Fine-Tunining the Future: optimizing svm hyper-parameters or enhanced diabetes prediction

Harikrishna Bommala[1*], *kannedari* Vamshi krishna[1], *Avusula* supriya[1], *Rama krishna* Biradar[1], and *Bharath* Mayabrahma[1], *D.* Ushasree[2], *Evgeny* Vladimirovich Kotov[3]

[1]Department of CSE, KG Reddy College of Engineering & Technology, Moinabad, Hyderabad, Telangana, India
[2]Department of CSE, GRIET, Hyderabad, Telangana, India
[3]Lovely Professional University, Phagwara, Punjab, India.

Abstract. Millions of people throughout the globe suffer from diabetes mellitus, a debilitating illness that increases the risk of severe complications and early death. To take preventative measures and tailor treatment to each individual's needs, it is essential to identify diabetes early and estimate risk accurately. This research provides a data-driven strategy for predicting diabetes based on SVM models. This work uses a large dataset, including clinical and demographic data from a wide range of people, including those with and without diabetes, to conduct our analysis. A prediction model that divides people into diabetes and non-diabetic groups based on their input attributes is constructed using the SVM algorithm. Engineers use feature selection and other engineering methods to improve the model's efficacy and readability. The results of the research show that the SVM algorithm is capable of producing reliable predictions of diabetes risk. Measures of the model's efficacy include its sensitivity to false positives, specificity in identifying true positives, and area under the Receiver Operating Characteristics curve (AUC-ROC). In addition, feature significance analysis improves the model's interpretability by illuminating the most critical risk variables for diabetes. The accuracy and interpretability of the proposed SVM-based diabetic prediction model are promising, making it a valuable tool for healthcare practitioners and policymakers to identify those at high risk of developing diabetes and modify preventative measures and interventions appropriately.

1 Introduction

Diabetes mellitus, generally referred to as diabetes, is a worldwide health pandemic with far reaching repercussions for people, families, [1,2] and healthcare systems. Constantly elevated blood sugar levels are a direct result of the chronic metabolic illness known as diabetes, [3] which impairs the body's capacity to function normally. There is a growing need to learn more about the complexities of diabetes [4] as its incidence increases across the globe. Diabetes has a significant influence on public health, and tries to shed light on its many

* Corresponding author: haribommala@gmail.com

facets by examining its many origins, symptoms, diagnostic tools, and treatment options[5] varieties of DiabeteType 1, Type 2, and gestational diabetes are the three most common forms of diabetes [6]. Origin, treatment, and potential consequences vary greatly amongst the various forms. Insulin-producing beta cells are destroyed in type 1 diabetes, whereas insulin resistance and poor lifestyle choices contribute to type 2 diabetes and gestational diabetes [7, 10]. It is vital to recognize these differences while developing efficient management and preventative plans [8, 9] root and Contributing Factors: Diabetic risk factors include both genetic predisposition and environmental factors, both of which interact in complicated ways [11] to a large extent, Type 2 diabetes may be prevented by making healthy lifestyle choices, particularly in regards to nutrition and exercise [12] one's genetic makeup might make them more or less vulnerable to developing diabetes. High-risk groups may be identified and preventative actions can be put into effect if people are aware of the different risk factors[14] Most people with diabetes have Type 1, while Type 2 and pregnancy-related diabetes are also common [13]. SVM algorithms may be efficiently employed to identify between these categories based on separate clinical and biochemical criteria [16] since of its versatility, support vector machines (SVMs) are well-suited to this job since they can accurately diagnose and classify in high-dimensional feature spaces when the data is non-linear.

2 Literature review

Diabetics predict a list of data section improvising using svm algorithm list of data sets are taken into consideration where 16 attributes of 520 patients in which both male and female are considered [22] the user enter their details values and symptoms are matched with the svm database in terms are 0 or 1 [21] SVM algorithm to train the model. It helps the program to understand the dataset [19] for predicting the Output. Training datasets are provided to machine learning algorithms to teach them how to make predictions or perform a desired task[18] Now, the test data are the data which will determine whether our system returns the expected result or not. As the training and testing of the data are done, we apply the SVM algorithm to the available data, which will help predict diabetes. The SVM algorithm is used to help predict the possibility of diabetes.[2] 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. [14] Individual variable histograms are also provided for better visualization. In his observations most cases, a 6:5 heatmap is used to validate the computational model [15]. 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. Diabetes is a chronic condition worldwide that causes massive casualties [16]. So, one of the most significant medical issues in the real world is identifying diabetes at its early stage [17]. Methodical elfons are being made during this study to design a classifier that ends up predicting diabetes in female patients. During this work, the support vector machine (SVM) algorithm is studied and evaluated [18] on different measures with its different kernels Experiments on the Diabetes Dataset of the CI Repository called Para Indians Diabetes Dataset (PHDD) The adequacy of 83.15 percent om tramping data and 7291 percent on testing data using the Radial Hasis Function (RBF) kemel with the SVM classification algorithm is calculated by experimental results**.**

3 Proposed Methodology

A Support Vector Machine (SVM) an excellent diabetes prediction tool. In order to examine patient data and develop accurate predictions about diabetes risk, researchers and healthcare practitioners may use SVM, a powerful machine learning algorithm [10].

