International Journal of Scientific & Engineering Research Volume 8, Issue 5, May-2017
ISSN 2229-5518

Predictive analysis model to determine predisposition to Type II Diabetes Mellitus and its complications

Aparna Deshmukh, Shweta Sharma ,Reshma Desai

Abstract—Diabetes Mellitus and associated complications is one of the most important public health challenges for all countries. It is a major cause of blindness, kidney failure, heart attacks, stroke and lower limb amputation. Diabetes can be treated and its complications can be avoided or delayed by means of dietary changes, physical activity, regular screening and appropriate treatment.

As per the recent data of WHO, incidence of diabetes has been rising rapidly. Identification of pre-diabetic state in healthy population is important in order to control progression of the metabolic disorder. Predictive analytical IT tools based on predisposing factors are of great value to determine the transition from healthy to pre-diabetic state and also from pre-diabetic to diabetic state. Pre-diabetic individuals can be more effectively subjected to preventive measures and therapeutic options to avoid or delay occurrence of Diabetes Mellitus. We present an overview of existing predictive tools and propose a composite multi-factorial IT based predictive model for risk assessment in Diabetes mellitus. The model will serve as a resourceful tool useful for Healthcare sector for strategic management of diabetes mellitus.

Index Terms— Diabetes mellitus, predictive model, management of metabolic disorder

1 INTRODUCTION

Type II Diabetes mellitus (T2DM) or Non Insulin Dependent Diabetes mellitus (NIDDM) is a metabolic disorder characterized by insulin resistance, impairment in glucose utilization and hyperglycemia. Other symptoms include loss of weight, fatigue, increased thirst, frequent urination, blurred vision, delayed wound healing etc. Diabetes mellitus is associated with severe complications such as diabetic foot disease, diabetic neuropathy, retinopathy, nephropathy, cardiovascular problems etc. Diabetes is a major cause of blindness, kidney failure, heart attacks, stroke and lower limb amputation. Over the past two decades the incidence of diabetes mellitus has increased significantly (Global report, WHO 2016) and have posed one of the important global challenges to public health sector. According the most recent WHO data, around 1.5 million deaths were caused due to diabetes and 2.2 million deaths were attributable to high blood glucose in 2014. WHO projects that diabetes will be the seventh leading cause of death in 2030 (Mathers 2006). India has been amongst the top countries with high number of diabetic population and increase in prevalence of Type II Diabetes mellitus (Global report WHO, 2016).

Diagnosis of the disorder is based on laboratory tests such as fasting and post-prandial (2hrs) serum glucose levels, insulin levels, glycated products etc., which are normally tested if any clinical symptoms are detected. However, development of diabetes is a gradual process, in which metabolism is impaired even before the appearance of clear clinical symptoms. Identification of the pre-diabetic state is important to arrest the process of disease development. Impaired glucose tolerance is one of the early sign of pre-diabetic state. Certain predisposing factors such as genetic constitution, lifestyle and unhealthy diet increase the risk of development of Type II Diabetes mellitus (Garber 2008).

Due to the complexity of the multi-factorial disease advancement, diabetes risk assessment is a critical area for improving medical decision support and prevention. Data mining is the field of computer science which is applied for processing of large data sets for data analysis and visualization for representing trends, using various techniques. Segmentation or Clustering are such techniques to identify groups and structures in the data that are correlated. Data mining tools and techniques can be used to discover new patterns, predict output of complex interdependent processes and interpret the data to provide meaningful and useful information for better understanding about diseases like diabetes mellitus. In the present paper, we propose use of data mining tools which utilize combination of biological, genetic and other factors to determine predisposition to the pre-diabetic state and to assess the risk of development of Type II Diabetes mellitus.

The paper is organized is as follows: Section I: A review of predictive tools by other scientists for Diabetes mellitus. Section II: Proposed model by the authors of the present paper.

Section I: Review of predictive tools and predisposing attributes for Diabetes mellitus

In the past few years, there has been a tremendous increase in the number of reports that have proposed models for prediction of type 2 diabetes mellitus. This increase is mainly due to the development of data mining tools and other advances made in the field of computer science and bio-informatics. Many of the Indian reports have proposed

prediction models based on the data obtained from Pima Indian diabetes dataset. They have used the eight attributes namely : Number of times pregnant, Plasma glucose concentration after 2 hours in an oral glucose tolerance test, Diastolic blood pressure, Triceps skin fold thickness, 2 Hour serum insulin, Body mass index, Diabetes pedigree function and Age. Based on these eight attributes, Adidela et al (2012) proposed a hybrid classification system and Iyer et al (2015) demonstrated the use of Decision Trees and Naïve Bayes algorithm to only identify a patient as tested positive or tested negative for diabetes, whereas Rupa Bagdi et al developed a decision support system which predicted the diabetes probability as low, high and medium.

