International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307 (Online), Volume-12 Issue-2, May 2022

# Predicting Diabetes using SVM Implemented by Machine Learning

**Srikar Sistla**

***Abstract*: Age, BMI, and insulin levels, which play important roles because they are not constant and do not follow any specific patterns, are some of the factors that can be used to identify the chronic disease of Diabetes. Besides the elements described above, a few additional will be studied in subsequent subjects in this study. Before cleaning the data, support vector machine (SVM) algorithms, pandas, NumPy, and sci-kit-learn libraries are used to predict the patient's diagnosis and classify the data into various categories. The output contains two parameters: DIABETIC and NON-DIABETIC. With the available dataset, the accuracy score of training data was 77.5 percent and the accuracy score of test data was 80.5 percent.***

***Keywords*: Medical Diagnosis; Diabetes; Medical Computing; Machine Learning; Support Vector Machines.***

## I. INTRODUCTION

Diabetes is a group of metabolic diseases marked by hyperglycemia caused by defects in insulin secretion, insulin action, or both. Diabetes' chronic hyperglycemia is linked to long-term damage, dysfunction, and failure of various organs, particularly the eyes, kidneys, nerves, heart, and blood vessels [2]. As a result, it is recognized as one of the most lethal diseases in medical history. Meanwhile, machine learning is transforming all aspects of our lives, including unsolvable diseases in the healthcare industry [1], enhancing a broader perspective to find a solution with the help of various medical records, making it more efficient to predict any disease in advance, thus assisting in the implementation of preventive measures. The use of machine learning has the potential to greatly expand the reach of diabetes care.

Diabetes usually does not show any symptoms in its early stages, and it is usually advised to get a test when one notices any of the symptoms because if it is too late to get tested, it will be too late to treat diabetes [3]. Machine learning provides tools that can be used to make predictions using data visualization, and the accuracy score of the prediction can be analyzed using the Support vector machine algorithm, which helps to revolutionize diabetes care in its early stages.

The benefits of improvements in healthcare systems tend to affect people who have prevailing ailments more directly, and this group comprises the majority of the group of people affected by many diseases such as diabetes, blood sugar, and blood pressure issues [4]. As health and healthcare form a critical pillar of a healthy society, it is necessary to use the capabilities of computational methods and artificial intelligence [5] to develop new methods for application in healthcare systems to promote a healthier society and reduce the risk of such diseases in our generations, further increasing the quality of life. The impact of technology in the medical world has revolutionized over a period. of time and despite many odds, technology has paced towards many solutions and predicted many lethal diseases in advance for advanced care and treatment.

Since the advent of artificial intelligence and related technologies, such computational methods have been applied to real-time detection models in almost every field. The use of data mining, machine learning, deep learning, and computer vision has drastically reduced the difficulty of studying newer techniques that can significantly improve existing methods. In the next section, the algorithms and methods are surveyed.

## II. LITERATURE REVIEW

This paper consists of gathering the data and then making predictions using the algorithms which can help us to identify diabetes in anyone. Referred papers provided a more detailed view and broader perspective on how this disease can be tackled and in how many ways the problem can identify the disease with the help of the upcoming technology and advancing algorithms which have been under research since over a decade. Referred papers, provided support which enhanced the understanding of decision making which machine learning usually does as shown in the table.

Manuscript received on 21 April 2022.
Revised Manuscript received on 27 April 2022.
Manuscript published on 30 May 2022.
* Correspondence Author
**Srikar Sistla***, Department of Computer Science, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya, Kanchipuram (Tamil Nadu), India. Email: srikars2201@yahoo.com

