# Medical Diagnosis of Diabetes Using Deep Learning Techniques and Big data Analytics

## ANITA SACHIN MAHAJAN

Assistant Professor, Department of Computer Engineering, G H Raisoni Institute of Engineering and Technology, Wagholi, Pune.

**Abstract-** Now days from health care industries large volume of data is generating. It is necessary to collect, store and process this data to discover knowledge from it and utilize it to take significant decisions. Diabetic Mellitus (DM) is from the Non-Communicable Diseases (NCD), and lots of people are suffering from it. Now days, for developing countries such as India, DM has become a big health issue. The DM is one of the critical diseases which has long term complications associated with it and also follows with various health problems. Diabetes is one of the threatening diseases to the entire mankind, though it is not fatal. Irrespective of the presence of several existing approaches for diabetes prediction, big data based diabetes prediction is quite rare. The applicability of the proposed work is wider because, medical records from different sources are extracted and the necessary attributes meant for predicting diabetes alone are processed. The goal of this work is attained by different phases such as data collection, preprocessing, attribute selection and prediction. The diabetes prediction is carried out by Extreme Learning Machine (ELM) classifier. The performance of the proposed approach is analyzed by varying the classifiers and the existing approaches in terms of disease prediction accuracy, precision, recall and time consumption. From the experimental results, the efficiency of the work is proven. Diabetes Mellitus (DM) is a metabolic diseases group where the person will have high blood sugar due to the pancreas unable to produce sufficient insulin or the cell's which are not responding to the insulin produced. Diabetes is a chronic disease and a major public health challenge worldwide. The main drawback is that there is lack of awareness of the people on eating habits. In our country, diabetes patient counts have increased steadily due to this reason. Diabetes is a metabolic disease affecting a multitude of people worldwide. Its incidence rates are increasing alarmingly every year. If untreated, diabetes-related complications in many vital organs of the body may turn fatal. Early detection of diabetes is very important for timely treatment which can stop the disease progressing to such complications. RR-interval signals known as heart rate variability (HRV) signals (derived from electrocardiogram (ECG) signals) can be effectively used for the non-invasive detection of diabetes. This research paper presents a methodology for classification of diabetic and normal HRV signals using deep learning mechanism. We employ long short-term memory (LSTM), convolutional neural network (CNN) and its combinations for extracting complex temporal dynamic features of the input HRV data. These features are passed into support vector machine (SVM) for classification. We have obtained the performance improvement of 0.03% and 0.06% in CNN and CNN-LSTM architecture respectively compared to our earlier work without using SVM. The classification system proposed can help the clinicians to diagnose diabetes using ECG signals with a very high accuracy of 95.7%.

Keywords: Deep learning, Diabetes, ECG, CNN, Hadoop, MapReduce, Predictive Analysis.

## I. INTRODUCTION

Computers have brought substantial improvements to technology that lead to the production of massive volumes of data. Healthcare industry contains very large and sensitive data. This data needs to be treated very carefully to get benefitted from it. There is need to develop some more accurate and efficient predictive models that helps in diagnosing a disease although it was revealed that diabetes mellitus is the diseases which becomes one of the global hazards. The problem about this project is it is not easy to do diagnosis whether it is positive or negative having diabetes. It is because of many reasons. Different people may be having different signs. So, it is not easily to assume that they have it or not [1]. The sign of the diabetes is always thirsty, always hungry, weight become decrease, feel weak, have problem of sight, headaches, always do urination and so on. However, the real diagnosis is still needed to assign the real result. Hence there it is needed to analyses the already available huge diabetic data sets to discover some incredible facts which may help in producing some prediction model. The focus is to develop the prediction models by using certain machine learning algorithms. The machine learning is a sort of artificial intelligence that enables the computers to learn without being explicitly Machine learning emphases on the development of computer programs that can teach themselves to change and grow when disclosed to new or unseen data. Machine learning algorithms are mostly categorized as being supervised or unsupervised [2].

### A. Types of Diabetes

**Type 1 Diabetes** is called insulin-dependent diabetes mellitus (IDDM) or juvenile-onset diabetes. Type1 mostly occurs in young people who are below 30 years. This type can affect children or adults, but majority of these diabetes cases were in children.

