PREDICTION OF DIABETES USING MACHINE LEARNING

Dita Mondal[1], Pranab Hazra[2], Megha Mandal[3], Debanjali Dutta[4], Swati Barui[5], Moupali Roy[6]

[1,3 and 4]Student, Electronics and Communication Engineering, Narula Institute of Technology, India

[2,5, and 6]Assistant Professors, Electronics and Communication Engineering, Narula Institute of Technology, India.

Abstract: Diabetes mellitus, a chronic metabolic disorder, has become a major public health concern globally due to its widespread prevalence and associated complications, including cardiovascular diseases, kidney failure, and nerve damage. Early detection and diagnosis of diabetes are crucial to minimizing these risks. This study investigates the application of machine learning (ML) techniques, specifically the Support Vector Machine (SVM), for predicting the likelihood of diabetes in individuals based on a set of health-related attributes. The dataset used in this research includes several clinical features such as age, body mass index (BMI), blood pressure, glucose levels, insulin levels, skin thickness among others.

The primary objective is to develop a robust predictive model that can classify individuals as diabetic or non-diabetic with high accuracy. SVM, known for its ability to handle high-dimensional data, is utilized for classification. The dataset is pre-processed, including normalization and handling of missing values, followed by the training of the SVM model using appropriate kernel functions. Model performance is evaluated through metrics such as accuracy, confusion matrix, and the area under the receiver operating characteristic (ROC) curve. Experimental results show that the SVM classifier achieves high accuracy and is capable of effectively distinguishing between diabetic and non-diabetic cases, demonstrating the feasibility of ML in healthcare applications. The study emphasizes the potential of SVM in providing timely and accurate predictions, thus enabling healthcare professionals to identify at-risk individuals early. This can lead to more effective interventions and personalized treatment plans, contributing to the global efforts in combating diabetes. The findings highlight the growing importance of artificial intelligence in the healthcare sector, offering innovative solutions to improve patient care and disease management.

Introduction

Diabetes is a well-known condition globally and poses significant challenges for both developed and developing nations. The pancreas produces the hormone insulin, which facilitates the movement of glucose from food into the bloodstream. A deficiency of this hormone, often due to pancreatic dysfunction, leads to diabetes[1,7]. This condition can cause severe complications, including coma, kidney and retinal failure, damage to pancreatic beta cells, cardiovascular and cerebral vascular issues, peripheral vascular diseases, sexual dysfunction, joint problems, weight loss, ulcers, and weakened immunity[5]. Studies on diabetes patients reveal a concerning trend: the prevalence of diabetes among adults aged 18 and above increased from 4.7% in 1980 to 8.5% in 2014, with rapid growth in developing regions. By 2017, approximately 451 million individuals were diagnosed with diabetes worldwide, a figure projected to rise to 693 million by 2045[4]. Another analysis highlights the gravity of the situation, reporting that over half a billion people live with diabetes, with this number expected to grow by 25% by 2030 and 51% by 2045[2]. While diabetes has no permanent cure, it can be effectively managed and even prevented if accurate early detection is achieved.

Diabetes has become one of the leading health concerns worldwide, with its prevalence rising steadily[19]. Early detection of diabetes is crucial for effective management and prevention of severe complications. In this project, we aimed to develop a system that predicts diabetes at an early stage, enabling individuals to take preventative actions and seek timely medical attention.

We used the Pima Indian Diabetes Dataset, which includes a variety of attributes such as age, BMI, insulin levels, blood pressure, and glucose concentration, all of which are significant factors in predicting diabetes[1]. The first step was to carefully analyse the data, exploring the relationships between the features and the likelihood of diabetes. This analysis helped us identify key patterns and correlations in the dataset[12]. We also examined the data for missing values and inconsistencies, ensuring the dataset was clean and ready for modelling. After cleaning the data, we proceeded with data preprocessing to prepare it for Machine Learning.

This involved handling missing values by using imputation techniques, normalizing numerical attributes such as BMI and glucose levels, and encoding categorical variables to make them suitable for the models.Data preprocessing is an essential step in ensuring that the Machine Learning algorithms can learn efficiently from the data[6,11]. Next, we applied various Machine Learning models, including classification algorithms like Support Vector Machines [5]. These models were trained on the pre-processed dataset, and we evaluated their performance using metrics like accuracy, precision, recall, and Z-score.

The final step involved testing the models on a dataset to ensure robust performance and reliability. After a thorough evaluation, we identified the most accurate approach for predicting diabetes[12]. The goal of this project is to provide an early detection system that can assist individuals in recognizing the early signs of diabetes, making it easier for them to take preventative measures and seek medical advice.