| Authors, Year of publication | Institute | Aim | Data Source | AI/ML approach | Result |
|---|---|---|---|---|---|
| [1]Rajiv Singla, [2]Ankush Singla, [3]Yashdeep Gupta, and [4]Sanjay Kalra | [1]Departmentof health informatics [2]Department of Endocrinology, All India Institute of Medical Sciences, Delhi, India [3]Department of Endocrinology, BRIDE, Karnal, Haryana, India | To provide personalized early care using decision-making techniques. | 21,796 patients from an EHR repository of a level 2 city in China | Hierarchical recurrent neural network (HRNN) | Successful use of HRNN but no clinical benefits elaborated. |
| AMERICAN DIABETES ASSOCIATION | N/A | Diagnosis and classification of diabetes mellitus. | Analyses of nationally representative data from the National Health and Nutrition Examination Survey (NHANES) | Medical research in the conventional sense. | Diagnosis success for diabetes detection at an early stage. |
| [1]Amelec Viloria, [2]Yaneth Herazo-Beltran, [3]Danelys Cabrera, [4]Omar Bonerge Pineda | 1,3 Universidad de la Costa, Barranquilla, Colombia<br>2 Universidad Simon Bolivar, Barranquilla, Colombia<br>4 Universidad Tecnológica Centroamericana (UNITEC), San Pedro Sula, Honduras | Diabetes Diagnostic Prediction Using Vector Support Machines | Body mass index (BMI), age, blood glucose concentration (CG) and prior medical diagnosis of DM (no diabetes,predisposition to diabetes and diabetes) of 500 patients from a public hospital in Colombia (for confidentiality reasons) were taken the name will not be provided in this paper). | Age, BMI, and blood glucose were set as indicators and therefore inputs to the SVM, while diagnosis is the variable to be predicted(classified). The kernel used for both training and prediction was radial-based. The 10-fold cross-validation method was used to validate the computational model. Accuracy, sensitivity, specificity, positive and negative prediction values, and confusion matrix, commonly used parameters in medical diagnostic prediction, were used as SVM performance metrics. | This classifier is a potential tool to help achieve good control over new DM cases. Using SVM model |
| [1]Takashi Nakahara, [2]Hideyuki Hyogo | [1]Department of Medicine and Molecular Sciences, Graduate School of Biomedical Sciences, Hiroshima University, Hiroshima, Japan<br>[2]Department of Medicine and Molecular Sciences, Graduate School of Biomedical Sciences, Hiroshima University, Hiroshima, Japan | Type 2 diabetes mellitus is associated with the fibrosis severity in patients with nonalcoholic fatty liver disease in a large retrospective cohort of Japanese patients | In a cross-sectional multicenter study conducted in Japan, 1,365 patients were examined with biopsy-proven NAFLD. | Medical research in the conventional sense. | DM appeared to be a significant risk factor for advanced fibrosis in patients with NAFLD, and would therefore need to be properly managed to prevent the progression of NAFLD |

## III. METHODOLOGY

Machine learning is a method for a computational system to learn the characteristics of input data. Such methods have been shown to be effective in the detection of diabetes. Many machine learning algorithms, including supervised, unsupervised, and reinforcement learning methods, have been developed. Because the entire model is data-driven, this is practically possible. Machine learning can save significant human effort when massive amounts of data are fed into the database. Models are trained on this data and output the most appropriate output based on the input data. The models can be trained on any parameters that are practical and meet medical requirements. Some may look at facial features, while others may look at blood test results obtained from patients. Because the disease manifests itself in a variety of ways, the parameters change accordingly. Researchers have probed various algorithms and tweaked numerous hyperparameters with many proposed methods to obtain results that appear most suitable for real-life applications.

The use of Support Vector Machines (SVM) in detecting diabetes based on factors such as Pregnancies, Glucose, blood pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age, and Outcome is discussed in this paper. Following the classification of the data, the number of patients with and without diabetes is shown below. Diabetes outcome of tested patients: 0 is positive and 1 is negative. Individual variable histograms are also provided for better visualization.

In most cases, a 6:5 heatmap is used to validate the computational model. SVM performance metrics included accuracy, sensitivity, specificity, positive and negative prediction values, and the confusion matrix, all of which are commonly used parameters in medical diagnostic prediction.

## IV. RESULT AND DISCUSSION

Based on the validation points and individual outputs, this can be further analysed as having a significant impact on diabetes care. The outputs of all the individual attributes are shown below.