**Type 2 Diabetes** is called non-insulin-dependent diabetes mellitus (NIDDM) or adult-onset diabetes. Risk factors for Type 2 diabetes includes older age, obesity, family history of diabetes, priorhistory of gestational diabetes, impaired glucose tolerance, physical inactivity, and race/ethnicity. Gestational Diabetes is the third main form and occurs when pregnant women without a previous history of diabetes develop a high blood glucose level. Congenital Diabetes occurs in human due to genetic defects of insulin secretion, cystic fibrosis-related diabetes, and high doses of glucocorticoids leads to steroid diabetes [3].

### 1.1 Background

In recent years the healthcare industry has generated large amounts of data. The value based treatment in hospitals and digitization of world likes to have the computerized data rather than hard copy form. The health care data includes Electronic Health Reports (EHR) of patients data, clinical reports, doctor's prescription, diagnostic reports, medical images, pharmacy information, health insurance related data, data from social Medias and medicinal journals [13]. All these information collectively forms Big Data in health care. By employing the analysis of big data will produce the predicted results for understanding the trends to improve the health care and life time expectancy, proper treatment at early stages at low cost. The analytics associated with big data is described by four characteristics: volume, velocity, variety and veracity [8]. The accumulation of health-related data continuously, resulting in an incredible volume of data; Velocity is accessing those data in real-time at a rapid speed; Variety includes diabetic glucose measurements, blood pressure readings, and various EHRs; Whereas veracity assumes the simultaneous scaling up in performance of the architectures and platforms, algorithms and tools to match the need of big data [8]. The healthcare industry is moving from reporting facts to discovery of insights, toward becoming data driven healthcare organizations. Big data holds great potential to change the whole healthcare value chain from drug analysis to patients caring quality. The probability of a 30-70 year old Indian dying from the four main non-communicable diseases diabetes, cancer, stroke and respiratory diseases is 26 percent at present, according to the World Health Organization. According to the Global Status Report, Non-Communicable Diseases (NCDs) would claim nearly 52 million lives globally by the year 2030. Nearly 8.5 million people died of NCDs diseases in the WHO's South-East Asia Region in 2012. In India, NCDs are estimated to have accounted for 60 percent of all deaths in 2014, while 26 percent between the ages of 30-70 years had a probability of succumbing to the four diseases. Diabetic Mellitus (DM) is one of the Non Communicable Diseases (NCD), is a major health hazard in developing countries such as India. The acute nature of DM is associated with long term complications and numerous of health disorders. There are three main types of this disease [6]. Type1 DM results from the body's failure to produce insulin, and presently requires the person to inject insulin. This form is referred as Insulin - Dependent Diabetes Mellitus (IDDM). Type 2 DM results from insulin resistance, a condition in which cells fail to use insulin properly, sometimes combined with an absolute insulin deficiency. This form was previously referred to as Non-Insulin - Dependent Diabetes Mellitus (NIDDM). The third main form, gestational diabetes occurs when pregnant women without a previous diagnosis of diabetes develop a high blood glucose level. It may precede development of type 2 DM. It was estimated that 61.3 million people aged 20-79 years live with diabetes at 2011 in India. This number was expected to increase to 101.2 million by 2030 [7].

Due to the growing unstructured nature of diabetic data form health industry or all other sources, it is necessary to structure and emphasis its size into nominal value with possible solution. With the help of technological developments, it is necessary to combine robust diabetic data sharing and electronic communication systems can facilitate better access to health services at all the levels of patients. So that all patient information needs to be in one repository. Deploying a Health Information Exchange (HIE) can extract clinical information from several disparate repositories and integrate that data within a single patient health record that all care providers can access securely. Predictive analysis is a method, that incorporates a variety of techniques from data mining, statistics, and game theory that uses the current and past data with statistical or other analytical models and methods, to determine or predict certain future events [14]. Significant predictions or decisions can be made by employing big data analytics in health care field. In this paper, we use the predictive analysis algorithm in Hadoop/Map Reduce environment to predict the diabetes types prevalent, complications associated with it and the type of treatment to be provided. Based on the analysis, this system provides an efficient way to cure and care the patients with better outcomes like affordability and availability.

