38%) and Parkinson's disease (40%), according to the established criteria. In our study we concentrated particularly on the major risk factors which are responsible for Alzheimer's disease, vascular dementia and Parkinson's disease. This paper proposes a new model for the classification of Alzheimer's disease, vascular disease and Parkinson's disease by considering the most influencing risk factors. The main focus was on the selection of most influencing risk factors for both AD and PD using various attribute evaluation scheme with ranker search method. Different models for the classification of AD, VD and PD using various classification techniques such as Neural Networks (NN) and Machine Learning (ML) methods were also developed. It is observed that increase in the vascular risk factors increases the risk of Alzheimer's disease. It was found that some specific genetic factors, diabetes, age and smoking were the strongest risk factors for Alzheimer's disease. Similarly, for the classification of Parkinson's disease, the risk factors such as stroke, diabetes, genes and age were the vital factors.**

***Index Terms*—Vascular Dementia, Alzheimer's disease, Parkinson's disease, Machine learning.**

I. Introduction

Data mining is a fast evolving technology, is being adopted in biomedical sciences and research. Data mining in medicine is an emerging field of high importance for providing prognosis and a deeper understanding of the classification of neurodegenerative diseases. Neurodegenerative diseases are now generally considered as a group of disorders that seriously and progressively impair the functions of the nervous system through selective neuronal vulnerability of specific brain regions. Alzheimer's disease (AD) is the most common neurodegenerative disease [1], which affects the brain and hence memory. It is a chronic, progressive organic brain disorder characterized by disturbance of multiple cortical functions, including memory, judgment, orientation, comprehension, learning capacity and language [2]. Parkinson's disease (PD) is a debilitating neurodegenerative disease characterized by bradykinesia, rigidity with cogwheeling, rest tremor, and postural instability. The prevalence of Parkinson's disease worldwide is known to range from 0.5 percent to 4 percent among the elderly aged 65 years or older [3]. Vascular dementia (VD) is the second most common cause of dementia after Alzheimer's disease. Accounting for up to one third of all dementias. The risk factors and the symptoms are same as that of Alzheimer's some times it is very difficult to differentiate between vascular dementia and that of Alzheimer's disease especially in a patient's with mixed dementia. The prevalence of Alzheimer disease (AD) and Parkinson's disease is increasing in the elderly. The prevalence of AD is projected to quadruple by the year 2047. The prevalence of diabetes and glucose intolerance in the elderly was over 40% in the Third National Health and Nutrition Examination Survey [4]. Persons in midlife have a 90% lifetime risk of developing hypertension [5]. Hyperlipidemia increases the risk of cardiovascular disease and it increases in adult life [6]. More importantly, these risk factors are all modifiable, representing an opportunity for the prevention of AD. Diabetes [7-9], hyperlipidemia [10], hypertension [11], heart disease [12], smoking [13, 14], homocysteine, and obesity are associated with a higher risk of AD. For the given dataset the diagnosis of AD and VD was established based on all available information gathered from the initial and follow-up assessments. AD was determined by consensus at a conference of physicians, neurologists, neuropsychologists, and psychiatrists. The diagnosis of vascular dementia was based on standard research criteria including memory impairment, on the neuropsychological test battery as well as evidence of impairment in social or occupational function. Where as for PD, the diagnosis was based on the National Institute of Neurological and Communication Disorders and Stroke/Alzheimer's Disease and Related Disorders Association criteria [15]. Our objective was to classify the AD, VD and PD based on the most influencing risk factors using various classifier techniques.

Sandhya Joshi is with Department of Computer Science and Engineering, MGR University, Chennai. INDIA. (E-mail: sanjoshi17@yahoo.com).

Vibhudendra simha G.G, P. Deepa Shenoy and Venugopal K.R are with Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, K. R Circle, Bangalore. INDIA. (e-mail:shenoypd@yahoo.com).

L.M. Patnaik is the Vice Chancellor of Defence Institute of Advanced Technology, Pune, India.

