

# Enhancing Healthcare Decision Making through Predictive Analytics Using Hybrid AI Techniques

## OPEN ACCESS

Volume: 13

Special Issue: 3

Month: February

Year: 2026

P-ISSN: 2321-788X

E-ISSN: 2582-0397

Citation:

Punithavel, R., et al.

“Enhancing

Healthcare Decision

Making through

Predictive

Analytics Using Hybrid

AI Techniques.”

*Shanlax International*

*Journal of Arts, Science*

*and Humanities*,

vol. 13, no. 3, 2026,

pp. 122–29.

DOI:

[https://doi.org/10.34293/](https://doi.org/10.34293/sijash.v13iS3-i1-Feb.10270)

[sijash.v13iS3-i1-](https://doi.org/10.34293/sijash.v13iS3-i1-Feb.10270)

[Feb.10270](https://doi.org/10.34293/sijash.v13iS3-i1-Feb.10270)

**Mr. R. Punithavel**

*Assistant Professor, Department of Computer Applications*

*KIT-Kalaigarkaranidhi Institute of Technology, Coimbatore*

**Dr. T. Ramesh**

*Associate Professor & Head*

*Department of Information Technology*

*Bharathiar University, Coimbatore*

**Mr. PP. Mithun**

*Department of Computer Applications,*

*KIT-Kalaigarkaranidhi Institute of Technology, Coimbatore*

**Ms. R. Anitha**

*Department of Computer Applications*

*KIT-Kalaigarkaranidhi Institute of Technology, Coimbatore*

## Abstract

*High levels of glucose in the bloodstream lead to the development of diabetes, which results in frequent urination, increased thirst, and increased hunger. It is crucial to address diabetes promptly as untreated cases may lead to severe complications in various body organs such as the heart, kidneys, blood pressure, and eyes. Predictive analytics over big data is a challenging task, particularly in healthcare. However, it can aid healthcare practitioners in making quick decisions about patients' health and treatment based on big data. The performance and accuracy of ML algorithms used in predictive Data analysis for predicting the occurrence of diabetes are compared and analyzed across various disciplines. In this study, different classification Computational methods, which may involve various algorithms, such as SVM, KNN, Logistic regression, and Randomforest, were considered, and their performance metrics such as Recall, FN Measure, Precision, and Accuracy were evaluated Derived from the confusion matrix. According to the experimental results, the SVM and ontology classifiers yielded the highest accuracy for diabetes prediction.*

**Keywords:** Support Vector Machine (SVM), K-NearestNeighbor(KNN), Logistic Regression, Random Forest, Performance Metrics, Accuracy, Precision, Recall, F-Measure, Accuracy, Confusion Matrix, Ontology Classifier, Diabetes Prediction, Machine Learning, Predictive Analytics, Big Data, Healthcare, Classification Algorithms

## Introduction

Diabetes, a condition that raises blood glucose levels due to insufficient or absent insulin, is one of the most well-known illnesses to have swept the globe in recent years. Because there are numerous requirements to be taken into account for a person to have this illness, its identification and prognosis may be laborious or occasionally ambiguous. However, even at an early stage, detection is not impossible.

Diabetes is increasing day by day in the world because of environmental, genetic factors. The numbers are rising rapidly due to several factors which includes unhealthy foods, physical inactivity and many more.

Diabetes is a hormonal disorder in which the inability of the body to produce insulin causes the metabolism of sugar in the body to be abnormal, thereby, raising the blood glucose levels in the body of a particular individual. Intense hunger, thirst and frequent urination are some of the observable characteristics. Certain risk factors such as age, BMI, Glucose Levels, Blood Pressure, etc., play an important role to the contribution of the disease. Diabetes is a very familiar word in the present world and crucial challenges in both developed and developing countries. The insulin hormone in the body produced by the pancreas allows glucose to pass from the food into the bloodstream. The lack of that hormone due to malfunctioning of the pancreas forms diabetes which can result in coma, renal and retinal failure, pathological destruction of pancreatic beta cells, cardiovascular dysfunction, cerebral vascular dysfunction, peripheral vascular diseases, sexual dysfunction, joint failure, weight loss, ulcer, and pathogenic effects on immunity. After cancer and heart disease, diabetes is the third most common cause of death. However, we can now solve this problem thanks to the development of machine learning techniques. The purpose of data mining and machine learning is to extract knowledge from the data in the dataset and produce an intelligible and concise description of the patterns.

