

## LITERATURE SURVEY ON DIABETES MELLITUS USING PREDICTIVE ANALYTICS OF BIG DATA

S.R.Surya Msc.,MCA.,M.Phil.,

Assistant Professor,  
Faculty of science and Humanities  
SRM Institute of Science and Technology.

**Abstract:** Predictive analytics is the practice of extracting information's from existing data sets in order to determine patterns and predict future .we use predictive analytics in a variety of several methods, from predictive marketing and data mining to applying machine learning (ML) and artificial intelligence (AI) algorithms to optimize business processes and uncover new statistical patterns. In this paper to give a detailed version of Predictive analytics (PA) uses technology to search through massive amounts of information about diabetic mellitus (DM). Its step to develop predictive model particularly in diabetes.

**Keywords:** Diabetic mellitus, data analytics, ANFIS, Machine learning.

### Introduction:

Predictive analytics which is help to healthcare organizations to evaluate data on the past behavior and predict likelihood of future behavior to enable better decisions and outcomes of their patient[1]. Predictive models can make human decisions more effective and highly automate an entire decision-making process. It increasingly, predictive analytics uses data from the IOT to improve safety and performance of patient outcomes.

Here, the Modernizing healthcare industry's move towards processing massive health records, and to access those for analysis and this will greatly increases the complexities. Due to the unstructured nature of Big Data form health industry, it is necessary to structure and emphasis their size into nominal value with possible solution. Healthcare industry faces many challenges that make us to know the importance to develop the data analytics of the diabetes mellitus.

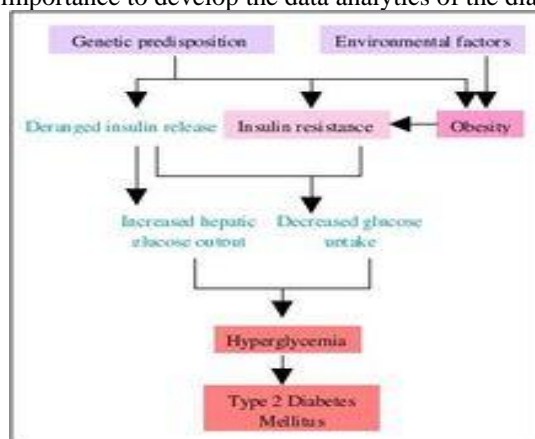


Fig 1: Architecture diagram for type 2 diabetes mellitus.

The rest of this paper is organizes as follows. Section 2 discusses some of the earlier proposed research work on Diabetes mellitus using big data analytics. Section 3 provides conclusion and some basic ideas on future research work.

### Related Works:

S.Shyni1 , R. Shantha Mary Joshitta et al[2] describe the over view of big data analytics in healthcare and to discuss about issues and chanlanges as scalability, data complexity, speed, accuracy, cost, storage, security and processing. P.Sampath ,S.Lavanya et al[3] proposed Diabetic mellitus(DM) is one the non communicable diseases (NCD). Here they can taken a raw diabetic data set and predictive analysis algorithm in hadoop/ map reduce environment to predict the types of prevalent

and complications associated with it. Based on this analysis, this provides an efficient way to cure and care of patients with the better outcomes like affordability and availability of the patients.