## 1.2 Motivation

Due to the technological advancement and the excessive utilization of data, today's digital world relies on 'Big Data Processing'. As the term indicates, the big data technology has the capability to manage huge data effectively [1]. The smart world is the reason for the rise in the volume of data, which can be in any digital form such as text, numerals, images, audio or video. While the volume of data grows over every moment of time, data storage and management are the crucial issues needed to be addressed. Data organization is the predominant requirement of any data management scheme. As the volume of data is greater, it is difficult for the analytical system to process the data, unless it is organized. The better the data organization, the better is the data utilization. Some of the sample data in this context are data shared in social media, business platforms, healthcare data and transactional data and so on. For instance, the social media data are observed in social networks such as facebook, twitter, instagram and so on. The data utilized by business platforms are enormous, as the term business is common. Healthcare data is utilized by medical industry for analyzing the historical medical records of the patients and to make decision making. Finally, transactional data involves all the electronic transaction based data which are usually the outcome of online shopping, banking sector and so on [2].

The data types and the data formats are not consistent and common. The big data technology relies on five different characteristics of data such as volume, velocity, variety, veracity and value [9]. All the terms are described one after the other. The term volume indicates the amount of data produced by an organization or any entity. Velocity represents the speed of data production and distribution. Variety denotes the varying data formats of the data and veracity indicates the data uncertainty. Finally, the value stresses on the data being formed out of certain business processes. All these characteristics of data place a crucial challenge in front of big data analysis. Hence, any big data analytical technique must consider these challenges to accomplish the goal. Understanding these challenges, this work attempts to present a big data analytical system for predicting the presence of diabetes mellitus in users. The motivation of this work is to present a big data analytical system for healthcare industry. The goal of this work is to predict whether or not the user is affected by diabetes mellitus, while facing many challenges as mentioned earlier [10].

## 1.3 Problem Statement

In conventional approach author cannot predict continuous outcomes as well as Vulnerable to overconfidence, i.e., the models can appear to have more predictive power than they actually do, resulting in over fitting. Also in Conventional approach requires quite a large sample size to achieve stable results. Optimizing the network can be challenging because of the number of parameters to be set in also for large neural networks, it requires high processing time. We found less of accuracy in conventional approach. In proposed approach presents a methodology for classification of diabetic and normal HRV signals using deep learning mechanism. In deep learning, a training process is based on analyzing large amounts of data. That is why we adapt their deep learning algorithms in the way neural networks can handle large amounts of continuous input diabetes data. Author employs long short-term memory (LSTM), convolutional neural network (CNN) and its combinations for extracting complex temporal dynamic features of the input HRV data. These features are passed into support vector machine (SVM) for classification. Author have obtained the performance improvement of 0.03% and 0.06% in CNN and CNN-LSTM architecture respectively compared to our earlier work without using SVM. The classification system proposed can help the clinicians to diagnose diabetes using ECG signals with a very high accuracy of 95.7

## 1.4 Existing Techniques

A literature review reveals many results on diabetes carried out by different methods and materials of diabetes problem in India. Many people have developed various prediction models using data mining to predict diabetes. Combination of classification-regression-genetic-neural network, handles the missing and outlier values in the diabetic data set, and also they replaced the missing values with domain of the corresponding attribute [1]. The classical neural network model is used for prediction, on the pre-processed dataset. In predictive analysis of diabetic treatment using regression based data mining techniques to diabetes data, they discover patterns using SVM algorithm that identify the best mode of treatment for diabetes across different age [2].

They concluded that drug treatment for patients in the young age group can be delayed whereas; patients in the old age group should be prescribed drug treatment immediately. Prediction and classification of various type of diabetes using C4.5 classification algorithm was carried out in Pima Indians Diabetes Database [3]. A detailed analysis of the Pima diabetic data set was carried out efficiently using of Hive and R. In this analysis we can derive some interesting facts, which can be used to develop the prediction models [4]. The soft computing based prediction model was developed for finding the risks accumulated by the diabetic patients. They have experimented with real time clinical data using Genetic Algorithm [5]. The obtained results pertaining to the level of risk which prone to either heart attack or stroke. The novel pre-processing phase with missing value imputation for both numerical and categorical data. A hybrid combination of Classification and Regression Trees (CART) and Genetic Algorithms to impute missing continuous values and Self Organizing Feature Maps (SOFM) to impute categorical values was improved in [6].