While doctors can easily predict diabetes through simple tests in a clinical setting, many people may not have access to such services, either due to geographical or financial constraints. Our aim is to bridge this gap by creating an accessible, low-cost solution for early prediction[14]. By simply entering basic data into the system, anyone can easily get results and understand their risk of developing diabetes, helping to promote early intervention and proactive health management [17].

Type of Diabetes

Diabetes is a chronic condition that affects the body's ability to regulate blood sugar (glucose) levels. There are several types of diabetes, each with distinct characteristics:

Type 1 Diabetes (TD1):

Description: TD1 is a chronic autoimmune disorder where the body's immune system mistakenly attacks the insulin-producing beta cells in the pancreas[13]. This results in little to no insulin production, making it difficult for the body to regulate blood sugar levels effectively[18].

Cause: Exact cause is unknown, but it's believed to involve genetic and environmental factors.

Autoimmune Reaction:

The immune system targets and destroys beta cells in the pancreas, mistaking them as harmful.

This loss of beta cells leads to a complete or near-complete lack of insulin production

Genetic Factors:

Certain genes, such as those in the HLA (human leukocyte antigen) region, increase the risk.

A family history of Type 1 Diabetes raises susceptibility.

Environmental Triggers:

Viral infections (e.g., enteroviruses or rubella) might trigger the autoimmune response in genetically predisposed individuals.

Symptoms:

The symptoms of Type 1 Diabetes often appear suddenly and may include:

Excessive Thirst (Polydipsia): Caused by high blood sugar levels pulling fluid from tissues.

Frequent Urination (Polyuria): The kidneys work to eliminate excess sugar, leading to dehydration.

Extreme Hunger (Polyphagia): Due to the lack of glucose entering cells for energy.

Unexplained Weight Loss: The body starts breaking down fat and muscle for energy in the absence of insulin[12].

Fatigue: Caused by insufficient energy supply to cells.

Blurred Vision: High blood sugar affects the eye's lenses, causing swelling and vision problems.

Slow Healing: Cuts and infections may take longer to heal.

Type 1 Diabetes is diagnosed using Blood Glucose Tests, HbA1c Test, C-Peptide Test and Autoantibody Tests.

Type 2 Diabetes (TD2)

Description: TD2 is a chronic metabolic disorder characterized by the body's inability to effectively use insulin or produce enough insulin to maintain normal blood sugar levels[19]. It is the most common form of diabetes, affecting millions of people worldwide.

Cause: The reason behind TD2 is

Insulin Resistance:

In Type 2 Diabetes, the body’s cells become less responsive to insulin, a hormone that helps glucose enter cells for energy[6].

Over time, the pancreas compensates by producing more insulin, but eventually, it cannot keep up with the demand.

Genetic Factors:
A family history of diabetes increases the likelihood of developing the condition.
Certain genetic variations may predispose individuals to insulin resistance or beta cell dysfunction[19].
Lifestyle Factors:
Obesity: Excess body fat, especially around the abdomen, contributes significantly to insulin resistance.
Physical Inactivity: Lack of exercise reduces insulin sensitivity in muscle cells[12].
Unhealthy Diet: Diets high in refined sugars, processed foods, and unhealthy fats can lead to weight gain and insulin resistance.

Symptoms:
Type 2 Diabetes develops gradually, and symptoms may be mild or go unnoticed for years. Common symptoms include[11]:
Increased thirst (polydipsia).
Frequent urination (polyuria).
Unexplained weight loss.
Fatigue or weakness.
Blurred vision.
Slow-healing wounds or frequent infections.
Darkened skin patches (acanthosis nigricans), often around the neck or armpits.

Type 2 Diabetes is diagnosed using Fasting Blood Sugar Test, Oral Glucose Tolerance Test (OGTT), HbA1c Test and Random Blood Sugar Test[6].

