



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 6, June 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Multi Disease Prediction using Machine Learning Algorithms

Bharath C¹, Deekshitha G P¹, Deepak M P¹, Gagan K R¹, Mohan Kumar K S²

Final Year UG Students, Department of Electronics and Communication Engineering BGSIT, B G Nagar, Mandya,
Karnataka, India¹

Assistant Professor, Department of Electronics and Communication Engineering BGSIT, B G Nagar, Mandya,
Karnataka, India²

ABSTRACT: Multi Disease Prediction using Machine Learning Algorithms is an advanced healthcare application that utilizes machine learning algorithms to accurately predict the likelihood of a patient having multiple diseases based on their medical history. The system employs a comprehensive dataset of medical records of various diseases, which are then analysed using machine learning techniques such as decision trees, support vector machines, and random forests. The system's predictions are highly accurate, and it can assist medical professionals in making more informed decisions and providing better treatment plans for patients. The Multiple Disease Prediction System utilizing Machine Learning holds promise for advancing healthcare outcomes and cost reduction through early disease prediction and prevention. The objective of this initiative is to create a sophisticated disease predicting model employing machine learning algorithms, including Random Forest, Decision Tree, and SVM. This model specifically targets prevalent diseases such as Breast cancer, Diabetes, Liver disease, and heart disease. By leveraging the capabilities of these algorithms, the system aims to provide accurate predictions, enabling proactive interventions and personalized healthcare strategies.

KEYWORDS: Brest cancer, Diabetes, Heart disease, Liver disease, Support Vector Machine, Decision Tree, Random Forest

I. INTRODUCTION

Recent years in the field of machine learning has witnessed remarkable advancements and applications in various domains, including healthcare. This proposed paper explores the utilization of the Support Vector Machines (SVM) model to predict the presence of prevalent diseases: heart disease, diabetes, Breast Cancer and Liver disease. These diseases are significant public health concerns that impose a considerable burden on individuals and healthcare systems worldwide. Prediction and accurate diagnosis of these diseases play a vital role in improving patient prognosis, optimizing treatment plans, and reducing healthcare costs. In the dynamic landscape of healthcare, the integration of advanced technologies has emerged as a pivotal force in transforming disease detection and management. Among these technological advancements, Machine Learning (ML) has shown remarkable potential in revolutionizing medical diagnostics. This advanced healthcare application enhances decision-making for medical professionals, promoting early disease detection and personalized treatment plans. The system's accurate predictions can improve healthcare outcomes and reduce costs through proactive interventions and optimized resource utilization.

II. LITERATURE SURVEY

The burgeoning field of machine learning (ML) and its applications in medical diagnosis have shown significant promise. Electrocardiography (ECG) plays a crucial role in heart rate analysis and cardiovascular disease diagnosis. The study by paper [1] leverages the MIT-BIH arrhythmia dataset to evaluate ML methods like Decision Tree, Logistic Regression, Random Forest, K-Nearest Neighbor, and Support Vector Machine (SVM) for ECG signal classification, finding Random Forest to achieve a superior accuracy of 98%. The significance of AI in disease prediction is emphasized in paper [2], which reviews statistical, machine, and deep learning methods. The authors of the paper [4], underlines the impact of AI on diagnostic accuracy and predictive capabilities, facilitated by extensive digital data collection in healthcare. An evaluation of eight ML algorithms for early breast cancer diagnosis found Deep Learning Model 2 and SVM with ReLU activation functions to achieve the highest test accuracy on the Kaggle dataset in paper

[3]. Research in paper [5] presents a multi-disease prediction system using CNN, Random Forest, and Logistic Regression, implemented via Flask API and Python ML libraries for rapid diagnosis. The surveyed studies underscore the transformative potential of ML in enhancing diagnostic accuracy and early disease detection across various medical fields.

III. METHODOLOGY AND SYSTEM DESIGN

Fig. 1 illustrates the use of various machine learning algorithms (Random Forest, SVM, and Decision Tree) to predict multiple diseases, enhancing the interaction between patients and doctors.