Deploying a health information exchange (HIE) repository promote and integrate the data within a single point of robust data sharing. This sharing of information and electronic communication systems enable access to health services and also promotes additional care over dual eligible patients. It recognizes which patient is requiring more care and attention than others. It gives needed data to determine which strategies should be put in place to maximize positive behavior modification [9]. The predictive analytics works in three areas such as Operations management, Medical management and biomedicine, and System design and planning. Healthcare predictive analytics system can help one of the issues that is to address the cost of patients being repeatedly admitted and readmitted to a hospital for chronic diseases which is similar or multiple. The survey of New England Journal of Medicine tells that one in five patients suffer from preventable readmissions. Therefore 1% of the population accounts for 20% of all US healthcare expenditures almost and 25% for over 80% of all expenditures [10]. Various big data technology stack and research over health care combined with efficiency. Cost savings, etc. are explained in better healthcare [11]. The hadoop usage in health care became more important to process the data and to adopt the large scale data management activities. The analytics on the combined compute and storage can promote the cost effectiveness to be gained using hadoop [12]. All the above researchers have been successful in analyzing the diabetic data set and developing good prediction models. In this paper, we use the predictive analysis technique in Hadoop/Map Reduce environment to predict and classify the type of diabetes. This system provides efficient way to care and cure the patients at low cost with better outcomes like affordability and availability.

## II. LITRATURE SURVEY

### 1. Research Gaps

1. Analysis of Various Data Mining Techniques to Predict Diabetes Mellitus Data mining approach helps to diagnose patient's diseases. Diabetes Mellitus is a chronic disease to affect various organs of the human body. Early prediction can save human life and can take control over the diseases. This paper explores the early prediction of diabetes using various data mining techniques.

2. Predictive Methodology for Diabetic Data Analysis in Big Data

We use the predictive analysis algorithm in Hadoop/Map Reduce environment to predict the diabetes types prevalent, complications associated with it and the type of treatment to be provided. Based on the analysis, this system provides an efficient way to cure and care the patients with better outcomes like affordability and availability.

3. Predicting Diabetes in Medical Datasets Using

Machine Learning Techniques Different machine learning techniques are useful for examining the data from diverse perspectives and synopsizing it into valuable information. The main goal is to determine new patterns and then to interpret these patterns to deliver significant and useful information for the users.

### 2. Literature Review

This section reviews the existing state-of-the-art literature with respect to diabetes prediction system based on big data analytics. In [2], a survey is presented by discussing about different feature selection algorithms for predicting diabetes mellitus. This work reviews certain basic feature selection algorithms such as k-Nearest Neighbor (k-NN), k-means, branch and bound algorithm. A basic diabetic dataset is chosen for carrying out the comparative analysis. The importance of feature analysis for predicting diabetes by employing machine learning technique is discussed in [3]. The critical reasons for the cause of diabetes are discussed in this paper. A diabetes prediction system based on association clustering and time series based data mining in continuous data is proposed in [4]. Machine learning approaches are utilized to handle the massive data and helps in disease prediction. The historical medical data of the patients are considered with respect to different parameters are considered. The parameters are analyzed for making final decision with respect to disease prediction. In [5], an automated insulin delivery based diabetes management is proposed. This work aims to maintain the perfect level of blood glucose levels at all times. In [6], a diabetes prediction model based on cloud analytics is proposed. This work employs classification and predictive analysis algorithm for predicting the occurrence of diabetes in patients. The proposed algorithm is implemented in the cloud environment and the probability of occurrence of diabetes is computed. A web application is proposed in [7] for predicting the diabetes with the help of machine learning algorithm. This work utilizes PIMA Indian database for predicting the diabetes by employing Artificial Neural Networks (ANN).

In [8], a rapid model detection scheme for online subcutaneous glucose concentration prediction system is proposed for candidates with type I diabetes. This work acquires a model and the parameters are modified by considering the data from new candidates for model updating. This prediction model is compared with the existing approaches and the results are analyzed. A breath analysis system meant for predicting glucose level and screening diabetes is presented in [9]. This work employs certain chemical sensors and the biomarkers in breath are detected. The major factors considered by this work are humidity and the ratio of alveolar air in breath. The prediction model is developed subject-specifically and

the accuracy of the work is claimed to be improvised. The sensitivity and specificity rates are computed for analyzing the performance of the proposed work.

