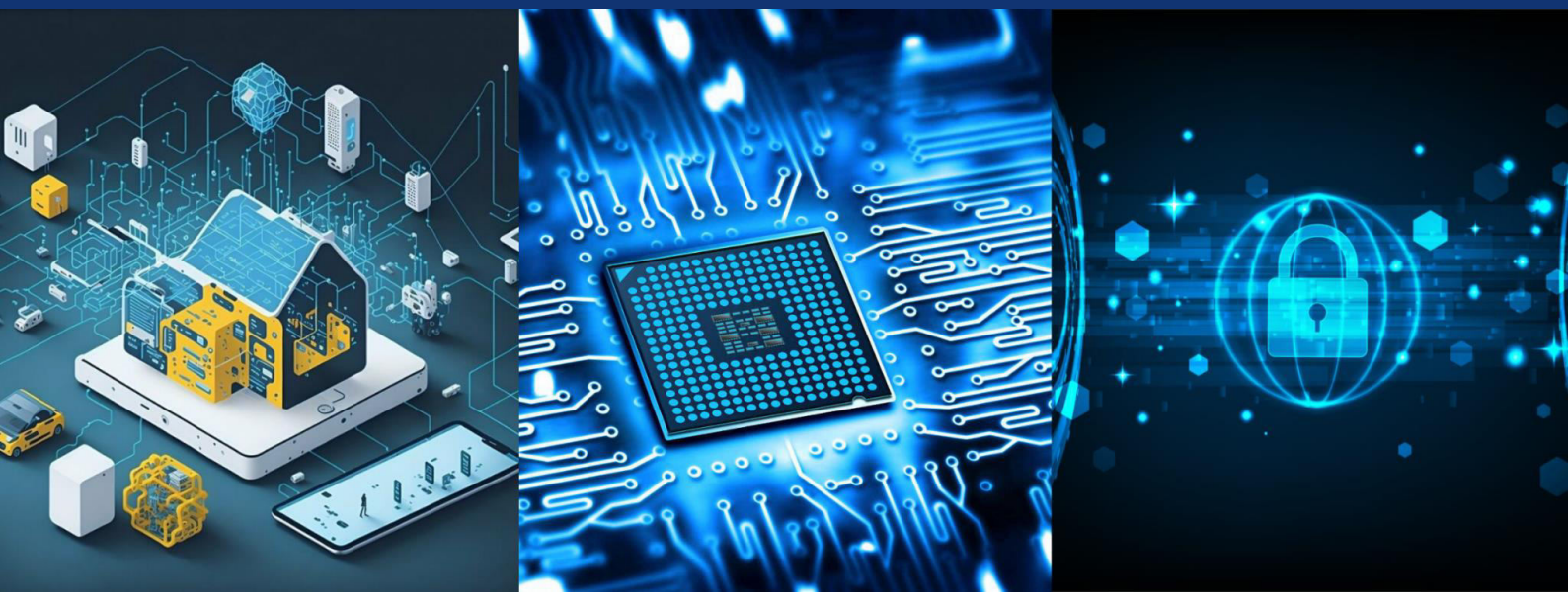
IJIRCCE

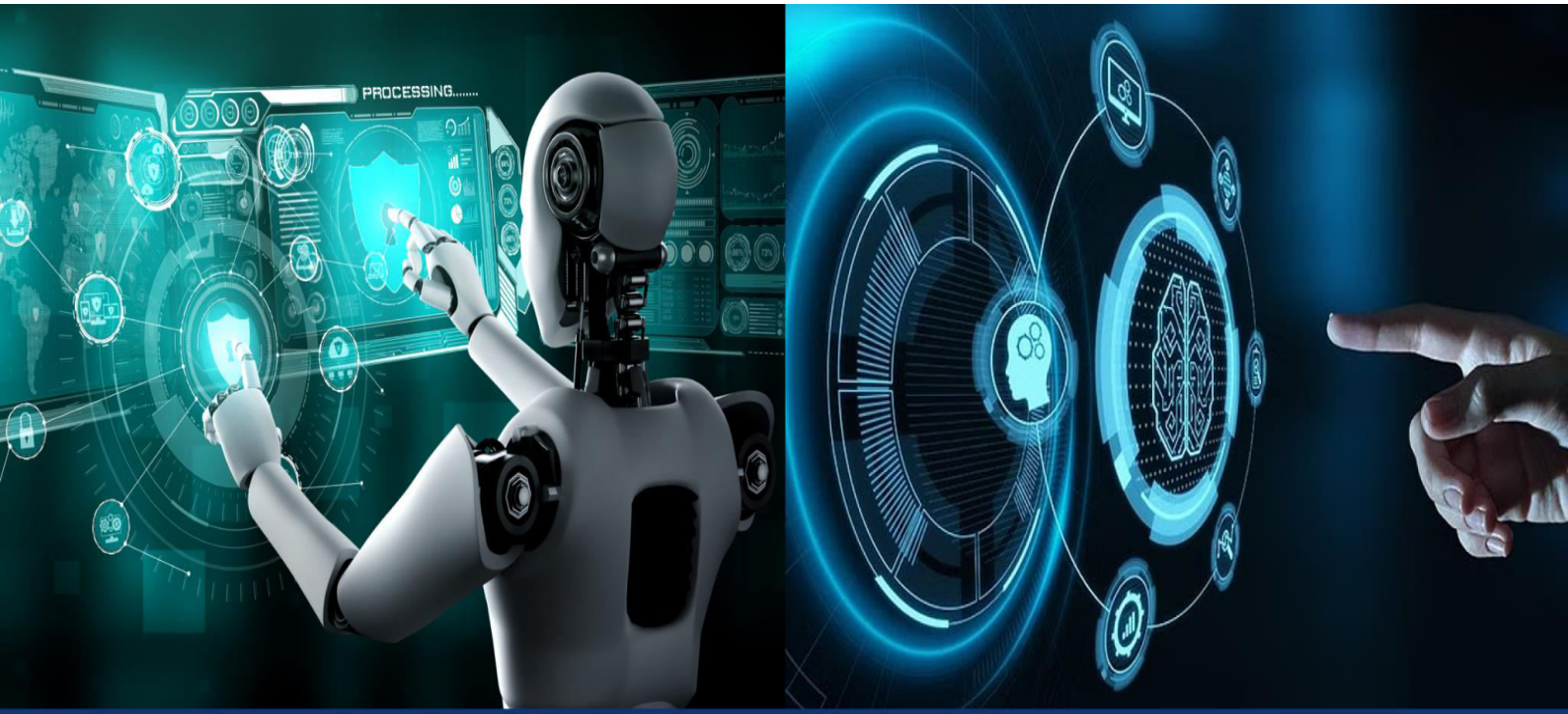
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 13, Issue 4, April 2025

