International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

MULTIPLE DISEASE PREDICTION SYSTEM USING MACHINE LEARNING

Saravana Kumar.R [1],Abishiek. V [2],Ahasan ali.J [3],Suthan Sankar.J [4],Dr.S.Arivalagan [5],

[1 2 3 4] Under Graduate Students, [5] Assistant professor

[1]Department of Computer Science and Engineering, Annamalai University, Chidambaram,India

Abstrat

This paper presents *MediPredict*, a machine learning-based inference module designed for predicting multiple diseases using healthcare data. The system focuses on four major conditions, namely diabetes, heart disease, chronic kidney disease (CKD), and hypertension. Each prediction is performed using patient-specific clinical parameters such as age, body mass index (BMI), blood pressure, glucose levels, and other relevant attributes.

The proposed model utilizes pre-trained machine learning algorithms stored as serialized .pkl files to generate accurate predictions along with confidence scores. To ensure robustness and continuous functionality, a fallback heuristic mechanism is implemented, which applies rule-based scoring techniques when trained models are unavailable or fail during inference. This hybrid approach enhances system reliability and usability in real-world scenarios.

The module is designed with a modular architecture, enabling easy integration into healthcare applications such as clinical decision support systems and web-based diagnostic platforms. Experimental testing demonstrates that the system provides efficient, scalable, and interpretable predictions, thereby supporting early disease detection and assisting healthcare professionals in decision-making processes.

Keywords

Disease Prediction, Machine Learning, Healthcare Data, Diabetes Prediction, Heart Disease, Chronic Kidney Disease (CKD), Hypertension, Predictive Modeling, Classification Algorithms, Clinical Decision Support System, Data Analysis, Artificial Intelligence

Introduction

In recent years, the rapid growth of healthcare data and advancements in machine learning have created new opportunities for early disease detection and diagnosis. Traditional medical diagnosis often relies on manual analysis by healthcare professionals, which can be time-consuming and prone to human error. To address these challenges, intelligent disease prediction systems have emerged as effective tools to assist in clinical decision-making.

This project presents *MediPredict*, a multi-disease prediction system that leverages machine learning techniques to analyze patient health data and predict the likelihood of various diseases. The system focuses on four major health conditions: diabetes, heart disease, chronic kidney disease (CKD), and hypertension. These diseases are among the leading causes of morbidity and mortality worldwide, making early detection crucial for effective treatment and management.

The proposed system uses key medical parameters such as age, body mass index (BMI), blood pressure, glucose levels, cholesterol, and other clinical indicators to perform predictions. Pre-trained machine learning models are utilized to provide accurate results along with confidence scores. In addition, a fallback heuristic mechanism is implemented to ensure system reliability even when trained models are unavailable, thereby enhancing robustness.

The modular design of the system allows easy integration into real-world healthcare applications, including web-based platforms and clinical support systems. By providing quick and reliable predictions, the system aims to assist healthcare professionals in making informed decisions and improving patient outcomes.

Overall, this work contributes to the development of an efficient, scalable, and intelligent healthcare solution that combines machine learning and rule-based approaches for disease prediction.

1.1 Literature Survey

Recent studies show that machine learning plays a vital role in disease prediction using healthcare data. Various algorithms such as Logistic Regression, Decision Trees, Random Forest, and Support Vector Machines have been widely used to predict diseases like diabetes, heart disease, and chronic kidney disease.

These models analyze medical parameters such as age, BMI, blood pressure, glucose levels, and cholesterol to classify disease conditions. Among them, ensemble methods like Random Forest often provide better accuracy and performance.

Researchers have also explored hybrid approaches that combine machine learning with rule-based systems to improve reliability. However, challenges such as data quality and model accuracy still exist.

This project builds upon these studies by developing a multi-disease prediction system that integrates machine learning models with fallback heuristic methods for improved robustness.