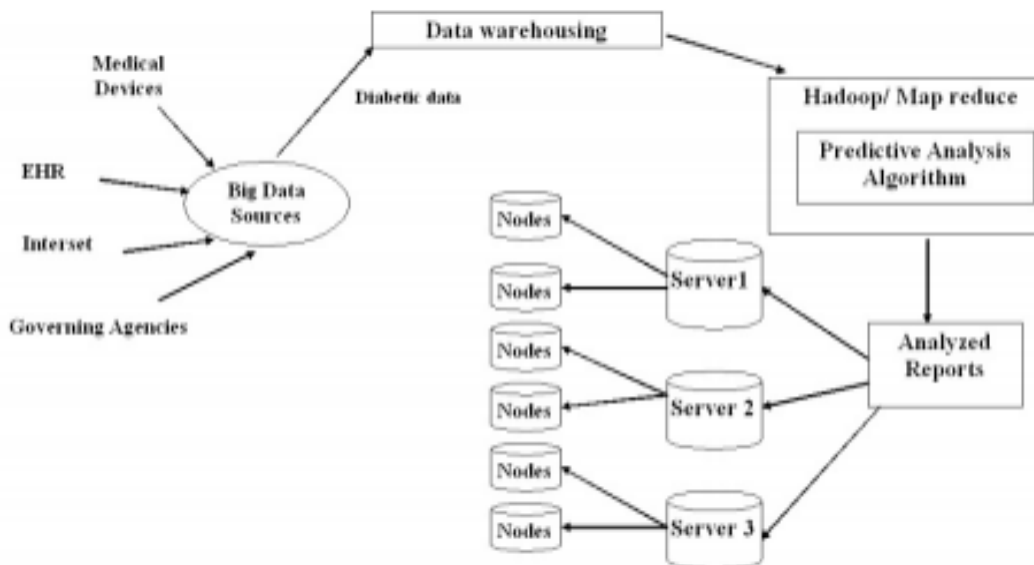


Fig 2: Architecture of the predictive analysis system-Health Care Applications

M. Durairaj, G. Kalaiselvi[4] discusses the different soft computing techniques for the prediction of diabetes. This propose an effective technique as artificial neural network for earlier prediction of the disease and for improving calculation accuracy they used decision tree algorithm. By using ANN shows very encouraging result compare with SVM.

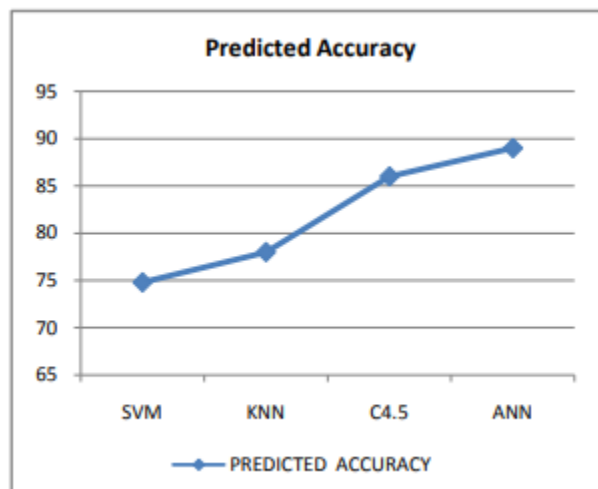


Fig 3: Predicated Accuracy of SVM.KNN,C4.5, ANN

Veena Vijayan V. Aswathy Ravikumar[5] discusses data mining algorithm such as EM algorithm, KNN algorithm for classifying the object, K-means algorithm for partitioning method based on some input parameter. ANFIS combines the features of adaptive neural network. For classification and experimental simulation use a Pima Indian diabetic dataset from university of California. Finally ANFIS provide better classification and prediction accuracy shows better accuracy when the k value is increased to large value. C. Kalaiselvi, G. M. Nasira[6] proposed to improve the classification accuracy and to achieve better efficiency a new approach like adaptive neuro fuzzy interface system(ANFIS). For classification Pima Indian diabetic dataset. The result shows that to reduce the cost at different medical test.

Saumya Salian<sup>1</sup>, Dr. G. Harisekaran[7] determine the risk predictors which can cause readmission among diabetic patients and analysis has been performed to predict risk of readmission of diabetic patients. They trained machine learning databases from UCI repository. Finally the result shows readmission in diabetic patient is successfully predicated using analysis. N.M. Saravanakumar, Dr.T.Eswari et.al[8] discuss to improve healthcare system through the reduction runtime and the optimal cost. To predict and classify the type of diabetic mellitus by using hadoop/ map reduce algorithm. The result shows the diabetes type 2 does not use insulin properly in our body at 24<sup>th</sup> week many women establish gestational diabetes.

Hal Kalechofsky[9] discusses about a detailed version of predictive models from base to state of art. they trained Koges dataset. It predicate multi stage adjustment model. Predictive analysis poised to [10] drive population health white paper suggest pima dataset. It predicate hybrid model and technique clustering . the result shows 92.23% accuracy in means of weka tool.

