Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 4249 - 4259
Received 05 March 2021; Accepted 01 April 2021.

A Comparative Study of Machine Learning Algorithms using Quick-Witted Diabetic Prevention

[1,*]T.P.Latchoumi, [2]J Dayanika, [3]G.Archana

[1,*]Assistant Professor, Department of CSE, VFSTR (Deemed to be university), Andhra Pradesh, India.

[2]Research Scholar, Department of CSE, VFSTR (Deemed to be university), Andhra Pradesh, India.

[3]Assistant Professor,VignansNirula Institute of Technology and Science for Women

Abstract

Wireless Sensor Networks (WSN) are the latest advances in major information technology. Cutting-edge multi-Gbps data accelerates networks and large-scale medical data analytics have enabled the development and implementation of new diabetes control systems applications. Advanced multi-Gbps data speeds allow a new type of network designed to connect virtually including machines, objects, and devices. It is important to develop effective methods for the diagnosis and treatment of diabetes because of its long-term and systemic effects on diabetic patients. The current diabetes screening system faces the following problems: the system is inconvenient and inconsistent, difficult to collect real-time data and predicting the disease severity based on the food habits of the patients which changes day today. The diabetes screening model lacks data-sharing mechanisms, patient behavior, and personal testing. As a result, there are no continuing proposals for the prevention and treatment of diabetes. To resolve these problems, proposed a next-generation diabetes solution called the Smart Diabetes Diagnosis System. This technology used to build this system includes machine learning, massive medical data, and a large cloud of health intelligence. An intelligent diabetes management system can provide a clearer and more focused personal risk assessment and treatment schedule, and provide patients with detailed daily guidance on how to improve their self-medication. A test model has been developed to test the feasibility of multiGbps peak data speeds. Machine Learning algorithms to assure the performance of our Smart Diabetes Test – Decision Tree, Support Vector Machine (SVM), Artificial Neural Networks (ANN), Naive Bayes approved. This paper has attempted to demonstrate that our system can provide patients with individual diagnoses and treatment recommendations. Smart Diabetes System utilizes advanced data classification and tracing techniques to visualize patient monitoring graphics to improve recovery rate periodically.

Keywords: Multi-Gbps peak data speeds, SVM, ANN, Advanced Diabetes System, Naïve Bayes.

I. Introduction

Diabetes is a highly toxic chronic health condition that can be prevented. Over the next half-century, the risk of diabetes is projected to rise significantly. Diabetes is a disease that causes impairment because of low levels of insulin in the blood. Signs of hyperglycemia can include frequent urination, thirst, and increased hunger [1-5]. If you don't take the medication, it will cause a lot of trouble. Severe complications may cause cardiovascular disease, redness of the legs, and redness of the eyes. Various data mining algorithms introduce various decision support systems to assist healthcare professionals. The effectiveness of the decision-making assistance system is recognized for its accuracy [6-14]. Consequently, the goal establishes a system to support decision-making and diagnosis of specific illnesses. Even though new treatments and technologies can help treat many serious diseases, the challenge of self-medication for diabetes has not been addressed. Self-management behaviors include glucose control, drug use, healthy eating, and regular exercise. The current system does not place a great deal of emphasis on self-management. There is a relatively high level of monitoring for all areas of self-management behavior. This non-compliance is explained by the fact that self-management behaviors require changes in the patient's everyday life. To make this change, patients need to be checked regularly every minute.

To resolve the above problem, we first call it the Intelligent Diabetes System, which incorporates new technologies such as machine learning and massive medical data. Then we present the information exchange mechanism and the Smart Diabetes personal test model. Finally, we created the Smart Diabetes Test Bed and we provide experimental results. Diabetes is a chronic disorder that affects about 8.5% of the population. 422 million people around the world are struggling with diabetes. It is important to point out that type 2 diabetes accounts for approximately 90% of cases. Adolescents and youth are more likely to develop diabetes, with more critical reports that could make it worse. Diabetes has an enormous impact on global wellbeing and the economy, so there is a need to improve diabetes prevention and treatment. Also, various factors can trigger illness, such as unhealthy and unhealthy lifestyles, emotional vulnerability, and accumulated social and workplace stress.

