Diabetes Prediction – Machine Learning (WebAPP)

**Bhavya Marupuraa, Sai Krishna Vaibhavb, Author Namec,,**

*a Computer Science Department, Manipal Institute of Technology, MAHE, Manipal, Karnataka, 576104*

*b Computer Science Department, Manipal Institute of Technology, MAHE, Manipal, Karnataka, 576104*

*c Computer Science Department, Manipal Institute of Technology, MAHE, Manipal, Karnataka, 576104*

**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.*

***Keywords:*** *Diabetes Prediction; Machine Learning; Support Vector Machine; Graphical User Interface; Web Application using Streamlit; Health Sector*

# 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.

Symptoms of Diabetes are:

* Blurred Vision
* Unexplained Weight Loss
* Mood Swings
* Recurring Infections
* Increased Thirst
* Frequent Urination
* Increased Thirst and Hunger
* Slowly Healing Wounds
* Fatigue
* Tiredness or Sleepiness

The International Diabetes Federation reports [9] 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.

Machine learning is a crucial component of artificial intelligence, allowing computers to learn from past experiences without the need for individual programming in each case. This remarkable technology empowers the development of computing devices that continuously improve and adapt based on their prior knowledge. It is believed that machine learning is an immediate necessity for the present scenario of events to enable automation with the least amount of possible flaws in order 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 incorporation of Support Vector Machines (SVM) facilitates the examination of data submitted by users, including medical history, lifestyle choices, and demographic data. Through empowering people to take control of their health, this initiative seeks to close the gap between technology and healthcare and promote a proactive approach.

# Literature Review

Over the years, multiple research efforts have delved into the realm of using ensemble and machine learning methods to accurately forecast diabetes. Notably, the widely accessible Pima Indian dataset [1] has been a common choice for the majority of these investigations. In this article, we will touch upon several studies that specifically focused on utilizing the Pima Indian dataset for automatic diabetes prediction.

Goyal and colleagues [4] created an advanced smart home health monitoring system to monitor diabetes. To support their research, the team utilized records from PIMA Native Americans. As part of their analysis, they employed SVM, KNN, and decision tree techniques to predict both diabetes and blood pressure status. Interestingly, the results showed that SVM performed the best among the methods, achieving an impressive 75% accuracy rate.

In their recent work [3], Mohan and Jain skillfully implemented the SVM algorithm to analyze the Pima Indian Diabetes Dataset and detect cases of diabetes. In order to achieve accurate results in the field of machine learning, they employed four distinct types of kernels: polynomial, sigmoid, RBF, and linear. Impressively, the authors achieved accuracies ranging from 0.69 to 0.82, with the highest accuracy of 0.82 achieved by using the radial basis kernel function within the SVM method. This noteworthy achievement highlights the effectiveness of the SVM approach in identifying diabetes.

Kumar's research, as described in [2], tackles the issue of diabetes prediction with a system designed to swiftly and precisely utilize the random forest algorithm. In the first stages, the team implemented typical techniques for data preprocessing, including reduction, integration, and cleaning. An impressive outcome of 90% accuracy was achieved with this approach, significantly surpassing the performance of alternative algorithms.

In their study, Kim [6] put forth a pioneering early detection system for type 2 diabetes utilizing machine learning. The team leveraged a private dataset of more than 253,000 volunteer data points from a local Korean hospital spanning six years. To tackle the issue of imbalanced data, they employed various algorithms such as synthetic oversampling, SMOTE, and undersampling. Through multiple machine learning techniques, both SVM and random forest classifiers demonstrated significant success.

In their study, Jackins [5] employed advanced machine-learning techniques and the Pima Indian dataset to create a highly effective model for predicting diabetes. The authors confidently assert that the Naive Bayes algorithm outshone the random forest method, with a notable increase in accuracy of 0.43%.

In a comprehensive study [10], Birjais delved into the examination of the PIMA Indian Diabetes (PID) dataset, which contained a staggering 768 instances and 8 attributes. This dataset, readily accessible through the renowned UCI machine learning repository, was chosen for its relevance to diabetes diagnosis - a pressing issue acknowledged by the World Health Organization (WHO) in 2014 as one of the most rapidly spreading chronic ailments worldwide. Employing a combination of state-of-the-art techniques such as gradient boosting, logistic regression, and naive Bayes classifiers, the study yielded impressive accuracies of 86%, 79%, and 77% respectively in predicting diabetes occurrence.