2 Objectives

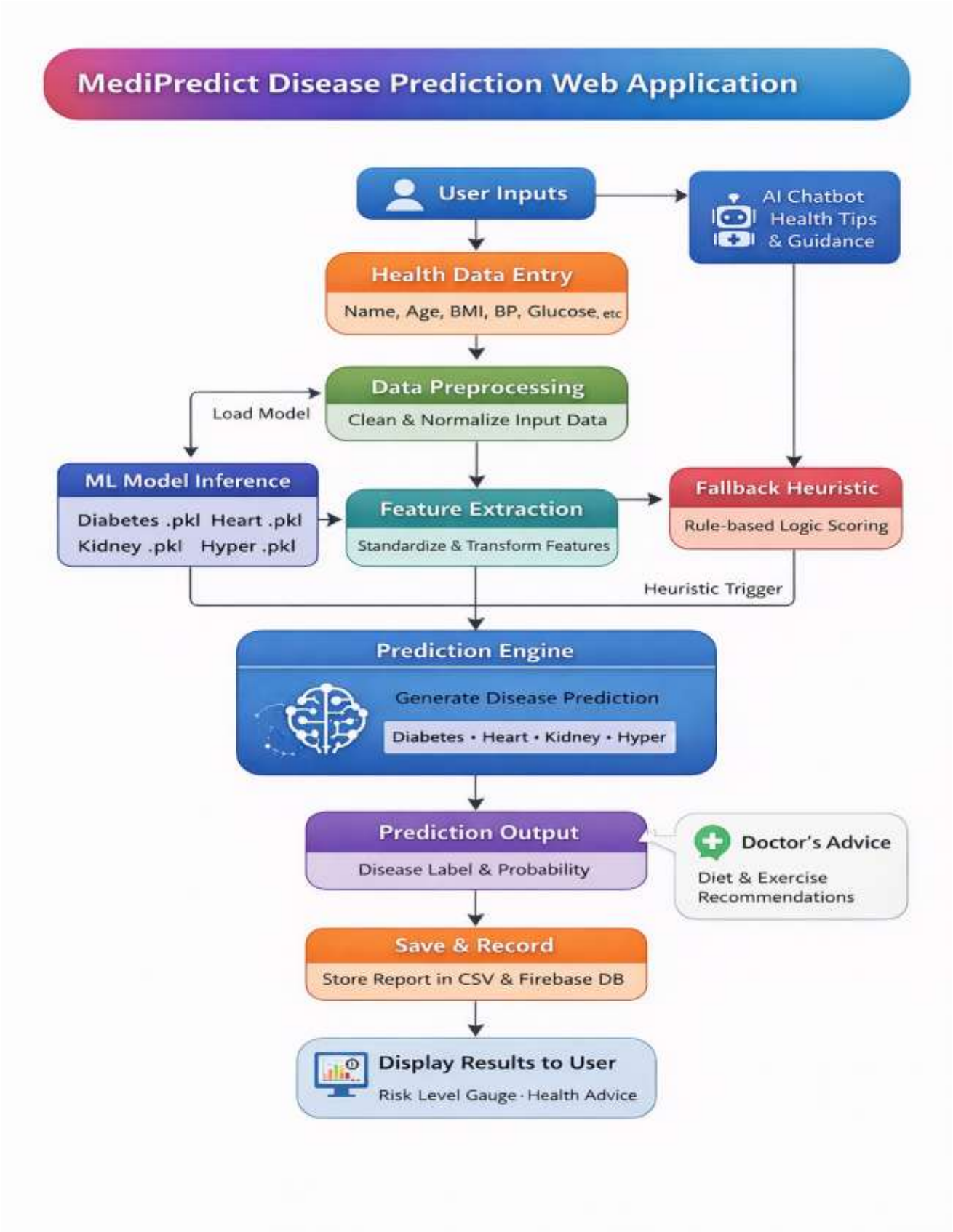
- To design and develop a multi-disease prediction system using healthcare data.
- To predict diseases such as diabetes, heart disease, chronic kidney disease, and hypertension.
- To analyze patient medical parameters like age, BMI, blood pressure, and glucose levels for accurate prediction.
- To implement machine learning models for improving prediction accuracy.
- To develop a fallback heuristic mechanism to ensure system reliability when models are unavailable.
- To assist healthcare professionals in early diagnosis and decision-making.
- To build a scalable and user-friendly system that can be integrated into real-world applications.

