JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

SMART E-HEALTH SYSTEM: Machine learning based e-health system for diseases prediction

[1]Ankush Gadge, [2]Monika Biradar, [3]Vishal Awate, [4] Nisha Bankar, [5] Prof. S. V. Shinde

[1-4]Student of Computer Engineering Department ,[5] Prof & Guide of Computer Engineering Department
[1-5] Department of Computer Engineering
[1-5] PDEA's College of Engineering Manjari Bk., Pune, Maharashtra, India
Savitribai Phule Pune University, Pune, Maharashtra, India

Abstract: Diabetes is a highly prevalent condition that affects people all around the globe. Long-term effects of diabetes include heart disease and renal failure, among other things. If this condition is discovered early, people may live longer and enjoy better lives. Nowadays, it seems that the healthcare business generates a large amount of data, which may be analyzed using proper machine learning algorithms and tools to deliver better insights. This kind of analysis aids in the early detection of rare and difficult-to-diagnose disorders, resulting in a higher cure success rate and lower medical costs. The goal of this study is to develop a model that will be able to predict the likelihood of diabetes in patients with the maximum accuracy. Three machine learning methods that may be utilized for data classification were selected for this research, and they are as follows: Multilayer Perceptron and Support Vector Machine On the diabetic data set, all Machine Learning Techniques were used to identify diabetes and detect stages of diabetes, and their results were analyzed using different metrics. In this study, an early-stage diabetes risk prediction dataset from the UCI machine learning repository was used. In a huge dataset, this research successfully demonstrated the capacity to detect individuals with early diabetes risks. Multilayer perceptron seems to categories the patient as diabetic or not with a better degree of accuracy than the other classification algorithms.

INTRODUCTION

Diabetes is a chronic condition that happens when the pancreas doesn't manufacture enough insulin or the body doesn't use it appropriately. Insulin regulates blood sugar. Type 1 diabetics have poor insulin production and must take insulin daily. There's no scientifically proven way to avoid type 1 diabetes. Type 1 diabetes symptoms include excessive urine excretion, thirst (polydipsia), continual starving, weight loss, visual abnormalities, and weariness [1]. Type 2 diabetes, also known as insulin-dependent or adult-onset diabetes, is the most common type of diabetes. Increased body weight and physical inactivity are mostly to blame for this kind of diabetes. The symptoms may resemble those of type 1 diabetes, although they are far less severe. Gestational diabetes is defined as hyperglycemia with a blood glucose level above normal but below the diabetes diagnostic threshold. Diabetes is a condition that develops during pregnancy. During and during pregnancy, women with gestational diabetes are more prone to problems. Women with gestational diabetes, as well as their children, have a higher chance of developing type 2 diabetes later in life. Prenatal screening, not reported symptoms, is used to identify gestational diabetes [2] [3].

Researchers such as [4] and [5] designed and proposed different kinds of machine learning based models to diagnose diabetes disease. However, only a few studies have concentrated on integrating the trained model into a user's app and designing a user interface so that consumers may monitor their health status on their smartphones. Furthermore, those models were trained using only one or two datasets, which does not guarantee that the model would perform as expected in real-world scenarios.

Polyuria, polydipsia, weariness, sudden weight loss, polyphagia, vision blurring, vaginal thrush, edema, delayed recovery, irritability, partial paresis, obesity, alopecia, and muscular stiffness are some of the indicators employed in this study's data collection. The key goals of these studies are to anticipate diabetes at an early stage so that patients may take appropriate efforts to manage it, as well as to determine the relationship between various symptoms and causes that cause diabetes. Finally, this study will aid us in determining the optimum machine learning classifier for diabetes prediction [6].

Authors Tigga and Garg in [8] conducted a survey and collected a dataset contains 950 records and 19 attributes that have measurable influence on diabetes such as Family Diabetes history, Blood Pressure, Exercise, BMI, Smoking level, Alcohol consumption, Sleeping hours, Food habits, Pregnancy, Urination frequency, Stress level and so on [7].