II. Literature survey

The prevalence of AD, VD and PD is increasing in the elderly. A detailed study on the various risk factors for Alzheimer's disease has been proposed by many researchers. J.A. Luchsinger et. al [16], the authors followed 1,138 individuals without dementia at baseline (mean age 76.2) for a mean of 5.5 years. In that study Participants were enrolled in a longitudinal cohort study by a random sampling of Medicare recipients age 65 or older residing in northern Manhattan Washington Heights, Hamilton Heights, Inwood, New york. Participants underwent in-person interviews of general health and function, medical history, physical and neurological examination, as well as a neuropsychological battery. They found that Four risk factors (diabetes, hypertension, heart disease, and current smoking) were associated with a higher risk of AD ($p< 0.10$) when analyzed individually. The risk of AD increased with the number of risk factors (diabetes, hypertension, heart disease, current smoking).

In a population based cohort of 2,574 Japanese-American men enrolled in the Honolulu-Asia Aging Study shows that the risk factors for AD is closely associated with that of Vascular dementia [17]. Longitudinal population-based studies [18-20] have shown that diabetes is associated with dementia as well as with the subtypes vascular dementia and Alzheimer's disease.

In another study Gang Hu et al [21], they prospectively followed 51,552 Finnish men and women of 25 to 74 years of age without a history of Parkinson's disease at baseline. Ascertainment of the Parkinson's disease status was based on the Nationwide Social Insurance Institution's Drug Register data. Hazards ratios of incident Parkinson's disease associated with the history of type 2 diabetes were estimated. It is also found that body mass index, systolic blood pressure, total cholesterol, education, leisure time physical activity, smoking, alcohol consumptions, are strongly associated with type 2 diabetes and they are also considered as the major risk of Parkinson's disease.

ML and NN methods were found to be beneficial to understand the disease progression and to asses prophylactic strategies which can distinguish the different states of dementia particularly AD [22].

III. Problem definition

The data set of consisted of 746 patients records collected from ADRC & ISTAART, USA. The base line data were collected from 1992 through 1994. Follow-up data were collected at intervals of approximately 18 months; this study includes data collected up to 2003. Around eight hundred and ninety patients were recruited to ADRC and diagnosed with AD (65%), VD (38%) and PD (40%), according to established criteria [23]. In our study we are concentrating mainly on the major risk factors which are responsible for Alzheimer's disease, vascular dementia and Parkinson's disease. The main objectives are:

1) To classify Alzheimer's disease, Vascular dementia and Parkinson's diseases by using various Machine Learning and Neural Network techniques.
2) To find the most influencing risk factors causing Alzheimer's disease, vascular dementia and Parkinson's disease.
3) To show that vascular risk factors increases the risk of Alzheimer's disease.
4) To show that stroke is the most important risk factor for both vascular dementia and Parkinson's disease.
5) Comparison of various classification techniques and finding the best classification technique for the given data.
6) To explore the effect of modification of one risk factor by another during the classification of AD, VD and PD. (e.g., diabetes by hypertension, heart disease, or smoking).

IV. Architecture and Model

The architecture and modeling of the present paper is shown in the Fig.1. It begins with the collection of patients records, which is followed by preprocessing of data set includes converting the data in to numeric values. The main focus of our work is on the selection of most influencing risk factors for AD, VD and PD. We used *chi-square* attribute evaluation scheme with ranker search method, along with this five more different attribute evaluation schemes then we developed different models for the classification of AD, VD and PD using various classification techniques.

A. Collection of patient's records

The patient's records are collected from Alzheimer's disease Research Center (ADRC) and International Society to Advance Alzheimer Research and Treatment (ISTAART), U.SA. Around eight hundred and ninety patients were recruited to ADRC and diagnosed with AD (65%), VD (38%) and PD (40%), according to established criteria. Patients were self referred or physician-referred and most often came for an initial diagnostic evaluation. Approximately half were from the greater Houston area, and the rest were from elsewhere in Texas or from other states. Diagnosis of Alzheimer's disease and Parkinson's disease was based on the criteria of the National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA). The description of the entire data set is shown in the Table 1

B. Pre-processing patient records

In this step the data was checked for missing and incorrect values, some of the data are in alpha numeric type which is converted in to numeric form. We found some inconsistencies in the patient's records which we corrected manually by taking some external reference.