Many of the authors have analyzed the data by using different clustering tools and made comparisons between them. An overview of the prediction models has been summarized as table number 1.

Shetty et al (2014) used hive and R to analyze the data, which is inconclusive with respect to the prediction model. Afrand et al (2012) suggested a different classifier system which they demonstrated to have greater accuracy as compared to other conventional data mining tools. Nithya et al (2015) compared clustering tools and suggested K-means algorithm as a good clustering tool that can be applied on diabetes data. Rajesh et al (2012) and Kumar et al (2012) compared classification algorithms and concluded C4.5 algorithm to be the most suited for clustering. Vijayan et al (2014) compared data mining algorithms and established that Co active ANFIS provided better classification and prediction accuracy as compared to the other tools. Even though these reports provide valuable information about the tools that can be applied for development of prediction models, the lacunae include usage of a few number and types of attributes and do not give a clear indication of the severity of the disease from the obtained results. Therefore they only propose the use of certain models but do not conclusively categorize patients as diabetic or non-diabetic from the prediction obtained.

It is very important to broaden the spectrum of the attributes which will increase the accuracy of the prediction models. Nagarajan et al (2015) included glycated haemoglobin, gender, insulin dependency, Living area, job type, food habits along with the earlier mentioned eight attributes on the basis of which they proposed a prediction model. Using clustering tools, they clustered the data for the attributes. This clustered dataset was given as input to the model which classified each patient's risk levels of diabetes as mild, moderate and severe. Kavitha et al (2012) developed a tool for diagnosis assessment which offered functions of DM based on the CART method. They also described the construction of a decision tree for diabetes diagnostics and its use as a basis for generating knowledge base rules. The system developed achieved 96.39% accuracy and 100.00% precision in detecting Diabetes.

Advances in molecular biology and genome sequencing have made it possible to characterize diabetes at the genetic level. Genome-wide association studies (GWAS) have identified genetic markers in the form of SNPs which predispose individuals to diabetes. They therefore, are extremely important attributes which should be considered when developing a prediction model for Diabetes mellitus. Lee et al (2011) identified 18 SNPs which were significantly associated with T2DM in their study subjects. They developed predictive models for T2DM using clinical and genotype data and were able to identify risk factors in the various clusters formed. However, the misclassification rates were found to be higher than expected.

The deregulation of many biological pathways precedes the development of type 2 diabetes. In recent years, research in diabetes has been able to identify many new molecules which are components of these biochemical pathways and therefore serve as biomarkers for a particular molecular stage of diabetes. This has made it possible to identify individuals who are at an early risk of developing diabetes (prediabetic stage) and can begin interventions as early as possible to reverse the condition. They have also helped to identify high risk individuals. Although many studies have assessed whether levels of a few molecules might predict future diabetes, none have quantitatively measured a large number of molecules simultaneously in a sufficient number of samples to robustly evaluate their utility for risk assessment. Kolberg et al., (2009) undertook a systematic analysis to search for patterns of biomarkers with more predictive power than individual biomarkers or previously examined biomarker combinations. By applying a variety of statistical methods for biomarker selection they developed a DRS (Diabetes risk score) model that incorporates six circulating biomarkers: adiponectin, C-reactive protein, ferritin, glucose, haemoglobin A1C (HbA1c), interleukin 2 receptor a (IL2Ra) and insulin. Golfine et al (2011) applied multiple regression selection techniques to identify the most informative biomarkers and developed multivariate models to estimate glucose tolerance, insulin sensitivity, and insulin secretion. The ability of the glucose tolerance model to discriminate between diabetic individuals and those with impaired or normal glucose tolerance was evaluated by area under the ROC curve (AUC) analysis. They identified fasting glucose, leptin, IGFBP-1, GPT, and HbA1c as the most informative markers for glucose tolerance; fasting glucose, insulin, Fas ligand, complement C3, and PAI-1 as most informative for insulin sensitivity; and fasting glucose, insulin, PAI-1, ACE, and IL-2R_ as most informative for insulin secretion.