Table 1. Diabetes Dataset Features [7]

Diabetes dataset		
Number of Records		950
Number of attributes		19
01	Age	Age in Year
02	Gender	Gender of the participant
03	Family Diabetes	Family history with diabetes
04	highBP	Diagnosed with high blood pressure
05	PhysicallyActive	Walk/run/physically active
06	BMI	Body Mass Index
07	Smoking	Smoking
08	Alcohol	Alcohol consumption
09	Sleep	Hours of sleep
10	SoundSleep	Hours of sound sleep
11	RegularMedicine	Regular intake of medicine?
12	JunkFood	Junk food consumption
13	Stress	Not at all, Sometimes, Often, Always
14	BPLevel	Blood pressure level
15	Pregancies	Number of pregnancies
16	Pdiabetes	Gestation diabetes
17	UriationFreq	Frequency of urination
18	Diabetic	Yes or No

This paper is arranged as follows: The works that are connected are described in Section 2. The methodology of this investigation is discussed in Section 3. We've covered the dataset and the study's implementation procedure in this section. Section 4 simulates the study findings. Finally, Section 5 brings this article to a close.

LITERATURE REVIEW

Singh et al. [9] suggested a diabetes prediction method based on machine learning. They built machine learning models to diagnose diabetes using probability-based Naive Bayes (NB), decision tree-based Random Forest (RF), and function-based Multilayer Perceptron (MLP).

Kavakiotis et al. [10] investigated Diabetes Mellitus (DM). This paper attempted to discuss and investigate machine learning and data mining methodologies in connection to DM science in great depth. DM is becoming one of the most serious public health issues of the twenty-first century.

To detect diabetes, Jerjawi et al. [11] employed an artificial neural network model. This study's goal was to find the most beneficial factors and their impact on diabetes.

Using machine learning methods, Maniruzzaman et al. [12] predicted and categorised diabetes. They gathered information from the National Health and Nutrition Examination Survey, which was place between 2009 and 2012. They tested a variety of machine learning classifiers, with the Random Forest (RF) classifier providing the greatest accuracy of 94.25 percent.

Sowjanya et al. [13] used mobile devices to construct a machine learning-based Diabetes risk prediction system. They employed four machine learning classifiers, including J48, SVM, MLP, and NB, and the best result came from the Decision Tree. They also created an android app to forecast diabetes risk.

Anand et al. [14] conducted research to predict diabetes in ICU patients. J48, NB, SVM, and MLP, among other machine learning models, were utilised. They built their models using WEKA software, and the J48 algorithm produced the greatest results. The J48 algorithm's specificity, sensitivity, and Receiver Operating Characteristics (ROC) were all 0.92, 0.89, and 0.92, respectively.

Using a machine learning algorithm, Dagliati et al. [15] predicted diabetic complications. Their research demonstrated how data mining and statistical approaches may be efficiently applied in clinical care to create models that employ patient-specific information to predict illness outcomes. They classified their data using Random Forest and Logistic Regression and achieved an accuracy of 0.83.

On diabetes prediction, Kumari et al. [16] used SVM with a radial basis function (RBF) kernel. The accuracy, sensitivity, and specificity attained in this study were 78 percent, 80 percent, and 76.5 percent, respectively.

A machine learning research of blood glucose (BG) level prediction was provided by Plis et al. [17]. The BG level was predicted using a Support Vector Regression (SVG) model.

Research on Diabetes analytics was conducted by Kaur et al. [18]. SVM with radial basis function kernel (SVM-RBF), SVM with linear kernel (SVM-linear), KNN, and ANN algorithms were used to create several diabetes detections models. With the aid of the Boruta Wrapper method, feature sets were chosen. SVM linear had the highest accuracy of 89 percent among these classifiers.

Ahmed et al. [19] pioneered a novel method for treating type 2 diabetic patients. The 318 medical data points collected included sex, smoking, age, renal problem, hypertension history, eye issue, and cardiac issue. They identified a 70.8 percent accuracy rate and a ROC value of 0.62 using the J48 method.

PROPOSED SYSTEM

Figure 1 shows the proposed System Architecture for diabetes prediction and detecting the stages of diabetes.

