Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Ml – Based Diabetes Foretell Using Svm & Logistic Regression in Healthcare

[1] AYESHA SIDDIQUA, [2] AYESHA FATIMA, [3] TAHNIYATH BEGUM,

[4] Dr. SYED ASADULLAH HUSSAINI

[1] BE Student, Dept. of Computer Science Engineering, ISL Engineering College

[2] BE Student, Dept. of Computer Science Engineering, ISL Engineering College

[3] BE Student, Dept. of Computer Science Engineering, ISL Engineering College

[4] Associate Professor, Dept. of Computer Science Engineering, ISL Engineering College

Article Info
Page Number: 1300-1308
Publication Issue:
Vol. 72 No. 1 (2023)

ABSTRACT
Diabetes is one of the most grievous health condition in the world which has no remedy to cure it after a particular stage. Based on the survey of the last 20 years, the number of people having diabetes tripled. Over 422 million people in the world are diagnosed with diabetes. It is caused due to increased blood sugar level because of imbalance in insulin processing by the body, which leads to varieties of disorders like Coronary failure, blood pressure, etc. This paper mainly focuses on the management of diabetes prediction, that will be approached using ML algorithms. It provides better results in diabetes detection by constructing models from patient datasets. The aim of this work is to make a prediction of diabetes more precisely with Logistic Regression (binary classification) and Support Vector Machine algorithm (SVM) in machine learning. It predicts the diabetes risk in early stages using symptoms and also predict using distinctive attributes of diabetes. Therefore, two different datasets of patients are used to train the models. This project work will function as an aid for the medical examiners in the diagnosis of diabetes of the patients. Thus, it can significantly help diabetes research and, ultimately, improve the quality of healthcare for diabetic patients.

Article History
Article Received: 15 October 2022
Revised: 24 November 2022
Accepted: 18 December 2022

Keywords: Diabetes, Support Vector Machine, Logistic Regression, Classification, Healthcare, Machine Learning.

I. Introduction

Diabetes is an illness caused because of high glucose level in a human body [1]. It occurs either when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces [1][2]. There are two types of diabetes, namely, Type 1 and Type 2. Type 1 diabetes is characterized by deficient insulin production and requires daily administration of insulin. Type 2 diabetes results from the body's ineffective use of insulin. Changing lifestyles require deliberate effort. Therefore, diabetics must take the ultimate responsibility for their care and treatment using available technology-related systems [1]. The advances in AI, ML and computer vision have made producing applications to automate tasks requiring intelligent behaviour, learning, and adaptation possible, hence, providing solutions to real-life problems such as diabetes management [8][11]. Many existing researches have handled for diabetes detection. Data mining approaches like clustering, classification using

KNN were studied in existing system. Lot of work has been carried out to predict diabetic diseases using dataset. Different levels of accuracy have been attained using various machine learning techniques. The accuracy of the existing system is estimated around 70-80% [2]. It required more memory and processing time. Then it does not provide the history of patient diabetic report. Finally, we built up a diabetes prediction system based on the required inputs, with high accuracy and overcoming all above mentioned problems, by using Support Vector Machine (SVM) and Logistic Regression ML algorithm. Also, the framework contain BMI, Insulin and Calorie calculation.

II. Literature Survey

Aishwarya and Vaidehi [2] presented Diabetes Prediction using Machine Learning Algorithms, in which various machine learning algorithms are applied on the dataset and the classification has been done using several machine learning algorithms such as Support Vector Machines(SVM), Random Forest Classifier, Decision Tree Classifier, Extra Tree Classifier, Ada Boost Machine Learning algorithm, Logistic Regression algorithm, K-NN algorithm, Linear Discriminant Analysis algorithm, Gaussian Naïve Bayes, Bagging algorithm and Gradient Boost Classifier. They used two different datasets- the PIMA Indian and another Diabetes dataset for testing the various models. Among all, the Logistic Regression algorithm gave them an accuracy value of 96%.

Tejas N. Joshi et al. [3] presented Diabetes Prediction Using Machine Learning Techniques which aim to predict the diabetes condition via three different supervised machine learning methods including: Support Vector Machines(SVM) algorithm, Logistic regression algorithm and Artificial Neural Networks(ANN). This work proposes an effective technique for earlier detection of the diabetes disease, which may also help the researchers to develop an accurate and effective tool that will reach at the table of clinicians to help them make better decision about the diabetes condition status.