Table 1:Overview of diabetes prediction models

Authors	Details of the data mining tool/algorithm
Shetty et al (2014)	Used Hive and R to analyze 8 using parameters: Number of times pregnant, Plasma glucose concentration, Serum Insulin, Diastolic BP, Diabetes pedigree, Body Mass Index , Age , Triceps skin fold thickness
Velide Phani	Analyzed diabetes data using data mining techniques: Naive Bayes, J48(C4.5) JRip ,

Kumar et al (2012)	Neural networks, Decision trees, KNN, Fuzzy logic and Genetic Algorithms based on accuracy and time.
Rupa Bagdi et al	Developed a decision support system which combined the strengths of both OLAP and data mining.
K. Rajesh et al (2012)	Prediction analysis for Diabetes using classification algorithms like C4.5, ID3, K-NN, LDA, NaiveBayes for diagnosing diabetes for the given dataset.
Abdulla et al (2013)	Predictive analysis of diabetic treatment using a regression based data mining technique.
Gao et al (2005)	Proposed CoLe for detecting diabetes in the initial stage. CoLe, a multi-agent system aims to achieve a mixture of knowledge that describes data in different perspectives.
Afrand et al (2012)	Used Artificial Intelligence for diagnose diabetes; used Extended Classifier System (XCS) which has greater accuracy than the conventional data mining techniques.
Adidela et al (2012)	Applied Fuzzy Id3 algorithm for predicting diabetes; suggested a combination of classification system developed using Em algorithm for clustering and fuzzy ID3 algorithm to attain decision tree for each cluster.
Mandal et al	Used hierarchical clustering algorithm to identify the trends for controlling diabetes mellitus.
Kavitha et al (2012)	Applied CART Method for monitoring Diabetes. The algorithm distinguishes between high risk and low risk patients. The system achieved an accuracy rate of 96.39%.
Ananthapadmanaban et al (2014)	Applied the NaiveBayes and SVM classification algorithms for predicting diabetic retinopathy with nearly 83% accuracy.
Nagarajan et al (2015)	Used clustering algorithm Simple K-Means and applied classification algorithms like RandomTree, NaiveBayes, SimpleCart and Simple Logistics for predicting the risk levels of diabetes.
Iyer et al (2015)	Used Decision Tree and Naïve Bayes algorithms for analyzing the patterns found in the data through classification analysis.
Lee et al (2011)	Analyzed genomic markers (SNPs) for significant association with T2DM using various classification algorithms including Quest (Quick, Unbiased, Efficient, Statistical tree), Support Vector Machine, C4.5, logistic regression, and K-nearest neighbour and computed the T2DM misclassification rates for each model. Concluded with identification of selected 18 markers with good correlation to decrease misclassification rate.
Joshitta et al (2015)	Applied Map Reduce with Hadoop framework on data of peoples' life styles, work pattern, family and health history and proposed Diabetes mellitus Prediction System.
Nithya et al (2015)	Evaluated the clustering algorithms such as Hierarchical Clustering, Density Based Clustering and Simple K means clustering algorithms. The algorithms are analyzed by using the trained set parameter based on its class attribute; reported that K means algorithm gives better performance when comparing with the other two algorithms by using the Diabetes dataset.

Thus, advances in diabetes research has made it possible to identify and detect significant circulatory biomarkers which may be better indicators of risk of development of diabetes as compared to the parameters presently used. We therefore propose that computer based prediction models should consider these biomarkers as the attributes for clustering in order to have a more accurate prediction of risk. The attributes like weight, BMI and fasting glucose levels can be used as baseline parameters. Based on these we propose the following architecture of the prediction model.

PROPOSED SYSTEM

In this paper, data mining techniques namely clustering and classification is proposed to diagnose the type of diabetes and its severity level from the data collected.

CLUSTERING ALGORITHM

The data mining technique of Clustering Algorithm is used to place data elements into related groups based on the group description. This technique groups the data into cluster such that each group has similar attributes or characteristics. We propose to use the Simple K Means clustering algorithm (R.Nithya, September 2015). It is an iterative algorithm in which, each iteration new cluster centers are computed and each data point is re-assigned to its nearest center. Also, k-means clustering algorithm is widely used in machine learning for clustering and quantization.

