International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

Big Data Processing System for Diabetes Prediction using Machine Learning Technique

B. Suvarnamukhi, M. Seshashayee

Abstract – 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, pre-processing, attribute selection and prediction. The diabetes prediction is carried out by Extreme Learning Machine (ELM) classifier. The performance of the proposed approach is analysed 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.

Keywords–Big data, electronic health records, machine learning, diabetes.

I. INTRODUCTION

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, 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 analysing 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.

The data types and the data formats are not consistent and common.

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
B. Suvarnamukhi*, Department of Computer Science, GITAM, Visakhapatnam, A.P, India
M. Seshashayee, Department of Computer Science, GITAM, Visakhapatnam, A.P, India

The big data technology relies on five different characteristics of data such as volume, velocity, variety, veracity and value [1]. 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.

The research goal is attained by segregating the work into three major phases, which are data acquisition, data pre-processing plus attribute selection and prediction. The data acquisition is carried out on a real-time basis and the data pre-processing phase attempts to eliminate unwanted information from the dataset. The attribute selection is the important for minimizing the time consumption of the proposed approach and the prediction is performed by a reliable machine learning algorithm, which is Extreme Learning Machine (ELM). The contributions of this work are as follows.

- This work attempts to propose a disease prediction system for diabetes, as it is very common now-a-days by means of big data analytics.
- The attribute selection is the most significant step, as it minimizes the computational and time complexity of the work.
- The attribute selection phase selects the optimal attributes from the data and passes the data to the machine learning algorithm.
- The performance of the system is tested in terms of sensitivity, specificity, accuracy and time consumption.

The remainder of this article is organised in the following way. Section 2 presents the related review of literature with respect to big data analytics. The proposed diabetes prediction system based on machine learning algorithm is described in section 3. Section 4 validates the performance of the proposed approach by employing standard performance measures and the paper is concluded in section 5.

II. REVIEW OF LITERATURE

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 Neighbour (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 analysed 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 updation. This prediction model is compared with the existing approaches and the results are analysed.

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 analysing 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 analyse 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 analysed 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 analysing 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 [16]. This work claims that multidimensional data analysis proves better functionality for disease prediction system and big data based prediction system is presented. A big data aware diabetes management scheme is proposed in [17]. This work discusses more on the big data requirements, solutions and reviews. Additionally, the existing approaches are analysed with respect to diabetes management.

Motivated by these existing works, the proposed work attempts to present a diabetes prediction system that is based on big data. This work is based on machine learning algorithm owing to its applicability and learning ability. The proposed approach is elaborated in the following section.

III. PROPOSED DIABETES PREDICTION SYSTEM WITH BIGDATA

This section discusses the proposed diabetes prediction system in detail along with the overview of the work.

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, 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 1.