N. Sneha1 and Tarun Gangil [4] has designed a model for Analysis of diabetes mellitus for early prediction using optimal features selection. They built the model to detect and prevent the complications of diabetes at the early stage through predictive analysis by improving the classification techniques. The dataset consists of 2500 entries and 15 attributes and 768 items used for testing and they have used 5 algorithms out of which support vector machine provides 77% accuracy. Hence, it is able to map the features effectively from low dimensions to high dimensions. It gives the best ft to the data with respect to the diabetic and non-diabetic patients.

III. System Architecture

The system provides a framework to predict the diabetes using symptoms for the patients and predict the diabetes by using various medical parameters. And maintains a database of patient personal information and records of diabetes data.

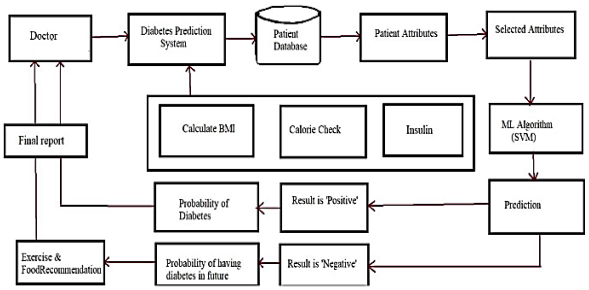

Fig. 1 Module 1

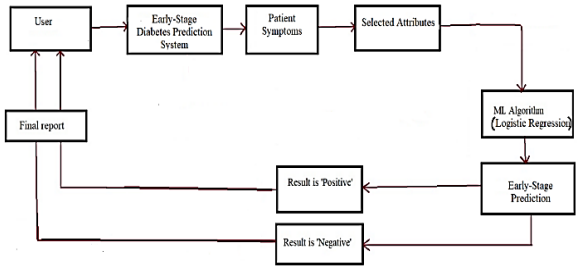

Fig. 2 Module 2

IV. Implementation

A. Data Collection

Dataset1: Diabetes Prediction Dataset

In the first step, we collect the data from UCI repository which is named as Pima Indian Diabetes Dataset. The 8 parameters used are the no. of times pregnant, BMI, plasma glucose, diastolic blood pressure, triceps skin fold thickness, diabetic pedigree function [11]. The dataset have many attributes of 768 patients.

S No.	Attributes
1	Pregnancy
2	Glucose
3	Blood Pressure
4	Skin thickness
5	Insulin
6	BMI(Body Mass Index)
7	Diabetes Pedigree Function
8	Age

Table 1: PIMA Diabetes Dataset Description [7]

Dataset 2: Early stage diabetes risk prediction dataset.

This dataset contains the sign and symptom data of newly diabetic or would be diabetic patient. This has been collected using direct questionnaires from the patients of Sylhet Diabetes Hospital in Sylhet, Bangladesh and approved by a doctor.

Data Set Characteristics:	Multivariate	**Number of Instances:**	520
Attribute Characteristics:	N/A	**Number of Attributes:**	17
Associated Tasks:	Classification	**Missing Values?**	Yes

Table 2: Early-Stage Diabetes Dataset Description [8]

B. Data Preprocessing

After loading the data, preprocessing is performed. Data preprocessing is the processing of a dataset in which data is transformed and encoded in a form such that the machine learning algorithm can parse it and only useful information is being extracted from the dataset. The values are then read sequentially for further training.

C. Feature Extraction

It is the process of converting raw data into numerical features that may be processed while maintaining the information in the original data set. It yields better results than simply applying machine learning to raw data. This is an important categorizing feature.

D. Model Creation

SVM algorithm, which stands for Support Vector Machine, is used in this project for module 1 and Logistic Regression is used for module 2. The Sci-kit Learn library has four SVM kernels. SVM creates a hyperplane that separate two classes [5][9].

Logistic regression(LR) is a statistical tool that can be used in classification modelling about the presence or absence of diabetes [6].

