Web Application for Diabetes Prediction using Machine Learning Techniques

BHAVYA MARUPURA, SAI KRISHNA VAIBHAV., NARENDRA V. G., SHIVAPRASAD G.
Department of Computer Science and Engineering,
Manipal Institute of Technology Manipal Academy of Higher Education,
Manipal Udupi Karnataka, 576104,
INDIA

Abstract: - The objective of this project is to predict a person's risk of having diabetes by utilizing Support Vector Machine (SVM) algorithms in an intuitive web application interface. This application attempts to provide accurate and reasonable predictions by using input health parameters (number of pregnancies, blood pressure, glucose level, insulin level, age, skin thickness, diabetes pedigree function, etc.) that users provide via a graphical user interface (GUI). By combining the power of SVM with user-friendly web technology, the project endeavors to enhance accessibility to predictive healthcare tools. The seamless integration of Machine Learning into a web application facilitates a simple and effective method for diabetes prediction, which could aid people in making accurate choices regarding their health. By promoting preventive measures and giving people early awareness, this initiative hopes to support proactive healthcare.

Key-Words: - Diabetes Prediction, Machine Learning, Support Vector Machine, Graphical User Interface, Web Application using Streamlit, Health Sector.

Received: March 11, 2024. Revised: September 15, 2024. Accepted: October 13, 2024. Published: November 27, 2024.

1 Introduction

Globally, the prevalence of diabetes, a chronic metabolic disease, is steadily increasing and presents serious health risks. Diabetes arises from various factors including age, sedentary habits, familial predisposition, hypertension, psychological factors like depression and stress, and unhealthy dietary choices. Diabetes puts a person at risk of heart disease, kidney disease, stroke, eye problems, blood vessel damage, nerve damage, and other conditions making the body incapable of producing insulin. Proactive management and early detection are essential to reducing its negative effects on people's health.

Diabetes can cause symptoms such as Frequent Urination, Increased Thirst and Hunger, Slowly Healing Wounds, Weight loss that goes unexplained, Mood Swings, recurrent Infections, Fatigue, Exhaustion, and drowsiness.

The International Diabetes Federation reports, [1], that globally, 382 million individuals are afflicted with diabetes, with projections indicating a rise to 592 million by 2035. Each day, numerous individuals are affected by this condition, with many unaware of their status. It predominantly impacts individuals aged between 25 and 74 years. Failure to detect and treat diabetes can result in a range of complications.

One of the most significant aspects of artificial intelligence is Machine Learning, which enables the development of computing devices with the capability to learn from past experiences without requiring programming in each instance. It is believed that Machine Learning is an immediate necessity for the present scenario of events to enable automation with the least number of possible flaws to eliminate human work. Present-day laboratory tests like oral glucose tolerance and fasting blood glucose are used to detect diabetes. Yet, this process takes a lot of time.

As a result, this project presents a novel method of predicting diabetes by utilizing Machine Learning, specifically Support Vector Machine (SVM) algorithms with four types of kernels: polynomial, sigmoid, RBF, and linear. The model is trained using data from both diabetic and nondiabetic instances (PIMA Indian Dataset) and is integrated into an easy-to-use web application interface using the Streamlit library in Python. The combination of Machine Learning algorithms and user interface (GUI) allows people to simply enter their information and receive personalized predictions about whether or not they have diabetes.

The application's combination of Support Vector Machines (SVM) enables the examination of data submitted by users, including medical history, lifestyle choices, and demographic data. By allowing people to take control of their health, this initiative seeks to close the gap between technology and healthcare and promote a proactive approach.

2 Literature Review

Numerous studies have been conducted to automatically predict diabetes through the use of ensemble and Machine Learning techniques. Most of these projects used the publicly available Pima Indian dataset, [2]. The following article briefly discusses some of these papers on automatic diabetes prediction using the PIMA Indian dataset.

A study, [3], created a system that can rapidly and accurately predict diabetes employing the random forest algorithm. Initially, the authors employed standard preprocessing methods for data, such as cleaning reduction, and integration. Compared to other algorithms used, the random forest accuracy was 90% obtained.