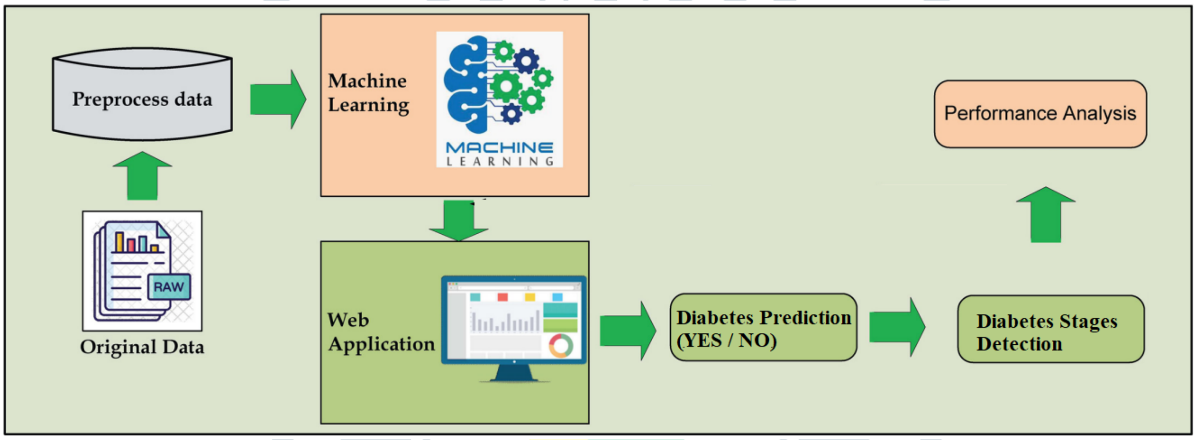

Fig 1. Proposed System Architecture

Following are various stapes involved in predicting the stages of diabetes

1) Data Input: We have used Diabetes Dataset from UCI machine learning repository [7].
2) Data Pre-processing: We transformed category characteristics such as Polyuria, Sudden weight loss, Polydipsia, Weakness, Gender, Genital thrush, Polyphagia, Visual blurring, Partial paresis, Itching, Irritability, Muscle stiffness, Delayed healing, Obesity, Alopecia, and Class into quantitative equivalents.
3) Data Splitting: In this phase, we divided our dataset into two sections: a training dataset and a research dataset. Seventy-five percent of the data was utilised for training, while twenty-five percent was used for testing.
4) Hyper-parameter Tuning: The hyper-parameters of the machine learning classifiers were tweaked in this stage. We used the Exhaustive Grid Search to do hyper-parameter optimization.
5) Applying Classifiers and Making Predictions: In this stage, we used our training data to train eleven various classifiers, including Multilayer Perceptron (MLP) and Support Vector Machine (SVM). Following training, each of these classifiers was used to predict the existence of diabetes in patients in the testing dataset.
6) Performance Evaluation and Model Selection: We examined the performance of these twelve classifiers in this stage. Five assessment criteria, namely accuracy, area under curve (AUC), precision, recall, and f-1 score, were utilised to evaluate the efficacy of these classifiers. We chose the best classifier for predicting a diabetes patient based on these assessment measures.
7) Web Application development: To develop a smart web application, we have used the JSP- Servlet framework and integrated the best model. To predict diabetes, a user is required to submit a form with necessary numbers of diabetes related parameters. The application uploaded in a server predicts

the results using the adopted machine learning model. We describe the adopted machine learning algorithms in the following sections.

A. Support Vector Machine

Support Vector Machines are used to handle classification and regression problems (SVM) [9]. The decision boundary that the SVM returns is defined by the following equation:

$$f(X) = wb + T$$

Here, w is the weight vector, X is the data dataset to be classified, and b is the linear coefficient.

RESULT AND DISCUSSION

Our proposed model is tested and evaluated in this section using a variety of machine learning algorithms, including SVM, and MLP. To find the effectiveness we have used UCI machine learning diabetes datasets.

A. Experimental setup

The proposed model is built in JSP-Servlet and executes on a computer having an Intel Core i5 processor with a 4 GB graphics card, 8GB RAM and a 64-bit Windows operating system running at 1.80 GHz.

B. Performance metrics

The performance of the proposed approach has measured using confusion matrix shown in Table 2. The confusion matrix has four different outcomes: true positive (TP), true negative (TN), false positive (FP), and false negative (FN), as follows:

Table 2: Confusion Matrix

Predicted Results	Actual Positive	Actual Negative
Yes	TP	FP
No	FN	TN

Accuracy: It measures the model's total number of accurate predictions and can be measured as a ratio between the number of correct prediction and total number of test cases of any model as follows:

$$Accuracy = (TP + TN) / (TP + FP + FN + TN)$$

Precision: The proportion of correct positive predictions to total positive predictions is known as precision:

$$Precison = (TP) / (TP + TN)$$

Recall: Total positive predictions vs. actual positive values is known as recall:

$$Recall = (TP) / (TP) + FN$$

F1-score: F1-score takes precision and recall into account and can be described as follows:

$$F1\text{-}Score = 2 * ((precion * recall) / (precion + recall))$$

C. Result

Fig. 2 represents the overall results of the experiment in terms of accuracy, precision, recall and f1-score. The accuracy of the SVM with kernel variation models is 66.0%, 73.0% of SVM-L and SVM-P for 20% dataset size and is 73.0%, 74.0% of SVM-L and SVM-P for 30% dataset size respectively, SVM-P outperforms the other methods while performing the diabetes prediction. Table 3 shows the precision, recall, f1-score and accuracy of SVM-L and SVM-P algorithms.