E. Training & Testing

Training: We split the data into training and testing datasets**.** During the training process we trained the machine from data source. We fit the SVM model for each kernel to our training set. We make predictions on our training set to see which kernel will give us the highest accuracy score. We call this Hyper-Parameter Optimization. The test data is transformed and predicts the accurate result.

Testing: Training data set will be validated using the test dataset model. The test data is transformed and predicted accurate result will be achieved, 90-92%.

F. Prediction

This module predicts the user is suffering from early-stage diabetes or not using SVM algorithm and predict the diabetes of patients using medical parameters using Logistic Regression and produce a final report.

System Flow Diagrams:

The system flow diagram for module 1 shows how the data flow when the doctor logins, perform the prediction of diabetes along with BMI, Insulin & calories calculation for required patients and produce the final medical report for the corresponding patients.

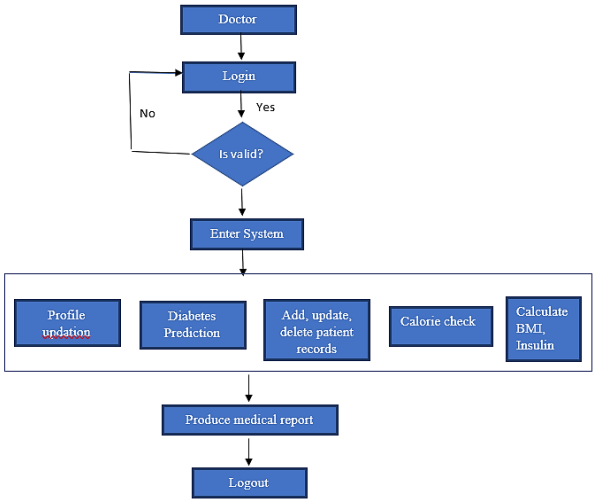

Fig. 3 Module 1

The system flow diagram for module 2 shows how the data flow when the user logins, perform early-stage diabetes prediction and get the result and medical report.

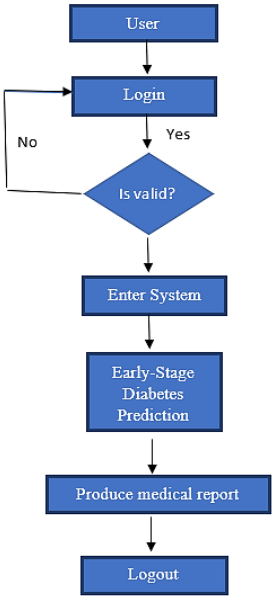

Fig. 4 Module 2

Screenshots:

A) Module 1

In fig. 5, symptoms data of user is filled for prediction & in fig. 6, the result shows positive case of diabetes along with the data filled.

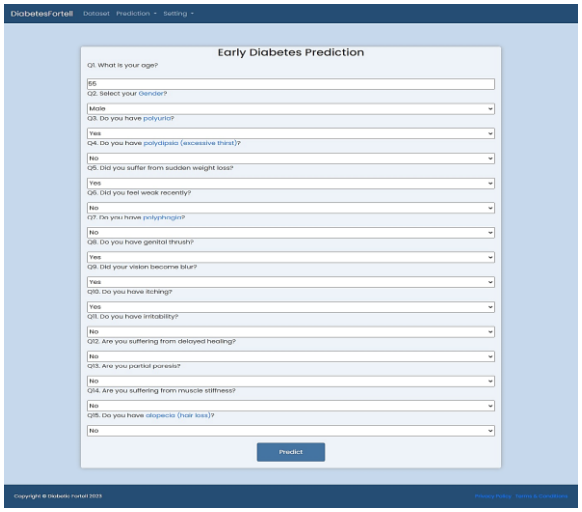

Fig. 5 Early Stage diabetes prediction

Fig. 6 Result

B) Module 2

In fig. 7, input data of patient is filled for prediction.