Literature Review

K. VijiyaKumar in 2019 introduced a system leveraging the Random Forest algorithm to predict diabetes. Their approach aims to enable early diagnosis of diabetes in patients with improved accuracy through the application of this machine learning technique. The developed model demonstrated high efficiency and effectiveness in forecasting diabetes, delivering results quickly while ensuring reliable predictions [21].
Nonso Nnamoko in 2018 proposed a method for predicting the onset of diabetes using an ensemble of supervised learning algorithms. In their approach, they employed five commonly used classifiers, with a meta-classifier designed to aggregate the outputs of these individual models. The performance of their proposed method was evaluated and compared with similar studies that utilized the same dataset, demonstrating improved prediction accuracy and highlighting the effectiveness of ensemble techniques in diabetes onset prediction [20].
Deeraj Shetty in 2017 proposed an intelligent diabetes disease prediction system using data mining techniques. Their system analyses diabetes-related data by applying algorithms such as Bayesian Networks and K-Nearest Neighbors (KNN) to patient databases. By examining various attributes associated with diabetes, the system provides a comprehensive prediction model to identify individuals at risk of developing diabetes [12].
Several comparative studies have evaluated multiple ML algorithms, including Random Forest, SVM, Logistic Regression, to determine the best-performing models. XGBoost often emerges as a top performer due to its ensemble nature and effective handling of imbalanced datasets. Random Forest is noted for its resistance to overfitting, while simpler models like Logistic Regression are preferred for rapid implementation and ease of use.
Building on the insights and methods explored in previous research, our work aims to refine and enhance diabetes prediction models by integrating various machine learning techniques. Through this studies which utilized supervised learning methods such as SVM, Logistic Regression, ANN, and data mining algorithms like Bayesian Networks and K-Nearest Neighbors (KNN), have highlighted the potential of these techniques in accurately diagnosing diabetes.
In our approach, we drew inspiration from these existing methods to understand the strengths and limitations of different algorithms in diabetes prediction. By analysing the effectiveness of early detection techniques used in prior studies, we sought to combine the most promising aspects of these models and apply them to our own work. Our goal is to create a more robust, accurate, and efficient system that can build on the knowledge gained from previous research and adapt it to address existing gaps in diabetes detection.

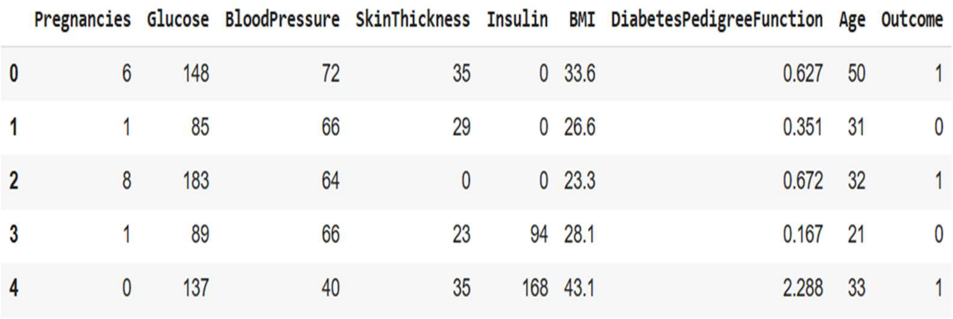
Methodology

Diabetes detection using machine learning involves a series of steps that are designed to process the available data, train a predictive model, and evaluate its performance. The objective is to develop a machine learning model that can classify individuals into diabetic or non-diabetic categories based on their medical data[10]. Below is a detailed methodology for how to approach this task:

Description of Data:

The dataset used for detection of diabetes typically contains various health-related attributes that are indicative of an individual's likelihood of having diabetes. In this case we will describe the PIMA Indian Diabetes Dataset, which is commonly used for this purpose. The PIMA Indian Diabetes Dataset is a widely used dataset in machine learning for binary classification tasks, specifically for predicting of diabetes. The dataset contains 768 cases with 9 attributes including target or outcome column. The objective is to predict based on the measures to predict if the patient is diabetic or not.

Table 1:



	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

➔The diabetes data set consists of 768 data points, with 9 features each.

➔ "Outcome" is the feature we are going to predict, 0 means No diabetes, 1 means diabetes.

Pregnancies: With multiple pregnancies, the levels of pregnancy-related hormones (like human placental lactogen and cortisol) are even higher, leading to greater insulin resistance. And women with multiple pregnancies are more likely to develop GDM compared to those with a single pregnancy.

Glucose: Glucose levels are crucial in diabetes management. Insulin, a hormone, helps cells absorb glucose for energy. In diabetes, insulin production is inadequate, or cells resist it, leading to high blood sugar (hyperglycemia), which can harm organs, nerves, and blood vessels. Low blood sugar (hypoglycemia) can cause immediate symptoms like confusion or fainting. Maintaining glucose within healthy ranges through diet, exercise, and medications is vital to avoiding complications and ensuring overall well-being.