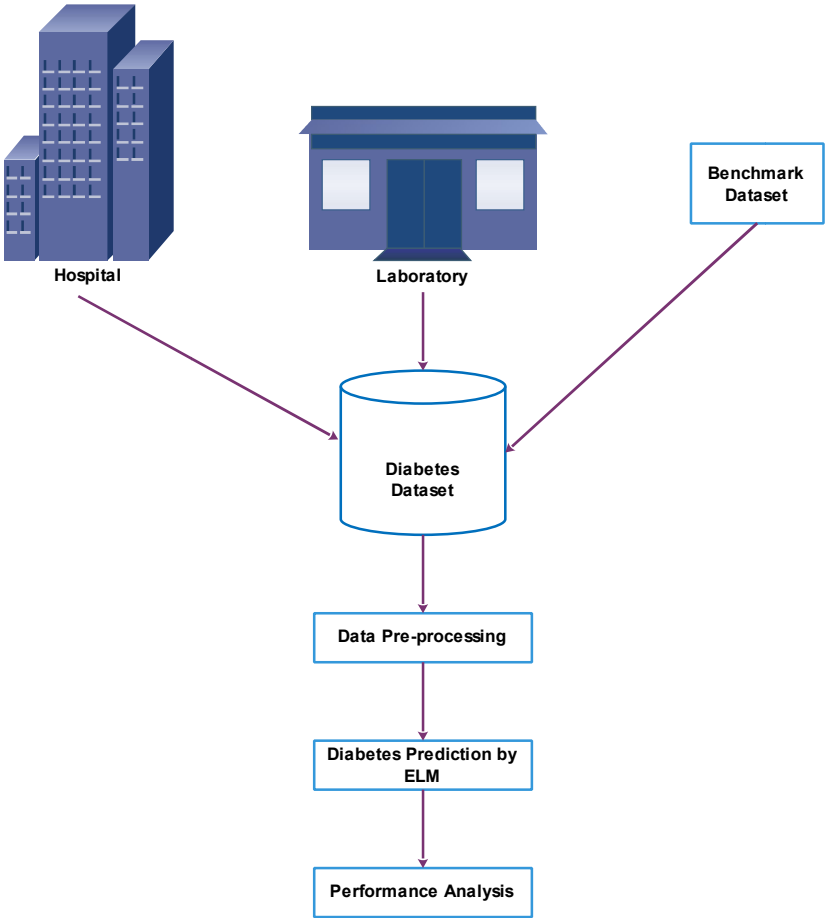

Fig.1. Overall flow of the proposed diabetes prediction system

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 analysing all the medical records associated with the patient. Diabetes is a by-product of several other ailments, which makes it necessary to analyse 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 pre-processing 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 [20]. 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 pre-processing 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. The utilized attributes along with their meanings are presented in Table 1.

Table 1. Dataset attributes with meaning

l.No	Attributes	Meaning	Data Type
1	Count of pregnancy	The total count of pregnancies encountered by the candidate	Numeric
2	Glucose level	The glucose concentration is measured by means of oral glucose tolerance test	Numeric
3	Systolic pressure	The systolic blood pressure indicates the pressure shown by blood vessels with respect to the heart beat.	Numeric
4	Diastolic pressure	The diastolic blood pressure denotes the pressure encountered by the arteries during the pause time between the heart beats.	Numeric
5	Thickness of skin	The thickness of skin is measured by considering the subcutaneous fat.	Numeric
6	Insulin	The insulin serum is considered for 2 hours.	Numeric
7	Body Mass Index (BMI)	The BMI is computed by considering both the height and weight of the patient. The ideal body weight of the patient is measured by this attribute.	Numeric
8	Diabetes Pedigree Function (DPF)	This attribute considers the diabetes disease history with respect to the blood relations.	Numeric
9	Age	The age of patient is considered as attribute.	Numeric
10	Class	Represents the diabetes type (I, II, Gestational or Normal)	Numeric

The dataset is prepared with all the ten attributes, which are extracted from the health records of multiple sources and the dataset is passed on to the machine learning algorithm. The following section describes the diabetes prediction algorithm based on ELM classifier.

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.

Let there are A training samples represented by (u_j, v_j), where $u_j = [u_{j1}, u_{j2}, \dots, u_{jn}]^T \in D^n$ and u_j represents the j^{th} training sample with n dimension. $v_j = [v_{j1}, v_{j2}, \dots, v_{jm}]^T \in D^m$ represents the j^{th} training label with m dimension, in which m is the total number of classes, which is three in this case. A Single hidden Layer Feed-Forward Neural Network (SLFN) is defined with a single activation function $a(x)$ and G neurons, is represented as follows.

$$\sum_{i=1}^{G} \gamma_i q(w_i . u_j + p_i) = r_j; j = 1,2, \dots, n \quad (1)$$

In equation 1, w_i represents the weights as vectors $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$, being responsible for mapping the i^{th} hidden neuron with the input neurons, where $i = [i1, i2, \dots, im]^T$. The vector with weights maps the i^{th} hidden neuron to the output neurons and the bias of the i^{th} hidden neuron is given by bs_i. ELM doesn't need to have any idea about the data in advance and so the $w_i$and bs_i are allotted in a random fashion. The SLFN is represented by

$$\sum_{i=1}^{G} \gamma_i a(w_i . u_j + bs_i) = v_j; j = 1,2, \dots, n \quad (2)$$

Consider HL as the ELM's hidden layer output matrix and the i^{th} column of HL denotes the i^{th} hidden neurons output vector by taking the inputs $u_{j1}, u_{j2}, \dots, u_{jn}$.

$$HL = \begin{bmatrix} a(w_1 . u_1 + bs_i) & \dots & a(w_G . u_1 + bs_G) \\ \vdots & \vdots & \vdots \\ a(w_1 . u_n + bs_i) & \dots & a(w_G . u_n + bs_G) \end{bmatrix} \quad (3)$$

$$\gamma = \begin{bmatrix} \gamma_1^T \\ \vdots \\ \gamma_G^T \end{bmatrix} \quad (4)$$

$$V = \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix} \quad (5)$$

The matrix form is represented as

$$HL\gamma = V \quad (6)$$

The output weights are calculated by means of norm least-square solution, which is represented by the following equation.

$$\gamma = HL^{\dagger}V \quad (7)$$

In the above equation, $HL^{\dagger}$ is the HL's Moore-Penrose generalized inverse. During the training, the ELM is given the total number of classes m as input, the activation function $a(x)$, count of hidden neurons G and ELM count in ensemble E. To achieve this, the ELM is treated with the training set $TS = \{(u_j, v_j) | u_j \in D^n, v_j \in D^m; j = 1,2, \dots, N\}$ and the ELM is trained by calculating γ for all TS by utilizing equation 7. The overall algorithm of this work is presented as follows.