C. Attribute Selection

The data set in our study consists of the eleven attributes, which are related to the various risk factors for AD, VD and PD. Among which the most prioritizing attributes were selected based on the attribute selection techniques.

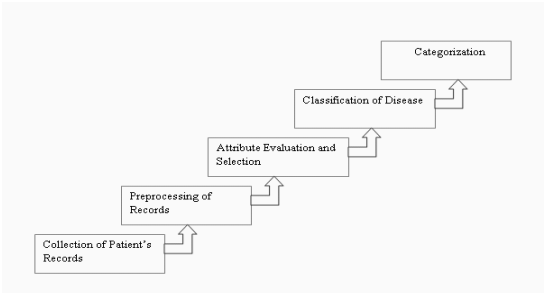

Figure 1. Architecture for the classification model

For selecting the attributes the ranker technique was used. For the given data set the five different methods were applied. Chi Squared Attribute Evaluation is the most widely used feature selection method. This method evaluates the various risk factors individually by measuring their chi-squared statistic with respect to the classes i. e. AD, VD and PD. In order to reduce the effect of the bias resulting from the use of information gain a variant known as gain ratio was introduced. The gain ratio adjusts the information gain for each attribute to allow for the breadth and uniformity of the attribute values. Gain ratio is defined by the formula:

Gain ratio = Information Gain / Split Information.

The Info Gain Attribute Evaluation method evaluates the worth of an attribute by measuring the information gain with respect to the class.

Info Gain (Class, Attribute) = H(Class) – H(Class/Attribute), Where H is the information entropy.

Similarly One R Attribute Evaluation, Symmetrical Uncertain Attribute Evaluation were applied and the results are shown in the Table 2. For each set of attributes further classification methods were applied.

D. Classification of diseases

Classification is the most commonly used data mining technique, which involves the separation of data into segments which are non-overlapping. Classification can be viewed as forecasting a discrete value. Any approach to classification assumes some knowledge about the data. Hence a training set is used to identify specific parameters. Training data requires sample input data, domain expertise, and a classification assignment to the data. In the present study we used three different methods for the classification of neurodegenerative disorders.

1) Machine Learning (ML)

Machine Learning and Knowledge Discovery from Databases (KDD) [24] are increasingly being applied in health care to build models, develop practice guidelines or refine guidelines for better medical decision making. They differ from traditional approaches by generating domain models such as decision trees, decision rules, graphs etc. from data. In the proposed system we used six different Machine Learning methods such as Decision tree [25], Bagging [26], BF tree, Random Forest tree [27], and RBF networks [28].

2) Neural Networks (NN)

Neural networks are very sophisticated modelling techniques capable of modelling extremely complex functions. The Neural network is developed based on the human nervous system and the neuron is the fundamental element [29]. A neuron has several inputs and only one output. The neural network consists of three layers 1.Input layer 2.Hidden layer 3.Output layer. The layers are connected with weighted edges. These edges are initialized and trained. We have used the crab classification in particular for the data set.

3) Multilayer Perceptrons (MLPs)

Multilayer perceptrons are feed forward neural networks trained with the standard backpropagation algorithm. They are supervised networks so they require a desired response to be trained. They learn how to transform input data into a desired response, so they are widely used for pattern classification. In our study we used three layered MLP. Inputs to the first layer (input layer) are the various risk factors, while the classification of the diseases i. e. AD, VD and PD forms the last layer, which is the output of the network.

E. Categorization

The knowledge extracted from decision tree is represented as classification using IF-THEN rules. One rule is being uniquely created for every path from root to leaf node. Each attribute value pair along a path forms the IF path i.e. the rule antecedent. Prediction is held in the leaf node which holds the rule consequent (THEN part). This concept is very easy for understanding especially when tree is large.

V. EXPERIMENTAL RESULTS

The Machine Learning system, WEKA software (Waikato Environment for Knowledge Analysis) developed at the University of Waikato in New Zealand. Weka workbench [30] is a collection of state-of-the-art machine learning algorithms and data preprocessing tools. It is designed so that the user can quickly try out existing methods on new datasets in flexible ways. It provides extensive support for the whole process of experimental data mining including preparing the input data, evaluating learning schemes statistically, and visualizing the input data and the result of learning.