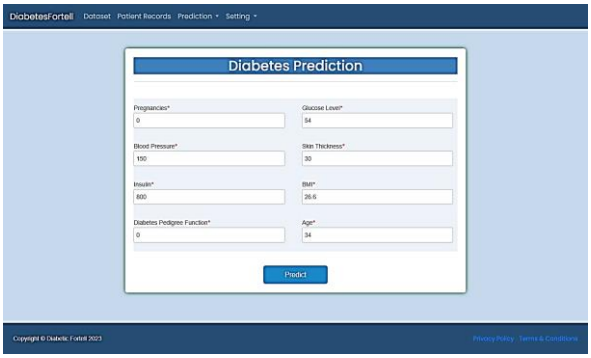

Fig. 7 diabetes prediction

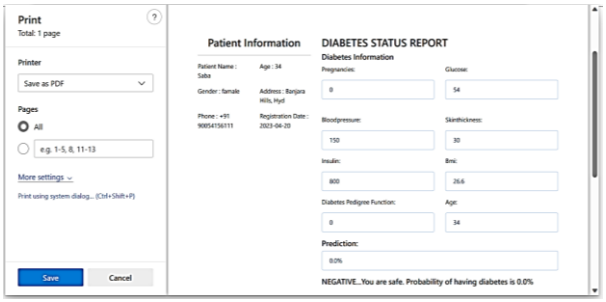

Fig. 8 Final Report

V. Conclusion & Future Work

In this work, a diabetes foretell is made in which we predicted probabilities of diabetes and generated the results based on the required input and at the last, provided a final diabetes report to be downloaded.

In future, features like exercise recommendation system, physical inactivity, family history of diabetes, and smoking habit can be added to the prediction environment. A structured dataset has been selected in the model but in the future, unstructured data will also be considered, and these methods will be applied to other medical domains for prediction, such as for different types of cancer, psoriasis, and Parkinson's disease.

VI. References

1. Mitushi Soni, Dr. Sunita Varma, “Diabetes Prediction using Machine Learning Techniques”, International Journal of Engineering Research & Technology, Volume 9, pp. 921-925, 2020.
2. Mujumdar, Aishwarya, and V. Vaidehi. “Diabetes prediction using machine learning algorithms”, International Conference on Recent Trends in Advanced Computing(ICRTAC), 165:292 299, January 2019.
3. Tejas N. Joshi, Prof. Pramila M. Chawan, "Diabetes Prediction Using Machine Learning Techniques".Int. Journal of Engineer- ing Research and Application, Vol. 8, Issue 1, (Part -II) Janu- ary 2018, pp.-09-13.
4. N.Sneha , Tarun Gangil, “Analysis of diabetes meelitus for early prediction using optimal features selection”, Journal of Big data, pp. 1-19, 2019.
5. Cortes, C., Vapnik, “Support-Vector Networks”, ML 20(3), 273–297 (1995).
6. Christodoulou E., Ma J., Collins G.S., Steyerberg E.W., Verbakel J.Y., Van Calster B, “A systematic review shows no performance benefit of ML over logistic regression for clinical prediction models” Journal of Clinical Epidemiology, Volume 110, June 2019.
7. Dhanikonda, S. R., Sowjanya, P., Ramanaiah, M. L., Joshi, R., Krishna Mohan, B. H., Dhabliya, D., & Raja, N. K. (2022). An efficient deep learning model with interrelated tagging prototype with segmentation for telugu optical character recognition. Scientific Programming, 2022 doi:10.1155/2022/1059004
8. Jain, V., Beram, S. M., Talukdar, V., Patil, T., Dhabliya, D., & Gupta, A. (2022). Accuracy enhancement in machine learning during blockchain based transaction classification.

Paper presented at the PDGC 2022 - 2022 7th International Conference on Parallel, Distributed and Grid Computing, 536-540. doi:10.1109/PDGC56933.2022.10053213 Retrieved from www.scopus.com