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304063

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Improving Diabetes Risk Prediction with Ensemble Machine Learning Techniques

Meiyalakan.K [1], Rajadurai P [2], Dharaneedharan V [3], Nirmalraj P [4], Sridhar T[5]

Assistant Professor, Department of Computer Science and Engineering, Mahendra Institute of Technology, Mallasamudram, Namakkal, Tamil Nadu, India[1]

Student, Department of Computer Science and Engineering, Mahendra Institute of Technology, Mallasamudram, Namakkal, Tamil Nadu, India[2,3,4,5]

ABSTRACT: One of the most expensive long-term conditions is diabetes, a metabolic disease characterized by elevated blood sugar levels brought on by the body's insufficient production of insulin. Diabetes can also lead to long-term damage to the heart, blood vessels, eyes, kidneys, and nerves. The risk of a heart attack or stroke is doubled for adults with diabetes compared to those without the disease. There is no cure for diabetes, despite the fact that it affects a huge portion of the world's population. Diabetics have a number of long-term health issues, even if the majority of drugs help them manage their symptoms to some degree. Therefore, diabetes can be controlled with machine learning approaches if it can be predicted early. Our goal is to use machine learning techniques and the ensemble approach to determine whether a patient has diabetes. Four algorithms SVM, KNN, Logistic Regression, and Random Forest classifier will be used. We will analyse each model to see which one has the highest accuracy and connect our top model to a web application that can determine the patient's risk of developing diabetes.

KEYWORDS: Random Forest, SVM, Machine Learning, KNN, and Logistic Regression

I. INTRODUCTION

Diabetes is a condition that is rapidly spreading, especially among young individuals. We must comprehend what occurs in the body in the absence of diabetes in order to comprehend the disease and how it develops. Our diets contain sugar (glucose), particularly those high in carbohydrates. Everybody, including those with diabetes, needs carbohydrates because they are the body's primary energy source. Bread, cereal, pasta, rice, fruit, dairy products, and vegetables—particularly starchy vegetables—are examples of foods high in carbohydrates. During digestion, these foods are transformed into glucose. The glucose is then circulated in the bloodstream to serve as a primary energy source for the body. A portion of the glucose is transported to the brain to support clear thinking and cognitive function. The remaining glucose is transported to our cells. The rest of the glucose is transported to our body's cells for energy and to our liver, where it is stored for future use by the body. Insulin is necessary for the body to use glucose as fuel. Pancreatic beta cells are responsible for the production and release of insulin, a hormone essential for glucose regulation. Insulin functions similarly to a door key. To enable glucose to pass from the bloodstream via the cell's doors and into the cell, insulin binds itself to them. Glucose accumulates in the bloodstream (hyperglycaemia)and diabetes occurs if the pancreas is unable to generate enough insulin (insulin shortage) or if the body is unable to use the insulin it produces (insulin resistance). High blood and urine sugar (glucose) levels are indicative of diabetes mellitus.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304063

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

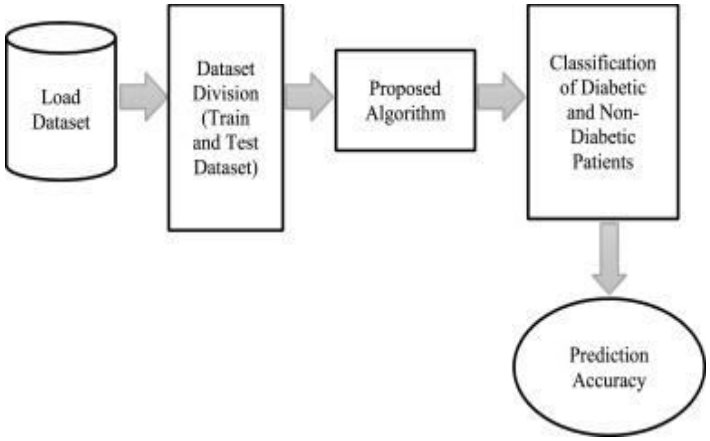


Figure 1.1: A comparative analysis on diagnosis of diabetes mellitus using different approaches