K-means algorithm (Veena Vijayan, 2014)

K means algorithm is one among the unsupervised learning algorithm. Unsupervised algorithms are those algorithms that operate on unlabelled samples. That means the output is unknown even if the input is known. They take input parameter, number of clusters and n object data set partition into k clusters. Algorithm select k objects randomly. Based on the closeness of each object with corresponding cluster, each object is assigned to one cluster. Next step is to find the points that are closest to each other. To assign the object to the closest center, Euclidean distance is preferred. Once the objects are distributed to k clusters, the new k cluster centers are found by taking the mean of objects of k clusters respectively. The process is repeated till there is no change in k cluster centers. K-means algorithm aims at minimizing an objective function.

D is input -data set; Output is k clusters.
Step 1: Initialize cluster centers as D.
Step 2: Randomly choose k objects from D.
Step 3: Repeat the following steps until no change in cluster means/ min error E is reached.
Step 4: Consider each of the k clusters. Compare the mean value of the objects in the clusters for initialization.
Step 5: Initialize the object with most similar value from D to one of k clusters.
Step 6: Take the mean value of the objects for each of k cluster.
Step 7: Update the cluster means with respect to object value.

CLASSIFICATION

Classification is a form of data analysis technique that can be used to extract models describing important data classes and predict future data trends (Srideivanai Nagarajan, April 2015). There are several popular techniques that dominate tools for classification and prediction including neural networks (NNs), naïve Bayes, Bayesian networks, decision trees (C4.5), association rules and support vector machines (SVMs). Some of these techniques may be used for both prediction and classification, while others have been used specifically for classification.

During Classification different groups are generated which are assigned different class. The training Dataset helps in generating the model by analyzing the dataset

The above model groups the dataset into three clusters namely 1) Healthy 2) Pre Diabetic and 3) Diabetic. After clustering the type of diabetes, the model can also apply classification algorithms like RandomTree, NaiveBayes, Simple Cart and Simple Logistics for predicting the risk levels. (VelidePhani Kumar et al, 2014). The obtained data can be classified based on various supervised machine learning algorithms, like Naive Bayes, Decision List and J48(C4.5)

Naïve Bayes: Naive Bayes classifier assumes that the presence or absence of a particular feature is unrelated to the presence or absence of any other feature, given the class variable. It can be trained very efficiently in a supervised learning.

J48 (C4.5): It is an open source algorithm in Weka data mining tool. A decision tree can be generated from the input data by C4.5 program. It is an algorithm used to generate a decision tree and is an extension of Quinlan's earlier ID3 Algorithm. The decision trees generated by this

can be used for classification and so referred to as statistical classifier.

Decision tree: Decision Tree are inexpensive to construct, easy to interpret, easy to integrate with database system and they have comparable or better accuracy in many applications.

Neural Network: An artificial neural network (ANN), often just called a "Neural network" (NN), is a mathematical model or computational model based on biological neural network. Neural networks process information in a similar way the human brain does.

The proposed system has attributes that are multi factorial. Hence, enhancements in the available algorithm are also proposed to enhance the classifiers.

Architecture of the proposed Predictive Analytical Model has been shown in figure 1. The attributes to be considered in the proposed data analysis are listed in table number 2.