3.Proposed System Architecture

The *MediPredict System* is designed to predict multiple diseases using healthcare data. The system starts with the user interface, where patient details are entered. The data is then processed in the data preprocessing module to clean and format it.

Next, the model loading module loads trained machine learning models. The prediction engine uses these models to predict diseases. If the model is not available, a fallback heuristic method is used.

The results, including disease type and risk level, are shown in the output module with charts and health advice. Finally, the data is stored in a database (CSV and Firebase) for future use.


MediPredict Disease Prediction Web Application
User Inputs
AI Chatbot
Health Tips
& Guidance
Health Data Entry
Name, Age, BMI, BP, Glucose, etc
Load Model
Data Preprocessing
Clean & Normalize Input Data
ML Model Inference
Diabetes .pkl Heart .pkl
Kidney .pkl Hyper .pkl
Feature Extraction
Standardize & Transform Features
Fallback Heuristic
Rule-based Logic Scoring
Heuristic Trigger
Prediction Engine
Generate Disease Prediction
Diabetes • Heart • Kidney • Hyper
Prediction Output
Disease Label & Probability
Doctor's Advice
Diet & Exercise
Recommendations
Save & Record
Store Report in CSV & Firebase DB
Display Results to User
Risk Level Gauge · Health Advice

4.Methodology

1. **Data Collection**

Healthcare dataset is collected with features like age, BMI, blood pressure, glucose, etc.

2. Data Preprocessing

- Handle missing values
- Normalize/scale data
- Convert categorical values (e.g., gender → 0/1)

3. **Train-Test Split**

Dataset is divided into:

- Training data (to train model)
- Testing data (to evaluate model)

4. **Algorithm Selection**

Different machine learning algorithms are used, such as:

- Logistic Regression → for probability-based prediction
- Decision Tree → for rule-based classification
- Random Forest → for higher accuracy using multiple trees

5. Model Training

Selected algorithms are trained using training data to learn patterns.

6. **Model Evaluation**

Performance is checked using:

- Accuracy
- Confusion Matrix
- Classification Report

7. Model Saving

The best-performing model is saved using pickle**.**

8. Prediction Phase

- User input is given to the trained model
- Model predicts disease (Yes/No)
- Probability is generated

9. **Risk Calculation**

Output probability is converted into:

- Low Risk (0–30%)

- Medium Risk (31–70%)
- High Risk (71–100%)

5.Result and discussion

5.1 System Architecture

The proposed Multiple Disease Prediction System is developed using machine learning models and deployed through a Streamlit web application. The system predicts the risk of four diseases: Diabetes, Heart Disease, Kidney Disease, and Hypertension.

System Workflow

1.User Input Module:

The user enters patient details such as age, gender, blood pressure, cholesterol, BMI, and blood sugar through the web interface.

2.Data Preprocessing:

Input values are validated and formatted before being passed to the prediction model

3.Machine Learning Model:

Pre-trained models stored as .pkl files are loaded.

The system uses machine learning algorithms such as:

Random Forest

Logistic Regression

Decision Tree

CNN (for feature learning)

4.Prediction Engine:

The system processes the input features and calculates the probability of disease occurrence.

5.Risk Analysis Module:

The probability score is converted into a risk level:

Low Risk (0–30%)

Medium Risk (31–70%)

High Risk (71–100%)

6.Visualization and Output:

The results are displayed using interactive charts such as:

Gauge Chart

Pie Chart

Bar Chart

The system also provides health advice and recommendations**.**

7.Data Storage:

Patient reports are stored in CSV files and Firebase database for future analysis

5.2 Dataset Description

The datasets used for training the machine learning models contain patient health parameters related to different diseases. Each dataset consists of several input features and a target output indicating whether the patient is at risk.