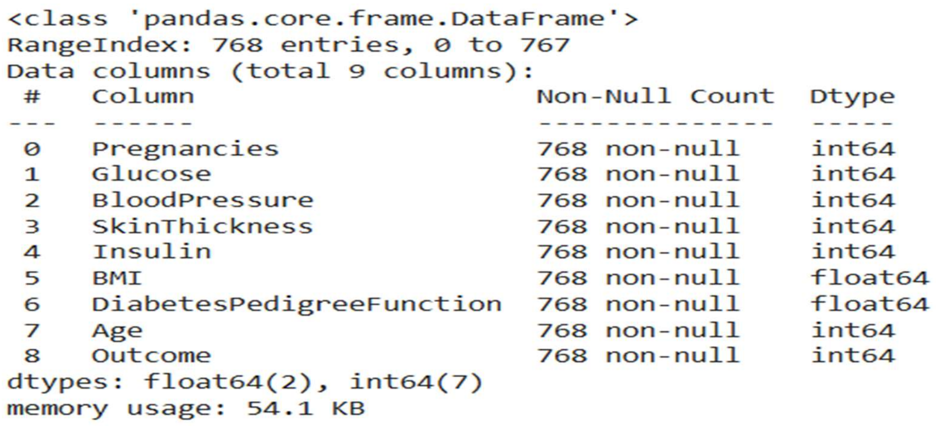
Blood Pressure: Blood pressure and diabetes are closely linked, as both can damage blood vessels and increase the risk of heart and kidney diseases. High blood sugar levels in diabetes can cause vascular damage, leading to hypertension. Similarly, high blood pressure worsens complications of diabetes by straining the heart and kidneys. Managing both is crucial to preventing severe health problems.

Skin Thickness: In diabetes, increased skin thickness can occur due to excess sugar affecting collagen and connective tissues. Conditions like diabetic dermopathy or scleroderma-like changes cause thickened, tight, or waxy skin, especially on fingers and hands. This reflects poor blood sugar control and vascular damage.

BMI: Body Mass Index (BMI) is closely linked to diabetes, as a higher BMI, particularly in the overweight or obese range, increases the risk of insulin resistance and type 2 diabetes. Excess body fat, especially around the abdomen, disrupts glucose metabolism, leading to elevated blood sugar levels.

Age: Age is a significant factor in the development of diabetes, as the risk increases with age, particularly after 45. As people age, insulin sensitivity often decreases, and the pancreas may produce less insulin, leading to a higher likelihood of developing type 2 diabetes.

Insulin: Insulin is a hormone that helps regulate blood sugar by allowing cells to absorb glucose for energy. In diabetes, either the body doesn't produce enough insulin (type 1) or the cells become resistant to it (type 2), leading to high blood sugar levels. Managing insulin is key to controlling diabetes and preventing complications.

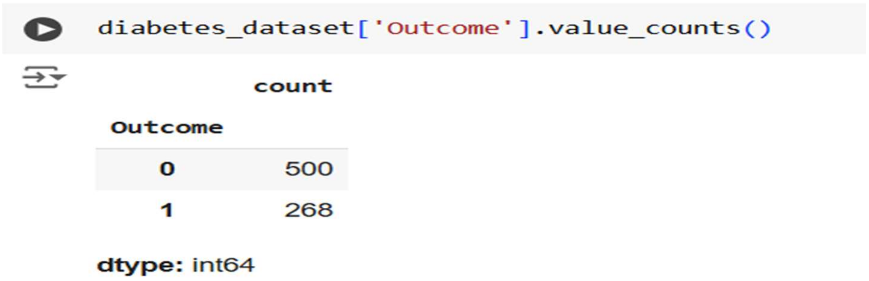


```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   Pregnancies               768 non-null    int64  
 1   Glucose                   768 non-null    int64  
 2   BloodPressure             768 non-null    int64  
 3   SkinThickness             768 non-null    int64  
 4   Insulin                   768 non-null    int64  
 5   BMI                       768 non-null    float64
 6   DiabetesPedigreeFunction  768 non-null    float64
 7   Age                       768 non-null    int64  
 8   Outcome                   768 non-null    int64  
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

Figure 1: Display the No-Null Count

➔ There is no null values in dataset

Distribution of Diabetic patient: In this case it is mentioned above that the dataset contains 768 samples and among them 500 were designated as 0, denoting the nonexistence of diabetes, while 268 were designated as 1, denoting the existence of diabetes.