But the current diabetes detection system faces the following challenges: the system is impractical and difficult to collect in real-time. There is also a lack of ongoing monitoring of the multiple physiological parameters of patients with diabetes. Diabetes screening models cannot conduct personal analyses of large data from a variety of sources, such as information-sharing mechanisms, lifestyles, sports, and diets. There are no on-going recommendations or surveillance strategies for diabetes prevention and treatment.

II. Literature Review

Electronic Medical Record (EMR) data collection for hospitalized diabetes patients. EMR data consists of both structured and unstructured data. In terms of structured data, they selected diabetes-related features according to doctors' advice. For unstructured data, including text and image data, the function was selected using the Convolutional Neural Network (CNN). Data analysis functions and deep learning algorithms are used to reach the audience with a diabetes diagnosis model. Using this model, users may receive a diabetes risk assessment [15-19]. They then developed a model for analyzing personal data using multiple sources and multivariate data. The personalized data includes information about the user's everyday life (for example, work, sleep, exercise, food consumption) collected by the smartphone. Along with the blood sugar index collected by the devices of the medical establishment. All this information is sent to the Big Data Cloud in Healthcare.

In the cloud, they initially used a diabetes community model to document diabetes risk assessments and organized training sessions. Then, based on Check for correction of the glycemic index and the label collected from medical devices. Once they have a justification for a diabetes risk assessment, they process the personal data to get a more sophisticated personal analysis template. Deprecated factor models were used to retrieve missing data from medical records collected at China Central Hospital. Secondly, statistical knowledge can be used to determine the main chronic diseases in the region [20-23]. Thirdly, to manage structured data, they consult healthcare professionals to obtain useful properties. Structurally, the functions that use them for text are automatically selected Outdated Factor Templates were used to retrieve lost information from medical records collected at the Chinese Central Hospital. Second, statistical knowledge can be used to determine key chronic diseases in the area. Third, to manage the structured data, they consult health professionals to obtain useful properties. For the structured textual data, they automatically select functions using the CNN algorithm. Finally, they proposed a structured and unstructured data algorithm for the multimodal disease risk prediction (CNNMDRP). Disease risk models are arrived at by combining structural and non-structural features. As proposed a data flow approach to track decision-making rules. Instead of updating the decision tree one by one by reloading all training data, additional searches only update the decision tree if the predictive results decline. This complimentary training method provides a timely prediction of diabetes conditions due to the continued introduction of blood glucose and insulin doses. Input data are not limited in time or data transmission. The Advanced Hoeffding Tree Adaptive (AHTA) training algorithm was obtained to mark the data sub-sections of the data flow to create the best performing classifier and stimulate quality decision-making rules. Finally, using a simple method called Bump Hunting (BH) can provide efficient decision rules by providing the proper amount of data. Of all the search spaces calculated from the input properties, BH found the smallest subregions to maintain the calculated output value at a minimum.

Focuses on assessing the impact of multiple compression and serialization methods on the performance of a significant data platform. And seeks to select the optimum method of compression and serialization of industrial data platform. The experiments reduced the platform's data compression time by 73.9% compared to the methods integrated into Hadoop and Spark. And a reduction lower than 96% in the 80.8% data. As the number of information increases, it will be shorter than standard methods. Inside and outside the plant, there is a great variety of data from various sources. An information platform must connect a range of data in hardware, production chains, products, and production software. Currently, there are no efficient standards or low-cost procurement mechanisms to interact with industrial data. This may entail significant costs in data collection and integration of equipment and systems. As a result, it is challenging to analyze industry information and develop application software. Medical robots can be implemented into the proposed front-end system if work to achieve a high user experience is provided. The cloud supported on the rear end is the entire brain of the system that allows the intelligent use of cognitive health. Bottom layer health data is stored in a mobile local cloud (for example, spot cloud, peripheral cloud, cloud at the edge of the network).