Diabetes is a condition that first manifests in pregnant women who suddenly develop high blood sugar. In two thirds of cases, it will reappear during subsequent pregnancies. type 1 or type 2 diabetes is most likely to develop after a pregnancy affected by gestational diabetes. Genetic factors are the main cause of diabetes, which is caused by at least two mutant genes in chromosome 6, the chromosome that affects the body's response to various antigens. Type 1 and type 2 diabetes may also be influenced by viral infection. Rubella, Coxsackievirus, mumps, hepatitis B virus, and cytomegalovirus infections have been linked to an increased risk of diabetes, according to studies. A class of metabolic disorders known as diabetes is characterized by hyperglycemia brought on by deficiencies in either the action or secretion of insulin, or both. Chronic hyperglycemia in diabetes is associated with long-term harm, malfunction, and failure of many organs, especially the kidneys, heart, blood vessels, nerves, and eyes [2]. It is therefore regarded as one of the worst illnesses in medical history. In the meantime, machine learning is revolutionizing every part of our lives, including the healthcare sector's inability to treat incurable diseases [1]. It improves a wider viewpoint to identify a solution with the aid of multiple medical records, making it more effective to anticipate any disease beforehand and facilitating the adoption of preventive measures. The application of machine learning could significantly increase the accessibility of diabetic treatment.

Figure 1.2: Diabetes prediction using SVM

Early-stage diabetes typically exhibits no symptoms, and it is generally recommended to get tested as soon as symptoms appear because it will be too late to treat diabetes if testing is delayed [3]. Early diabetes care can be revolutionized with the use of machine learning, which offers tools for making predictions using data visualization and analysing the prediction's accuracy score using the Support Vector Machine algorithm. These computational techniques have been used for real-time detection models in practically every field since the introduction of artificial intelligence and associated technology. Studying emerging strategies that can greatly improve existing ones has become more easier because to the usage of data mining, machine learning, deep learning, and computer vision. The algorithms and techniques are reviewed in the next section.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304063

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. METHODS

Data pre-processing is the most important step in the process. The efficacy of data can be diminished by the presence of missing values and other impurities in the majority of healthcare-related data. From vast volumes of diabetes-related data, data mining can uncover hidden insights. It uses a variety of techniques to analyze vast amounts of data in order to uncover knowledge that is buried. It involves methods at the nexus of artificial intelligence, machine learning, insights, and database systems and is the exacting process of finding instances in large data sets. To extract a data mining model from data, a set of calculations and heuristics are employed.

In order to transform information obtained from research papers, medical reports, flow charts, and evidence tables into useful information for decision-making, medical data mining is utilized in knowledge collecting and analysis. Analyzing and uncovering hidden patterns in vast amounts of data to draw conclusions is known as data analytics. The healthcare sector benefits from big data analytics. In the healthcare sector, databases are enormous. In order to learn from the data and predict outcomes appropriately, one can use big data analytics to explore enormous databases and find hidden information and trends. The process of identifying and eliminating erroneous or corrupt records from a table, data set, or record set is known as data cleaning. It involves replacing, altering, or eliminating the coarse or dirty data in order to distinguish between knowledge that is incomplete, inaccurate, or tangential.

Data preprocessing is carried out to enhance the quality and efficacy of the results obtained following the mining process and data cleaning. This procedure is crucial for good prediction and accurate outcomes when applying machine learning techniques to the dataset.

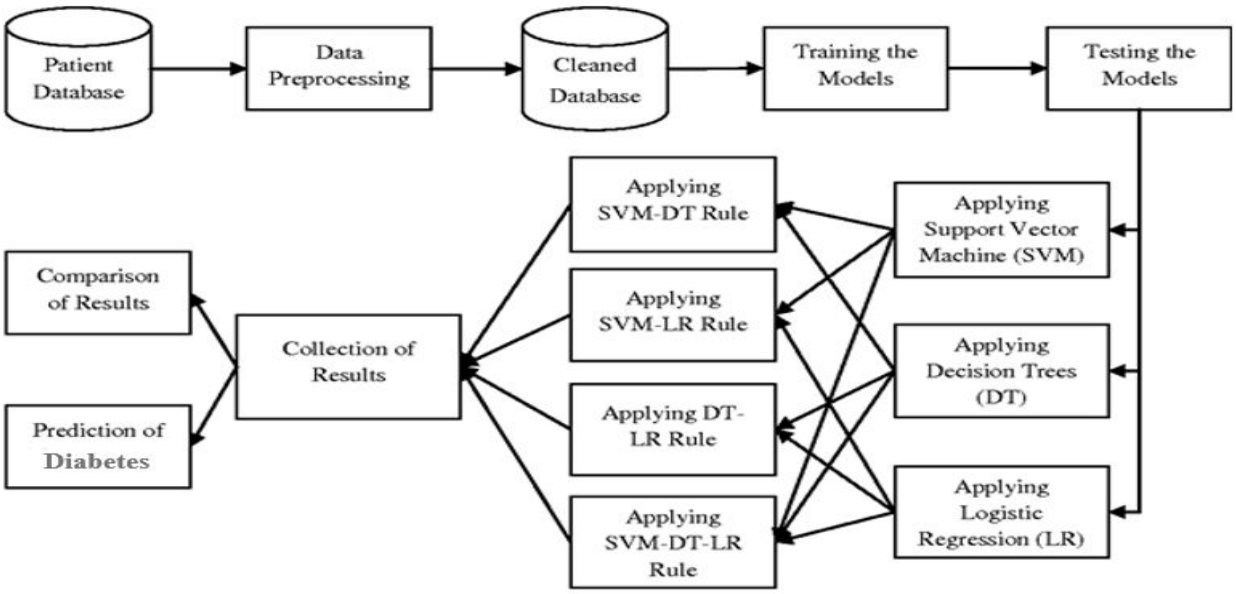


Figure 2.1: Machine learning based diabetes prediction

Data can be grouped according to its attributes. By building a model using sample data and existing information, classification is achieved. Enhancing the consistency of the data-based results is one of the objectives of classification. Machine learning techniques are applied after the data is ready. In the context of diabetes prediction, various classification and ensemble algorithms are utilized. One common technique for building ensembles is to manipulate the input data given to a single classifier. Running the classifier with a training set that consists of a randomly chosen sample and a replacement from the original dataset can help achieve this. The primary objective of applying machine learning techniques is to evaluate the performance and accuracy of various algorithms, as well as to identify the most significant features that play a crucial role in diabetes prediction.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304063