The algorithm SVM, [4], examines and identify diabetes using the Dataset, Pima Indian Diabetes. This work used four different kernel types: linear, polynomial, RBF, and, sigmoid to identify diabetes in the Machine Learning platform. Between 0.69 and 0.82, the authors' accuracies varied depending on the kernel. The maximum accuracy of 0.82 was attained by the SVM method using a radial basis kernel functionA smart home health monitoring system, [5], for tracking diabetes. The researchers also sued PIMA Native American records during their study. To predict diabetes, KNN, SVM, and decision trees. And decision-based decision-making to predict blood pressure status. In comparison, SVM produced better results, with an accuracy rate of 75%.

Used Machine Learning methods and the dataset of Pima Indian to develop a diabetes prediction model, [6]. The authors claim that with accuracy increments of 0.43%, the Naive Bayes method outperformed the random forest technique.

The [7] presents a Machine Learning-based early type 2 diabetes prediction method. Over 253,000 volunteer data points from a nearby Korean Hospital were included in a confidential dataset that the scientists used for six years. Synthetic oversampling, SMOTE, and Under-sampling methods are used to address the problem of data imbalance. A number of machine-learning approaches are used. The random forest and SVM classifiers obtained the best accuracy.

To create an automatic diabetes prediction system, [8] used a private hospital dataset, which is located nearby in Bangladesh along with Pima Indian. Using the datasets, multiple Machine Learning techniques were trained in this work. On the private dataset, the decision tree and K-Nearest Neighbor models yielded 79.2% and 81.2% accuracy, respectively.

This study [9] focuses on developing effective Machine-Learning classifiers to detect diabetes using clinical data. Various algorithms including Gradient Boosting, Logistic Regression, Naive Bayes, K-Nearest Neighbor, Support Vector Machine, and Decision Tree, Random Forest are trained and evaluated. Pre-processing Techniques such as normalization and label encoding are used to improve model accuracy. Feature selection methods are applied to identify significant risk factors. The models are tested on multiple datasets, outperforming previous studies by 2.71% to 13.13% depending on the dataset and algorithm. The most accurate algorithm is selected for further development, and the model is integrated into a web application using Python Flask. Overall, the findings demonstrate the potential of preprocessing and Machine Learning classification in accurately predicting diabetes from clinical data.

In study [10] experimented with the PIMA Indian Diabetes (PID) dataset, which is available through the UCI Machine Learning repository and, consists of 768 instances with 8 attributes. Also, in order to diagnose diabetes, the World Health Organization (WHO) identified it as one of the chronic diseases with the fastest global growth rate in the year 2014. The study used Gradient Boosting (77%), Logistic Regression (79%), and Bayes classifier (86%) to predict the occurrence of diabetes.

Study [11], used the Diabetes dataset-which had 520 events and 16 characteristics obtained from the UCI repository to conduct an experiment. They concentrate on diabetes early diagnosis. The dataset was validated using seven different classification classifiers and obtained accuracies: Multilayer Perceptron (97.5%), Logistic Regression (93%), Naïve Bayes (91%), SVM (94%), Decision Tree (94%), and Random Forests (98%). With an accuracy of 98%, the results indicated that the Random Forest classifier worked as best.

Study [12] conducted an experiment on diabetes data from the UCI repository, which included 520 patients and 17 features. Focusing on early diagnosis of diabetes, they used Learning techniques such as SVM, Naïve Bayes classifiers, and LightGBM and examined data from 5320 diabetics and people with diabetes aged 16-90. SVM shows the best performance in classification and recognition accuracy, with the highest accuracy reaching 96.54%. The widely used Naïve Bayes classifier achieves an accuracy of 93.27%, while LightGBM has a lower accuracy of 88.46%. These findings suggest that SVM is the best classification algorithm for diabetes prediction. Using 520 patients and 17 features from the UCI repository, [12] ran an experiment with diabetes data. With an emphasis on early diabetes diagnosis, they analyzed data from 5320 diabetics and individuals with diabetes aged 16 to 90 using learning approaches like SVM (96.54%), Naïve Bayes classifier (93.27%), and LightGBM (88.46%). SVM performs the best in terms of recognition and classification accuracy and is the most effective classification method for predicting diabetes.