## Literature Review

Arwathi Chen Lyngdoh et al. conducted research on predicting diabetes disease using 5 supervised ML Algo : KNN, Naive Bayes, Decision Tree Classifier, Random Forest, and SVM. by including current risk variables and performing cross-validation, they achieved consistent accuracy with the KNN classifier achieving a high accuracy of 76%. The main objective of the study was to identify the best outcomes for accurately predicting diabetes disease, considering accuracy and computing time.

MitushiSoni et al. ML classification and ensemble techniques were employed to make predictions about diabetes using a dataset. They employed K-Nearest Neighbors, Logistic Regression, Decision Tree, Support Vector Machine, Gradient Boosting, and Random Forest algorithms, and found Random Forest outperformed the others in terms of accuracy.

Shejal Kale et al. applied ML Classification & Using ensemble techniques to make predictions about diabetes on a given dataset. Sivaranjani S et al. used SVM and Random Forest (RF) methods for identifying potential risks of Diabetes Related Diseases. WORKING OF SYSTEM

## System Architecture

Using a variety of machine learning techniques, including SVM, KNN, Logistic Regression, Random Forest, and Decision Tree, this project seeks to create an early diabetes prediction system. These methods mostly concentrate on passively imparting theoretical knowledge through lectures, presentations, or videos, with the hope that workers will subsequently use what they have learned in the workplace.

## **Unsupervised Machine Learning**

As the name implies, unsupervised learning does not require supervision, which sets it apart from supervised learning. In unsupervised machine learning, the machine is trained on an unlabeled dataset and makes predictions about the output without any oversight.

In unsupervised learning, the models are trained with the data that is neither classified nor labelled, and the model acts on that data without any supervision.

## **Machine Learning Algorithms**

We have also use various algorithms in our project to train the model and they are following:

- K-Nearest Neighbors.
- Logistic Regression.
- Random Forest.
- Support Vector Machine.

## **Implementation**

### **A. Hardware Platform Used**

The hardware requirement may serve as the basis for a contract for the implementation of the system and should therefore be complete and consistent in specification.

The hardware used for the system is mentioned below.

- PROCESSOR: Intel CORE i3 or above
- RAM: minimum 4.00GB
- HARD DISK: minimum 100GB

### **B. Libraries and Software Platform Used**

The software requirement document is the specification of the system. The software requirement provides a basis for creating the software requirements specification.

- OPERATING SYSTEM: Windows
- SYSTEM TYPE: 64-bit, intel CORE i5
- SOFTWARE: Jupyter Notebook, VS Code, Anaconda
- TECHNOLOGIES: Python
- LIBRARIES: Flask, pandas, NumPy, pickle, sklearn, xgboost,

## **Methodology**

### **A. Dataset**

The National Institute of Diabetes and Digestive and Kidney Diseases provided the dataset. There are nine attributes total in the diabetes dataset. Every individual in the records is female, and the dataset's first attribute is the number of pregnancies they have had Development of Systems.

### **Attributes**

- Pregnancy Attribute
- Glucose Attribute
- Blood Pressure Attribute

### **B. Data Pre-Processing**

Preprocessing data is crucial, particularly for healthcare data, which may include missing values and other impurities that could compromise data mining's efficacy. With the aid of ML methodology

on the CSV file, this procedure is necessary to produce accurate results and successful predictions. Throughout the experiment, we handled the dataset effectively and easily by using the Pandas and NumPy libraries to make it useful.

### **C. Missing Value Elimination**

Since it is impossible to have a value of zero for some features, values of zero are eliminated. By removing unnecessary features or instances, lowering the dimensionality of the data, and enabling feature subset selection.

### **D. Splitting Data**

To improve model performance, the dataset is normalized after data cleaning to make sure all features are on the same scale. After that, the data is separated into training and testing sets; normally, two-thirds of the data are used for training and one-third for testing. The training data is used to train the model, and its accuracy is assessed using the test data. This aids in detecting overfitting, which occurs when a model exhibits strong performance on training data but weak performance on unseen data, suggesting that it has learned patterns by heart rather than by heart.

### **Goal of the Design**

Design is an important step in software engineering where system requirements are turned into a structured framework for implementation. It is a meaningful engineering representation of the system that needs to be built, and it must meet customer needs while also meeting quality standards that have already been set.