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

KNN: This method can be applied to regression as well as classification. The main purpose of this algorithm is to classify a new item according to its attributes and training data. This method predicts the level of the test sample and yields the result with the highest level possibility based on the k training samples, which are the test person's closest neighbouring neighbours.Based on the concept of probability, logistic regression is a classification technique that can be applied as a predictive analytical technique. It is employed to address classification issues. This method produces a discrete binary answer between 0 and SVM. SVM is a supervised machine learning method that performs better than other machine learning algorithms and has been thoroughly evaluated in nonlinear function forecasting tasks and real-world classification issues. Random forest is a technique that can help prevent overfitting in ensemble construction and bootstrapping.

A technique for teaching a computing system to recognize the properties of input data is called machine learning. These techniques have been demonstrated to be successful in identifying diabetes. Numerous machine learning algorithms, such as reinforcement learning, supervised learning, and unsupervised learning techniques, have been created. Practically speaking, this is feasible because the entire paradigm is data-driven. Large volumes of data can be loaded into a database, and machine learning can save a great deal of human laboUr. Using this data, models are trained to produce the best result possible given the input data. Any realistic parameters that satisfy medical standards can be used to train the models. A patient's blood test results may be examined by some, while facial traits may be examined by others. The parameters vary since the disease might present itself in different ways. To get findings that seem most appropriate for real-world applications, researchers have experimented with several algorithms and adjusted a lot of hyperparameters using a variety of suggested techniques. Based on variables like pregnancy, blood pressure, glucose, skin thickness, insulin, body mass index, and diabetes, Support Vector Machines (SVM) are used to identify diabetes.

III. MOTIVATION

Diabetes is one of the fastest-growing global health concerns, affecting millions worldwide. According to the World Health Organization (WHO) and International Diabetes Federation (IDF), diabetes cases have significantly increased in recent decades, leading to severe complications such as heart disease, kidney failure, nerve damage, and vision loss. The key challenge with diabetes is that its symptoms often go unnoticed until it reaches an advanced stage, making early detection crucial. Traditional diagnostic methods, such as fasting blood sugar tests, oral glucose tolerance tests, and HbA1c tests, are effective but have limitations: **Time-consuming and costly:** Requires multiple clinical visits and expensive laboratory tests. **Limited accessibility:** Many remote and underdeveloped areas lack healthcare facilities.
Reactive approach: Diagnosis often happens after symptoms appear rather than as a preventive measure.

IV. ABRIDGEMENT

This project focuses on developing a machine learning-based diabetes prediction system, specifically using Support Vector Machines (SVM) along with other ML models like Random Forest and Logistic Regression. The system is implemented using Python and employs popular ML libraries such as Scikit-learn, TensorFlow, and Keras. A well-structured dataset containing various health indicators (blood glucose levels, BMI, blood pressure, insulin levels, age, and family history) is used to train and evaluate the models.
The main objectives of this project are: To develop an efficient and accurate diabetes risk prediction model. To provide a cost-effective and accessible healthcare solution. To assist medical professionals in early diagnosis and intervention. The system is expected to improve early detection rates, reduce healthcare costs, and enhance personalized treatment plans for at-risk individuals.

V. RELATED WORKS

Several research studies and projects have explored diabetes prediction using machine learning techniques. Some key related works include:
Pima Indian Diabetes Dataset (PIDD) Research: The **PIDD dataset**, provided by the National Institute of Diabetes and Digestive and Kidney Diseases, has been widely used in diabetes prediction. Researchers have applied various ML algorithms to this dataset, such as:
Decision Trees : Simple but prone to overfitting.
Naïve Bayes : Good for small datasets but assumes feature independence.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Artificial Neural Networks (ANNs): High accuracy but computationally expensive.
Support Vector Machines (SVMs) : Proven effective due to their ability to handle high-dimensional data.
Real-World Applications
Mobile Apps and Wearable Health Devices Many healthcare applications now integrate ML-based diabetes risk assessment using real-time health monitoring.
Clinical Decision Support Systems (CDSS): Hospitals use ML models to assist doctors in predicting disease risks based on patient history. By analysing these related works, this project aims to enhance accuracy by optimizing feature selection, hyperparameter tuning, and integrating multiple ML models to create a more robust diabetes prediction system.

VI. EXISTING SYSTEM

In the traditional system, patients are often unaware of their diabetes risk until symptoms worsen, prompting a hospital visit. The diagnosis process typically involves long waiting times, multiple consultations, and various lab tests, such as blood glucose and insulin level analysis. This approach is time-consuming, costly, and may delay early detection. Additionally, patients often miss opportunities for timely intervention due to a lack of awareness and accessibility to preventive screenings. This reactive system limits proactive healthcare measures, leading to potential complications and poorer health outcomes.
Conventional Diagnostic Approaches
Fasting Blood Sugar Test (FBS) : Measures blood glucose levels after an overnight fast.
Oral Glucose Tolerance Test (OGTT): Measures body response to sugar intake.
Hemoglobin A1C Test: Measures average blood sugar levels over the past 2-3 months.