Fig. 1 Architecture of Proposed Predictive Analytical Model

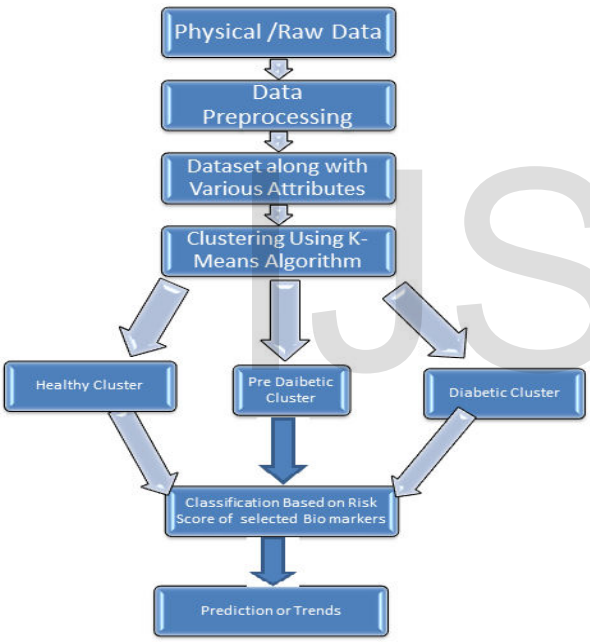

Table 2: List of attributes for the proposed prediction model

Sr. No.	Attributes	Deterministic values		
		Healthy cluster	Pre-diabetic cluster	Diabetic cluster
1.	Age1	<40	> 40	> 40
2.	Body Mass Index2	< 25 kg/ m2	≥ 25 kg/ m2	≥ 25 kg/ m2
3.	Abdominal fat -Waist circumference -Waist / hip ratio	Men: <102cm Women: <80cm < 0.92	Men: 102-112cm Women: 80-88cm 0.92-0.98	Men: >112cm Women: >88cm >0.98
4.	Lifestyle score* (based on diet, exercise and other habits)	< 5	>5	>5
5.	Diabetes pedigree function (Probability	0%	25-75%	100%

	calculation)			
6.	Blood pressure	<130/ 85	>130/ 85	>130/ 85
7.	H/ oPregnancy Induced increase in plasma sugar	No	Yes	Yes
8.	Blood glucose - Fasting - Post-prandial	<100mg/ dl <140mg/ dl	100-125 mg/ dl 140-199mg/ dl	>126 mg/ dl ≥ 200 mg/ dl
9.	Glycosylated Haemoglobin (Hb1Ac)	<5.5%	5.5-6.5%	>6.5%
10.	Fasting Insulin level	< 9.0 microIU/ mL	> 9.0 microIU/ mL	> 9.0 microIU/ mL
	Score of Genetic markers (SNPs identified by GWAS)	< 5	5-9	>9
	DRS based on circulatory biomarkers [including adiponectin, C-reactive protein, ferritin, IL-2 receptor]	0-4	4-7	7-10
	Lipid profile HDL triglyceride	<35 mg/ dL (0.90 mmol/ L) <150 mg/ dL	>35 mg/ dL (0.90 mmol/ L) 150-250 mg/ dL	>35 mg/ dL (0.90 mmol/ L) >250 mg/ dL (2.82 mmol/ L)

* Lifestyle score will be calculated using questionnaire based on evaluation of exercise and diet and other habits. Lower scores indicate healthier lifestyle.

DISCUSSION

The field of medical sciences is at the upsurge of immense data associated with various diseases, complications and elements contributing to development and progression of diseases. The advent of data mining tools and techniques can be used for the data analysis to generate representing trends, discover new patterns and interpret the data to provide useful information for better understanding for detection and treatment of diseases.

Type II Diabetes mellitus, one of the leading causes of death globally, has posed a serious challenge to face up to India. Successful outcome of the national healthcare strategies towards the challenge requires an integrated multidisciplinary approach for prevention of causative and developmental factors of the disease. To address the problem, mass screening programs with computer based prediction models will help early detection of high risk individuals.

Although, prediction models have been developed earlier, they use a restricted number of attributes, hence may not give errorless predictions. To overcome the limitations, we propose a new composite predictive system to conclusively categorize individuals as diabetic, pre-diabetic and non-diabetic (healthy). Our method shows a significant advantage over other methods, due to its ability to identify pre-diabetic state.

Pre-diabetic state can be reversed to healthy state if diagnosed in time before the development of NIDDM. Pre-diabetic state can be corrected with the help of proper diet, physical activity coupled with maintaining normal body weight and regular monitoring. Similarly, use of the predictive model will help screening of diabetic patients to avoid or delay its consequences and treatment for complications. Early detection of affected individuals with diabetic complications would allow implementation of timely and effective therapies.

CONCLUSION AND FUTURE WORK

In the present paper, we propose a novel data prediction model which uses an array of attributes influencing predisposition to Diabetes mellitus. The appropriate selection of diverse attributes will help correct prediction. The unique advantage of the model is to identify pre-diabetic state individuals. We aspire to use the model on the clinical data from collaborating medical institutes in future. The scope for using the proposed model is high due to large population of estimated diabetic patients.

REFERENCES

[1] 2014 USRDS annual data report: Epidemiology of kidney disease in the United States. United States Renal Data System. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2014:188–210.

[2] Adidela DR, Lavanya DG, Jaya SG, Allam AR. „Application of fuzzy ID3 to predict diabetes. International Journal of Advanced Computational Mathematical Sciences“. 2012; 3(4):541–5.