Here four different machine learning algorithms were applied, namely Decision table, BF tree, Random Forest tree, and Bagging methods, and the simulation study was conducted using WEKA. Along with the ML and the Neural Network techniques such as crab classification and Multilayer Perceptron were applied by making use of Matlab software. The Matlab short for matrix laboratory is a useful tool which provides a few important data mining tools which include the Neural Network tool box.

The proposed model was tested on patients data consisting of 746 training cases and 180 testing cases, with 11 attributes such as genes, age, Alcohol(ml/day), LDL (mg/dl), BMI (Kg/m2), hypertension, smoking, diabetes, Hist_of_heart_disease, Family_hist_of_Neuropsychiatry. Our study consists of two stages; in the first stage we applied various attribute selection methods for on the given data set, the results are shown in the Table 2. Secondly for each outcome of the selection method different models was simulated using various ML and NN methods such as decision tree, BF tree, Random forest tree, Bagging method, RBF Networks and Multilayer Perceptron for the classification of Alzheimer's disease, Vascular dementia and

Parkinson's disease.

The classification accuracies obtained for various ML and NN methods for each attribute evaluation techniques are shown in the Table 3.

TABLE I. DESCRIPTION OF PATIENTS RECORD DATASET DATA SET

Characteristics	Alzheimer's Disease (AD)	Vascular Dementia	Parkinson's Disease (PD)
Age	≥68	≥59	≥74
History of heart disease			
Presence	59	79	68
Absence	65	69	85
Family History of			
Neuropsychiatry: AD	65	72	32
PD	-	24	19
VD	76	35	49
History of Hypertension			
High	132	98	165
Low	68	47	21
Normal	72	68	31
Smoking per day			
0-9	98	35	52
10-15	73	58	19
>=20	33	43	27
Alcohol (ml per day)			
250-500	89	55	110
>=500	69	67	87
History of Diabetes			
Type1	116	89	52
Type 2	112	68	110
Presence of Apolipoprotein	187	24	80
BMI (kg/m2) Above Normal	89	89	78
LDL (mg/dl) Above Normal	90	68	75
History_of Stroke	53	77	145

TABLE II. RESULTS OF VARIOUS ATTRIBUTE EVALUATION METHODS ON THE MAJOR RISK FACTORS FOR AD, VD AND PD

Ranks	*Chi-Square AttributeEval (Weights)*	*GainRatio AttributeEval (Weights)*	*Info Gain AttributeEval (Weights)*	*ReliefF AttributeEval (Weights)*	*Symmetrical Uncert Attribute Eval (Weights)*
1	Genes (0.85)	Diabetes (0.91)	Genes (15.79)	Genes (12.08)	Genes (6.85)
2	Smoking (0.74)	Genes (0.89)	Age (14.85)	Diabetic (10.67)	Diabetes (6.64)
3	Age (0.72)	Stroke (0.88)	Diabetes (10.85)	Stroke (10.68)	Stroke (6.62)
4	Diabetes (0.71)	Age (0.72)	Stroke (10.66)	Smoking (8.96)	Smoking (6.02)
5	Stroke (0.68)	Smoking (0.71)	Smoking (9.01)	Age (6.82)	Age (6.00)
6	Family_Hist (0.63)	Family_Hist (0.63)	Family_Hist (7.65)	Family_Hist (5.63)	Family_Hist (5.59)
7	Alcohol (0.48)	Alcohol (0.51)	Alcohol (5.49)	Alcohol (4.89)	Alcohol (5.29)
8	LDL (0.32)	Hist_heart_ Disease (0.31)	Hypertension (4.06)	Hypertension (4.12)	Hypertension (4.08)
9	Hypertension (0.12)	Hypertension (0.22)	Hist_heart_ Disease (1.09)	Hist_heart_ Disease (2.09)	Hist_heart_ Disease (2.85)
10	Hist_heart_ Disease (0.09)	LDL (0.18)	LDL (1.01)	LDL (1.16)	LDL (1.85)
11	BMI (0.06)	BMI (0.06)	BMI (0.09)	BMI (0.06)	BMI (0.11)

From the table 3, it was observed that the classification accuracy is high for Random forest tree for all the attribute evaluation techniques, when compared with RBF network and Decision table (Fig. 2). For neural network method, multilayer Perceptron and RBF network the classification accuracy is almost same for all the five different attribute evaluation methods.