DISADVANTAGES OF EXISTING SYSTEM
Patients remain unaware of their diabetes risk until symptoms become severe. Diagnosis involves long waiting times and multiple consultations. Lab tests and procedures are time-consuming and expensive.
Delayed detection reduces opportunities for early intervention. Limited access to preventive screenings hampers proactive care. Reactive healthcare leads to higher risk of complications and poor outcomes.

VII. PROPOSED SYSTEM

The proposed system leverages machine learning algorithms, such as Support Vector Machines, Random Forest, and Logistic Regression, to predict diabetes risk accurately and efficiently. By analyzing key health indicators like blood glucose levels, BMI, blood pressure, and insulin levels, the system enables early detection without the need for extensive hospital visits. Built on a Python-based framework using libraries like scikit-learn, TensorFlow, and Keras, the system offers timely and cost-effective predictions. It empowers healthcare professionals to implement proactive interventions, facilitating personalized care and reducing the burden of diabetes. This approach promotes better health outcomes through early diagnosis and efficient management.
Machine Learning Algorithms: SVM, Random Forest, and Logistic Regression will be used to classify individuals as diabetic or non-diabetic based on input features.
Comprehensive Feature Selection: Health parameters such as blood glucose, BMI, blood pressure, age, and insulin levels will be analyzed.
Automated Prediction System: The model will process user data and provide **instant risk assessment**.
Improved Accuracy & Efficiency: The system will undergo hyperparameter tuning to enhance prediction performance.
Scalability: Can be **integrated into mobile applications, wearable devices, or hospital systems** for real-time diabetes screening.
Workflow of the Proposed System
Data Collection : Gather medical records and health indicators.
Data Preprocessing : Handle missing values, normalize data, and perform feature selection.
Model Training & Evaluation : Train SVM, Random Forest, and Logistic Regression models.
Performance Optimization: Hyperparameter tuning and cross-validation.
Deployment : Deploy the trained model in a user-friendly interface.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304063

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

ADVANTAGES OF PROPOSED SYSTEM

Provides accurate and timely diabetes risk predictions using machine learning. Reduces the need for lengthy hospital visits and costly lab tests. Enables early detection, allowing for proactive intervention and care.

Analyzes key health indicators efficiently for personalized patient management. Improves accessibility to diabetes risk screening through non-invasive methods. Enhances patient outcomes by reducing complications and promoting better health.

VIII. ARCHITECTURE

The architecture of a machine learning-based diabetes prediction system consists of multiple stages, from data collection to model deployment. Below is a detailed breakdown of the system's architecture: The architecture follows a modular pipeline consisting of the following components:

Data Acquisition Layer

Dataset Collection: Medical datasets (e.g., Pima Indian Diabetes Dataset (PIDD) or real-world hospital data). Features include glucose levels, BMI, blood pressure, insulin, age, etc.

Data Sources: Electronic Health Records (EHRs). Wearable health devices (smartwatches, glucose monitors) Manual input by users (health apps, surveys).

Data Preprocessing Layer

Handling Missing Data: Missing values are filled using mean/mode imputation or removed if necessary.

Feature Scaling & Normalization: Since different features have varying units, Min-Max Scaling or Standardization is applied.

Feature Selection: Correlation Analysis, PCA, or Recursive Feature Elimination (RFE) is used to select the most important features.

Data Splitting: The dataset is split into training (80%) and testing (20%) sets for model evaluation.

Machine Learning Model Layer Algorithm Selection: Support Vector Machine (SVM): Primary model for prediction.

Comparison Models: Random Forest, Logistic Regression, and KNN.

Training the Model: The training dataset is used to train the ML models. Hyperparameter tuning (Grid Search CV, Randomized Search) is performed to optimize performance.

Model Evaluation: The trained models are tested using the testing dataset. Metrics like accuracy, precision, recall, F1-score, and AUC-ROC are used to evaluate performance.

Prediction & Classification Layer

User Input: A patient enters their medical details (e.g., glucose level, BMI, etc.) via a user interface (web/app).

Model Processing: The trained SVM model processes the input features and predicts whether the patient is diabetic or not.

Result Output: The system displays the diabetes risk score (Low/Medium/High) or a binary classification (Diabetic / Non-Diabetic).

Deployment Layer

Model Deployment: The trained model is deployed using Flask / FastAPI (for web applications) or integrated into mobile apps.

Integration with Healthcare Systems: The system can be linked to hospital databases for real-time patient monitoring.

IX. RESULT ANALYSIS

The result analysis of the proposed diabetes prediction system highlights the critical role of the Support Vector Machine (SVM) algorithm as the core classifier. SVM is highly effective in handling high-dimensional and non-linear data, making it well-suited for medical datasets that typically contain diverse and heterogeneous features. The core strength of SVM lies in its margin maximization technique, which constructs an optimal hyperplane to classify individuals as diabetic or non-diabetic based on selected health parameters. The model is trained using a well-structured dataset, which is divided into training, validation, and testing sets. Various kernel functions—such as linear, polynomial, radial basis function (RBF), and sigmoid—were evaluated to identify the best-performing model. Among these, the RBF kernel achieved the highest accuracy of 77.27%, indicating its superior ability to manage complex, non-linear decision boundaries in the dataset.To ensure the robustness of the model, hyperparameter tuning was performed by optimizing kernel selection, regularization strength, and margin width. Comparative analysis with other classifiers such as Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, and Naive Bayes demonstrated that while Random Forest offered slightly