### **Design Strategy**

The process of planning a new system or replacing an old one is called system design. In simple terms, system design is like a blueprint for a building. It tells you what all the parts of the finished product should be. The next step is system design.

### **Abstraction**

By using abstraction, problems can be solved without having to consider the finer points of their numerous component subproblems. System designers can make incremental improvements at every stage of design thanks to abstraction

### **Modularity**

The goal of modularity is to break down the main module into manageable chunks with clearly defined interfaces between them. This makes the design more clear, which makes it easier to implement, debug, test, document, and maintain the software product. In this context, modularity is an essential tool for building large software projects.

### **Module Diagram**

#### **Architecture Prototype**

##### **1. Prototype Features**

- Industry-specific interactive 3D work environments.
- A system of instant feedback for improving performance in real time.
- A data analytics dashboard to track staff development.

### **System Overview**

- The goal of the proposed Virtual Reality (VR) training prototype is to train employees in an immersive, interactive environment that mimics real-world workplace situations.

### **Workflow Overview**

- The employee uses the LMS or a specific portal to access the VR training system.
- The designated simulation scenario is loaded by the VR Interface Module.
- The cloud backend receives the data for analysis, synchronization with the LMS, and storage.

### **Purpose of Feature Extraction**

- Performance evaluation: Calculate how well workers complete tasks in virtual reality simulations.
- Adaptive Learning: Identify skill gaps and adjust the difficulty or complexity of scenarios in real time.
- Analyze learner concentration and immersion to maximize training materials..

### **Feature Utilization in VR Training**

- Real-Time Feedback: During training, give users prompt direction or remedial instructions.
- Adaptive Scenario Adjustment: Depending on user engagement and skill level, change the task's difficulty.
- Explainable AI (XAI): Highlights Why content is flagged as misinformation, aiding transparency.

### **Visualization & Analytics**

- Dynamic Dashboard: Plots misinformation volume over time.
- Heatmaps: Identify geographical spread if location data is available.
- Network Diagrams: Visualize how misinformation spreads through social networks.

### **Alert System**

- Sends alerts to platform moderators or users when suspicious content is detected.
- Incorporates severity levels to prioritize responses.

### **Evaluation & Testing**

- Dataset: Benchmarked using labeled misinformation datasets (e.g., Fake News Net, PHEME).
- Metrics: Precision, recall, F1-score to assess detection accuracy.
- User Studies: Prototype tested with user feedback to refine usability.

### **Technical Stack**

- Backend: Python/Flask or Node.js for server logic.
- ML Models: TensorFlow, PyTorch, or Scikit-learn.
- Database: NoSQL (MongoDB) for flexibility, plus relational DB (PostgreSQL) for structured data.
- Front-end: React.js or Angular for interactive, real-time data visualization.

## Frame Work Overview

### 1. Core Components

- VR Interface Module: uses VR headsets and controllers to create realistic 3D environments..
- Allows for interactive task simulations that mimic actual situations.
- Training Content Engine: Hosts industry-specific training modules and scenarios.
- Adapts task difficulty and guidance based on user performance.

### 2. Workflow

- Login and Profile Access: Through an LMS or a specific portal, staff members can access training.
- Scenario Loading: The designated simulation is loaded from the content engine via the VR interface.
- Interactive Training: Users complete scenarios , carry out tasks and get instructions

### 3. Feature Extraction

- Interaction Frequency: Indicates how frequently a user engages with virtual tools, interfaces, or objects while performing a task.
- Navigation Patterns: Monitors movement in the virtual reality environment, including routes traveled, coverage of the area, and duration of time spent in particular areas.

### 4. Decision Points

Documents decisions, both right and wrong, made in branching scenarios.

### 5. Model & Detection

- Supervised Models: Training classifiers (SVM, Random Forest, Deep Learning models like CNN, RNN, Transformers) on labeled data to detect misinformation.
- Unsupervised Models: Clustering algorithms to detect anomalous patterns or suspicious communities.
- Knowledge Graphs: Incorporating factual knowledge bases to verify claims.

### 6. Visualization & Dashboard

- Interactive Dashboard: Real-time monitoring of detected misinformation.
- Trend Analysis: Visualizing misinformation spread and impact.
- Source Attribution: Mapping key sources or influencers involved.