VI. DISCUSSION

In longitudinal analyses of 746 subjects we found that there was a strong association between the risk of neurodegenerative disorders such as AD, VD and PD. The classification model was developed employing major risk factors such as age, diabetes mellitus, heart disease, hypertension, smoking, LDL, alcohol, BMI and genetical factors. We sought to identify those risk factors that are common for all the three diseases and then classifying the diseases based on those risk factors. There are some common risk factors in AD, VD and PD which play a vital role in the classification.

The most important risk factors for AD, VD and PD are APOE gene, Diabetes, Stroke and Age. The ranks and the weights of these risk factors are shown in table 2. It is evident that APOE gene, diabetes, smoking and age are the most influencing risk for AD. Stroke and diabetes play a vital role for both VD and PD, since stroke is ranked between third and fifth positions.

The most dominant risk factor in all the three neurodegerative disorders is APOE gene, which occupies the topmost position in almost all the methods except for Gain ratio. It was also observed that family history with neurodegenerative disorder is more likely to get either of these diseases. Other factors like LDL, BMI and hypertension though being risk factors, are commonly found among all patients, thus is ranked below all other risk factors. Most of the influencing risk factors are common for all diseases and if a patient suffers from Vascular Dementia there is every chance of getting Alzheimer Disease. Hence the risk factors for both the diseases are common and stroke increases the growth rate of vascular dementia.

TABLE III. RESULTS OF VARIOUS CLASSIFICATION ACCURACIES OBTAINED USING ML METHODS.

	Decision Table	BF Tree	Random Forest Tree	Bagging	RBF Network	Multilayer Perceptron	Neural Network
Chi Squared	96.8356	99.4251	99.5438	97.3857	97.9643	99.6782	98.5471
Gain Ratio	96.8356	98.3295	99.3656	97.8641	97.9643	99.6782	99.4976
Info Gain	96.8356	98.3295	99.3656	98.4961	97.9643	99.6782	98.4692
One R	96.8356	98.3295	99.4818	98.4961	97.9643	99.6782	99.3948
Relief F	96.8356	98.3295	99.5438	98.4961	97.9643	99.6782	98.2715
Symmetrical Uncertain	96.8356	98.3295	99.5438	98.364	97.9643	99.6782	99.1629

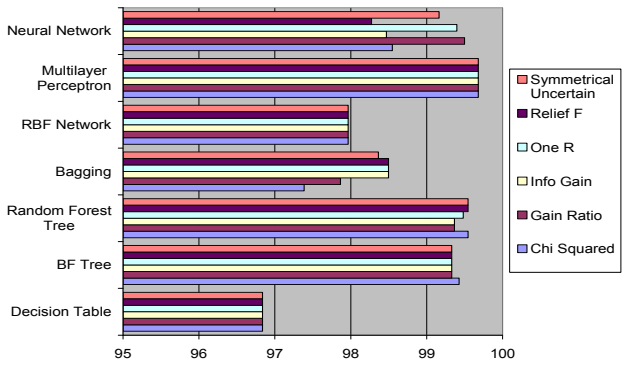

Figure 2. Classification accuracies of neurodegenerative disorders using ML methods.

AD is the most common form of dementia in the elderly. Recent epidemiological evidence suggests an association between AD and vascular risk factors such as hypertension, diabetes mellitus, BMI, LDL, APOE gene and general atherosclerosis. The role of life-style factors such as smoking and alcohol consumption in causing AD has been extensively studied. The APOE e4 allele has been repeatedly found to be associated with both AD and vascular dementia [31, 32]. The prevalence of Alzheimer disease (AD) is projected to quadruple by the year 2047 [33]. Vascular risk factors may increase the risk of AD [34] and are highly prevalent in the elderly. Diabetes, hyperlipidemia, hypertension, heart disease, smoking, homocysteine, and obesity are associated with a higher risk of AD as well as PD. Hypertension and type 2 diabetes are commonly associated conditions, and their concordance is increased in populations. Hypertension

affects up to 40% or more of diabetic patients [35].