III. Proposed System

The next-generation diabetes solution is called the Hypertension in Diabetes (HDS) -Smart Diabetes System, which integrates new technologies such as fifth-generation HDS high-speed mobile networks, machine learning, and big medical data. Introduction of the next HDS-

Smart Diabetes Personal Data Analysis Model. Finally, based on a broad cloud of health information, we carry out HDS-Smart Diabetes tests and provide test results. And 5G-Smart is twice as important among people with HDS diabetes. On the one hand, it uses technology HDS as a communication infrastructure aimed at providing high-quality, continuous monitoring of the physiological state of diabetic patients and providing medical services without restricting their freedom. On the other hand, the following HDS goals are profitability, convenience, privacy, sustainability, and intelligence.

IV. **Methodology**

The research is mainly for data analysis Recommended Dietary Allowances (RDA) was conducted to identify the various data sources, their attributes and collected Pima Indians diabetes content from the UCI database as a step-by-step. The database was cleaned up, i.e. duplicate and missing values were deleted as of a data processing step. After that, the candidate package was clarified. Performance-based models were assessed to explore precision, memory, f1 score, incorrect classification level, and ROC-AUC score. To analyze the accuracy of the database, the above models were compared to a variety of performance measures. The evaluation of the performance and improvement of the candidate classification models is based on the measurement graph.

Figure 1 shows the process and sequence of experiments that followed this work.

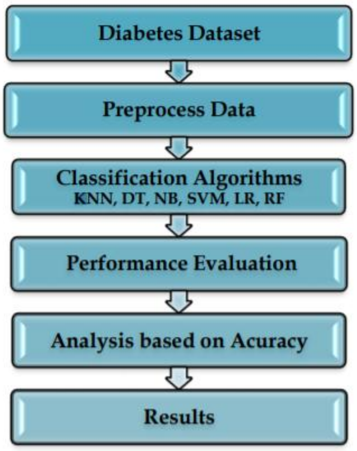

Figure-1: Proposed Framework

This paper mainly represents the description of various classifiers

The advanced KNN classification using the logistic regression technique comprises the supervised machine learning model. K-Nearest neighbors are mostly useful techniques across many machine learning systems, and this technique is also used for regression and classification problems. This model saves all possible situations based on similar measurements It has a battery function "K". It can be a measure of the distance between Euclid, Hamming, Manhattan, and Minkowski for the class of independent variables. Neighbors classify the data point by the KNN space as a measured function and will be used as an average, not an average vote of the closest neighbors. With a score of X = x1, x2,... . xk), and Y = (y1, y2, ...yk), Euclidean-distance between them is given by the Equation 1.

$$D(X^i, X^k) = \sqrt{\sum_j \left(x_j^i - x_j^k\right)^2}. \qquad \text{------} \quad (1)$$

K is an integer +ve and assumes that its value observes a particular data set. Separate these specified limits. The value of 'k' for each class. To apply classification techniques to the data in the first place, the limits separate the two classes with different meanings from 'k'. If the value of 'k' is equal to 1, then it belongs to the category of nearest neighbors. KNN is one of the challenges for modeling the selection of the 'k' value. The KNN classifier plays well. The issue of non-acceptance, to be honest, KNN gives perfect results applications such as data compression and economic advance notice. However, KNN can do the calculations the program is expensive to predict in real-time there are many examples of training and noise data and including Logistic-regression. The logistics regression model uses the technique to analyze data maps in known categories based on properties of existing data. By setting a threshold value for classifying the data elements of the model. It categorizes the data elements according to the likelihood that the expected information will exceed the threshold. for example, the predicted value is 0.5, classifying one category to another. Decision limits may be linear and non-linear. In many difficult cases, The target variable of logistical regression is the binary classification. Mathematical representation of logistics in light of resistance.