In the pursuit of creating a highly efficient diabetes prediction system, Pranto [7] conducted a study using both a private dataset from a local hospital in Bangladesh and the renowned Pima Indian dataset. Through extensive training, various machine learning methods were employed in this study. The results showed that both the K-Nearest Neighbor and decision tree models achieved remarkable accuracy rates of 81.2% and 79.2%, respectively, when applied to the private dataset.

In this study, Nazin (8) delves into the development of powerful machine-learning classifiers for detecting diabetes using clinical data. A range of algorithms, including Decision Tree, Naive Bayes, k-nearest neighbor, Random Forest, Gradient Boosting, Logistic Regression, and Support Vector Machine, are meticulously trained and evaluated. To further improve the accuracy of the models, pre-processing techniques such as label encoding and normalization are implemented. Additionally, feature selection methods are employed to identify significant risk factors. These models are then tested on various datasets and show superior performance compared to previous studies, with an increased accuracy ranging from 2.71% to 13.13%, depending on the dataset and algorithm used. The most precise algorithm is chosen for further development and successfully integrated into a web application.

In their study [11], Sadhu, A. and Jadli A. utilized a diabetes dataset from the UCI repository, which contained 520 instances and 16 attributes. Their main objective was to predict early-stage diabetes. To achieve this, they applied seven classification techniques - k-NN, logistic regression, SVM, naive Bayes, decision tree, random forests, and multilayer perceptron - to the validation set of the dataset. The results showed that the random forests classifier had the highest accuracy score of 98%, followed by logistic regression at 93%, SVM at 94%, naive Bayes at 91%, decision tree at 94%, random forests at 98%, and multilayer perceptron at 98%, after training multiple machine learning models.

In a study by Xue [12], a diabetes dataset from the well-known UCI repository was utilized to explore early detection methods. The dataset contained data from 520 patients with 17 different attributes. With an aim to develop an accurate diabetes prediction model, Xue and their team trained on data from 520 diabetic and potential diabetic patients, ranging from 16 to 90 years of age, using off-the-shelf supervised machine learning techniques such as SVM, naive Bayes classifiers, and LightGBM. After thorough analysis, it was found that SVM outperformed the other algorithms with an impressive classification and recognition accuracy rate of 96.54%. The commonly used naive Bayes classifier achieved a lower accuracy of 93.27%, while LightGBM showed even lower accuracy at 88.46%. These compelling results indicate that SVM is the most effective algorithm for diabetes prediction.

In their study, Agarwal [13] incorporated the PID dataset comprising of 738 patients. To gauge the dataset's effectiveness in detecting diabetic patients, the researchers employed a range of models such as SVM, k-NN, NBC, ID3, C4.5, and CART. Results revealed that the SVM and LDA algorithms exhibited the highest accuracy, achieving an impressive 88% success rate.

Saravananathan and Velmurugan [14] conducted a thorough investigation into the performance of J48, CART, SVM, and k-NN algorithms on a medical dataset. Through the use of various metrics such as accuracy, specificity, sensitivity, precision, and error rate, the authors compared the results of each algorithm. The results showed that J48 had the highest accuracy of 67.15%, followed by SVM with 65.04%, CART with 62.28%, and finally k-NN with 53.39%. These findings shed light on the strengths and weaknesses of each algorithm and provide valuable insights for researchers and practitioners in the medical field.

In conducting his research [15], Kavakiotis employed various methods, including NBC, RFC, k-NN, SVM, DT, and LR, to predict diabetes. These techniques were expertly applied using a ten-fold cross-validation approach. The most noteworthy finding was that SVM proved to be the most accurate, achieving an impressive 84% accuracy score.

In a notable and engaging study, Rawat (2016) investigated the effectiveness of five different ML algorithms for predicting diabetes based on eight distinct attributes. These algorithms included AdaBoost, LogicBoost, RobustBoost, naive Bayes, and bagging. The tests were conducted on a group of diabetic PIMA Indians, resulting in remarkably accurate classifications of 81.77% and 79.69% for the bagging and AdaBoost methods, respectively. This impressive performance solidifies the attractiveness, effectiveness, and efficiency of these DM prediction strategies.