In the general population, high blood pressure is one the most important risk factors for stroke. Stroke is considered as the major risk factor for PD. Dementia is increasingly recognized as an important feature of Parkinson's disease in the elderly.

VII. Conclusion

This paper proposes a new model for the classification of neurodegenerative disorders such as Alzheimer's disease, vascular dementia and Parkinson's disease by considering the most influencing risk factors. The concept of attribute evaluation method has been adopted in this model for evaluating the association between risk factors, and it was observed that the classification accuracy varies with the effect of modification of one risk factor by another.

By observing the classification accuracies, genes, diabetes, age and smoking seems to be strongest risk factors for pathogenesis of Alzheimer's disease. Similarly, for the classification of Parkinson's disease, the risk factors such as stroke, diabetes, genes and age play a very major role in the development of various symptoms of Parkinson's disease. It was also found that the major risk factors for vascular dementia are the diabetes, APOE gene and stroke which are also considered as the major risk factors for AD and PD. The risk of Alzheimer's disease increases with the number of vascular risk factors. The entire work demonstrates the effectiveness of considering most influential risk factor for the correct classification of AD, VD and PD. The classification model was validated with the test sample of 180 patient's records and the model achieved an average classification accuracy of 99.33% with Random forest tree and the Multilayer Perceptron. Further more, the classification model can also be tested out on different data sets for the classification of other neurodegenerative disorders such as Pick's disease and Huntington's disease.

Acknowledgment

We would like to thank Dallas Anderson, National Institute of Aging, USA, J Roy, Alzheimer's Disease Research Centre (ADRC), Hanumanthachar Joshi, International Society to Advance Alzheimer Research and Treatment (ISTAART) for providing useful information

References

[1] R. Scatena, Martorona, P. Bottani, G. Botta, P. Pastove and Giardina., Expert Opin. Investig. Drugs, vol. 16, 2007, pp. 59-72.

[2] M. E. Jay, "Cholinesterase inhibitors in the treatment of dementia," JAOA, vol. 3, 2005, pp. 145-158.

[3] L. M. De Lau, M. M. Breteler, "Epidemiology of Parkinson's disease," Lancet Neurol, vol. 5, 2006, pp. 525–35.

[4] M.I. Harris, K. M. Flegal, C. C. Cowie, "Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in U.S. adults: The Third National Health and Nutrition Examination Survey 1988–1994," Diabetes Care, vol. 21, 1998, pp. 518–524.

[5] R. S. Vasan, A. Beiser, S. Seshadri, "Residual lifetime risk for developing hypertension in middle-aged women and men: The Framingham Heart Study," JAMA, 2002, vol. 287, pp.1003–1010.

[6] J. M. Morgan, D. M. Capuzzi, Hypercholesterolemia. The NCEP Adult Treatment Panel III Guidelines. Geriatrics, vol. 58, 2003, pp. 33–38.

[7] A. Ott, R. P. Stolk, F. Harskamp, M. M. Breteler, "Diabetes mellitus and the risk of dementia: the Rotterdam Study," Neurology vol. 53, 1999, pp. 1937–1942.

[8] C. L. Leibson, W. A. Rocca, V. A. Hanson, "Risk of dementia among persons with diabetes mellitus: a population-based cohort study," Am J Epidemiol, vol. 145, 1997, pp. 301–308.

[9] R. Peila, B. L. Rodriguez, L. J. Launer, "Type 2 diabetes, APOE gene, and the risk for dementia and related pathologies: the Honolulu–Asia Aging Study," Diabetes, vol. 51, 2002, pp.1256–1262.

[10] H. Jick, G. L. Zornberg, S. S. Jick, S. Seshadri, D. A. Drachman, "The risk of dementia," Lancet, vol. 356, 2000, pp. 1627–1631.