In [10], a short-time prediction of glucose concentration is presented on the basis of neural networks by considering meal information. This work utilizes continuous glucose monitoring devices, which works in association with the meal information. A neural network model with a first order polynomial extrapolation algorithm is utilized to analyses the linear and non-linear components of glucose dynamics. A real-time non-invasive detection and diabetes classification system based on Convolutional Neural Network (CNN) is proposed in [11]. This work employs 1D CNN which is based on real-time breath signals, which are acquired from gas sensors. In [12], a metabolic syndrome and development of diabetes mellitus is presented by predictive modelling on the basis of machine learning techniques. This work studies the relationship between diabetes and risk factors associated with it. The diabetes is predicted by employing J48 decision tree and Naïve bayes techniques. In [13], different machine learning and data mining methods in diabetes research is discussed. This work reviews various applications of machine learning and data mining techniques in diabetes research with respect to prediction and diagnosis, diabetes complications and healthcare management. This work takes several clinical datasets into consideration and the knowledge about the data is gained by several supervised and unsupervised learning approaches. The diabetes disease is analyzed and detected with the help of data mining techniques based on big data in [14]. The data mining techniques are employed over healthcare systems with the help of an automatic tool, which could detect the disease by analyzing the severity of the disease and the suitable treatment type is predicted. This work analyses the performances of both supervised and unsupervised techniques. In [15], a diabetic data analysis scheme with prediction is proposed for bigdata. This work utilizes Hadoop and MapReduce environment to predict the presence of diabetes and the type is classified. Systems and precision medicine approaches to diabetes heterogeneity is explained on the perspectives of big data in.
.

## III. PROPOSED SYSTEM

### 3.1 Overview of the Work

Due to the advancement of technology and the development of medical science, the healthcare domain manages the medical records in digital format rather than physical records. Though the data management is made easier, it is difficult to manage the unpredictably growing data. Besides this, the data growth is directly proportional to the time and the growth of data is inevitable. Now-a-days, the medical data such as patient's historical health information, medical records, diagnostic reports, and medication related records are all maintained by means of big data via Electronic Health Records (EHR). The disease prediction systems meant for big data are quite rare in literature. The overall flow of the work is depicted in figure.
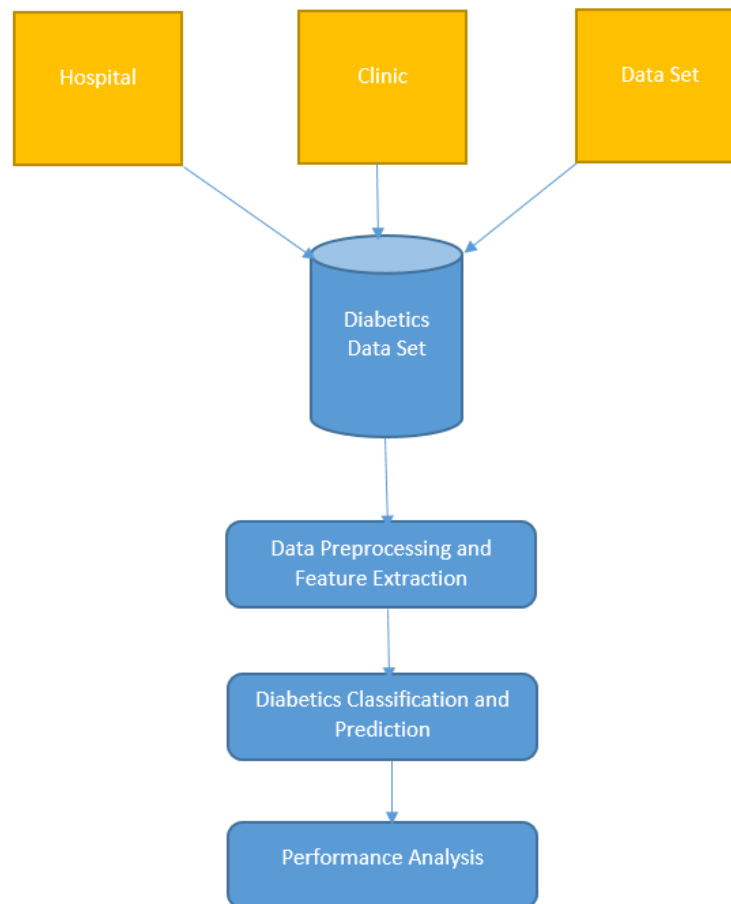
**Fig.1. Overall flow of the proposed diabetes prediction system**

As per International Diabetes Federation (IDF), India is one of the participants with more diabetic people. It is reported that about 72 million people are affected by Diabetes in 2017 [18]. Though the diabetes is not fatal, it is not curable and the ill-effects continue till the lifetime of the patient.