### Key challenge

- Scalability: Handling large volumes of data from multiple platforms.
- Multi-modal Data: text, images, and videos.
- Evolving Misinformation: Adapting to new forms of misleading content.

### Result & Analysis

The results of the study show that, in comparison to conventional training techniques, Virtual Reality (VR) simulations greatly improve employee training results. Employees who received VR training showed increased levels of engagement, knowledge retention, and task performance accuracy, according to data gathered from surveys, interviews, and performance evaluations across a range of industries.

### **Industry-Specific Findings**

1. Healthcare: VR simulations of medical procedures in the healthcare industry decreased mistakes in challenging tasks and enhanced compliance with safety regulations.
2. Manufacturing: Workers who received VR training demonstrated improved machine operation, hazard identification, and process sequence adherence.
3. Customer service: Communication, problem-solving, and stress management skills were improved through simulated customer interactions.

### **Limitations & Future Improvements**

- Expensive setup fees and restricted availability of VR gear.
- Create portable and reasonably priced VR training programs.
- For more lifelike simulations, incorporate AI and cutting-edge graphics.

### **Conclusion**

The primary objective of this research project was to design and implement a Diabetes Prediction System using Machine Learning methods and to perform a comparative performance analysis of these methods. The proposed approach employed various classification and ensemble learning algorithms, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), and Gradient Boosting. Among these, Logistic Regression achieved the highest classification accuracy of 83%, outperforming other classifiers. The experimental results demonstrate that the proposed model can effectively assist healthcare professionals in making early predictions and timely decisions, thereby improving diabetes management and potentially saving human lives. Future work can include integrating deep learning approaches, expanding the dataset with real-time patient data, and deploying the system as a web or mobile application to assist in large-scale healthcare monitoring. Overall, the study demonstrates that combining multiple machine learning algorithms can result in a robust and accurate system for early diabetes prediction, contributing to better healthcare outcomes and life-saving decisions.

### **References**

1. R. Punithavel, Dr. T. Ramesh “Artificial Intelligence Based Hybrid Predictive Analysis of Diabetes Mellitus”, International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211, Volume 12, Issue 1, January-2024.
2. RituVerma, Kishwor Bhandari, Sanjay Prasad Sah, VaraganthamAnitha, Yogita Deepak Mane, Punithavel.R, Sudipta Banerjee, “AI in Healthcare: Predicting Patient Outcomes Using Machine Learning Techniques” Afircan journal of Biological Science, Volume 6, Issue Si4, 2024
3. Arwatki Chen Lyngdoh, Nurul Amin Choudhury, Soumen Moulik, “Diabetes Disease Prediction Using Machine Learning Algorithms”, IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES) 2020, DOI: 10.1109/IECBES48179.2021.9398759.
4. Ashwini r, s m aieshaafshin, kavya v, deepthiraj “diabetes prediction using machine learning”ijrti | volume 7, issue 7 |, 2022.
5. Shejal Kale, PritiRahane, Mansi Ghumare, SnehalPatilB “Diabetes Prediction Using Different Machine Learning Approaches”IJSDR | Volume 7 Issue 5 ,2022.
6. Tejas N. Joshi, Prof. Pramila M. Chawan, “Diabetes Prediction Using Machine Learning Techniques”.Int. Journal of Engineering Research and Application, Vol. 8.
7. Nonso Nnamoko, Abir Hussain, David England, “Predicting Diabetes Onset: an Ensemble Supervised Learning Approach “. IEEE Congress on Evolutionary Computation (CEC), 2018

8. Deeraj Shetty, Kishor Rit, Sohail Shaikh, Nikita Patil, “Diabetes Disease Prediction Using Data Mining “.International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 2017.
9. Nahla B., Andrew et al,”Intelligible support vector machines for diagnosis of diabetes mellitus. Information Technology in Biomedicine”, IEEE Transactions. 14, (July. 2010), 1114-20.
10. A.K., Dewangan, and P., Agrawal, “Classification of Diabetes Mellitus Using Machine Learning Techniques,” International Journal of Engineering and Applied Sciences, vol. 2, 2015.
11. AzraRamezankhani, Omid Pournik, Jamal Shahrabi, FereidounAzizi and FarzadHadaegh, ”An Application of Association Rule Mining to Extract Risk Pattern for Type 2 Diabetes Using Tehran Lipid and Glucose Study Database”, Int J Endocrinol Metab, April 2015.