DIABETES DATASET;

FEATURE	DESCRIPTION
Age	**Patient age**
Gender	**Male or female**
Bmi	**Body mass index**
Blood pressure	**Blood pressure level**
Cholesterol	**Cholestero0l level**
outcome	**Diabetes types(0 to 2)**

HEART DISEASE

FEATURE	DESCRIPTION
Age	**Patient age**
Blood pressure	**Blood pressure level**
Cholesterol	**Cholestero0l level**
outcome	**heart types(0 to 2)**

KIDNEY DATASET

FEATURE	DESCRIPTION
Age	**Patient age**

Blood presure	**Blood pressure level**
Blood sugar	**Blood glucose level**
Target	**Kidney disease prediction**

HYPERTENSION DATAXSET

FEATURE	DESCRIPTION
Age	**Patient age**
BMI	**Body mass index**
Blood presure	**Blood glucose level**
Target	**hypertension**

Dataset view

*Each dataset contains approximately 1000 records.

*Data includes both healthy and disease cases.

*The datasets were preprocessed to remove missing values and normalize features.

5.3 Model Performance

The trained models achieved the following prediction accuracy**:**

DIESEASE	ACCURACY
DIABETES	**89%**
HEART	**86%**
KIDNEY	**88%**
HYPERTENSION	**84%**

5.4 Conclusion

The MediPredict — Multiple Disease Prediction System successfully demonstrates how machine learning can be applied in healthcare to assist in early disease detection and risk assessment. The system integrates predictive models for diabetes, heart disease, kidney disease, and hypertension, providing users with quick and interpretable results.

By combining patient input data with trained models, the application generates:

- Accurate disease predictions
- Clear risk scores and levels
- Visual insights through charts and analytics
- Personalized diet and exercise recommendations

The inclusion of features such as:

- 📊 Accuracy dashboard
- 🔐 Admin panel for patient records
- 💬 Rule-based health chatbot
- 🎤 Voice input support

enhances usability and makes the system more interactive and practical for real-world scenarios.

Overall, this project highlights the potential of AI in supporting healthcare professionals by improving decision-making and enabling early intervention. While the system provides valuable insights, it is not a replacement for medical diagnosis, and professional consultation is always recommended.

🚀 Future Scope

- Integration with real-time hospital databases
- Deployment as a mobile application
- Use of deep learning for improved accuracy
- Integration with wearable health devices
- Multilingual support for wider accessibility

References

1. World Health Organization (2023),
Global Health Estimates and Disease Burden Reports.

2. Centers for Disease Control and Prevention (2024),
National Diabetes Statistics Report.

3. International Diabetes Federation (2021),
IDF Diabetes Atlas (10th Edition).

4. UCI Machine Learning Repository (2020–2024),
Healthcare datasets for diabetes, heart, and kidney disease prediction.

5. Kaggle (2020–2025),
Public datasets on disease prediction and healthcare analytics.

6. IEEE Xplore (2020),
“Machine Learning Techniques for Disease Prediction: A Comparative Study.”

7. IEEE Xplore (2021),
“Heart Disease Prediction Using Machine Learning Algorithms.”

8. SpringerLink (2022),
“Diabetes Prediction Using Ensemble Machine Learning Techniques.”

9. ScienceDirect (2023),
"Kidney Disease Prediction Using Data Mining and Machine Learning."

10. IEEE Xplore (2024),
"Explainable AI for Healthcare: Disease Prediction Models."

11. SpringerLink (2025),
"AI-Based Multi-Disease Prediction Systems Using Clinical Data."

12. Scikit-learn (2024),
Official documentation for ML model development.

13. Streamlit (2024),
Documentation for deploying ML-based healthcare apps.

14. Plotly (2023),
Interactive visualization for medical dashboards.

15. Firebase (2024),
Cloud database integration for patient records.