## **JCS** International Journal of Computer Science

Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS

ISSN: 2348-6600



http://www.ijcsjournal.com Reference ID: IJCS-194 Volume 5, Issue 1, No 8, 2017

PAGE NO: 1190-1194

**ISSN: 2348** 

#### Sri Vasavi College, Erode Self-Finance Wing

3<sup>rd</sup> February 2017

National Conference on Computer and Communication NCCC'17

http://www.srivasavi.ac.in/

nccc2017@gmail.com

#### EXPLORING ON CLASSIFICATION AND PREDICTION OF DIABETICS USING DATA MINING AND NEURAL NETWORK

P. Anitha M.C.A., M.Phil., B.Ed.,
Part Time Research Scholar,
Vellalar Collge for Women (Autonomous),
Erode – 638012.
E-Mail: anithavcw@gmail.com
Mobile: 9842660060

**Dr. P.R. Tamilselvi M.Sc., M.C.A., M.Phil, Ph.D.,** Research Guide & Assistant Professor, Govt. Arts and Science College, Komarapalayam – 638183. Mobile:9942055733

**ABSTRACT-** Data mining is the process of pattern discovery and extraction where huge amount of data is involved. Its applications can greatly benefit all parties involved in the healthcare industry. The huge generated by amounts of data healthcare transactions are too complex and voluminous to be processed and analyzed by traditional methods. It provides the methodology and technology to transform these mounds of data into useful decision making. information for Different techniques are used to find interesting patterns for medical diagnosis and treatment. Diabetes is a group of meta bolic disease in which there are high blood sugar levels over a prolonged period. Diabetes has affected over 246 million people worldwide with a majority of them being women. According to the WHO report, by 2025 this number is expected to rise to over 380 million. This paper concentrates on the overall literature survey related to various data mining techniques for predicting diabetes. This would help the researchers to know various data mining algorithm and method for the prediction of diabetes mellitus. In regard to this emerge, we have reviewed the various paper involved in this field in terms of method, algorithms and results.

**Keywords:** Data Mining, Classification Techniques, Diabetes Mellitus, Neural Network and Health Care.

#### **I INTRODUCTION**

The development of Information Technology has generated large amount of databases and huge data in various areas. The research in databases and information technology has given rise to an approach to store and manipulate this precious data for further decision making. Data mining is the extraction of hidden predictive information from large data bases. It is the computational process of discovering patterns in large datasets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. Classification is the processing of finding a set of models which describe and distinguish data classes or concepts. The derived model is based on the analysis of a set of training data. The derived model may be represented in various forms, such as classification (IF-THEN) rules, decision trees, mathematical formulae, or neural networks. The diabetes prediction and awareness system is developed and implemented using classification based data mining algorithm. It helps the user to know whether they are diabetic or non-diabetic. It

# **IJCS** International Journal of Computer Science

Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS

ISSN: 2348-6600

**ISSN: 2348** 

**PAGE NO: 1190-1194** 

3<sup>rd</sup> February 2017

nccc2017@gmail.com

http://www.ijcsjournal.com **Reference ID: IJCS-194** 

Volume 5, Issue 1, No 8, 2017

### Sri Vasavi College, Erode Self-Finance Wing

National Conference on Computer and Communication NCCC'17

#### http://www.srivasavi.ac.in/

also raises awareness among the user and helps to keep track of their health status.

#### **II DIABETES DISEASE**

Diabetes describes a group of metabolic diseases in which the person has high blood glucose either because insulin production is inadequate, or because the body's cells do not respond properly to insulin, or both. Patients with high blood sugar will typically experience polyuria; they will become increasingly thirsty and hungry. Diabetes can be classified into three types.

- Type 1 diabetes (insulin-dependent diabetes)
- Type 2 diabetes (insulin resistance)
- Gestational diabetes
- Determination of diabetes disease:

Doctors can determine whether a patient has a normal metabolism, prediabetes or diabetes in one of three different ways - there are three possible tests:

- The A1C test
- The FPG (fasting plasma glucose) test

The OGTT (oral glucose tolerance test) Complications of diabetes are:

Foot complications - Skin complications - Heart problems-Hypertension \_ Mental health-Neuropathy

#### **III LITERATURE SURVEY**

In this year 2014 Soumadip Ghosh et.al. [1], "A NEURO-FUZZY NOVEL CLASSIFICATION TECHNIQUE FOR DATA MINING", in this study, they proposed a novel Neuro-fuzzy classification technique for data mining. The inputs to the Neurofuzzy classification system were fuzzified by applying generalized bell shaped membership

function. The proposed method utilized a fuzzification matrix in which the input patterns were associated with a degree of membership to different classes. Their objective was to analyze the proposed method and, therefore compare its performance with two powerful supervised classification algorithms Radial Basis Function Neural Network (RBFNN) Adaptive Neuro-fuzzy Inference System and (ANFIS). They assessed the performance of these classification methods in terms of different performance measures such as accuracy, root-mean square error, kappa statistic, true positive rate, false positive rate, precision, recall, and f-measure. The results suggest that the proposed NFS classifier has the potential to significantly improve the conventional classification methods for use.