Here y = 1 is a case of yes / true. The LHS is called a logit (P (y = 1)), Figure 2 shows the log (P (y = 1)) in reverse. sigmoid function.

Received 05 March 2021; Accepted 01 April 2021.

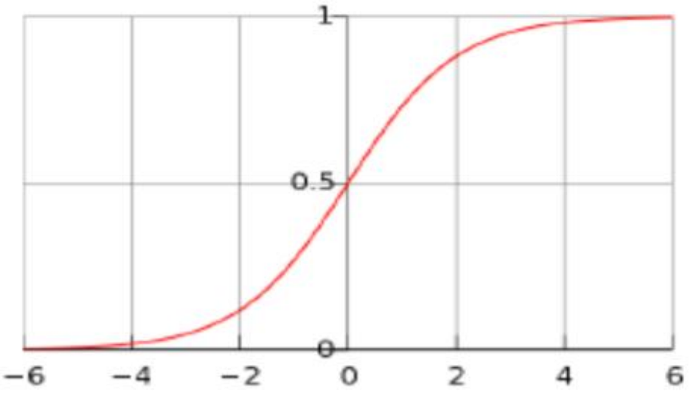

Figure 2: Sigmoid Function

Diabetes database characteristics, data types, and their statistics shown in Table 1 and Table 2.

Table 1: Basic Characteristics of dataset used

Number of samples	2532
Missing values	None
Number of numeric features	6
Number of nominal features	7
Number of classes	2

Table 2: The features and their data types

Feature Name	**Feature Type**
Age	Numeric
Sex	Nominal
BMI	Numeric
Height	Numeric
Weight	Numeric
Family history of diabetes	Nominal
History of use drugs for high blood pressure	Nominal
History of high blood pressure	Nominal
History of aborted baby	Nominal
History of pregnancy	Nominal
Diastolic blood pressure	Numeric
Systolic blood pressure	Numeric

Based on information the defined attribute predicts the patient's independence do you have diabetes or not. The database used is secondary Diabetes data from the UCI database ML database. Data were collected from a female patient who is 21 years of age and older and lives in Phoenix, Arizona. This is a two-class problem with a 1st grade meaning "Positively evaluated in the treatment of diabetes," the data package said there were 768 observations in the 500 and 1st-grade samples class 2 268 has no missing value. This is also being observed the data set does not have a real value, such as zero body mass index and zero plasma glucose. There are a total of eight of the nine properties are independent properties and one dependent. All characteristics are different or continuous and digital. The attribute definition and the statistics are shown in Tables 1 and 2. Range Index: 768 entries, from 0 to 767 which give the diabetes information based on the given dataset.

Table - 3 : Accuracy of different classification techniques

Classification technique	**Accuray**
KNN	73.08%
DT	69.8%
NB	75.525%
SVM	65.62%
LR	77.6%
RF	74.09%

Comparison of different machine-learning classifier models evaluated in the diagnosis of diabetes. performance of the accuracy of the classifiers has been underestimated correctly classified cases from the total number example. The performance of the adjuster class is Compare to the accuracy and value of percentage values areshown in Figure 3.