In their analysis of the PID dataset, Rathore employed classification techniques such as SVM and DTs to predict diabetes mellitus. Their study, specifically focused on women's health and its relation to PIMA India. Impressively, SVM achieved an impressive accuracy rate of 82% in this specific prediction task. [17]

Hassan utilized a variety of classification techniques, such as DT, k-NN, and SVM, to forecast the onset of diabetes mellitus in [18]. After thorough comparison, it was evident that SVM outperformed both DT and KNN, boasting a remarkable accuracy of 90.23%. This suggests the potential of utilizing SVM as the preferred method for predicting diabetes mellitus.

In their study, Nai-Arun and Moungmai [19] focused on developing a web application that could accurately predict diabetes. Through their investigation, they evaluated a range of prediction techniques including DTs, NNs, LR, NBC, RFC, bagging, and boosting. Their results revealed that RFC stood out as the primary performer, achieving an impressive accuracy level of 85.558% and an ROC value of 0.912. Clearly, their findings highlight the potential of RFC in accurately predicting diabetes.

The main objective of Jyoti's research [20] was to create a highly accurate system for early detection of diabetes by incorporating multiple machine learning techniques. Through a thorough analysis of various algorithms such as K-nearest neighbor, Logistic Regression, Random Forest, Support Vector Machine, and Decision Tree, the most precise model was selected for predicting diabetes. Remarkably, with a precision rate of 99%, the Decision Tree algorithm proved to be the optimal choice. This success was demonstrated through experimentation on the John Diabetes Database, confirming the effectiveness of Jyoti's method.

Branimir's groundbreaking study [21] tackled two crucial difficulties head-on: the variability found in existing methods and the opacity in their feature selection. Following the comprehensive PRISMA approach, 18 models were thoroughly evaluated, with a particular emphasis on tree-based algorithms. The results unequivocally pointed to KNN and SVM as the top contenders for accurate prediction tasks.

# Research Gap and Objectives

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.

When launching a web application that gathers and analyzes health data, particularly for predicting diabetes, safeguarding data privacy and security should be the top priority. Adequate precautions should be taken to prevent unauthorized access or breaches of user data. Additionally, it is important to provide comprehensive explanations of the predictions provided by the web application as users may not fully comprehend them or know how to interpret them. This will enable users to take necessary precautions to detect and manage their diabetes at an early stage.

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.

# Methodology

There are a multitude of machine learning algorithms that have been created, such as Naive Bayes, Decision Trees, Linear Regression, K Nearest Neighbors, Random Forest, Support Vector Machines, and Logistic Regression. In this particular study, we utilize support vector machines (SVM) with four distinct kernel types - sigmoid, polynomial, RBF, and linear - to effectively detect diabetes and accurately evaluate individual cases. What sets our project apart is the inclusion of a user-friendly web interface that collects medical data to make precise predictions about a user's diabetic status. The following outlines the methodology for each step:

## Importing Libraries and Data Collection

Library Imports: Utilize Python libraries like Pandas for data alteration, NumPy for computational operations, and Scikit-learn for machine learning tools.

Dataset Loading: Obtain 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 with the assistance of the imported Pandas library.

The data contains key characteristics such as number of pregnancies, blood pressure, age, glucose level, diabetes pedigree function, insulin level, and skin thickness. These elements serve as the building blocks for determining if an individual has diabetes.

## Data Preprocessing and Standardizing

Data Wrangling: Tackling missing values,/outliers, and inconsistencies within the dataset.

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

## Data Splitting

By dividing the pre-processed dataset into training and testing sets using an 80:20 ratio, we can ensure that the model is properly trained and accurately evaluated. The larger portion, 80%, will be used for training the model, while the remaining 20% will be used to evaluate its performance.

## Training Predictive Model

Support Vector Machines (SVM) are a crucial tool in Machine Learning, as they can handle both linear and non-linear data with ease. These powerful algorithms can be used for both regression and classification tasks, making them incredibly versatile. The way they work is by identifying the most optimal hyperplane that separates different classes within a dataset.

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.

By incorporating a range of kernels, SVM can effectively manage complex datasets that are not easily separated in their original feature space. In our study, we examined four specific SVM kernels, each with carefully optimized hyperparameters for the regularization parameter (C), kernel type, and kernel coefficient. Our goal was to achieve maximum performance for our model.