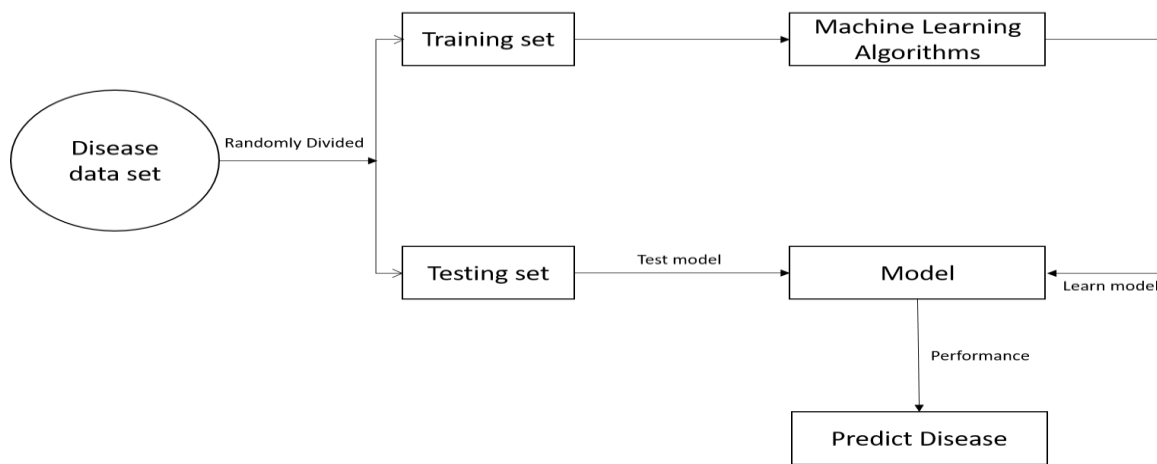


FIG. 1 Block diagram of Proposed System

The Advantages of the Proposed System include the following:

- Early Disease Prediction: Leverages machine learning for predicting the likelihood of multiple diseases, enhancing diagnostics.
- Comprehensive Dataset Analysis: Utilizes a broad dataset of medical records and symptoms for thorough analysis and accurate predictions.
- Machine Learning Techniques: Employs decision trees, SVM, and random forests to handle complex data relationships, boosting prediction accuracy.

A. Workflow of proposed work

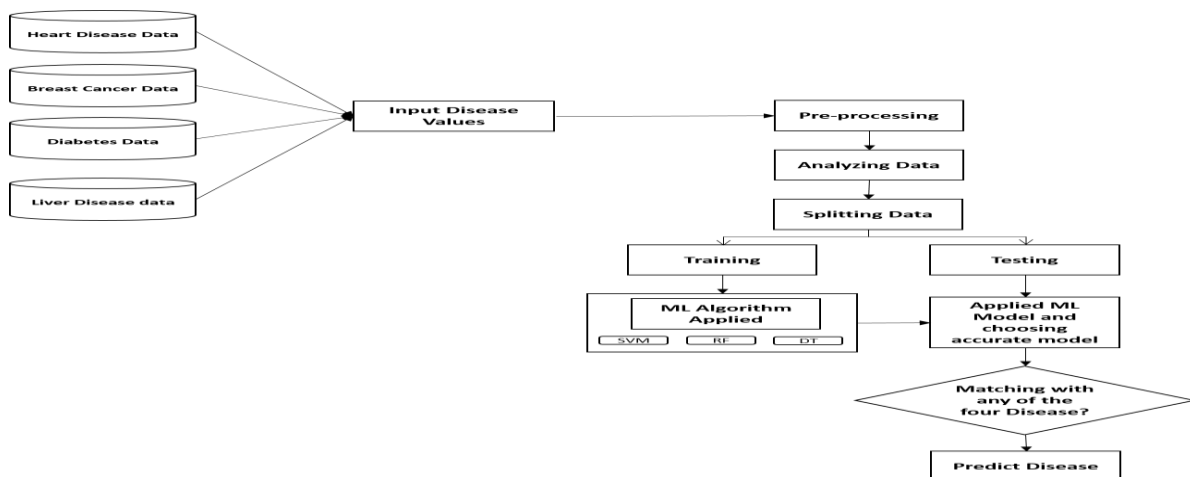


FIG. 2 Flow Chart

Step 1: Input Disease Values

- This step involves collecting data for various diseases. The data could include things like a patient's age, weight, blood pressure, and family history.

Step 2: Pre-processing

- The data collected in step 1 might not be usable in its raw format. This step involves cleaning and organizing the data so that the machine learning algorithm can understand it. For instance, this might involve scaling numerical data or encoding categorical data.

Step 3: Splitting Data

- Once the data is pre-processed, it's divided into two sets: a training set and a testing set.

Step 4: Training

- The machine learning algorithm is trained on the training data. During this process, the algorithm learns to identify patterns in the data that can be used to predict diseases.