[11] I. Skoog, B. Lernfelt, S. Landahl, "15-year longitudinal study of blood pressure and dementia," Lancet, vol. 347, 1996, pp. 1141–1145.

[12] A. Otto, M. M. Breteler, M. C. de Bruyne, F. Van Harskamp, "Atrial fibrillation and dementia in a population-based study. The Rotterdam Study," Stroke, vol.28, 1997, 316–321.

[13] A. Ott, A. J. Slooter, A. Hofman, "Smoking and risk of dementia and Alzheimer's disease in a population-based cohort study: the Rotterdam Study," Lancet, vol.351, 1998, pp.1840 –1843.

[14] C. Merchant, M. X. Tang, S. Albert, J. Manly, Y. Stern, "The Influence of Smoking on the risk of Alzheimer's disease," Neurology, vol. 52, 1999, pp. 1407-1408.

[15] G. McKhann, D. Drachman, M. Folstein, R. Katzman, "Clinical diagnosis of Alzheimer's disease: report of the NINCDS/ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease," Neurology, vol. 34, 1984, pp. 939–944.

[16] J. A. Luchsinger, M. X. Tang, Y. Stern, S. Shea, R. Mayeux, "Diabetes mellitus and risk of Alzheimer's disease and dementia with stroke in a multiethnic cohort," Am J Epidemiol, vol. 154 2001, pp. 635–641.

[17] Rita Peila, L. Beatriz, Rodriguez, "Type 2 Diabetes, APOE Gene, and the Risk for Dementia and Related Pathologies," Diabetes, vol. 51, 2002, pp. 44-52.

[18] J. Kuusisto, K. Koivisto, L. Mykkanen, Helkala, "Association between features of the insulin resistance syndrome and Alzheimer's disease independently of apolipoprotein e4 phenotype: cross sectional population based study," BMJ, vol. 315, 1997, pp. 1045–1049.

[19] C. L. Leibson, W. A. Rocca, V. A. Hanson, R. Cha, "Risk of dementia among persons with diabetes mellitus: a population-based cohort study," Am J Epidemiol, vol. 145, 1997, pp.301–308..

[20] A. Ott, R. P. Stolk, F. Van Hardkamp, A. Hofman, M. M. B. Breteler, "Diabetes mellitus and the risk of dementia," Neurology, vol. 53, 1999, pp. 1937–1942.

[21] Gang Hu, Pekka Jousilahti, Siamak Bidel, Riitta Antikainen, Jaakko Tuomilehto, "Type 2 diabetes and the risk of Parkinson's disease," Diabetes Care In Press, published online January 24, 2007.

[22] Sandhya Joshi, P. Deepa Shenoy, K. R. Venugopal , L. M. Patnaik, "Evaluation of Different Stages of Demntia Employing Neuropsychological and Machine Learning Techniques", IEEE, ICAC 2009, PP. 154 – 160.

[23] G. McKhann, D. Drachman, M. Folstein, R. Katzman R, D. Price, "Clinical diagnosis of Alzheimer's disease: report of the NINCDS/ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease," Neurology, vol. 34, 1984, pp. 939–944.

[24] U. Fayyad, "Data Mining and Knowledge Discovery in databases: Implications from Scientific Databases," In: Proc. Of the 9th Int. Conf. on scientific and Statistical Database Management, pp. 2-11. Olympia, Washington, USA, (1997).

[25] J. R. Quinlan, "C4.5: Programs for machine learning", Morgan. Kaufmann, Los Altos, Eds., Academic, California, pp.271- 350, 1993.

[26] J. Han and M. Kamber, "Data mining-concepts and techniques". St. Louis, Elsevier, 2nd Edition, 2006.

[27] Suri Pushpa, Prasad Vinod, Carsten Maple, "Creating a Random Forest Trees for a Multiprocessor System," parelec, pp.290-295, International Symposium on Parallel Computing in Electrical Engineering (PARELEC'06), 2006

[28] Buhmann, D. Martin, (2003), Radial Basis Functions: Theory and Implementations, Cambridge University Press, ISBN 978-0-521-63338-3.