```
diabetes_dataset['Outcome'].value_counts()
```

```
         count
Outcome
   0      500
   1      268

dtype: int64
```

Figure 2: Diabetic Patient Distribution

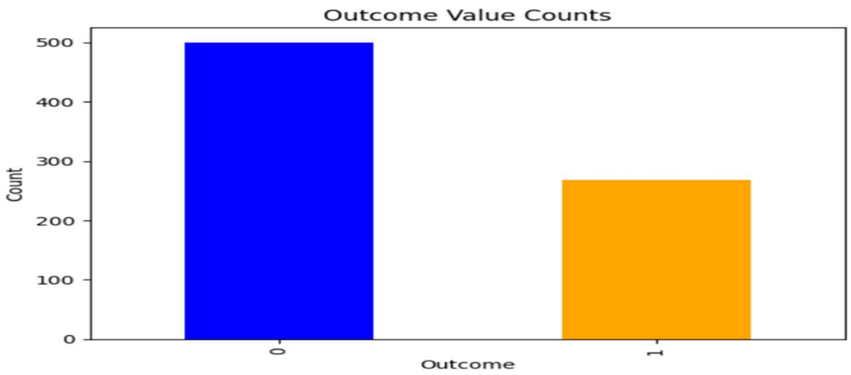


Figure 3: Diabetic Patient Distribution Bar Graph

Support Vector Machine (SVM):

Choroidal Support Vector Machines (SVM) are versatile and powerful machine learning algorithms that are widely used for tasks such as classification, regression, and outlier detection. SVMs can efficiently handle both linear and nonlinear problems, making them suitable for a variety of real-world applications, including text classification, image recognition, spam filtering, handwriting analysis, gene expression studies, face detection, and anomaly detection. The strength of SVM lies in its ability to identify the optimal decision boundary, known as the maximum margin hyperplane, which separates data points from different classes. By maximizing the margin between the data classes, SVMs demonstrate high performance in both binary and multiclass classification tasks. Additionally, SVMs can be extended to nonlinear scenarios using kernel functions, which map the data into higher-dimensional spaces where it becomes linearly separable.

The core goal of the SVM algorithm is to determine the best possible hyperplane in an N-dimensional space that can distinctly separate data points belonging to different classes. It achieves this by maximizing the margin, which is the distance between the hyperplane and the nearest data points from each class, referred to as support vectors. The nature and dimension of the hyperplane are determined by the number of input features. For example, with two features, the hyperplane is a straight line, whereas with three features, it becomes a two-dimensional plane. When the number of features exceeds three, the hyperplane exists in higher dimensions, making it challenging to visualize but still mathematically well-defined.

How SVM Works:

Hyperplane: SVM aims to find the best dividing line (or hyperplane in higher dimensions) that maximizes the margin between the closest points of two classes (called support vectors).

Support Vectors: These are the data points that lie closest to the hyperplane. The margin is calculated using these points.

Margin: The distance between the hyperplane and the closest points (support vectors). SVM tries to maximize this margin.