[13], used a PID dataset consisting of 738 patients in their study. The authors used different models such as K-NN, NBC, SVM, CART, C4.5, and ID3, to know the efficacy of the dataset in identifying diabetic patients. SVM and LAD were found to be the most accurate methods giving an accuracy rate of 88%.

In the study [14] K-NN, SVM, J48, and CART algorithms were used on a medical dataset. Authors have used metrics such as sensitivity, specificity, precision, Accuracy, and error rate. According to them, J48 algorithms presented the maximum accuracy at 67.15%, the SVM at 65.04%, CART at 62.28%, and K-NN at 53.39%.

Study [15], used the learning approaches: LR, K-NN, NBC, SVM, DT, and RFC for diabetes prediction. The author implemented these algorithms using a 10-fold cross-validation technique. He reported that the SVM presented the highest accuracy among all proposed approaches with a source of 84%.

The categorization of “Diabetes Prediction” according to eight attributes was studied in [16]. To analyze and predict diabetes patients, the study introduced five Machine learning algorithms: Naïve Bayes, AdaBoost, RobustBoost, LogicBoost, and Bagging. A group of PIMA Indian Diabetes datasets were used to evaluate the strategies, and the findings showed the bagging (81.77%) and AdaBoost (79.69%).

Rathore utilized classification techniques such as SVM and DTs for predicting diabetes mellitus, utilizing the PID dataset for their analysis. The study focused on women's health, particularly in the context of PIMA India. SVM achieved an accuracy rate of 82% in this prediction task, [17].

In [18] used classification methods including DT, k-NN, and SVM to predict diabetes mellitus. Among these approaches, SVM demonstrated superior performance compared to DT and KNN, achieving a maximum accuracy of 90.23%.

An online tool with a focus on diabetes prediction accuracy was created, [19]. They evaluated a number of prediction techniques, including LR, NNs, NBC, DTs, RFC, Bagging, and Boosting. According to their analysis, RFC performed the best in terms of accuracy and ROC score, obtaining a 0.912 ROC value and an accuracy level of 85.55%.

In this work [20], created a system that could combine the outcomes of many Machine Learning algorithms to provide a more accurate early diagnosis of diabetes in patients. Several techniques were used including DT, RFC, K-NN, LR, and SVM. Each algorithm’s accuracy was assessed and the model for diabetes prediction was chosen from those that showed the highest level of accuracy. Trials were carried out using the John Diabetes database, and the outcomes demonstrated the suitability of the system’s design by employing the DT algorithm to achieve an astounding 99% accuracy.

Study [21] introduced a system to address two primary challenges: the heterogeneity observed in previous techniques and the lack of transparency in features. Employing the PRISMA methodology, the study conducted comparisons among 18 different models, focusing on algorithms based on trees. The findings highlighted KNN and SVM as the primary choices for prediction tasks.

3 Problem Solution

The research exploring diabetes prediction through Machine Learning tools has shown significant promise, yet a considerable gap exists in accessible predictive healthcare tools. Although the model mentioned above works well, some issues need to be addressed.

There is a lack of readily available tools for diabetes prediction, despite advances in Machine Learning for healthcare. Widespread adoption may be impeded by the lack of user-friendly interfaces in many of the current models.

Research involving the PIMA Indian Dataset has been the focus of most researchers, potentially leading to differences. It's possible that the model's applicability to a broader range of demographics than just the Pima Indian community was limited by the dataset's lack of diversity in population representation.

The model may lose its efficacy over time if fresh data are not introduced to it regularly or if it is not adjusted to reflect the evolving health trends.

To address these gaps and limitations, the objectives are twofold:

To enhance accessibility through a user-friendly interface, ensuring easy input of health parameters for diabetes prediction, and improving model generalizability by exploring methods that account for diverse populations.

Furthermore, implementing strategies for continuous Learning and updates to the model, along with fostering user engagement and feedback mechanisms, aims to enhance the web application's usability and efficacy in predicting diabetes across diverse demographics, thus filling the existing gaps and mitigating inherent limitations.