* Linear Kernel: The linear kernel computes the dot product between data points, making it ideal for datasets with clearly separable classes that can be defined by a straight line or plane. It operates efficiently for large datasets and tends to be less susceptible to overfitting due to its simplicity.
* Polynomial Kernel: The polynomial kernel transforms data into higher dimensions using polynomial functions. This kernel is beneficial when datasets demand more complex boundaries beyond linear separation. By introducing non-linearity through higher dimensions, it accommodates more intricate relationships between data points.
* RBF Kernel: The Radial Basis Function (RBF) kernel, a highly adaptable choice, measures the similarity of data points to landmarks in a higher-dimensional space. Widely utilized due to its versatility, the RBF kernel excels in capturing complex relationships within datasets that lack easily definable boundaries, offering a robust solution across various data complexities.
* Sigmoid Kernel: The sigmoid kernel employs hyperbolic tangent functions to map features into higher dimensions. While less computationally intensive compared to other kernels, it might be more sensitive to feature scaling. It serves as an alternative for specific datasets where other kernels might not perform optimally.

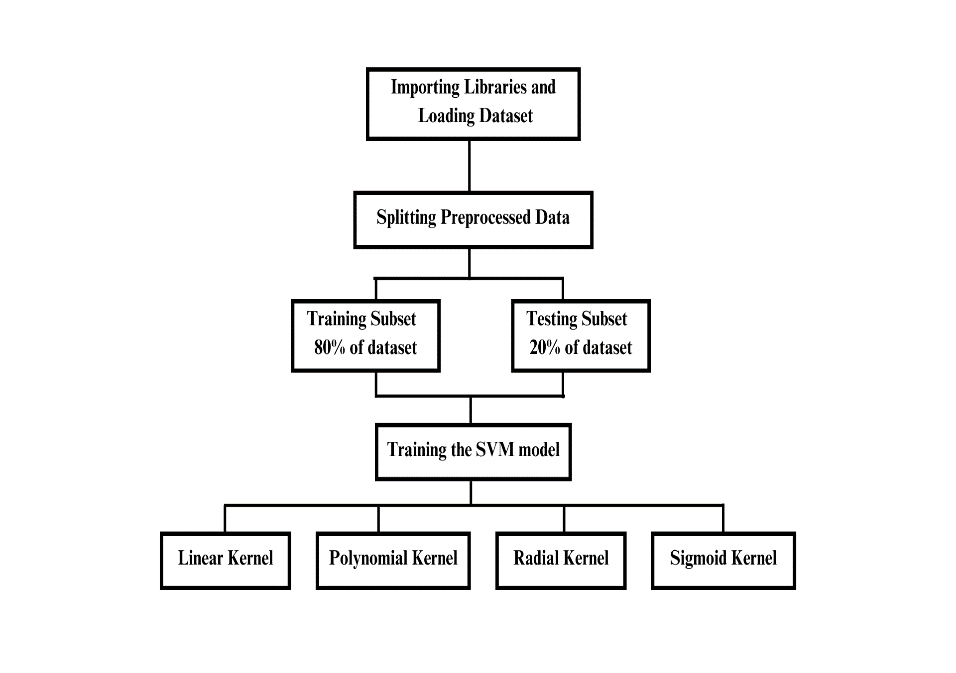
Ultimately, the accuracy is computed for each of the four implementations.

## Web Application – Streamlit Library

By importing the pickle library from Python, we load our trained SVM models in binary format and use the Streamlit library to create an intuitive web interface.

Streamlit is a popular open-source Python library specifically designed to simplify and accelerate the process of creating dynamic web applications for machine learning and data projects. It enables developers and data scientists to build user-friendly interfaces with minimal code, allowing for rapid prototyping and deployment of data-driven applications.

In this study, we incorporate input fields in the web application to collect user medical data, passing it through the loaded SVM model will display the predicted diabetic status of the user.

This methodology combines data collection, preprocessing, model training, evaluation, and deployment into a web application using Python libraries like Scikit-learn for machine learning, Pickle for model serialization, and Streamlit for building the user interface.

**Fig. 1.** Research Approach

# Discussion and Results

The Pima Indian Diabetes dataset typically includes information about various health parameters for individuals, such as glucose levels, blood pressure, BMI, age, and whether the person is diagnosed as diabetic or not. The classification of whether a person is diabetic or not is often represented as a binary outcome in the dataset, with values like 0 (indicating non-diabetic) and 1 (indicating diabetic).

The table displays a noteworthy trend indicating the likelihood of a user having diabetes based on the recorded values of 789 instances in the dataset. It is evident that if a user's records exceed the standard values, their chances of having diabetes are significantly higher, as seen in the table below. (Table 1).