Ian Duncan[11] determine the pima dataset , to predict hybrid model to implement this process of diabetes in weka tool and to get the accuracy 92.38%. The promise of big data in diabetes management [12] discuss diabetes is also one of the most expansive challenges facing in healthcare industry today. It represent a huge potential for value based care models in preventing diseases, better patient outcomes.

Aruna Pavate, Nazneen Ansari[13] suggest soft computing based algorithm would enable a probable model in the classification process. So as to reduce the mortality rate in the population data. The outcomes in this prediction of diabetes patients risk level which prone to heart attack along with the risk factors grade point by means of genetic algorithm. Veena H. Bharat et al[14] proposed an efficient imputation method using a hybrid combination of CART and Genetic algorithm as a preprocessing step. The classical neural network model is used for prediction on the preprocessed dataset. The result shows preprocessing which involves imputation using CART along with optimization of the error function using a genetic algorithm , the maximum accuracy 82.33% as against the accuracy of the neural network prediction model build on the unprocessed data which is 75.86%.

### Conclusion and Future Work

In this survey paper mainly focused on predict Diabetes mellitus by using big data analytics. According to this analysis the most of the paper covers the algorithm Hadoop/Map reduce environment to predict the diabetes types prevalent and complications of patient with it. In each of the paper they used different dataset for analytics. For the calculation they used decision tree algorithm. So, the result is quite good. It is possible to future improve the diabetes mellitus to use any machine learning algorithm / for efficiency of the prediction. we will use ANN because it shows better result when compare with SVM.

### References

1. N. Jayanthi B. Vijaya Babu, N. Sambasiva Rao “Survey on clinical prediction models for diabetes prediction” journal of bigdata, december 2017.
2. S.Shyni, R. Shantha Mary Joshitta , Dr. L. Arockiam ” Applications of Big Data Analytics for Diagnosing Diabetic Mellitus: Issues and Challenges” International Journal of Recent Trends in Engineering & Research (IJRTER) Volume 02, Issue 06; June - 2016 [ISSN: 2455-1457].
3. T.Eswari, P.Sampath, S.Lavanya “Predictive Methodology for Diabetic Data Analysis in Big Data” Procedia Computer Science 50 ( 2015 ) 203 – 208
4. M. Durairaj, G. Kalaiselvi ” Prediction Of Diabetes Using Soft Computing Techniques- A Survey” INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 4, ISSUE 03, MARCH 2015
5. Veena Vijayan V. Aswathy Ravikumar “Study of Data Mining Algorithms for Prediction and Diagnosis of Diabetes Mellitus” International Journal of Computer Applications (0975 – 8887) Volume 95– No.17, June 2014.
6. C. Kalaiselvi, G. M. Nasira “A New Approach for Diagnosis of Diabetes and Prediction of Cancer Using ANFIS” World Congress on Computing and Communication Technologies-2014.
7. Saumya Saliyanl , Dr. G. Harisekaran “Big Data Analytics Predicting Risk of Readmissions of Diabetic Patients” International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064.
8. N.M. Saravanakumar ,Dr.T.Eswari, P.Sampath, S.Lavanya “Predictive Methodology for Diabetic Data Analysis in Big Data” Procedia Computer Science Volume 50, 2015, Pages 203-208
9. Hal Kalechofsky,” A Simple Framework for Building Predictive Models” A Little Data Science Business Guide September 2016.
10. Predictive analysis: Poised to drive population health white paper.
11. Ian Duncan ”Introduction to predictive modeling” November 15, 2017.
12. <https://www.scalablehealth.com/> “The Promise of big data in diabetes management”.
13. Aruna Pavate ; Nazneen Ansari ” Risk Prediction of Disease Complications in Type 2 Diabetes Patients Using Soft Computing Techniques” 2015 Fifth International Conference on Advances in Computing and Communications.
14. Veena H. Bharat, Prasanth G. Roe “ An efficient prediction model for diabetic dataset using soft computing techniques” pp-328-345, 2009 @ springer.verlag Berlin Heidelberg 2009.
15. Sabibullah M , Shanmugasundaram V , Raja Priya K ”Diabetes Patients Risk through soft computing model”. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Volume 2, Issue 6, November – December 2013.