4 Methodology

Numerous Machine Learning algorithms have been developed, including Naive Bayes, Decision Trees, Linear Regression, K Nearest Neighbors, Random Forest, Support Vector Machines, and Logistic Regression. In this paper, we employ support vector Machines (SVM) with four different kernel types: sigmoid, polynomial, RBF, and linear to identify diabetes and assess each case's accuracy. Our project's sole advantage is that it features a web interface that collects users' medical information to accurately predict whether the user is diabetic. The methodology for each step is as follows:

A. Importing libraries and Data Collection

Library Imports: Utilize Python libraries like Pandas (i.e. data alteration), NumPy (i.e. computational operations), and Scikit-learn for Machine Learning tools.

Dataset Loading: Access the PIMA Indian Diabetes dataset from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) website or repository and load it into a data frame using the imported Pandas library.

The dataset consists of features like number of pregnancies, age, blood pressure, glucose level, diabetes pedigree function, insulin level, skin thickness, etc. These attributes serve as the foundation for predicting whether the user is diabetic or not.

B. Data Preprocessing and Standardizing

Data Cleaning: Handle missing values, outliers, and inconsistencies within the dataset.

Feature Standardizing: Normalize or scale features to ensure all have a similar impact during modeling.

C. Data Splitting

Using an 80:20 ratio, split the pre-processed dataset into training and testing sets. Model training will take place on the training set (80%), and model performance evaluation will take place on the testing set (20%).

D. Training Predictive Model

Machine Learning models are trained using Support Vector Machines (SVM). This is a powerful supervised Learning algorithm capable of handling both non-linear and linear data used for both regression and classification tasks. It works by figuring out which hyperplane divides the classes in a dataset the best (Figure 1).

In a binary classification scenario, SVM aims to find a hyperplane that maximizes the margin between two classes (either diabetic or non-diabetic), effectively creating a linear separator. It aims to classify data points by their position relative to this hyperplane.

SVM can utilize various kernels to handle complex datasets that are not linearly separable in their original feature space. Here are the four SVM kernels that have been used in our study. Hyperparameters control the constant parameter(C), kernel mode, and kernel coefficient are optimized to optimize model performance.

Linear Kernel: Linear kernel calculates point features of data points and is suitable for datasets that can be grouped by straight lines or planes. It can handle large datasets efficiently and is less prone to overfitting due to its simplicity.

Polynomial Kernel: The polynomial kernel uses the polynomial function to transform data into a longer dimension. This kernel is useful when the dataset requires more complex boundaries than discrete boundaries. It can establish the relationship between data points by displaying nonlinearity in more dimensions.

RBF Kernel: The Radial Basis Function (RBF) kernel is the best option that evaluates the similarity of data referring to the field in high space. Widely used for its performance, the RBF kernel excels at capturing hard-to-identify relationships in datasets and provides robust solutions across a wide range of hard materials domains.

Sigmoid Kernel: The sigmoid kernel uses the hyperbolic function to map features to larger dimensions. Although it will be less computationally intensive than other cores, it will be more responsive in benchmarking. It can be used as an alternative to special files that other kernels cannot handle well.

Ultimately, the accuracy is computed for each of the four implementations.

E. Web Application – Streamlit Library

We import Python's pickle library to load our trained SVM models in binary format and also the Streamlit library to put up an intuitive web interface.

Streamlit is one of the most popular open-source Python libraries built to ease and speed up the development of web applications for Machine Learning and data projects. Streamlit allows one to build user-facing applications with minimal lines of code; hence, developers and data scientists will find it easy, making rapid prototyping and deployment of data-driven applications possible.

Add input fields in the web application that will capture user medical data. This will be passed through the loaded SVM model to display the user's predicted diabetic status.

This section has a methodology that merges data collection with its preprocessing, model training and evaluation, and finally, wrapping these in a web application using Python libraries such as Scikit-learn for Machine Learning, Pickle for model serialization, and Streamlit for building the user interface.