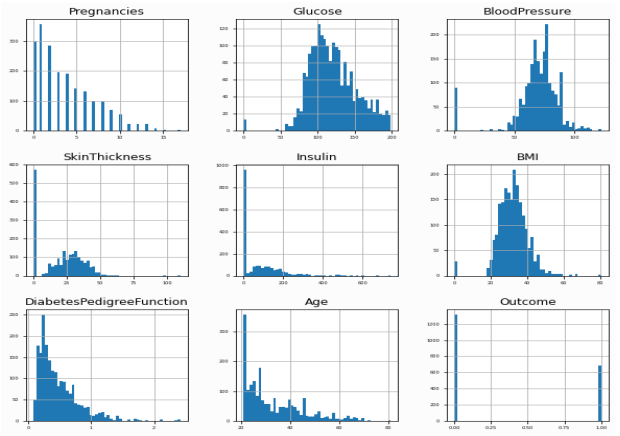


**Fig 4.1. Histogram of factors based on diabetes prediction**

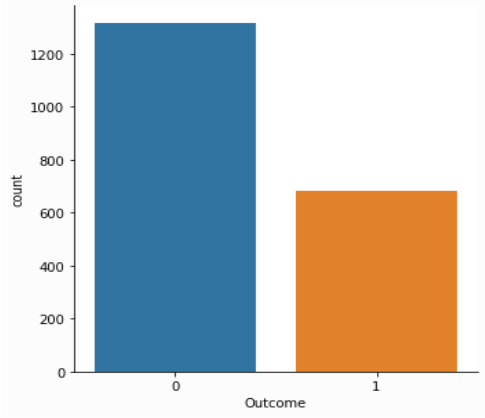


**Fig 4.2. Histogram tested for positive and negative for diabetes**

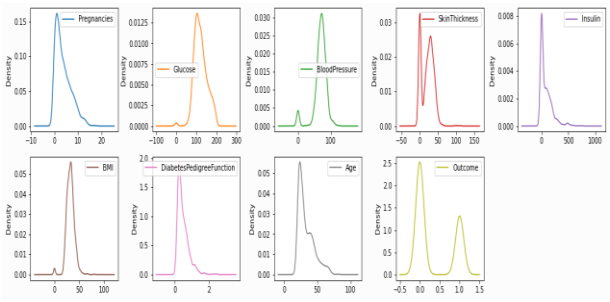


**Fig 4.3. Graph of parameters used in diabetes**


**Fig 4.4. Heatmap of diabetes along with attributes**

## V. CONCLUSION AND FUTURE ENHANCEMENTS

Followed by a detailed examination and testing with data, the outputs and findings show that machine learning and AI are really a suitable fit for advanced and early diabetes therapy for diabetes in the future. To forecast the patient's diagnosis and categorize the data, support vector machine (SVM) techniques, pandas, NumPy, and sci-kit-learn packages are utilized. There are two parameters in the output: DIABETIC and NON-DIABETIC. The accuracy score of training data was 77.5 %, while the accuracy score of test data was 80.5 %, using the provided dataset. Moreover, since the project is concluded, no additional enhancements are anticipated for the time being.

## REFERENCES

1. Singla, Rajiv, et al. "Artificial intelligence/machine learning in diabetes care." Indian journal of endocrinology and metabolism 23.4 (2019): 495.
2. American Diabetes Association. “Diagnosis and classification of diabetes mellitus.” Diabetes care vol. 33 Suppl 1,Suppl 1 (2010): S62-9. doi:10.2337/dc10-S062.
3. Amelec Viloria, Yaneth Herazo-Beltran, Danelys Cabrera, Omar Bonerge Pineda, Diabetes Diagnostic Prediction Using Vector Support Machines, Procedia Computer Science, Volume 170, 2020, Pages 376-381, ISSN 1877-0509.B. Smith, “An approach to graphs of linear forms (Unpublished work style),” unpublished.
4. Nakahara, T., Hyogo, H., Yoneda, M. et al. Type 2 diabetes mellitus is associated with the fibrosis severity in patients with nonalcoholic fatty liver disease in a large retrospective cohort of Japanese patients. J Gastroenterol 49, 1477–1484 (2014). https://doi.org/10.1007/s00535-013-0911-1.
5. Parth Solanki, Dhruv Baldaniya, Dhruvikkumar Jogani, Bhavesh Chaudhary, Manan Shah, Ameya Kshirsagar, Artificial intelligence: New age of transformation in petroleum upstream, Petroleum Research, Volume 7, Issue 1, 2022, Pages 106-114, ISSN 2096-2495, https://doi.org/10.1016/j.ptlrs.2021.07.002.

## AUTHORS PROFILE

**Srikar Sistla:** 3rd year B.E CSE Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya University, research work on Diabetes predictions using ML algorithms

*Retrieval Number: 100.1/ijsce.B35570512222*
*DOI: 10.35940/ijsce.B3557.0512222*
*Journal Website: www.ijsce.org*



*Published By:*
*Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)*