Hence, it is better to predict rather than to treat the disease. The diabetes mellitus disease can take anyone of the two types, which are type 1, 2 and 3. When a patient is affected by Diabetes of type 1, then the insulin is not secreted by the body such that the patient has to inject insulin. In case of type 2 diabetes, the body develops resistivity against insulin and the insulin is not exploited in usual way. The type 3 diabetes is commonly called as gestational diabetes, which occurs due to the excessive blood glucose levels. Hence, it is highly appreciable when the diabetes is predicted in advance. The prediction can be done by analyzing all the medical records associated with the patient.

Diabetes is a by-product of several other ailments, which makes it necessary to analyses all the health records of the patient. Considering the adverse effects caused by diabetes, this work intends to predict diabetes based on the machine learning algorithm. Prediction can be achieved by incorporating data mining techniques over the data, which analyses and studies the historical data for predicting the future outcome.

The goal of this work is attained by segregating the complete work into multiple phases and they are data acquisition, pre-processing and disease prediction. The initial phase collects the data from several EHRs and different laboratories. All the so collected data are standardized by the second phase, which is the data preprocessing phase. The third phase is the heart of the work, which predicts the chance of diabetes occurrence in the patient. Finally, the performance of the proposed work is tested in terms of prediction accuracy, precision and recall measures. The proposed approach is elaborated in the following sub-sections.

### 3.2 Data Acquisition

Data acquisition is the most basic step, which collects the data from multiple sources such as medical laboratories, hospitals and public benchmark datasets. The health records from medical laboratories and hospitals are collected. Additionally, the publicly available PIMA Indian Diabetes database is utilized [12]. The collected data contains several attributes, which are irrelevant to determine diabetes and hence, the relevant attributes alone are needed to be processed. This task is performed by the data preprocessing phase, which is as follows.

### 3.3 Data Processing and Attribute Selection

The main aim of data pre-processing is to prepare the data to become suitable for the forthcoming processes of disease prediction. The pre-processing phase eliminates the duplicate records being present in the database and makes sure that all the attributes are present in the dataset. As the health records are collected from multiple sources, different attributes are utilized in the dataset. Hence, the required attributes meant for predicting diabetes alone are taken into consideration for preparing the data to be processed. The considered attributes of this work to predict diabetes are count of pregnancy, glucose level, systolic pressure, diastolic pressure, thickness of skin, insulin, Body Mass Index (BMI), plasma, Diabetes Pedigree Function (DPF), age, class. These attributes are effective in predicting between the diabetes. Hence, the altered dataset contains twelve significant attributes for predicting the diabetes [10].

### 3.4 Diabetes Prediction by ELM

In order to predict diabetes on the candidates, ELM is employed as the machine learning algorithm owing to its faster learning ability and efficiency [20]. The prediction phase is known to have two phases, which are training and prediction phases. The training phase provides a way for the classifier to gain knowledge from the training samples. Totally, this work manipulates 800 records collected from different laboratories, hospitals and benchmark PIMA dataset. Out of the collected records, this work employs forty percent of the data for training and the remaining sixty percent is employed for prediction. Hence, 320 records are utilized for training and the remaining 480 samples are utilized for prediction [13].

### CONCLUSIONS

Big Data Analytics in Hadoop's implementation provides systematic way for achieving better outcomes like availability and affordability of healthcare service to all population. Non-Communicable Diseases like diabetes, is one of a major health hazard in India. By transforming various health records of diabetic patients to useful analyzed result, this analysis will make the patient understand the complications to occur. The goal of this research deals with the study of diabetic treatment in healthcare industry using big data analytics. The design of predictive analysis system of diabetic treatment may give enhanced data and analytics yield the greatest results in healthcare. By employing location aware healthcare service, anyone from rural area can get proper treatment at low cost. Treatment can be offered when it is identified in advance. This paper presents a big data processing system diabetes prediction by employing machine learning algorithm. The goal of the work is to reduce the false positive and false negative rates as much as possible, so as to boost up the precision and recall rates. The ELM classifier is utilized for diabetes prediction, owing to its faster learning capability. The performance of the work is analyzed by varying the classifiers and tested against existing techniques. The experimental results prove the efficacy of the proposed approach and in future, this work is planned to be extended such that the medical images are processed.