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304063

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

higher precision, SVM provided better generalization and stability across varying input conditions. Evaluation metrics such as accuracy, precision, recall, F1-score, and AUC-ROC were used to assess performance. Furthermore, feature importance analysis revealed that glucose levels, BMI, and age were the most significant predictors of diabetes, followed by diabetes pedigree function and blood pressure. Visual aids such as confusion matrices and feature contribution plots provided deeper insights into model performance. Additionally, a user-friendly web application was developed, enabling individuals to input medical parameters and receive real-time diabetes risk assessments. This enhances healthcare accessibility, particularly in remote areas lacking proper diagnostic facilities. The proposed model offers a cost-effective and efficient solution for early-stage diabetes detection, supporting proactive medical intervention. The study suggests further enhancement of the system by incorporating more diverse datasets, deep learning models, and real-time data from wearable devices for continuous monitoring.

This future direction aims to improve accuracy and facilitate more personalized healthcare services.

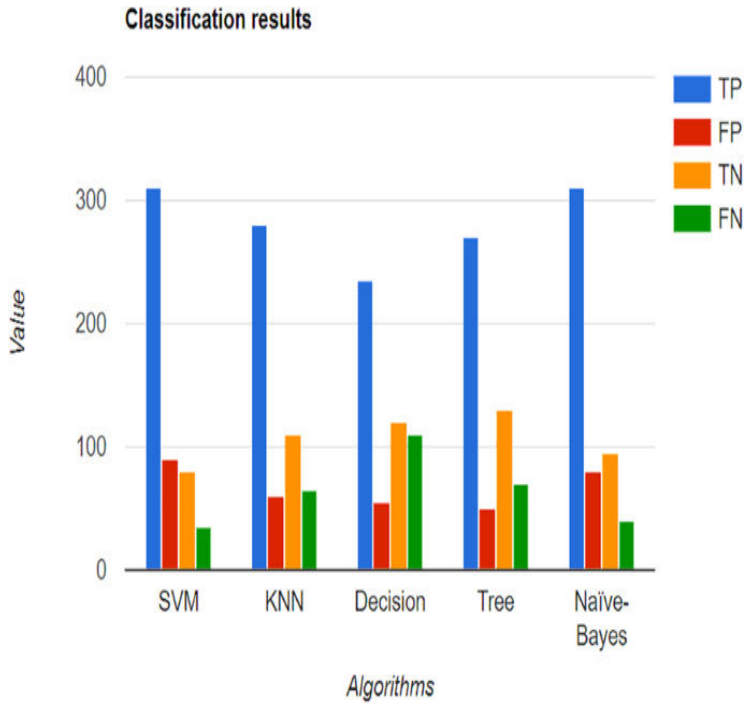


Figure 9.1: Analysis of diabetes mellitus for early prediction using optimal features selection

By varying the distance between the data points and the hyper plane, this classifier seeks to create a hyper plane that can effectively divide the classes. The hyper plane is chosen depending on a number of kernels.

I experimented with four different kernels: sigmoid, linear, poly, and rbf. The goal of this initiative is to use pertinent medical data to forecast when a person may develop diabetes. To find out if they have diabetes or not, users can enter their medical data onto an online site. An intelligent model then processes the data. Users can enter certain medical data fields, which are essential for determining the risk of getting diabetes, using the interface, as shown in the following figure. Additionally, our research incorporates a user-friendly web application designed to be accessible to individuals with limited technical expertise. This web app enhances the effectiveness of the prediction model by ensuring users can easily interact with the system and obtain accurate health insights.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304063

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

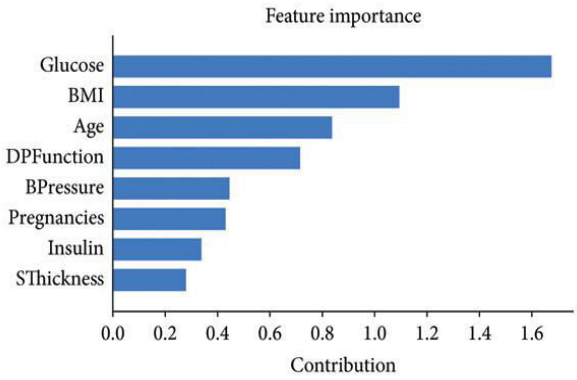


Figure 9.2: Technique Using Machine Learning for the Prediction of Diabetes Disease

The chronic illness known as diabetes severely compromises the body's capacity to control blood sugar levels. The existence of this disease can be predicted using a variety of machine learning techniques, such as Support Vector Machines (SVM), Logistic Regression, K-nearest neighbours (KNN), and XG Boost. Our research focuses on creating a system that employs SVM to predict the occurrence of diabetes. At first, only a few characteristics associated with gestational diabetes were included in the dataset. Our enhanced method increased the accuracy of the SVM model to 77.27% by using a larger dataset with a wider variety of characteristics. This grew a wider range of criteria and aspects will be incorporated into this project's future scope, according to the outlook. Increasing the number of parameters will boost the accuracy of the model. Traditional machine learning methods and algorithms can be used to further improve the accuracy of diabetes predictions by iteratively improving the data and adding more pertinent variables.