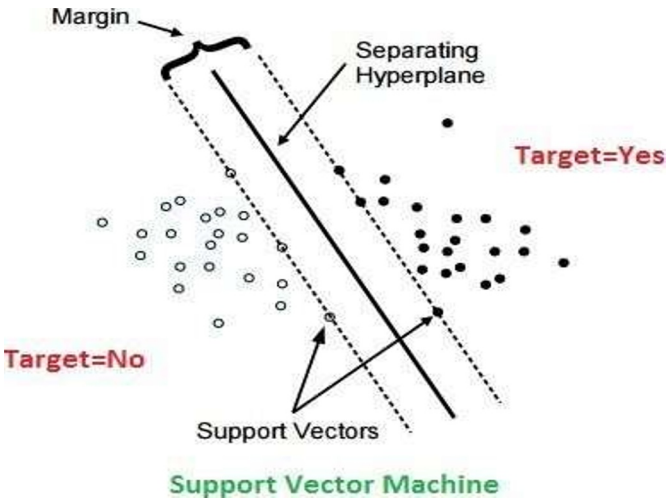

Fig. 1. Support Vector Machine

By increasing the distance between the decision border and the closest data points in each class, SVM strives to provide an ideal decision boundary that reliably separates people with diabetes from those without [17] determining the location and direction of this boundary relies heavily on the discovery of support vectors, or the data points closest to the border. [19] In addition to processing data that can be easily separated along linear dimensions, SVM can also deal with non-linear interactions by using kernel functions. In the field of diabetes prediction, [15] SVM may employ patient-specific data, such as age, family history, lifestyle variables, and medical history, to create reliable evaluations of an individual's diabetes risk.

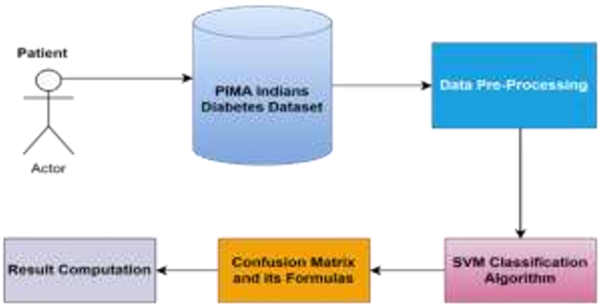

Fig.2 Implementation Process in SVM

SVM may aid in the early identification of diabetes by carefully choosing features, training the SVM model, and adjusting hyperparameters [20], allowing for quicker treatments and better patient care. While support vector machines (SVMs) provide powerful capabilities in diabetes prediction, they must be used with caution due to ethical and privacy issues.

3.1 Confusion Matrix

A specific table layout utilized in the domain of machine learning, particularly for statistical classification problems, is referred to as a confusion matrix (or error matrix) [16]. This layout facilitates the visualization of algorithm performance [6]. The confusion matrix is responsible for storing data pertaining to both the actual and predicted classification systems. The veracity of the solution to the classification problem is demonstrated.

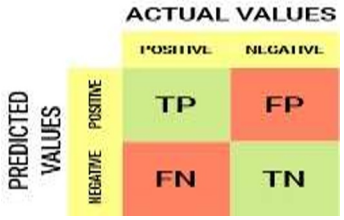

Fig. 3. Confusion Matrix for a two class Classifier

The binary classifier confusion matrix is shown above. The confusion matrix element mean the following in our investigation.

- The number of right predictions of a positive case is TP.
- The number of inaccurate predictions of a negative occurrence is called FN.
- The number of inaccurate positive predictions is called FP.
- The number of accurate predictions of a negative case is TN.