Diabetes Prediction Algorithm
Input : Dataset
Output : Diabetes Prediction
//Training
Input – Training samples
Output – Knowledge gaining
Begin
For all training samples
Pre-process the data;
Train ELM w.r.t diabetes classes;
Acquire knowledge;
End for;
End;
//Prediction
Input – Patient's health record
Output – Diabetes Prediction
Begin
For the patient's health record
Extract necessary health attributes;
Apply ELM;
If diabetes detected
Determine the class of diabetes;
Else
Return Normal;
End for;
End;

Let $TS' = \{u_j | u_j \in D^n; j = 1,2,\dots,N'\}$ be the prediction set. At first, the output matrix $v_{test}(k) = HL_{test}(k) \times \gamma_k$ with dimensionality $N' \times m$. Finally, the output matrices are added together and the greatest value in the row is marked. This work fixes the k value as 7, as it produces optimal results. By this way, the ELM predicts the class of diabetes and the performance of the proposed work is evaluated in the following section.

IV. RESULTS AND DISCUSSION

The performance of the proposed approach is tested on a stand-alone computer with 8 GB RAM. The proposed algorithm is implemented in MATLAB R2016b, which presents a solution to handle big data in varying applications. The performance of the proposed approach is compared with the existing approaches such as modified CNN [11], J48 decision tree + Naïve bayes [12], data mining [14] in terms of prediction accuracy (AC), precision (P) and recall (R). Additionally, the time consumption of the work is also analysed. The formulae for computing these performance measures are presented as follows.

All the utilized performance measures are based on True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) rates. The accuracy rate is the most important factor of any disease prediction system. The accuracy rate is computed by considering the TP and TN rates divided by all the attempts made by the prediction system. Precision rate is computed by dividing the TP rates by the summation of TP and FN rates. Hence, greater the FN rates lesser is the precision rates. On the other hand, recall rate is the ratio of TN and the summation of FP and TN rates.

$$AC = \frac{TP+TN}{TP+TN+FP+F} \times 100 \quad (8)$$

$$P = \frac{TP}{TP+FN} \times 100 \quad (9)$$

$$R = \frac{TN}{FP+T} \times 100 \quad (10)$$

In the above equation, AC, P and R indicate accuracy, precision and recall rates. TP rates indicate that the diabetes affected patient is correctly predicted as positive similarly, TN rates indicate that a normal candidate is predicted as negative to diabetes. FP rates imply that the normal candidate is wrongly predicted with diabetes and FN rates represent that the diabetic patient is wrongly predicted as normal. Hence, both FP and FN rates are equally sensitive, as it affects the health of the candidate directly. Hence, a disease prediction system must prove minimal FP and FN rates as much as possible for enhancing the reliability of the system.

The performance of the work is tested in two scenarios, which are varying classification technique and comparison with existing approaches. The experimental results of the proposed work are presented as follows.

4.1 Performance analysis by varying classifier

The objective of this section is to justify the choice of ELM over other classifiers. The classifiers being considered for comparison are Naïve bayes, SVM and k-Nearest Neighbour (k-NN). The performance of the proposed diabetes prediction is analysed by varying the classifier and the attained results are as follows.

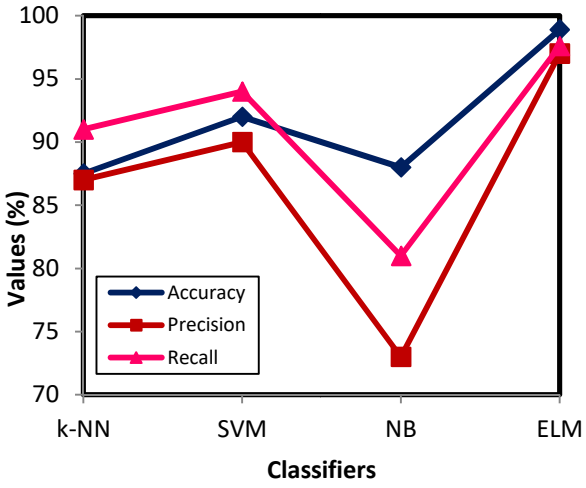

Fig.2. Performance analysis w.r.t classification techniques

From the experimental results, it is evident that the proposed disease prediction system with ELM classifier performs better than the analogous classifier. The learning ability and efficiency of ELM paves way better accuracy, precision and recall rates. The precision and recall rates determine the reliability of the disease prediction system. The greater the precision and recall rate, the more reliable is the disease prediction system. The precision and recall rates are improved by reducing the FP and FN rates. The time consumption of different classifiers for disease prediction is tabulated in table 2.