All scientific fields are adopting machine learning techniques in large quantities, which is revolutionizing companies around the world. The implementation of automated learning processes and techniques has resulted in notable breakthroughs, particularly in the healthcare industry. In the treatment of long-term conditions like diabetes, these methods have shown promising outcomes. With its wide range of applications, the web has also shown promise in the medical domain. A subset of web technology called ontology improves the processing of concepts and relationships in a way that is comparable to human cognition, which makes it very helpful for interpreting complex data. The results of our comparison between ontology-based machine learning and popular classification machine learning techniques show that, even in the absence of feature selection, the ontology classification approach had the best accuracy. This points to an interesting direction for further study, where even better outcomes may be obtained by combining machine learning principles with ontology classification. In order to produce more precise forecasts, suggestions, and decision-making instruments, we urge scholars to investigate and develop in this area. According to our viewpoint, we intend to improve this comparison analysis by utilizing cutting-edge techniques that combine the ontology classification method with machine learning principles. In order to further enhance the system's prediction capabilities, we also hope to include regression machine learning methods. Better patient outcomes and more efficient disease management will result from this future effort, which will guarantee a more thorough and accurate diabetes prediction system.

X. CONCLUSION

Early identification of diabetes is one of the major real-world medical challenges. This study aims to forecast diabetes by means of a system that is designed with methodical efforts. Five machine learning classification algorithms are examined and assessed using a variety of metrics in this paper. The John Diabetes Database is used to conduct experiments. Using the Decision Tree algorithm, experimental findings show that the designed system is adequate with an accuracy of 99%. Other diseases may be predicted or diagnosed in the future using the system that was created and the machine learning classification methods that were employed. The work can be expanded and enhanced to include other machine learning methods for the automation of diabetes analysis.

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304063

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

XI. FUTURE ENHANCEMENT

The Diabetic Disease Risk Prediction System has great potential for future enhancements to improve its accuracy, usability, and impact on healthcare. One possible enhancement is the integration of additional health data sources such as genetic information, lifestyle factors, and medical history, which could provide a more comprehensive risk assessment and further personalize predictions. The system could also benefit from incorporating real-time data from wearable devices, allowing for continuous monitoring of a user's health metrics like heart rate, physical activity, and sleep patterns, thereby offering more accurate predictions and proactive interventions. Another significant improvement would be the use of deep learning techniques, such as neural networks, to enhance the predictive power of the system, especially in recognizing complex patterns in large datasets. Additionally, expanding the system to include mobile app support would make it more accessible to a broader user base, allowing patients to input data and receive risk predictions on-the-go.
Furthermore, implementing a recommendation system for personalized lifestyle changes, including diet and exercise suggestions, based on the predicted risk level, could provide added value to users.
Collaboration with healthcare providers and integrating the system with electronic health records (EHR) could ensure that the predictions are actionable and that appropriate follow-up care is provided.

REFERENCES

1.R. Thomas, S. Halim, S. Gurudas, S. Siva prasad, and D. Owens, "Idf diabetes atlas: A review of studies utilizing retinal photography on the global prevalence of diabetes related retinopathy between 2015 and 2018," Diabetes research and clinical practice, vol. 157, p. 107840,2019.
2.B. Nithya and Dr. V. Ilango," Predictive Analytics in Health Care Using Machine Learning Tools and Techniques", International Conference on Intelligent Computing and Control Systems, 978-1-5386-2745-7,2017.
3.T. Vos, A. D. Flaxman, M. Naghavi, R. Lozano, C. Michaud, M. Ezzati, K. Shibuya, J. A. Salomon, S. Abdalla, V. Aboyans et al., "Years lived with disability (ylds) for 1160 sequelae of 289 diseases and injuries1990–2010: a systematic analysis for the global burden of disease study 2010," The lancet, vol. 380, no. 9859, pp.2163–2196,2012.
4.W. H. Organization et al., "Diabetes fact sheet n 312. October 2013," Archived from the original on, vol. 26, 2013.
5.U. Diabetes and H. Lobby, "What is diabetes," Diabetes UK,2014.
6.W. H. Organization et al., "The top 10 causes of death fact sheet no 310," Geneva, Switzerland: World Health Organization, 2013.
7.A. D. Association et al., "Economic costs of diabetes in the us in 2017," Diabetes care, vol. 41, no. 5, pp. 917–928
8.C. for Disease Control, Prevention et al., "National diabetes statistics report, 2020," Atlanta, GA: Centers for Disease Control and Prevention, US Department of Health and Human Services, pp. 12–15, 2020.
9. World Health Organization etal., Guidelines for the prevention, management and care of diabetes mellitus, 2006.
10.A. E. Kitabchi, G. E. Umpierrez, J. M. Miles, and J. N. Fisher, "Hyperglycemic crises in adult patients with diabetes," Diabetes care, Vol. 32, no. 7, pp. 1335–1343, 2009.
11.E. Saedi, M. R. Gheini, F. Faiz, and M. A. Arami, "Diabetes mellitus and cognitive impairments," World journal of diabetes, vol. 7, no. 17, p. 412, 2016.
12.P. Suresh Kumar and S. Pranavi "Performance Analysis of Machine Learning Algorithms on Diabetes Dataset using Big Data Analytics", International Conference on Infocom Technologies and Unmanned Systems, 978-1-5386-0514-1, Dec. 18-20, 2017.
13.Gauri D. Kalyankar, Sivananda R. Poojara and Nagaraj V. Dharwadkar," Predictive Analysis of Diabetic Patient Data Using Machine Learning and Hadoop", International Conference On I-SMAC,978-1-5090-3243-3,2017
14.Aiswarya Iyer, S. Jeyalatha and Ronak Sumbaly," Diagnosis of Diabetes Using Classification Mining Techniques", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.1, January 2015.
15.Meiyalakan.KOnline Multi-Crop Procurement and Loan System. Volume 10, Issue 5, pp: 32-35,
16.Meiyalakan K. Knowledge-Based Approach to Detect Potentially Risky. Websites. Volume 10, Issue 6, pp: 1353-1357
17.Meiyalakan K," Estimation of the Available Bandwidth in Inter-Cloud Links for Task Scheduling in Hybrid Clouds", DOI: 10.15680/IJIRCCE.2022.1007086
18.Meiyalakan K," A Trends of Event Detection by Analyzing Social Media Platforms Data", 10.15680/IJIRSET.2023.1203081,
19. K.Meiyalakan Blockchain Based Anonymous Authentication of Cloud Data Using Stochastic DiffusionSearchAlgorithm,https://www.dsengg.ac.in/pdf/cells/ICIRIST-EBook-2024.pdf.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