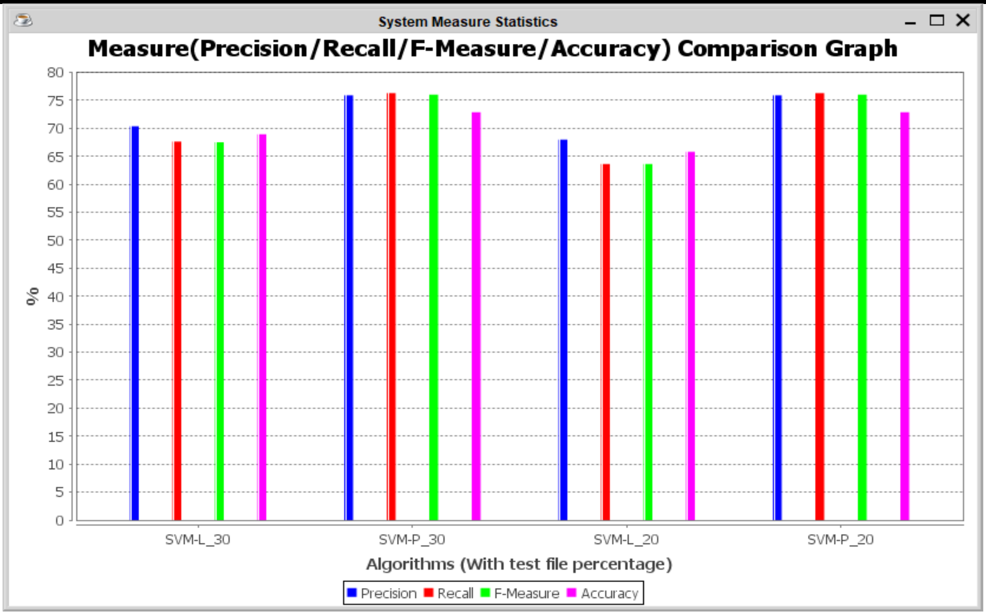

Fig. 2 Diabetes Prediction Performance Parameters Comparison

Test File %	SVM - L				SVM - P			
	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score	Accuracy
20	68	63	63	66	75	76	76	73
30	75	76	76	73	76	77	77	74

Table 3 Performance Parameters Comparison of SVM – L and SVM - P

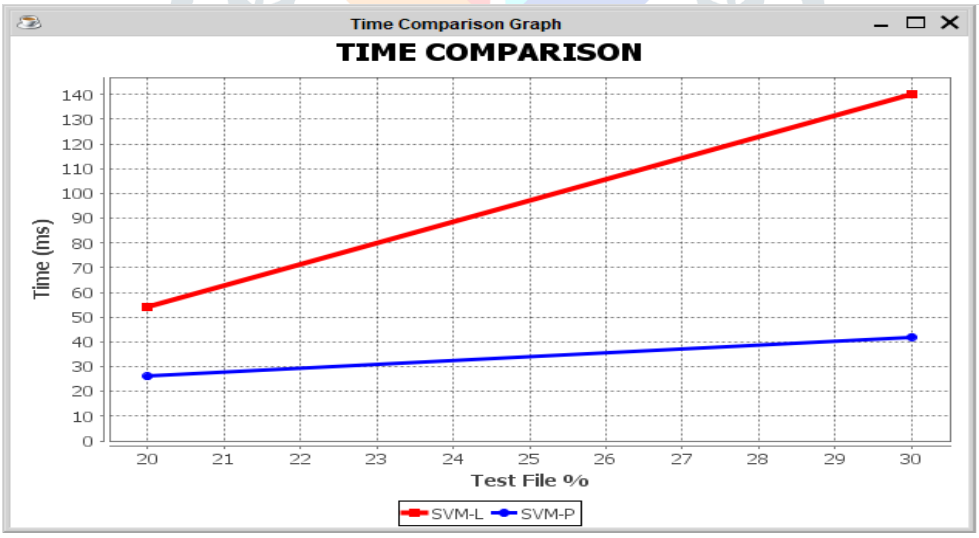

Fig. 3 Diabetes Prediction Time Comparison Graph

Test File %	SVM – L Time (ms)	SVM – P Time (ms)
20	55	26
30	139	31

Table 4: Time Comparison of SVM – L and SVM – P

CONCLUSION

In our research, firstly, we have adopted several machine learning algorithms and evaluated their performances to predict the diabetes of individuals. Secondly, we have conducted several experiments and evaluated the performances of the proposal. We found that MLP outperforms the other algorithms. Finally, based on our observed results, a smart web application is developed for predicting the diabetes accordingly.