9. Kathole, A. B., Katti, J., Dhabliya, D., Deshpande, V., Rajawat, A. S., Goyal, S. B., . . . Suciu, G. (2022). Energy-aware UAV based on blockchain model using IoE application in 6G network-driven cybertwin. Energies, 15(21) doi:10.3390/en15218304
10. Keerthi, R. S., Dhabliya, D., Elangovan, P., Borodin, K., Parmar, J., & Patel, S. K. (2021). Tunable high-gain and multiband microstrip antenna based on liquid/copper split-ring resonator superstrates for C/X band communication. Physica B: Condensed Matter, 618
11. Kothandaraman, D., Praveena, N., Varadarajkumar, K., Madhav Rao, B., Dhabliya, D., Satla, S., & Abera, W. (2022). Intelligent forecasting of air quality and pollution prediction using machine learning. Adsorption Science and Technology, 2022 doi:10.1155/2022/5086622
12. Ramalingaswamy Cheruku, Damodar Reddy Edla, "Diabetes Classification using Radial Basis Function Network by Combining Cluster Validity Index and BAT Optimization with Novel Fitness Function". International Journal of Computational Intelligence Systems, January 2017.
13. Islam, M.M.F., Ferdousi, R., Rahman, S., Bushra, H.Y. "Likelihood prediction of diabetes at early stage using data mining techniques." Computer Vision and Machine Intelligence in Medical Image Analysis. Springer, Singapore, vol 992, August 2019.
14. S. G. Rabiha, A. Wibowo, Lukas and Y. Heryadi, "Diabetes Classification Using Support Vector Machine : Binary Classification Model," 2021 4th International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 2021.
15. V. Ganesh, J. Kolluri and K. V. Kumar, "Diabetes Prediction using Logistic Regression and Feature Normalization, " 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2021.
16. Zehra, A., Asmawaty, Aznan, M.A.M, "A comparative study on the pre-processing & Pima Indian Diabetes Dataset mining", 3rd International Conference on Software Engineering & Computer Systems, Pahang, August 2013.
17. M.A.Bari & Shahanawaj Ahamad, "Code Cloning: The Analysis, Detection and Removal", in International Journal of Computer Applications(IJCA), Vol:20, New York, U.S.A., April 2011.
18. Hafsa Fatima, Shayesta Nazneen, Maryam Banu, Dr. Mohammed Abdul Bari, "Tensor flow-Based Automatic Personality Recognition Used in Asynchronous Video Interviews", Journal of Engineering Science (JES), ISSN NO:0377-9254, Vol 13, Issue 05, May 2022.
19. Ijteba Sultana, Mohd Abdul Bari and Sanjay, "Impact of Intermediate Bottleneck Nodes on the QoS Provision in Wireless Infrastructure less Networks", Journal of Physics: Conference Series, Conf. Ser. 1998 012029 , CONSILIO Aug 2021.
20. Syed Shehriyar Ali, Mohammed Sarfaraz Shaikh, Syed Safi Uddin, Dr. Mohammed Abdul Bari, "Saas Product Comparison & Reviews Using Nlp", Journal of Engineering Science (JES), ISSN: 0377-9254, Vol 13, Issue 05, MAY/2022.
21. Mohammed Shoeb, Mohammed Akram Ali, Mohammed Shadeel, Dr. Mohammed Abdul Bari, "Self-Driving Car: Using Opencv2 and Machine Learning", The International journal

of analytical & experimental modal analysis (IJAEMA), ISSN: 0886-9367, Volume XIV, Issue V, May/2022.

22. Mr. Pathan Ahmed Khan, Dr. M.A Bari,: "Impact Of Emergence With Robotics At Educational Institution & Emerging Challenges", International Journal of Multidisciplinary Engineering in Current Research(IJMEC), ISSN: 2456-4265, Vol 6, Issue 12, December 2021,Page 43-46.
23. Mohammed Abdul Bari, Shahanawaj Ahamad, Mohammed Rahmat Ali, "Smartphone Security & Protection Practices", International Journal of Engineering and Applied Computer Science (IJEACS), Vol: 03, Issue: 01, December 2021.
24. Shahanawaj Ahamad, Mohammed Abdul Bari, "Big Data Processing Model for Smart City Design: A Systematic Review" , ISSUE 08 ISSN: 0011-9342; Design Engineering (Toronto) Elsevier SCI Oct.
25. Mohammed Abdul Bari, Shahanawaj Ahamad, Mohammed Rahmat Ali, "Smartphone Security and Protection Practices", International Journal of Engineering and Applied Computer Science (IJEACS); ISBN: 9798799755577 Volume: 03, Issue: 01, December 2021 (International Journal,U K) Pages 1-6.