Kernel Trick: When data is not linearly separable, SVM uses a kernel function to transform data into a higher-dimensional space where it becomes separable.

In a situation with two independent variables, x1 and x2 and a dependent variable that belongs to one of two classes (e.g., represented as red and blue circles), the decision boundary between the classes is a straight line because the data exists in a two-dimensional space.

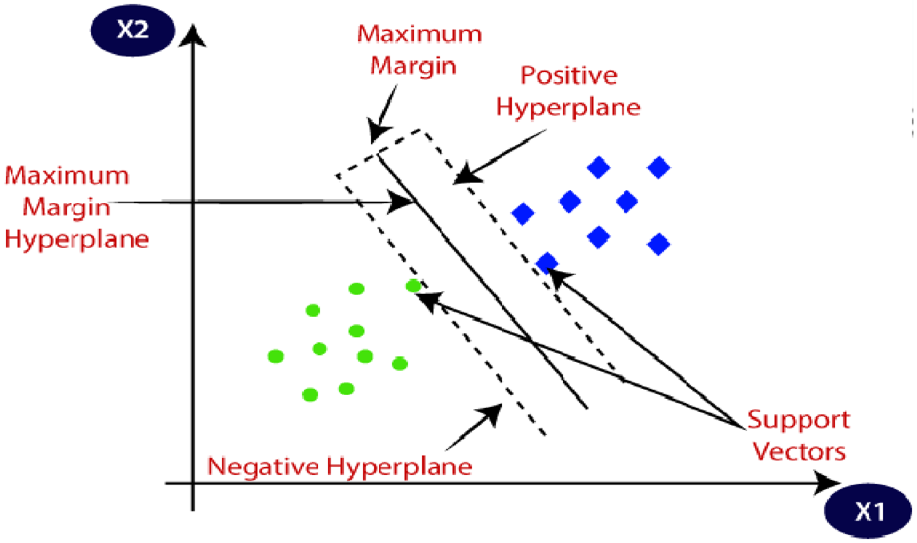


Figure 4: Support Vector Machine

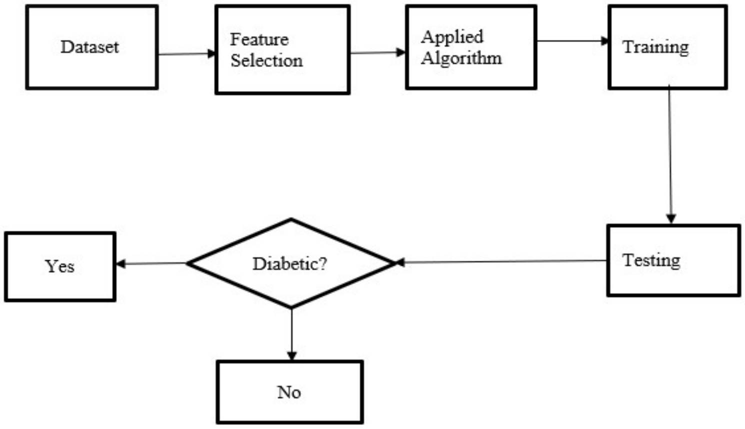
While many potential lines (or hyperplanes) can separate the two groups of points, the goal is to identify the optimal hyperplane. This is the line that achieves the maximum margin, which refers to the greatest distance between the hyperplane and the nearest data points from each class. These closest points are known as support vectors and are crucial in defining the optimal boundary.

The process of maximizing this margin ensures that the separation is as clear as possible, which helps improve the model's ability to classify new, unseen data accurately. This is the key principle behind Support Vector Machines (SVM) in machine learning.

Advantages of SVM
Effective in high-dimensional spaces.

Works well with a small to medium-sized dataset.
Robust to overfitting using regularization.
Can handle non-linear data using kernel functions.
Kernel in Support Vector Machines (SVM)
In Support Vector Machines (SVM), a kernel is a mathematical function that transforms data into a higher-dimensional space, enabling SVM to solve problems where the data is non-linearly separable. Instead of working with the original input space, kernels allow SVM to operate in this transformed space to find an optimal hyperplane that separates the classes effectively.
The kernel trick is the core idea behind SVM that avoids explicitly computing the transformation, which can be computationally expensive. Instead, it uses a kernel function to compute the dot product of transformed features directly.
Why is a Kernel Needed
In some datasets, the classes cannot be separated using a straight line (linear hyperplane) in the original feature space.
Kernels map the input features into a higher-dimensional space where the classes become linearly separable.
This allows SVM to classify data with complex decision boundaries efficiently.

Workflow:



Result:

In this study, we used the Support Vector Machine (SVM) model to detect diabetes based on a dataset containing features such as age, BMI, insulin levels, and blood pressure, glucose and pregnancies. The SVM classifier was trained and tested using a 80-20 train-test split.
The model achieved 78% training accuracy and 77% testing accuracy, that indicating an overall balance between detecting positive cases (diabetes) and avoiding false positives.
The confusion matrix further demonstrates the performance, showing true positives, true negatives, false positives, and false negatives.
Table 2:

Support Vector Machine (SVM)	**Model accuracy**
Training	0.78664
Testing	0.77272

Confusion Matrix

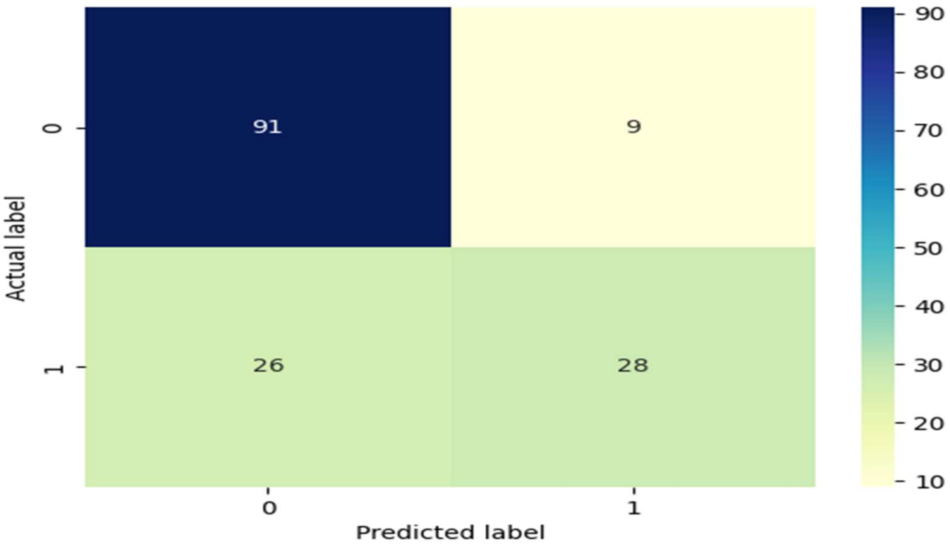


Figure 5: Confusion Matrix

The confusion matrix for the diabetes detection project using the Support Vector Machine (SVM) model is a tool to evaluate the performance of the model by comparing the predicted results with the actual outcomes. It helps in identifying the true positives, true negatives, false positives, and false negatives.
Here is a general structure for the confusion matrix:
Table 3:

	Predicted: Positive (Diabetic)	Predicted: Negative(Non-Diabetic)
Actual: Positive (Diabetic)	True Positive (TP)	False Negative (FN)
Actual:Negetive(Non-Diabetic	False Positive (FP)	True Negative (TN)

True Positive (TP): The number of diabetic patients correctly identified by the model.
False Positive (FP): The number of non-diabetic individuals incorrectly classified as diabetic.
False Negative (FN): The number of diabetic individuals incorrectly classified as non-diabetic.
True Negative (TN): The number of non-diabetic individuals correctly classified as non-diabetic.
For our model the confusion matrix description is as followed:

Table 4:

	Predicted: Positive (Diabetic)	Predicted: Negative (Non-Diabetic)
Actual: Positive (Diabetic)	91 (TP)	9 (FN)
Actual: Negative (Non-Diabetic	26(FP)	28 (TN)

From this matrix, we can calculate the following performance metrics:
Accuracy = (TP + TN) / (TP + TN + FP + FN)
Precision = TP / (TP + FP)
Recall (Sensitivity) = TP / (TP + FN)
This confusion matrix will help assess the effectiveness of the SVM model in correctly identifying diabetic and non-diabetic individuals.

ROC Curve:
The Receiver Operating Characteristic (ROC) curve (Figure 7) showed an Area Under the Curve (AUC), indicating a high ability to distinguish between diabetic and non-diabetic cases.

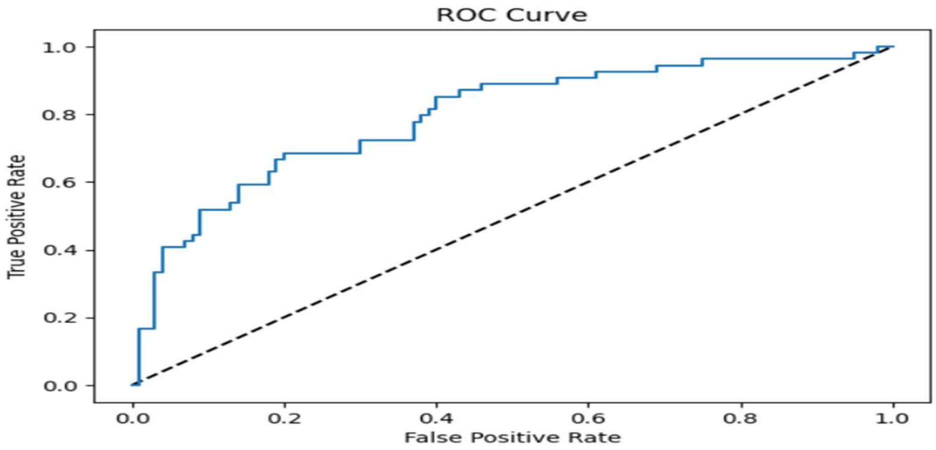
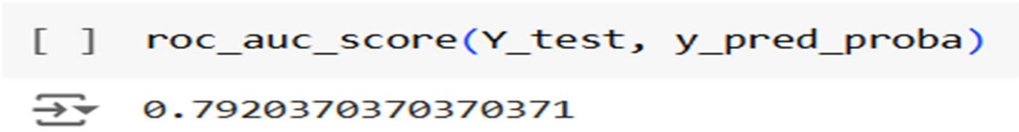


Figure 6: ROC Curve