It's important to note that determining whether a person is diabetic or not solely based on ranges of individual health parameters (like glucose levels, blood pressure, etc.) is not straightforward and might vary based on multiple factors. Healthcare professionals typically consider a combination of factors and conduct specific tests to diagnose diabetes accurately.

**Table 1.** Elevated records suggest diabetes likelihood

|  |  |
| --- | --- |
| 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 |

The following are the accuracies in each case that were obtained after training our SVM model using the dataset involving the four kernels (Table 2):

**Table 2.** SVM model accuracies with kernels

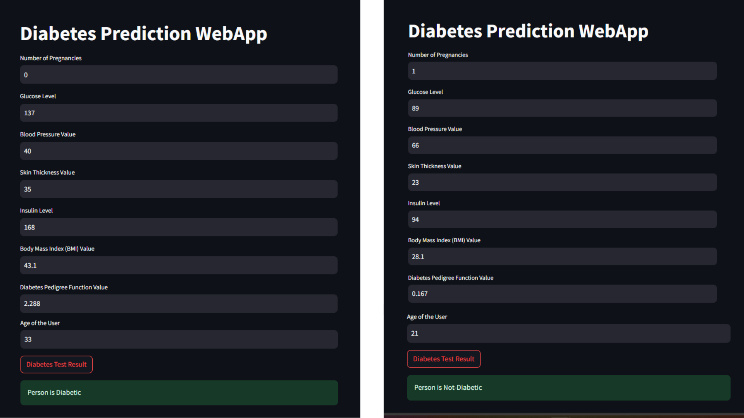
|  |  |  |
| --- | --- | --- |
| Kernel Type | Train Accuracy | Test Accuracy |
| Linear Kernel | 0.7654 | 0.8246 |
| Polynomial Kernel | 0.7850 | 0.7792 |
| RBF Kernel | 0.8469 | 0.7922 |
| Sigmoid Kernel | 0.6840 | 0.7402 |

Considering these results, the RBF Kernel emerges as the most suitable choice for the model. Despite a slightly lower test accuracy compared to the Linear Kernel, its robust performance on both training and test datasets signifies a good balance between complexity capture and generalization. The RBF Kernel's adaptability to diverse data patterns, coupled with its competitive accuracy metrics, positions it as the optimal kernel choice for the diabetes prediction model on the PIMA Indian Diabetes dataset.

This model, equipped with the RBF Kernel, offers a reliable predictive tool for users seeking to assess their risk of diabetes. Streamlit's interactive features will enable easy integration of the model, allowing users to receive predictions regarding their diabetic status based on their input parameters, fostering proactive health management and informed decision-making. Deploying this model through Streamlit will ensure accessibility, usability, and accuracy, empowering users to make informed health-related choices based on predictive insights.

By deploying this predictive model through Streamlit, we aim to democratize access to accurate health insights, empowering individuals to prioritize their well-being through proactive health management. As we continue to refine and expand the capabilities of our predictive tool, our overarching goal remains to support individuals in making informed health-related decisions and fostering a culture of proactive self-care.

The application now seamlessly incorporates patient details, providing precise predictions on their diabetic status. Users can input their information effortlessly, receiving instant insights into whether they have diabetes or not. With this accessible and accurate tool, individuals can take proactive steps towards managing their health, fostering informed decision-making, and promoting a healthier lifestyle.

As we continue to refine and expand our tool's capabilities, we're committed to democratizing access to accurate health insights and empowering individuals to prioritize their well-being.

# Conclusion and Future Work

In conclusion, the development and evaluation of the diabetes prediction model utilizing Support Vector Machine (SVM) kernels on the PIMA Indian Diabetes dataset have yielded valuable insights. Among the kernels tested, the RBF Kernel emerged as the optimal choice, demonstrating robust predictive performance with a training accuracy of 84.69% and a test accuracy of 79.22%.

This model, poised to be deployed within a Streamlit-based web application, stands as a valuable tool for individuals seeking proactive health management. The integration of Streamlit's user-friendly interface with the RBF Kernel-equipped model offers a user-centric approach, allowing individuals to input their medical details and receive predictions regarding their diabetic status, fostering informed decision-making and proactive health measures.