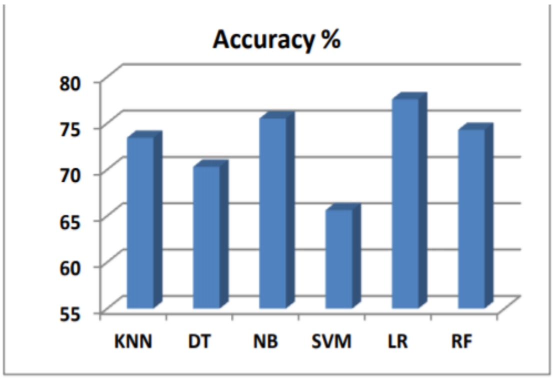

Figure 3: Accuracy Comparison classifiers

The classifier specification is based on classified cases and cabs are calculated by Eqn (1). change the spread of diabetes, Introduces glucose, insulin, pregnancy, and skin thickness. The relationship between Pearson's age and the glucose level is 0.26 and the p is 1.2e-13, as shown in Table 3 shows that logistical regression is the first to be observed (LR) with 77.6% accuracy, followed by Naive Bayes (NB) the highest accuracy is 75.525 and SVM the lowest accuracy is 65.62%. Hence the accuracy of the probability of logistic regression is higher than in comparison to other classification techniques. Figure 4shows graphs of accuracy comparison classification models that is shown above.

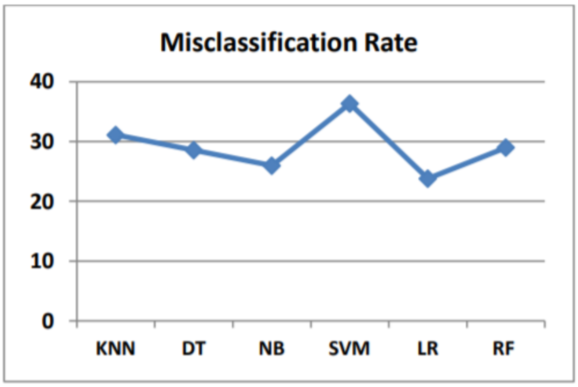

Figure-4 : Comparison of Misclassification Rate

The result Logistic regression (LR) is considered the best classification method for this test the highest accuracy compared to 77.6% is given along with other classes of techniques.

V. Conclusion

The detection and prognosis of diabetes are one of the most prevalent real-life medical issues. Including consistency, the human body leads to micro-vessels for a long time complications of diabetes. Systematized in this document the experimental study was performed using various Machines Learn the classifier to predict the likelihood of type A diabetes humanity. These models are suitable for training the package was compared and its performance was evaluated accuracy, memory, precision, and ROC-AUC-scores. Нь the results show that the Logistic Regression (LR) classifier 77.6% best performance with the highest accuracy, f1-score 75, ROC-AUC score of 73.6, and minimal compared to others, the classification level is 23.8 algorithms. This work can be extended to improve the accuracy of the hypotheses using advanced supervised machine learning techniques.

References

1. Pradhan P., “Genetic programming methods for detection Diabetes ", writes the International Journal of Computational Engineering Research, 2012, vol. 2, pp. 91–94
2. Priam A., “Comparative analysis of solution tree classification Algorithms. "Included in the current International Journal of Engineering Technologists, 2013, bot. Vol.3, pp. 334–337.
3. Latchoumi, T. P., Balamurugan, K., Dinesh, K., &Ezhilarasi, T. P. (2019). Particle swarm optimization approach for waterjet cavitation peening. Measurement, 141, 184-189.
4. Latchoumi, T. P., &Parthiban, L. (2017). Abnormality detection using weighed particle swarm optimization and smooth support vector machine.
5. Ranjeeth, S., Latchoumi, T. P., & Paul, P. V. (2020). A survey on predictive models of learning analytics. Procedia Computer Science, 167, 37-46.
6. Orabi, "The Early Diabetes Prevention System." DataMining Industry Conference, 2016, Springer.pp.420–427.
7. Ranjeeth, S., Latchoumi, T. P., &Victer Paul, P. (2019). Optimal stochastic gradient descent with multilayer perceptron based student's academic performance prediction model. Recent Advances in Computer Science and Communications. https://doi. org/10.2174/2666255813666191116150319.
8. Latchoumi, T. P., Ezhilarasi, T. P., &Balamurugan, K. (2019). Bio-inspired weighed quantum particle swarm optimization and smooth support vector machine ensembles for identification of abnormalities in medical data. SN Applied Sciences, 1(10), 1-10.
9. Latchoumi, T. P., &Sunitha, R. (2010, September). Multi agent systems in distributed datawarehousing. In 2010 International Conference on Computer and Communication Technology (ICCCT) (pp. 442-447). IEEE.
10. Loganathan, J., Janakiraman, S., &Latchoumi, T. P. (2017). A Novel Architecture for Next Generation Cellular Network Using Opportunistic Spectrum Access Scheme. Journal of Advanced Research in Dynamical and Control Systems,(12), 1388-1400.
11. Ranjeeth, S., Latchoumi, T. P., & Paul, P. V. (2020). Role of gender on academic performance based on different parameters: Data from secondary school education. Data in brief, 29, 105257.
12. Latchoumi TP, Reddy MS, Balamurugan K. Applied Machine Learning Predictive Analytics to SQL Injection Attack Detection and Prevention. European Journal of Molecular & Clinical Medicine.;7(02):2020.
13. Garikapati, P., Balamurugan, K., Latchoumi, T. P., &Malkapuram, R. (2020). A Cluster-Profile Comparative Study on Machining AlSi 7/63% of SiC Hybrid Composite Using Agglomerative Hierarchical Clustering and K-Means. Silicon, 1-12.