[3] Afrand P, Yazdani NM, Moetamedzadeh H, Naderi F, Panahi MS. „Design and implementation of an expert clinical system for diabetes diagnosis“. Global Journal of Science, Engineering and Technology; 2012. p. 23–31. ISSN:2322-2441

[4] Aljumah A, Ahamad MG, Siddiqui MK,“Application of data mining: Diabetes health care in young and old patients“ Journal of King Saud University – Computer and Information Sciences (2013) 25, 127–136.

[5] Ananthapadmanaban KR, Parthiban G. „Prediction of chances - diabetic retinopathy using data mining classification techniques. Indian Journal of Science and Technology“. 2014; 7(10):1498–503.

[6] Bagdi R, Patil P, „Diagnosis of Diabetes Using OLAP and Data Mining Integration“ in International Journal of Computer Science & Communication Networks,Vol 2(3), 314-322.

[7] Bays HE, Chapman RH, Grandy S, The Sheild Investigators' Group,Relatioship of Body Mass Index to Diabetes mellitus, Hypertension and dyslipidaemia: Comparison Of Data From Two National Surveys, Int J Clin Pract. 2007 may 61(5): 737–747, Doi: 10.1111/j.1742-1241.2007.01336.x

[8] Bourne RR, Stevens GA, White RA, Smith JL, Flaxman SR, Price H et al Causes of vision loss worldwide, 1990-2010: a systematic analysis. . Lancet Global Health 2013;1:e339-e349

[9] Caveney E, Cohen O Diabetes and Biomarkers Journal of Diabetes Science and Technology (2011) Volume 5, Issue 1, 192-197

[10] Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus.
World Health Organization, Geneva, 1999. Report Number: WHO/NCD/NCS/99.2.

[11] Diagnostic criteria and classification of hyperglycaemia first detected in pregnancy (WHO/NMH/ MND/13.2). Geneva: World Health Organization; 2013

[12] Diagnostic criteria and classification of hyperglycaemia first detected in pregnancy. World Health Organization, 2013. Report Number: WHO/NMH/MND/13.2.

[13] Gao J, Denzinger J, James RC. „CoLe: A cooperative data mining approach and its application to early diabetes detection.“ Proceedings of the 5th International Conference on Data Mining (ICDM'05); 2005.

[14] Garber AJ, Handelsman Y, Einhorn D, et al. Diagnosis and management of prediabetes in the continuum of hyperglycemia: when do the risks of diabetes begin? A consensus statement from

the American College of Endocrinology and the American Association of Clinical Endocrinologists. *Endocr Pract*. 2008;14:933-946.

[15] Global Action Plan for the Prevention and Control of Noncommunicable diseases 2013-2020 World Health Organization, Geneva, 2013

[16] Global report on diabetes.World Health Organization, Geneva, 2016.

[17] Goldfine A, Gerwien R, Kolberg J, O'Shea S, Hamren S, Hein G, Xu X, Patti M Biomarkers in Fasting Serum to Estimate Glucose Tolerance, Insulin Sensitivity, and Insulin Secretion Clinical Chemistry (2011) 57:2 , 326–337

[18] Handelsman Y, Bloomgarden ZT, Grunberger G, et al. American Association of Clinical Endocrinologists and American College of Endocrinology: clinical practice guidelines for developing a diabetes mellitus comprehensive care plan—2015. *Endocr Pract*. 2015;21:1-87.

[19] Iyer A, Jeyalatha S,Sumbaly R „Diagnosis Of Diabetes Using Classification Mining Techniques." International Journal of Data Mining & Knowledge Management Process (IJDKP) January 2015,Vol.5, No.1, 1-14

[20] Jellinger PS, Smith DA, Mehta AE, et al. American Association of Clinical Endocrinologists' guidelines for management of dyslipidemia and prevention of atherosclerosis. *Endocr Pract*. 2012;18(suppl 1):1-78.

[21] Johnson JL, Duick DS, Chui MA, Aldasouqi SA; Identifying prediabetes using fasting insulin levels Endocr Pract. 2010 Jan-Feb 16(1):47-52. doi:10.4158/EP09031.

[22] Joshitta RSM , Arockiam L „A Predictive Model to Forecast and Pre-Treat Diabetes Mellitus using Clinical Big Data in Cloud" International Journal of Applied Engineering Research 2015, ISSN 0973-4562 Vol. 10 No.82, 55-59

[23] K. Rajesh, V. Sangeetha, „Application of Data Mining Methods and Techniques for Diabetes Diagnosis" in International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 3, September 2012.