```
[ ]  roc_auc_score(Y_test, y_pred_proba)

     0.7920370370370371
```

The X-axis shows the False Positive Rate, indicating the proportion of negatives misclassified as positives. They-axis shows the True Positive Rate, which measures the proportion of positives correctly identified. A diagonal line ([0, 0] to [1, 1]) represents random guessing. A model performing along this line is no better than chance. The ROC curve demonstrates the trade-off between sensitivity (True Positive Rate) and specificity (1 - False Positive Rate) as the decision threshold changes.

Complication We Faced: While working on this diabetes detection project using SVM, several challenges were encountered. Managing data was a major issue. The accuracy of train data is average in this case and we are trying to modify the dataset to resolve this issue in future. Although the test accuracy is good. Finally, preparing the model form deployment in a web application required careful optimization for scalability and integration. Future Scope: In the future, we have planned to expand this project into a fully functional web application that allows users to assess their risk for diabetes through an interactive interface. The web app can incorporate the following features:
User Input Interface.
Real-Time Prediction.
Continuous Learning.
Data Security and Privacy.
Mobile Compatibility.

Conclusion

The use of Support Vector Machines (SVM) for diabetes detection demonstrates the transformative potential of machine learning in modern healthcare. SVM is particularly well-suited for this application due to its ability to handle complex and multidimensional datasets, making it effective in distinguishing between diabetic and non-diabetic individuals even when the data exhibits non-linear patterns. By implementing feature scaling and carefully selecting relevant features, the model achieves high accuracy and reliability, proving its value in early detection and prevention efforts.

This project highlights the significant role of early diagnosis in managing diabetes, which is crucial in mitigating its long-term complications. The SVM-based approach is not only efficient and scalable but also adaptable to different datasets and healthcare scenarios. Its robust classification performance underscores the importance of leveraging advanced computational tools in addressing the global health challenge posed by diabetes.

Furthermore, the findings pave the way for future enhancements, such as incorporating additional clinical and lifestyle data or integrating other machine learning techniques to build hybrid models. Such advancements could improve both the precision and interpretability of predictions, making the solution more practical for deployment in diverse healthcare environments. Overall, this study reaffirms the potential of SVM as a powerful tool for predictive analytics, offering a proactive solution to aid healthcare professionals in combating diabetes effectively.

Reference

1. Mujumdar, A., & Vaidehi, V. (2019). Diabetes prediction using machine learning algorithms. Procedia Computer Science, 165, 292-299.
2. Rani, K. J. (2020). Diabetes prediction using machine learning. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 6, 294-305.
3. Saru, S., & Subashree, S. (2019). Analysis and prediction of diabetes using machine learning. International journal of emerging technology and innovative engineering, 5(4).
4. Alehegn, M., Joshi, R., & Mulay, P. (2018). Analysis and prediction of diabetes mellitus using machine learning algorithm. International Journal of Pure and Applied Mathematics, 118(9), 871-878.
5. Kumari, V. A., & Chitra, R. (2013). Classification of diabetes disease using support vector machine. International Journal of Engineering Research and Applications, 3(2), 1797-1801.
6. Tigga, N. P., & Garg, S. (2020). Prediction of type 2 diabetes using machine learning classification methods. Procedia Computer Science, 167, 706-716.
7. Khanam, J. J., & Foo, S. Y. (2021). A comparison of machine learning algorithms for diabetes prediction. Ict Express, 7(4), 432-439.
8. Indoria, P., & Rathore, Y. K. (2018). A survey: detection and prediction of diabetes using machine learning techniques. International Journal of Engineering Research & Technology (IJERT), 7(3), 287-291.
9. Bhat, S. S., Selvam, V., Ansari, G. A., Ansari, M. D., & Rahman, M. H. (2022). Prevalence and early prediction of diabetes using machine learning in North Kashmir: a case study of district bandipora. Computational Intelligence and Neuroscience, 2022(1), 2789760.
10. Collins, G. S., Mallett, S., Omar, O., & Yu, L. M. (2011). Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. BMC medicine, 9, 1-14.
11. Ayon, S. I., & Islam, M. M. (2019). Diabetes prediction: a deep learning approach. International Journal of Information Engineering and Electronic Business, 13(2), 21.
12. Deeraj Shetty, Kishor Rit, Sohail Shaikh, Nikita Patil, "Diabetes Disease PredictionUsing Data Mining ".International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 2017.
13. Panda, M., Mishra, D. P., Patro, S. M., & Salkuti, S. R. (2022). Prediction of diabetes disease using machine learning algorithms. IAES International Journal of Artificial Intelligence, 11(1), 284.
14. Soni, M., & Varma, S. (2020). Diabetes prediction using machine learning techniques. International Journal of Engineering Research & Technology (IJERT), 9(09), 2278-0181.
15. Kaur, H., & Kumari, V. (2022). Predictive modelling and analytics for diabetes using a machine learning approach. Applied computing and informatics, 18(1/2), 90-100.
16. Deberneh, H. M., & Kim, I. (2021). Prediction of type 2 diabetes based on machine learning algorithm. International journal of environmental research and public health, 18(6), 3317.
17. Yu, W., Liu, T., Valdez, R., Gwinn, M., & Khoury, M. J. (2010). Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. BMC medical informatics and decision making, 10, 1-7.
18. Katsarou, A., Gudbjörnsdottir, S., Rawshani, A., Dabelea, D., Bonifacio, E., Anderson, B. J., ... & Lernmark, Å. (2017). Type 1 diabetes mellitus. Nature reviews Disease primers, 3(1), 1-17.
19. Chatterjee, S., Khunti, K., & Davies, M. J. (2017). Type 2 diabetes. The lancet, 389(10085), 2239-2251.
20. Nnamoko, N., Hussain, A., & England, D. (2018, July). Predicting diabetes onset: an ensemble supervised

learning approach. In 2018 IEEE Congress on evoluti-onary computation (CEC) (pp. 1-7). IEEE

21. VijiyaKumar, K., Lavanya, B., Nirmala, I., & Caroline, S. S. (2019, March). Random forest algorithm for the prediction of diabetes. In 2019 IEEE international conference system, computation, automation and networking (ICSCAN) (pp. 1-5). IEEE.
22. Shetty, D., Rit, K., Shaikh, S., & Patil, N. (2017, March). Diabetes disease prediction using data mining. In 2017 international conference on innovations in information,embedded and communication systems (ICIIECS) (pp. 1-5). IEEE.