In this year 2014 G.Keerthana et al. [2]. "PERFORMANCE ENHANCEMENT OF CLASSIFIERS USING **INTEGRATION** OF **CLUSTERING** AND CLASSIFICATION TECHNIQUES", in this paper, performance comparison of simple classification algorithms and integrated clustering and classification algorithms are carried out. It was found that the integrated clustering-classification technique was better than the simple classification technique. In this paper four different classifiers are integrated with the simple k-means clustering algorithm and density clustering algorithm. based This integration technique was applied on "Diabetes" data set. From the observation and analysis it was concluded that the performance of Density Based + NavieBayes is better than other algorithms.

In this year 2015 Aurea Celeste Ribeiro et.al. [3], CLASSIFICATION "DIABETES USING А REDUNDANCY REDUCTION



Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS

ISSN: 2348-6600

**ISSN: 2348**-

http://www.ijcsjournal.com **Reference ID: IJCS-194** 

Volume 5, Issue 1, No 8, 2017

3<sup>rd</sup> February 2017

**PAGE NO: 1190-1194** 

Sri Vasavi College, Erode Self-Finance Wing

National Conference on Computer and Communication NCCC'17

http://www.srivasavi.ac.in/

nccc2017@gmail.com

PREPROCESSOR" This work proposes а classification methodology based on efficient coding of the input data, which is carried out by decreasing input data redundancy using well-known ICA algorithms, such as Fast ICA, JADE and INFOMAX. The classifier used in the task to discriminate diabetics from non-diabetics is the one class support vector machine. Classification tests were performed using noninvasive and invasive indicators. The results suggest that redundancy reduction increases one-class vector support machine performance when discriminating between diabetics and non diabetics up to an accuracy of 98.47% while using all indicators. By using only noninvasive indicators, an accuracy of 98.28% was obtained. The ICA feature extraction improves the performance of the classifier in the data set because it reduces the statistical dependence of the collected data, which increases the ability of the classifier to find accurate class boundaries.

M.Mounika et.al. [4], In this year 2015 "PREDICTIVE ANALYSIS OF DIABETIC TREATMENT USING **CLASSIFICATION** ALGORITHM", The data mining tool and classification algorithm is applied to clinical datasets for the prediction of blood glucose level. he successful diabetic treatment effective and classification calculation for predictions are discussed. It is cleared that drug prescription is effective for old age groups with type-2 diabetes, drug prescription and dietary controls is effective for adult age groups where as young age groups need to concentrate on other factors such as dietary controls, physical exercise, smoke cessation and insulin. Therefore this prediction gives a positive mode of treatment for different age groups.

Classification algorithm such as Naïve Bayes, OneR and ZeroR is applied to diabetes datasets collected from the clinic and blood sugar level for young, old and adult patients is predicted using the rule generated by each models. This experiment shows that Naïve Bayes is the fastest and ZeroR is the slowest. The performance comparison is found using weka data mining tool.

In this year 2016 S.Poonkuzhali et.al. [5], "DESIGN OF A DATA ANALYTICS MODEL TO PREDICT DIABETES USING MACHINE LEARNING TECHNIQUES" the proposed data analytics system will use a hybrid classifier model that uses two different data mining classification techniques to classify the disease. The prediction of individual instance of data is done by a machine learning framework that uses Artificial Neural Networks. A Multilayer Perceptron Neural Network is created and it is trained using Back Propagation Algorithm that will classify the patients tested positive as 1 and patients tested negative as 0. This trained neural network gave a total mean square error of 0.1213 on test. The system thus produced a recognition rate of 80%.

This research work aims at providing an automated diagnosis system which works at a very high recognition rate than the previous systems. Early diagnosis of the disease can help the patient avoid the severe health complications due to diabetes. This will help the patients to take continuous care of their health by properly following the physician's advice. This system helps in early diagnosis of this disease. This Artificial Neural Network trained with back propagation resulted in 80% recognition rate which is found to be slightly higher.