[24] Karmakar T, Mallick S, Chakraborty A, Maiti A, Chowdhury S, Bhattacharya M Signature biomarkers in Diabetes Mellitus and associated Cardiovascular diseases Clinical Hemorheology and Microcirculation, 2015 59 (1) 67-81

[25] Kavitha K, Sarojamma RM. „Monitoring of diabetes with data mining via CART Method. International Journal of Emerging Technology and Advanced Engineering". 2012; 2(11):157–62.

[26] Kivimaki M, Hamer M, Batty GD, et al. Antidepressant medication use, weight gain, and risk of type 2 diabetes: a population-based study. *Diabetes Care*. 2010;33:2611-2616.

[27] Kolberg JA, Jorgensen T, Gerwien RW, Hamren S, McKenna MP, et al. (2009) Development of a type 2 diabetes risk model from a panel of serum biomarkersfrom the Inter99 cohort. Diabetes Care 32: 1207–1212.

[28] Kumar VP, Velide L, „A data mining approach for prediction and treatment of diabetes disease" in international journal of science inventions today Volume 3, Issue 1, January-February 2014.

[29] Lee J, Keam B, J Jang E, Park MS,Lee JY, Kim DB, Lee C, Kim T, Oh B Park HJ, Kwack KB, Chu C, Kim H, „Development of a Predictive Model for Type 2 Diabetes Mellitus Using Genetic and Clinical Data" Public Health Res Perspect 2011 2(2), 75e82

[30] Lorenzo C, Williams K, Hunt KJ, Haffner SM. The National Cholesterol Education Program - Adult Treatment Panel III, International Diabetes Federation, and World Health Organization definitions of the metabolic syndrome as predictors of incident cardiovascular disease and diabetes. *Diabetes Care*. 2007;30:8-13.

[31] Mandal S, Dubey V. „Implementation and evaluation of diabetes management system using clustering technique." Special Issue of International Journal of Computer Science and Informatics. 2(2):33–6.

[32] Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med, 2006, 3(11):e442.

[33] Nagarajan S , Chandrasekaran RM. Design and Implementation of Expert Clinical System for Diagnosing Diabetes using Data Mining Techniques Indian Journal of Science and Technology 2015, Vol 8(8), 771–776

[34] National Cholesterol Education Program Expert Panel on Detection E, Treatment of High Blood Cholesterol in A. Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation*. 2002;106:3143-421.

[35] Nithya R, Manikandan P, Ramyachitra D „Analysis of clustering technique for the diabetes dataset using the training set parameter" International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 9,166-169.

[36] Párrizas M, Brugnara L, Esteban Y. Circulating miR-192 and miR-193b are markers of prediabetes and are modulated by an exercise intervention J Clin Endocrinol Metab. 2015 Mar;100(3):E407-15. doi: 10.1210/jc.2014-2574. Epub 2014 Dec

[37] Sarwar N, Gao P, Seshasai SR, Gobin R, Kaptoge S, Di Angelantonio et al Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies.Emerging Risk Factors Collaboration.. Lancet. 2010; 26;375:2215-2222.

[38] Shafizadeh TB, Moler EJ, Kolberg JA, Nguyen UT, Hansen T, et al. (2011) Comparison of Accuracy of Diabetes Risk Score and Components of the Metabolic Syndrome in Assessing Risk of Incident Type 2 Diabetes in Inter99 Cohort. PLoS ONE 6(7): e22863. doi:10.1371/journal.pone.0022863

[39] Shetty S, Shetty S, „ Analysis of Diabetic Data Set Using Hive and R" in International Journal of Emerging Technology and Advanced Engineering Volume 4, Issue 7, July 2014, 626-629

[40] Vazquez G , Duval S ,JacobsD, Jr.Silventoinen K, Comparison of Body Mass Index, Waist Circumference, and Waist/Hip Ratio in Predicting Incident Diabetes: A Meta-Analysis Epidemiol Rev (2007)29(1):115-128.DOIhttps://doi.org/10.1093/epirev/mxm008

[41] Vijayan VV, Ravikumar R „Study of Data Mining Algorithms for Prediction and Diagnosis of Diabetes Mellitus" International Journal of Computer Applications 2014 (0975 – 8887) Volume 95– No.17, 12-16

[42] Wang X, Strizich G, Hu Y, Wang T,. Kaplan R Qibin Q Genetic markers of type 2 diabetes: Progress in genome-wide association studies and clinical application for risk prediction Journal of Diabetes 8 (2016) 24–35