REFRENCES

[1] Simmons, Kimber M, and Aaron W Michels. “Type 1 diabetes: A predictable disease.” World journal of diabetes vol. 6,3 (2015): 380-90. doi:10.4239/wjd.v6.i3.380

[2] Wikipedia contributors. (2020, October 26). Diabetes. In Wikipedia, The Free Encyclopedia. Retrieved 06:20, October 31, 2020, from https://en.wikipedia.org/w/index.php?title=Diabetes&oldid=985520269

[3] World Health Organization, (2020, October 28). Diabetes. Retrieved 06:20, October 31, 2020, from https://www.who.int/health-topics/diabetes

[4] Albahli, S. (2020). Type 2 machine learning: An effective hybrid prediction model for early type 2 diabetes detection. Journal of Medical Imaging and Health Informatics, 10, 1069–1075

[5] Kaur, H., & Kumari, V. (2020). Predictive modelling and analytics for diabetes using a machine learning approach. Applied Computing and Informatics. 10.1016/j.aci.2018.12.004.

[6] Emon, M. U., Keya, M. S., Meghla, T. I., Rahman, M. M., Al Mamun, M. S., Kaiser, M. S. (2020, November). Performance Analysis of Machine Learning Approaches in Stroke Prediction. In 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 1464-1469). IEEE

[7] UCI Machine Learning Repository. (1998). Diabetes data set. https://archive.ics.uci.edu/ml/datasets/diabetes

[8] Tigga, N. P., & Garg, S. (2020). Prediction of type 2 diabetes using machine learning classification methods. Procedia Computer Science, 167, 706–716

[9] Singh, D. A. A. G., Leavline, E. J., & Baig, B. S. (2017). Diabetes prediction using medical data. Journal of Computational Intelligence in Bioinformatics, 10(1), 1-8.

[10] 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.

[11] El Jerjawi, N. S., & Abu-Naser, S. S. (2018). Diabetes prediction using artificial neural network.

[12] Maniruzzaman, M., Rahman, M. J., Ahammed, B., Abedin, M. M. (2020). Classification and prediction of diabetes disease using machine learning paradigm. Health Information Science and Systems, 8(1), 7.

[13] Sowjanya, K., Singhal, A., & Choudhary, C. (2015, June). MobDBTest: A machine learning based system for predicting diabetes risk using mobile devices. In 2015 IEEE International Advance Computing Conference (IACC) (pp. 397-402). IEEE.

[14] Anand, R. S., Stey, P., Jain, S., Biron, D. R., Bhatt, H., Monteiro, K., ... & Chen, E. S. (2018). Predicting mortality in diabetic ICU patients using machine learning and severity indices. AMIA Summits on Translational Science Proceedings, 2018, 310.

[15] Dagliati, A., Marini, S., Sacchi, L., Cogni, G., Teliti, M., Tibollo, V., ... & Bellazzi, R. (2018). Machine learning methods to predict diabetes complications. Journal of diabetes science and technology, 12(2), 295- 302.

[16] Kumari, V. A., & Chitra, R. (2013). Classification of diabetes disease using support vector machine. International Journal of Engineering Research and Applications, 3(2), 1797-1801.

[17] Plis, K., Bunescu, R., Marling, C., Shubrook, J., & Schwartz, F. (2014, June). A machine learning approach to predicting blood glucose levels for diabetes management. In Workshops at the Twenty-Eighth AAAI conference on artificial intelligence.

[18] Kaur, H., & Kumari, V. (2020). Predictive modelling and analytics for diabetes using a machine learning approach. Applied computing and informatics.

[19] Islam, M. F., Ferdousi, R., Rahman, S., & Bushra, H. Y. (2020). Likelihood prediction of diabetes at early stage using data mining techniques. In Computer Vision and Machine Intelligence in Medical Image Analysis (pp. 113-125). Springer, Singapore.