Fig. 1: Methodology: Training a predictive model

5 Results and Discussion

The majority of the information in a dataset of the Pima Indian Diabetes relates to several health metrics, such as BMI, age, glucose levels, and blood pressure. The categorization of an individual as having diabetes or not is typically represented in the dataset as a binary outcome with values such as 1 for diabetic and 0 for non-diabetic.

Based on the values of the 789 instances of the dataset that are present, the Table 1 shows a trend that demonstrates the usual range of parameters that indicate the possibility that a user has diabetes. If a user's record exceeds the standard values, as indicated in Table 1.

Table 1. Diabetes range of parameters

Features	Standard Values
Number of Pregnancies	4
Glucose Level	120.89
Blood Pressure	69.10
Skin Thickness	20.53
Insulin Level	79.79
BMI	31.99
Diabetes Pedigree Function	0.47
Age	34

It's noteworthy that establishing a person's status as diabetic or not only by looking at ranges of personal health metrics (such as BMI, Age, Blood pressure, etc.) is not always easy and can depend on several variables. To diagnose diabetes properly, medical professionals usually take into account several variables and carry out particular tests.

The following Table 2 illustrates the accuracies in each case that were obtained after training our SVM model using the dataset involving the four kernels. Figure 2 illustrates diabetes prediction accuracies of SVM different kernels.

Table 2. Train and Test accuracies of SVM kernels

Kernel Type	Train Accuracy	Test Accuracy
Linear Kernel	0.7654	0.8246
Polynomial Kernel	0.7850	0.7792
RBF Kernel	0.8469	0.7922
Sigmoid Kernel	0.6840	0.7402

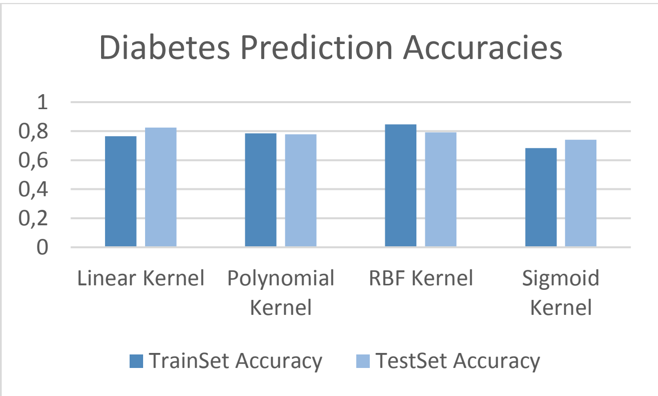

Fig. 2: Diabetes prediction accuracies

The RBF Kernel comes out to be the most appropriate considering these results. This is despite it having slightly less test accuracy compared to the Linear Kernel since, at the end of the day, this kernel is able to perform well on both train and test datasets, capturing a good level of complexity while generalizing well, too. With its flexibility in dealing with a wide variety of data patterns, the RBF kernel is most suited as a choice for the best kernel of the diabetes prediction model on the PIMA Indian Diabetes dataset, added to its quite competitive accuracy metrics.

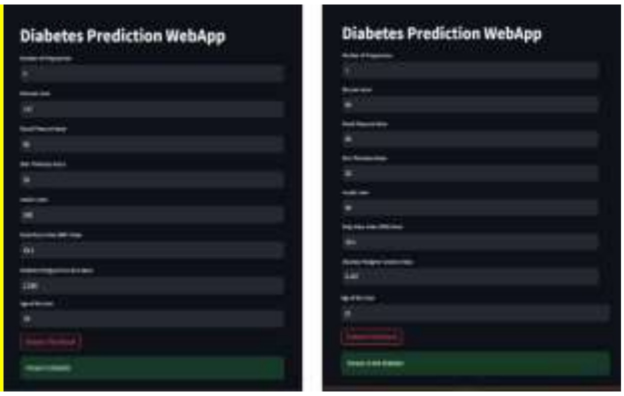

Fig. 3: WebApp for diabetes prediction

Informed by the RBF kernel, this model presents a reliable predictive tool for users to gauge against risk factors of diabetes. Streamlit's interactivity allows easy integration, whereby the model can output predictions in user inputs against parameters and facilitate proactive health management with informed decision-making. As such, this model should be deployed using Streamlit to help users be empowered with predictive insights into making informed choices related to health.