14. Aroulanandam VV, Latchoumi TP, Balamurugan K, Yookesh TL. Improving the Energy Efficiency in Mobile Ad-Hoc Network Using Learning-Based Routing, Revue d'IntelligenceArtificielle, Vol 34(3), pp. 337-343, 2020.DOI: https://doi.org/10.18280/ria.340312
15. Ezhilarasi, T. P., Dilip, G., Latchoumi, T. P., &Balamurugan, K. (2020). UIP—A Smart Web Application to Manage Network Environments. In Proceedings of the Third International Conference on Computational Intelligence and Informatics (pp. 97-108). Springer, Singapore.
16. Loganathan, J., Janakiraman, S., Latchoumi, T. P., &Shanthoshini, B. (2017). Dynamic Virtual Server For Optimized Web Service Interaction. International Journal of Pure and Applied Mathematics, 117(19), 371-377.
17. NaiArun, “Comparison of Diabetes Risk Classifiers Pro Prediction in Computer Science "Predictions". 2015, bot. 69, pp. 132–142.
18. A.Chinnamahammadbhasha, and Balamurugan, K. "Fracture analysis of fuselage wing joint developed by aerodynamic structural materials." Materials Today: Proceedings, Volume 38, Part 5, 2021, Pages 2555-2562, 2020.
19. Pavan MV, Balamurugan K, Balamurugan P. Compressive test Fractured Surface analysis on PLA-Cu composite filament printed at different FDM conditions. InIOP Conference Series: Materials Science and Engineering 2020 Dec 1 (Vol. 988, No. 1, p. 012019). IOP Publishing.
20. Balamurugan K. Metrological changes in surface profile, chip, and temperature on end milling of M2HSS die steel. International Journal of Machining and Machinability of Materials. 2020;22(6):443-53.
21. K. Balamurugan and Dr. M. Uthayakumar, PREPARATION AND MACHINING STUDIES OF LaPO4 Y2O3 CERAMIC MATRIX COMPOSITE, http://hdl.handle.net/10603/166221.
22. Balamurugan K, Uthayakumar M, Sankar S, Hareesh US, Warrier KG. Process optimisation and exhibiting correlation in the exploitable variable of AWJM. International Journal of Materials and Product Technology. 2020;61(1):16-33.
23. Loganathan, J., Latchoumi, T. P., Janakiraman, S., &parthiban, L. (2016, August). A novel multi-criteria channel decision in co-operative cognitive radio network using E-TOPSIS. In Proceedings of the International Conference on Informatics and Analytics (pp. 1-6).