[29] C. Kwan, F. L. Lewis, "Robust backstepping control of nonlinear systems using neural networks," IEEE Trans. on Systems, Man, and Cybernetics Part A, vol. 30, 2000, pp. 753–766.

[30] I. H. Witten and E. Frank. "Data mining- practical machine learning tools and techniques". St. Louis, Elsevier, 2005.

[31] L. A. Farrer, A. Cupples, L. Haines, B. Hyman, "Effects of age, sex, and ethnicity on the association between apolipoproteinE genotype andAlzheimer disease," JAMA, vol. 278, 1998, pp. 1349– 1356.

[32] J. Pedro-Botet, M. Senti, X. Nogues, "Lipoprotein and Apolipoprotein profile in men with ischemic stroke: role of lipoprotein (a), triglyceride-rich lipoproteins, and apolipoprotein E polymorphism," Stroke, vol. 23, 1992, vol. 1556–1562.

[33] R. Brookmeyer, S. Gray, C. Kawas, "Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset," Am J Public Health, vol.88, 1998, pp. 1337–1342.

[34] J. Luchsinger, R. Mayeux, "Cardiovascular risk factors and Alzheimer's disease," Curr Atheroscler Rep, vol.6, 2004, pp. 261–266.

[35] B. J. Materson, S. Oparil, J. T. Wright, "Diabetes Association. Treatment of hypertension in adults with diabetes," Diabetes Care. Vol. 26 , 2003, pp. 80 –82..

Sandhya Joshi is a research scholar pursuing her doctoral research work on Data mining applications to neurodegenerative disorders with special reference to Alzheimer's disease in M G R University, Chennai. She completed her master's degree in computer application from Visveshwaraiah Technological University. Her research areas of interest include data mining, Neural networks, Behavioral disorders and aging associated diseases. She has published number of research papers in scientific journals and international conferences.

G. G Vibhudendra Simha is pursuing his Bachelors in Computer Science and Engineering at University Visvesvaraya College of Engineering, Bangalore University, Bangalore. His research interests include Data Mining.

P Deepa Shenoy obtained her Ph.D in Computer Science and Engineering from the Bangalore University. Currently she is working as Professor, Department of Computer Science and Engineering, at University Visvesvaraya College of Engineering, Bangalore University, Bangalore. She received Bachelors and Masters Degree from Bangalore University. She has published more than thirty research papers in International Conference and Journals. Her research interests are Soft Computing, Data Mining and Bio-metrics.

K R Venugopal is currently Principal, University Visvesvaraya College of Engineering (U.V.C.E), Bangalore University, Bangalore. He obtained his Bachelor of Engineering from U.V.C.E. He received his Masters degree in Computer Science and Automation from Indian Institute of Science, Bangalore. He was awarded Ph.D in Economics from Bangalore University and Ph.D in Computer Science from Indian Institute of Technology, Chennai. He has a distinguished academic career and has degrees in Electronics, Economics, Law, Business Finance, Public Relations, Communications, Industrial Relations, Computer Science and Journalism. He has authored 27 books on Computer Science and Economics which include Petro Dollar and World Economy, Mastering C, Microprocessor Programming, Mastering C++ etc. He has also been serving as Professor and Chairman, Department of Computer Science and Engineering, U.V.C.E, Bangalore University, Bangalore. During his three decades of service at U.V.C.E, he has over 200 research papers to his credit. His research interests include Computer Networks, Parallel and Distributed Systems, Digital Signal Processing and Data Mining.

L M Patnaik is Vice Chancellor, Defence Institute of Advanced Technology, Pune, India. He was a Professor with Department of Computer Science and Automation, Indian Institute of Science (IISc) - Bangalore since 1986. During the past 35 years of his service at IISc, he has published more than 500 search publications in refereed International Journals and Conference Proceedings. He is a Fellow of all the four leading Science and Engineering Academics in India, Fellow of IEEE and the Academy of Science for the Developing World. He has received twenty national and international awards, notable among them are the IEEE Technical Achievement Award for his significant contributions to High Performance Computing and Soft Computing. His areas of research interests have been Parallel and Distributed Computing, Mobile Computing, CAD for VLSI circuits, Soft Computing and Computational Neuroscience.