It integrated patient details and gave the application precise predictions about their diabetic status. A user is free to enter their details and get an instant outlook on their diabetic condition. With this accessible and accurate tool, people will be able to adopt more proactive ways of health management, informed decision-making, and a healthy lifestyle.

6 Conclusion and Future Work

Finally, the development and evaluation of the diabetes prediction model with Support Vector Machine kernels according to the PIMA Indian Diabetes dataset have returned some valuable insight. Of the considered kernels, the RBF Kernel turned out to be the best in terms of predictive performance, which was found to be of a robust nature both during the training process—84.69% and in test accuracy—79.22%.

This is about to be deployed into a Streamlit-based web application, which becomes a very essential tool for people who want to manage their health proactively. Integration of Streamlit's user-friendly interface with this model, equipped with the RBF kernel, offers a use case-oriented approach, thereby allowing the input of medical details and receiving predictions related to their diabetic status for informed decision-making and proactive health measures.

Some of the next steps that can be done for the completion of this project include:

Expanding the dataset to include a wide variety of demographics could make the model even more generalizable. Second, some fine-tuning of the model parameters and using different ensembling methods would further improve predictive accuracy.

Finally, continuous Learning from the model, with real-time data updates, would be much more relevant to health trends.

More on the web application (Figure 3) is enhancing it with features such as providing health recommendations based on predictions and personalized and creating more parameters to health. This shall give users a full health check.

Collaboration with medical professionals to validate model predictions against their clinical diagnoses shall help build the reliability and applicability of the model in real-world healthcare.

For the most part, this project lays a solid base of predictive healthcare tools; hence, future endeavors will move forward in aspects of refinement, scalability, and improved accuracy to help people in proactive health management.

Acknowledgment:
The authors are grateful to the Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India-576104, for providing the necessary resources and facilities.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work the authors used QuillBot/Grammatically reconstruct the sentences, Grammarly/Grammar check in order to check grammar as well as reconstruct the sentence. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References:

[1] Firdous S, Wagai GA, Sharma K (2022). A survey on diabetes risk prediction using machine learning approaches. Journal of Family Medicine and Primary Care, vol.11(11), pp. 6929-6934. doi: 10.4103/jfmpc.jfmpc_502_22.

[2] Kaggle. Pima Indians Diabetes Database, [Online]. https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database (Accessed Date: August 5, 2023).

[3] Vijiya Kumar, K., Lavanya, B., Nirmala, I., Caroline, S.S. Random forest algorithm for the prediction of diabetes. In: *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, Pondicherry, India, pp. 1–5 (2019). doi: 10.1109/ICSCAN.2019.8878802.

[4] Mohan, N., Jain, V.: Performance analysis of support vector Machine in diabetes prediction. In: *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India pp. 1–3 (2020). doi: 10.1109/ICECA49313.2020.9297411.

[5] Goyal, Ayush & Hossain, Gahangir & Chatrati, Saiteja & Bhattacharya, Sayantan & Bhan, Anupama & Gaurav, Devottam & Mishra Tiwari, Sanju. (2020). Smart Home Health Monitoring System for Predicting Type 2 Diabetes and Hypertension. *Journal of King Saud University - Computer and Information Sciences*. 34. doi: 10.1016/j.jksuci.2020.01.010.

[6] Jackins, V., Vimal, S., Kaliappan, M., & Lee, M.Y. (2020). AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. *The Journal of Supercomputing*, 77, 5198 - 5219. doi: 10.1007/s11227-020-03481-x.

[7] Deberneh HM, Kim I (2021). Prediction of Type 2 Diabetes Based on Machine Learning Algorithm. *Int. J. Environ Res Public Health*, 18(6),3317. doi: 10.3390/ijerph18063317.

[8] Pranto B, Mehnaz SM, Mahid EB, Sadman IM, Rahman A, Momen S (2020). Evaluating Machine Learning Methods for Predicting Diabetes among Female Patients in Bangladesh. *Information*, vol. 11(8):374. https://doi.org/10.3390/info11080374.