20.K. Meiyalakan," Predicting Emotions from TextusingComputingTechnique"https://www.dsengg.ac.in/pdf/cells/ICIRIST-EBook-2024.pdf
21.K. Meiyalakan, "Cross-Analysis of Network Intrusion Detection Systems Based on Machine Learning",https://www.dsengg.ac.in/pdf/cells/ICIRIST-EBook-2024.pdf.
22.K.Meiyalakan,"MachineLearningContribution to the Field of Security and Privacy Using Android", ICSIEM 2024,16 -17, April 2024.
23. J.Stanly Jayaprakash, M. Jasmine Pemeena Priyadarsini, B. D. Paramesh Chari Hamid Reza Karimi, and Sasikumar Guru Moorthy. (2022). Deep Q-Network with Reinforcement Learning for Fault Detection in Cyber-Physical Systems. Journal of Circuits, Systems and Computers.
24.Mr.J.Stanly Jayaprakash Dr.S. Arumugam (2014). Efficient Biometric Security System Using Intra-Class Finger-Knuckle Pose Variation Assessment. International Journal of Computer Science & Engineering Technology (IJCSET).
25.Stanly Jayaprakash, Kishore Balasubramanian, Rossilawati Sulaiman, Mohammad Kamrul Hasan, *, B. D. Paramesh Chari Cloud Data Encryption and Authentication Based on Enhanced Merkle Hash Tree Method J and Celestine Iwendi 2021.
26.Multimodal finger biometric score fusion verification using coarse grained distribution function JS Jayaprakash, S Arumugam2015
27.A novel approach for fingerprint sparse coding analysis using k-svd learning technique S Arthi, J Stanly Jayaprakash 2024.
28."A Fusion Attention Mechanism with Bi-LSTM-Based Sarcasm Detection for Selecting High-Profit Products". Journal of Circuits, Systems and Computers C. Gayathri., R. Samson Ravindran.(2025).https://doi.org/10.1142/S0218126625501439.
29.Machine learning based internet browsers in malicious website detection, international journal of innovative research in computer and Communication engineering, Sowmiya R, June 2021
30.Virtual Human Resource Management with Recruitment in Software Engineering Roles" International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), Sowmiya R Volume 12, Issue 3, March 2024.
31."Advanced Drowsiness Detection system using OPENCV and KERAS" International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE) Sowmiya R, Volume 12, Issue 5, May2024.
32."Intelligent Phishing Website Detection model with Deep Learning- based Innovative Technique" International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET) Sowmiya.R, Volume13, Issue 3, March 2024.
33."Predicting Emotions from Text Using Computing Technique" "International conference on Integrating Recent Innovations in Science and Technology: Shaping the future (ICIRIST-2024)" Sowmiya R,2024.
34.Real time cyber physical false data attack Detection in machine learning methods, International Journal of Innovative Research in Computer and Communication Engineering Sowmiya R, June 2021
35."Soil parameter analyzing using curse of dimensionality for accuracy and prediction, international journal of innovative research in science, engineering and technology, Ms. R. Sowmiya Feb 2020.
36. Durairam.R.. Machine Learning Approaches for Brain Disease Diagnosis. Volume 10, Issue 6, pp: 1092-1097.
37."Predicting Emotions from Text Using Computing Technique" "International conference on Integrating Recent Innovations in Science and Technology: Shaping the future (ICIRIST-2024)" Sowmiya.R,2024.
38.C.Anusuya et al., "Credit Card Fraud Detection using Machine Learning-Based Random Forest Algorithm," Int. J. Sci. Adv. Res. Technol., vol. 9, no. 3, Mar. 2023.
39.C.Anusuya et al., "Alzheimer Disease with Blood Plasma Proteins detected using Convolutional Neural Network (CNN)," Int. J. Innov. Res. Compute. Commun. Eng., vol. 11, no. 3, Mar. 2023.
40.C.Anusuya et al., "Facial Recognition Services for E-Voting System by Using Blockchain Technology," Int. J. Innov. Res. Compute. Commun. Eng., vol. xx, no. xx, pp. xxx-xxx, year.
41.C.Anusuya et al., "A Comparative Feature Extraction Study Using Textural Features to Extract Vital Information from Lung Images," Int. J. Adv. Inf. Sci. Technol., vol. 12, no. 02, Feb. 2023.
42.C.Anusuya et al., "Diagnostics Decision Support System for Tuberculosis Using Fuzzy Logic," IRACST Int. J. Compute. Sci. Inf. Technol. Sec., vol. 2, no. 3, Jun. 2012.
43.C. Anusuya et al., "Classification of Uncertain Data Using Selection Algorithm," *Int. J. Mod. Eng. Res.*, vol. 2, no. 3, pp. 1066-1072, May-Jun. 2012.
44.Parvathi M "Sensing of Near Duplicates in Large Image Database", Volume12, Issue 3, March2023,DOI:10.15680/IJIRSET.2023.120316

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details