Moving forward, there are several avenues for further exploration and enhancement of this project:

* Firstly, expanding the dataset to encompass a more diverse demographic would improve the model's generalizability.
* Additionally, fine-tuning the model parameters and exploring ensemble methods might further boost predictive accuracy.
* Incorporating real-time data updates and continuous model learning mechanisms would ensure the model's relevance and adaptability to evolving health trends.
* Furthermore, enhancing the web application's functionalities by integrating features such as personalized health recommendations based on predictions and incorporating additional health parameters could provide a more comprehensive health assessment for users.
* Collaborations with healthcare professionals to validate the model's predictions and ensure alignment with clinical diagnoses would bolster its reliability and applicability in real-world healthcare settings.

Overall, the project lays a solid foundation for predictive healthcare tools, and future endeavors could focus on refinement, scalability, and increased accuracy to better serve individuals in proactive health management.

**References**

1. <https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database> This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases.
2. Vijiya Kumar, K., Lavanya, B., Nirmala, I., Caroline, S.S.: Random forest algorithm for the prediction of diabetes. In: International Conference on System, Computation, Automation and Networking, pp. 1–5 (2019).
3. Mohan, N., Jain, V.: Performance analysis of support vector machine in diabetes prediction. In: *International Conference on Electronics*, Communication and Aerospace Technology, pp. 1–3 (2020)
4. Chatrati, S.P., Hossain, G., Goyal, A., et al.: Smart home health monitoring system for predicting type 2 diabetes and hypertension. *J. King Saud Univ. Comput. Inf. Sci.* 34(3), 862–870 (2020)
5. Jackins, V., Vimal, S., Kaliappan, M., Lee, M.Y.: AI‐based smart prediction of clinical disease using random forest classifier and Naive Bayes. *J. Supercomput.* 77, 5198–5219 (2021).
6. Deberneh, H.M., Kim, I.: Prediction of type 2 diabetes based on a machine learning algorithm. *Int. J. Environ. Res. Public Health* 18, 1–14 (2021)
7. Pranto, B., et al.: Evaluating machine learning methods for predicting diabetes among female patients in Bangladesh. *Information* 11, 1–20 (2020).
8. Nazin Ahmed, Rayhan Ahammed, Md. Manowarul Islam, Md. Ashraf Uddin, Arnisha Akhter, Md. Alamin Talukder, Bikash Kumar, Machine learning based diabetes prediction and development of smart web application, International Journal of Cognitive Computing in Engineering 2 (2021)
9. Firdous, Shimoo1; Wagai, Gowher A.2, Sharma, Kalpana1. A survey on diabetes risk prediction using machine learning approaches. Journal of Family Medicine and Primary Care, November 2022.
10. R Birjais, AK Mourya, Chauhan R, Kaur H Prediction and diagnosis of future diabetes risk - A machine learning approach. SN Appl Sci 2019
11. Sadhu A, Jadli A Early-stage diabetes risk prediction - A comparative analysis of classification algorithms. Int Adv Res J Sci Eng Technol (IARJSET).
12. Xue J, Min F, Ma F Research on diabetes prediction method based on machine learning. J Phys Conf Ser 2020.
13. Agrawal P, Dewangan AK A brief survey on the techniques used for the diagnosis of diabetes-mellitus. Int Res J Eng Tech IRJET 2015.
14. Saravananathan K, Velmurugan T Analyzing diabetic data using classification algorithms in data mining. Indian J Sci Technol 2016
15. Kavakiotis I, Tsave O, Salifoglou A, Maglaveras N, Vlahavas I, Chouvarda I Machine learning and data mining methods in diabetes research. Comput Struct Biotechnol J 2017.
16. Rawat V, Suryakant S A classification system for diabetic patients with machine learning techniques. Int J Math Eng Manag Sci 2019.
17. Rathore A, Chauhan S, Gujral S Detecting and predicting diabetes using supervised learning:An approach towards better healthcare for women. Int J Adv Res Comput Sci 2017
18. Hassan AS, Malaserene I, Leema AA Diabetes mellitus prediction using classification techniques. Int J Innov Technol Explor Eng 2020
19. Nai-Arun N, Moungmai R Comparison of classifiers for the risk of diabetes prediction. Procedia Comput Sci 2015
20. KM Jyoti Rani, Diabetes Prediction Using Machine Learning, Department of CSE, IMG Engineering College, Ghaziabad, Uttar Pradesh, July-August-2020.
21. B. Ljubic, A. A. Hai, M. Stanojevic et al., “Predicting complications of diabetes mellitus using advanced machine learning algorithms,” Journal of the American Medical Informatics Association 2020.

.