Table 2. Time consumption analysis w.r.t classifier

Classifier	Training Time (s)	Prediction Time (s)
k-NN	2.6	1.4
NB	2.4	1.1
SVM	1.9	0.98
ELM	1.2	0.64

The efficiency of ELM is proven with respect to faster learning capability and quicker disease prediction when compared to other classifiers.

The prediction time consumed by ELM is far lesser than the comparative classifiers. Hence, the proposed work employs ELM for disease prediction and the following section compares the performance of the proposed approach against the existing techniques.

4.2 Performance comparison with the existing approaches

The performance of the proposed diabetes prediction system is compared with the existing approaches and the experimental results are presented as follows. The following figure presents the performance results of the proposed work.

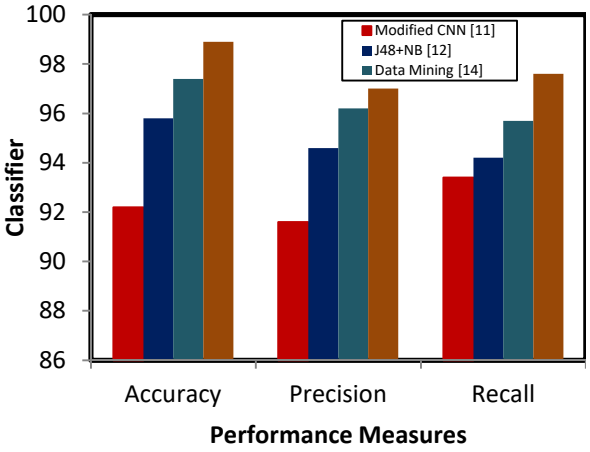

Fig.3. Performance analysis w.r.t existing approaches

The experimental results show that the proposed work proves better performance and the second best competitor to the work is data mining [14]. The following table 3 presents the time consumption of the proposed approach.

Table 3. Time consumption analysis w.r.t classifier

Classifier	Training Time (s)	Prediction Time (s)
Modified CNN [11]	1.6	0.8
J48+NB[12]	2.4	0.9
Data mining [14]	2.1	1.1
Proposed	1.2	0.64

Thus, it is observed that the proposed approach performs better prediction of diabetes and is evident through the experimental results. The following section concludes the article.

V. CONCLUSIONS

This article 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 analysed 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. 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.
2. 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.
3. 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.
4. 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.
5. Mertz, L. (2018). Automated Insulin Delivery: Taking the Guesswork out of Diabetes Management. IEEE pulse, 9(1), 8-9.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. Lekha, S., & Suchetha, M. (2017). Real-time non-invasive detection and classification of diabetes using modified convolution neural network. IEEE journal of biomedical and health informatics, 22(5), 1630-1636.
12. Perveen, S., Shahbaz, M., Keshavjee, K., & Guergachi, A. (2019). Metabolic Syndrome and Development of Diabetes Mellitus: Predictive Modeling Based on Machine Learning Techniques. IEEE Access, 7, 1365-1375.
13. Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. Computational and structural biotechnology journal, 15, 104-116.
14. Bai, B. M., Nalini, B. M., & Majumdar, J. (2019). Analysis and Detection of Diabetes Using Data Mining Techniques—A Big Data Application in Health Care. In Emerging Research in Computing, Information, Communication and Applications (pp. 443-455). Springer, Singapore.
15. Prasad, S. T., Sangavi, S., Deepa, A., Sairabanu, F., & Ragasudha, R. (2017, February). Diabetic data analysis in big data with predictive method. In 2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET) (pp. 1-4). IEEE.
16. Capobianco, E. (2017). Systems and precision medicine approaches to diabetes heterogeneity: a Big Data perspective. Clinical and translational medicine, 6(1), 23.
17. Ünalir, M. O., Can, Ö., Sezer, E., Bursa, O., & Ak, H. (2017, September). Big data aware diabetes management: Requirements, solutions and reviews. In 2017 International Artificial Intelligence and Data Processing Symposium (IDAP) (pp. 1-6). IEEE.
18. https://www.idf.org/our-network/regions-members/south-east-asia/members/94-india
19. https://www.kaggle.com/uciml/pima-indians-diabetes-database
20. Guang-Bin Huang, Hongming Zhou, Xiaojian Ding, and Rui Zhang, Extreme Learning Machine for Regression and Multiclass Classification, IEEE Transactions on systems, Man and Cybernetics - Part B, Vol.42, No.2, pp.513-529, 2012.