[9] Nazin Ahmed, Rayhan Ahammed, Md. Manowarul Islam, Md. Ashraf Uddin, Arnisha Akhter, Md. Alamin Talukder, Bikash Kumar Paul (2021). Machine learning based diabetes prediction and development of smart web application, *International Journal of Cognitive Computing in Engineering*, Vol. 2, pp. 229-241. https://doi.org/10.1016/j.ijcce.2021.12.001.

[10] Birjais, Roshan, Mourya, Ashish Kumar, Chauhan, Ritu, Kaur, Harleen (2019). Prediction and diagnosis of future diabetes risk: a machine learning approach. *SN Applied Sciences*, vol. 1, 1112. https://doi.org/10.1007/s42452-019-1117-9.

[11] Apratim Sadhu, Abhimanyu Jadli (2021). Early-Stage Diabetes Risk Prediction: A Comparative Analysis of Classification Algorithms, *International Advanced Research Journal in Science, Engineering and Technology*, vol. 8 (2), pp. 193-201. doi: 10.17148/IARJSET.2021.8228.

[12] Jingyu Xue, Fanchao Min, Fengying Ma (2020). Research on Diabetes Prediction Method Based on Machine Learning, *Journal of Physics: Conference Series*, vol. 1684. doi: 10.1088/1742-6596/1684/1/012062.

[13] Pragati Agrawal, Amit kumar Dewangan (2015). A brief survey on the techniques used for the diagnosis of diabetes-mellitus, *International Research Journal of Engineering and Technology (IRJET)*, vol. 2(3), pp. 1039-1043.

[14] K. Saravananathan, T. Velmurugan (2016). Analyzing Diabetic Data using Classification Algorithms in Data Mining, *Indian Journal of Science and Technology*, vol. 9(43). doi: 10.17485/ijst/2016/v9i43/93874.

[15] Ioannis Kavakiotis, Olga Tsave, Athanasios Salifoglou, Nicos Maglaveras, Ioannis Vlahavas, Ioanna Chouvarda (2017). *Machine Learning and Data Mining Methods in Diabetes Research, Computational and Structural Biotechnology Journal*, vol. 15, pp. 104-116. https://doi.org/10.1016/j.csbj.2016.12.005.

[16] Rawat, Vandana & Suryakant,. (2019). A Classification System for Diabetic Patients with Machine Learning Techniques. *International Journal of Mathematical, Engineering and Management Sciences*, vol. 4, pp. 729-744. doi:10.33889/IJMEMS.2019.4.3-057.

[17] Sakshi Gujral, Aakansha Rathore, Simran Chauhan (2017). Detecting and Predicting Diabetes Using Supervised Learning: An Approach towards Better Healthcare for Women, *International Journal of Advanced Research in Computer Science*, vol. 8(5), pp.

1192-1194. https://doi.org/10.26483/ijarcs.v8i5.3674.

[18] Hassan, A. S., Malaserene, I., & Leema, A. A. (2020). Diabetes mellitus prediction using classification techniques. *Int. J. Innov. Technol. Explor. Eng.,* vol. 9(5), pp. 2080-2084. doi: 10.35940/ijitee.E2692.039520.

[19] Nongyao Nai-arun, Rungruttikarn Moungmai (2015). Comparison of Classifiers for the Risk of Diabetes Prediction, *Procedia Computer Science*, vol. 69, pp. 132-142. https://doi.org/10.1016/j.procs.2015.10.014

[20] Aishwarya Mujumdar, V Vaidehi (2019). Diabetes Prediction using Machine Learning Algorithms, *Procedia Computer Science*,vol. 165, pp. 292-299. https://doi.org/10.1016/j.procs.2020.01.047.

[21] Branimir Ljubic, Ameen Abdel Hai, Marija Stanojevic, Wilson Diaz, Daniel Polimac, Martin Pavlovski, Zoran Obradovic (2020). Predicting complications of diabetes mellitus using advanced machine learning algorithms, *Journal of the American Medical Informatics Association*, vol. 27(9), pp. 1343–1351, https://doi.org/10.1093/jamia/ocaa120.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study**.**

Conflict of Interest

The authors have no conflicts of interest to declare.