### REFERENCES

[1] Dr Saravanakumar,Eswari, Sampath, Lavanya "Predictive Methodology for Diabetic Data Analysis in Big Data," ELSEVIER, ISBCC 2015.

[2] AiswaryaIyer, S. Jeyalatha, Ronak Sumbaly "Diagnosis of Diabetes Using Classification Mining Techniques," IJDKP Vol.5, No.1, January 2015.

[3] Sabibullah M, Shanmugasundaram V, Raja Priya K, "Diabetes Patient's Risk through Soft Computing Model,"International Journal of Emerging Trends Technology in Computer Science, vol 2(6), 2013.

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

[5] Rajnik L. Vaishnav, Dr. K. M. Patel, "Analysis of Various Techniques to Handling Missing Value in Data set," International Journal of Innovative and Emerging Research in Engineering Volume 2, Issue 2, 2015

[6] Wei Dai, Wei Ji, "A MapReduce Implementation of C4.5 Decision Tree Algorithm," International Journal of Database Theory and Application Vol.7, No.1 (2014), pp.49-60

[7] Machine Learning tutorials and examples https://www.toptal.com/machinelearning/ machinelearningtheory- an-introductory-primer

[8] Gauri D.Kalyankar, Shivananda R Poojara, N V Dharwadkar,"Weblog Analysis Using Hadoop," National Research Symposium on Computing – RSC 2016, ISBN: 978-81931456-1-8, Dec. 19-20, 2016.

[9]S.Salian and G. Harisekaran, "Big Data Analytics Predicting Risk of Readmissions of Diabetic Patients," International Journal of Science and Research, vol. 4, April 2015.

[10] K. Sharmila and S. Manickam, "Efficient Prediction and Classification of Diabetic Patients from bigdata using R,"International Journal of Advanced Engineering Research and Science, vol. 2, Sep 2015.

11. Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological Forecasting and Social Change, 126, 3- 13.

12. Lomte, R., Dagale, S., Bhosale, S., & Ghodake, S. (2019, April). Survey of Different Feature Selection Algorithms for Diabetes Mellitus Prediction. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-5). IEEE.

13. Dutta, D., Paul, D., & Ghosh, P. (2018, November). Analysing Feature Importances for Diabetes Prediction using Machine Learning. In 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 924-928). IEEE.

14. Rani, S., & Kautish, S. (2018, June). Association Clustering and Time Series Based Data Mining in Continuous Data for Diabetes Prediction. In 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1209-1214). IEEE.

15. Mertz, L. (2018). Automated Insulin Delivery: Taking the Guesswork out of Diabetes Management. IEEE pulse, 9(1), 8-9.

16. Manna, S., Maity, S., Munshi, S., & Adhikari, M. (2018, September). Diabetes Prediction Model Using Cloud Analytics. In 2018 International Conference on Advances in Computing,

Communications and Informatics (ICACCI) (pp. 30-36). IEEE.

17. Dey, S. K., Hossain, A., & Rahman, M. M. (2018, December). Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm. In 2018 21st International Conference of Computer and Information Technology (ICCIT) (pp. 1-5). IEEE.

18. Zhao, C., & Yu, C. (2015). Rapid model identification for online subcutaneous glucose concentration prediction for new subjects with type I diabetes. IEEE Transactions on Biomedical Engineering, 62(5), 1333-1344.

19. Yan, K., Zhang, D., Wu, D., Wei, H., & Lu, G. (2014). Design of a breath analysis system for diabetes screening and blood glucose level prediction. IEEE transactions on biomedical engineering, 61(11), 2787-2795.

20. Zecchin, C., Facchinetti, A., Sparacino, G., De Nicolao, G., & Cobelli, C. (2012). Neural network incorporating meal information improves accuracy of short-time prediction of glucose concentration. IEEE transactions on biomedical engineering, 59(6), 1550-1560.