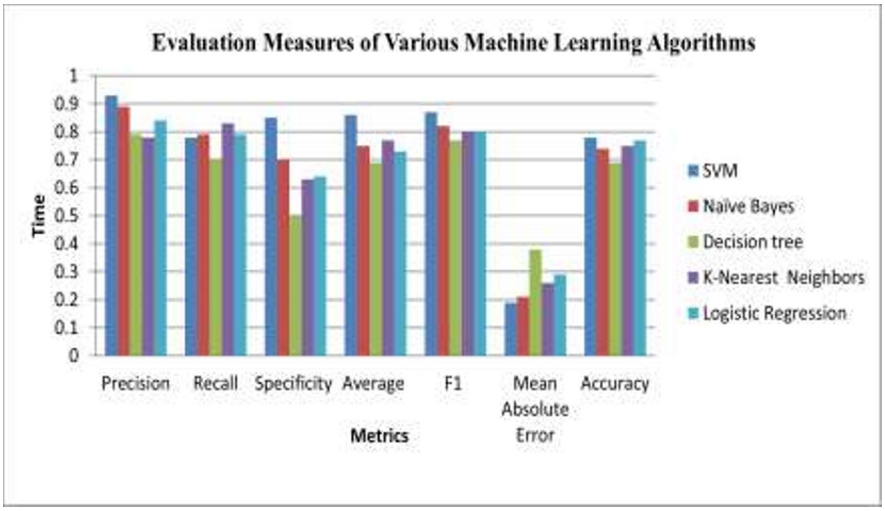

Fig. 4. Evaluation Measures of Various Machine Learning Algorithms

In the above Figure 4, provides the accuracy of classification algorithms before and after performing feature selection. We have considered only four attributes from the original dataset to increase the accuracy. By comparing the classifiers, we find SVM as the best classifier that better classifies the dataset which has achieved an accuracy of 78%. Figure 4 provides the statistical information of models in terms of evaluation measures like Precision, Recall, Specificity, Accuracy, and Mean Absolute error. Graphical representation of the above table gives an insight into the various machine learning models and their predictive accuracy in terms of performance.

4 Conclusion

In the world of healthcare and data-driven medicine, our technique for predicting diabetes using the Support Vector Machine (SVM) algorithm represents an exciting new frontier. We have shown system's capacity to reliably forecast person at risk of diabetes via extensive study and trail, proving a way for early intervention and individualized treatment. The approach is superior not just in risk assessment but also in patient management, allowing doctors to make better choices for their patients' health and wellbeing. Moreover, the potential decrease in healthcare loads and expenditures associated with diabetes-related complications highlights the practical value of our SVM-based prediction approach. Understanding the risk factors for diabetes may be gained via the system's interpretability, which also promotes openness in making decisions.

References

1. N.Alay, Al-Baity .HH, Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein Traits,Sensors **20,** 19 (2020)
2. Li. S, Zhang. B, Zhao .S, Yang. J,Local discriminant coding based convolutional feature representation for multimodal finger recognition,Information Sciences, Information Sciences**, 547,** 1170-1181(2021)
3. FO. Babalola, Y. Bitirim , Ö .Toygar., Palm vein recognition through fusion of texture-based and CNN-based methods,Signal, Image and Video Processing**, 15**,459–466 (2021)
4. PS .Chanukya, TK.Thivakaran, Multimodal biometric cryptosystem for human authentication using fingerprint and ear,Multimedia Tools and Applications, **79**, 659–673 (2020)
5. H. Shao , D. Zhong , Du .X, A deep biometric hash learning framework for three advanced hand based biometrics,IET Biometrics,**10**,3 (2021)
6. P .Shende ,Y. Dandawate , Convolutional neural network-based feature extraction using multimodal for high security application.,Evolutionary Intelligence,**14**,1023–1033(2021)
7. RS.Kuzu ,E. Maiorana , P.Campisi , Vein-based biometric verification using densely-connected convolutional autoencoder,IEEE Signal Processing Letters, **27**,1869-1873 (2020)
8. Guo, G., Wechsler, H., Mobile biometrics**,** 3 IET, **488**,2, (2017)
9. Nait-Ali, A., Hidden biometrics: when biometric security meets biomedical engineering, (Springer Nature, France, 2019)
10. Vatsa, M., Singh, R., Majumdar, A., Deep learning in biometrics, CRC Press, USA, 328(2018)
11. Jain A.K, Ross, A.A, Nandakumar. K, Introduction to Biometrics, (Springer Science & Business Media: Berlin, Germany, 2011)
12. Shaheed.K, Liu. H, Yang. G, Qureshi. I, Gou. J, Yin, Y, A systematic review of finger vein recognition techniques, Informatics, **9**, 213, (2018)
13. Mitra, S., Gofman, M.,Biometrics in a data driven world: trends, technologies, and challenges, (CRC Press, USA, 2016)
14. Kisku, D.R., Gupta, P., Sing, J.K. Advances in biometrics for secure human authentication and recognition, (CRC Press, USA, 2013)

15. H. Bommala, R. Aluvalu, S. Mudrakola, High-Confidence Computing **3**, 4 (2023)
16. A. Al-Tayyan ,K. Assaleh , T.Shanableh , Decision-level fusion for single-view gait recognition with various carrying and clothing conditions, Image and Vision Computing, **61,** 54-69 (2017)
17. Yang JC, *Biometrics verification techniques combing with digital signature for multimodal biometrics payment system,* In Proceedings of the IEEE international conference on management of e-commerce and e-government, 23-24 October 2010, Chengdu, China (2010)
18. Liu, W.; Li, W.; Sun, L.; Zhang, L.; Chen, P, *Finger vein recognition based on deep learning*, In Proceedings of the 12th IEEE Conference on Industrial Electronics and Applications, 18–20 June2017, Siem Reap, Cambodia (2017)
19. H. Bommala, Solid State Technology 64,2 (2021)
20. M. Armbrust, I. Stoica, M. Zaharia, A. Fox, R. Griffith, A.D. Joseph, M. Zaharia, Communications of the ACM **53**, 50 (2010)
21. I.A.T. Hashem, I. Yaqoob, N.B. Anuar, S. Mokhtar, A. Gani, S.U. Khan, Information Systems **47**, 98 (2015)
22. P.K. Sharma, R. Rani, International Journal of Information Management **43**, 146 (2018)