# **IJCS** International Journal of Computer Science

Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS

ISSN: 2348-6600

**ISSN: 2348-6** 

http://www.ijcsjournal.com **Reference ID: IJCS-194** 

Volume 5, Issue 1, No 8, 2017

3<sup>rd</sup> February 2017

nccc2017@gmail.com

**PAGE NO: 1190-1194** 

Sri Vasavi College, Erode Self-Finance Wing

National Conference on Computer and Communication NCCC'17

http://www.srivasavi.ac.in/

In this year 2016 Mehrbakhsh Nilashi et.al. [6], "A SOFT COMPUTING APPROACH FOR DIABETES DISEASE CLASSIFICATION" The aim of this study is to classify diabetes disease by developing an intelligence system using machine learning techniques. In this method is developed through clustering, noise removal and classification approaches. Accordingly, we use expectation maximization, principal component analysis and support vector machine for clustering, noise removal and classification tasks, respectively. They also develop the proposed method for incremental situation by applying the incremental principal component analysis and incremental support vector machine for incremental learning of data. Experimental results on Pima Indian Diabetes dataset show that proposed method remarkably improves the accuracy of prediction and reduces computation time in relation to the non-incremental approaches. The hybrid intelligent system can assist medical practitioners in the healthcare practice as a decision support system.

#### In year 2016 Dr. Meena Arora et.al. [7],

"RANDOM FOREST V/S SCALED CONJUGATE PREDICT GRADIENT TO DIABETES MELLITUS", this study takes after the procedures utilizing random forest tree as a base learner alongside standalone information mining procedure scaled conjugate gradient to characterize patients with diabetes mellitus utilizing diabetes hazard variables. This characterization is done crosswise over three diverse ordinal grown-ups bunches in PIMA Indian dataset. Test result demonstrates that, general execution of adaboost group strategy is superior to anything sacking and in addition standalone random forest tree. They can be utilized

for the investigation of vital clinical parameters, expectation of different maladies, estimating errands in solution, extraction of medicinal learning, treatment arranging backing and patient administration. Various calculations were proposed for the expectation and determination of diabetes. These calculations give more precision than the accessible customary frameworks. This paper incorporates calculations of random forest tree and scaled conjugate gradient. From the perception SCG have the slightest characterization exactness and Random forest tree give the better grouping precision results.

#### **IV CONCLUSION**

Data mining has played an important role in diabetes research. Data mining would be a valuable asset for diabetes researchers because it can unearth hidden knowledge from a huge amount of diabetesrelated data. We believe that data mining can significantly help diabetes research and ultimately improve the quality of health care for diabetes patients. Improving the quality of patient care and reducing healthcare costs are the ideal goals of many programs. Data mining has helped these programs succeed. Healthcare facilities and groups use data mining tools to reach better patient-related decisions. Patient satisfaction is improved because data mining provides information that will help staff with patient interactions by recognizing usage patterns, current and future needs, and patient preferences.

### REFERENCES

1.Neha Shukla and Dr. Meena Arora, "Random Forest v/s Scaled Conjugate Gradient To Predict International Journal of Diabetes Mellitus",

International Journal	of Computer Science
Scholarly Peer Review	ewed Research Journal - PRESS - OPEN ACCESS
ISSN: 234	88-6600
http://www.ijcsjournal.com Volume 5,	Issue 1, No 8, 2017 ISSN: 2348-6600
Reference ID: IJCS-194	PAGE NO: 1190-1194
Sri Vasavi College, Erode Self-Finance Wing National Conference on Computer and Communication NCCC'17 http://www.srivasavi.ac.in/3 <sup>rd</sup> February 2017 nccc2017@gmail.com	
Computational IntelligenceResearchISS0973-1873, ResearchIndiaPublications: http://www.ripublication.com 2.Research on BioMedical engineering rbejournal.org Volume 31, Number 2, p. 97-106, 2015, "Diabetes classification using a redundancy reduction preprocessor". 3.M Mounika et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol, "Predictive Analysis of Diabetic Treatment Using Classification Algorithm".	<ul> <li>Mining Model to Predict the Occurrence of Type-2 Diabetes using Neural Network".</li> <li>6.Centers for disease control and prevention, national diabetes, fact sheet, www.cdc.gov/diabetes (2011).</li> <li>7.IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 3, January 2012 ISSN (Online): 1694-0814 www.IJCSI.org. "Implementation of Genetic Algorithm in Predicting Diabetes" S.Sapna1, Dr.A.Tamilarasi2 and M.Pravin Kumar3.</li> <li>8.Han, J.,&amp;Micheline, K. (2006). Data mining:</li> </ul>
4.Dr. V.Srividhya G.Keerthana et al. / International	Concepts and Techniques, Morgan Kaufmann
Journal of Computer Science Engineering (IJCSE)	Publisher.
"Performance Enhancement of Classifiers using	9.Jun12doi:10.4028/www.scientific.net/JERA.24.1
Integration of Clustering and Classification	37 Accepted: 2016-04-13. FFBAT- Optimized
Techniques"	Rule Based Fuzzy Logic Classifier for Diabetes.

5.International Conference on Recent Trends in Computational Methods, Communication and Controls (ICON3C 2012) " An Improved Data 10.Egyptian Informatics Journal (2014) 15, 129– 147 2014 " A novel Neuro-fuzzy classification technique for data mining" Soumadip Ghosh .