Step 5: Applying ML Algorithm

- Once the model is trained, it's applied to the testing data. The model makes predictions about the diseases for the patients in the testing data.

Step 6: Choosing an Accurate Model

- The performance of the model is evaluated on the testing data. This might involve metrics like accuracy, precision, and recall. These metrics are used to determine how well the model performed.

Step 7: Matching with any of the four Diseases?

- Once a suitable model is obtained, new data can be input into the model. The model will then predict which of the four diseases (heart disease, breast cancer, liver disease, or diabetes) the new data point corresponds to.

Step 8: Predict Disease

- Finally, the model outputs a disease prediction based on the patterns it learned from the training data. It's important to note that these are just predictions, and a medical professional should always be consulted for diagnosis.

B. Machine Learning Models

1. Support Vector Machine: SVM is a robust and efficient algorithm that is well-suited for a wide range of applications. Its ability to handle high-dimensional data and non-linear relationships through kernel tricks makes it a powerful tool in the machine learning toolbox.
2. Random Forest: Random Forest is a powerful machine learning technique that creates multiple decision trees using random subsets of data and features, reducing overfitting and improving prediction accuracy.
3. Decision Tree: A decision tree is a flowchart-like structure used for both classification and regression, where internal nodes represent features, branches represent rules, and leaf nodes denote outcomes.

IV. HARDWARE AND SOFTWARE REQUIREMENT**A. Hardware Requirements:**

Modern Operating System:

- Windows 7 or 10
- Mac OS X 10.11 or higher, 64-bit
- Linux: RHEL 6/7, 64-bit (almost all libraries also work in Ubuntu)
- x86 64-bit CPU (Intel / AMD architecture). ARM CPUs are not supported.
- 4 GB RAM
- 5 GB free disk space

B. Software Requirement:

- Python Programming Language
- Python IDLE

V. RESULT

The graph here shows the Features Importances. where these 11 elements are considered for processing the proposed system.

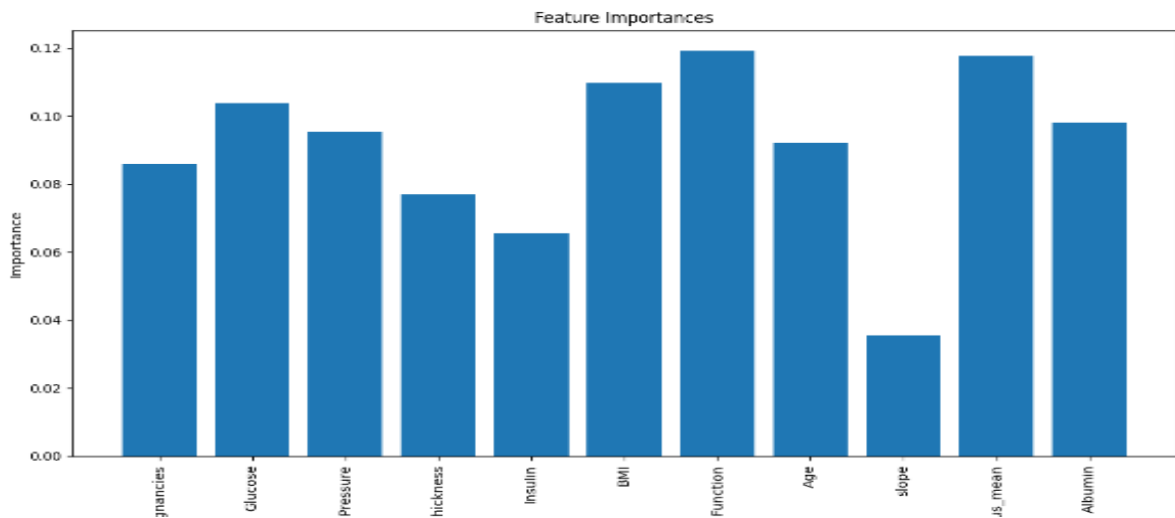


FIG. 3 Features Importances of dataset

The chart appears to show the feature importance of dataset of a machine learning model for predicting multi disease. The features are listed on the X-axis, and their importance scores are listed on the Y-axis.

The features listed on the X-axis include: "Pregnancies", "Glucose", "Pressure", "Thickness", "Insulin", "BMI", "Function", "Age", "slope", "radius_mean", "Albumin".

Following are the accuracy tables for each algorithm:

Table 1. Accuracy and other metrics for Algorithms

| Algorithm | Accuracy | Precision Rate | Recall Rate | F1 Score |
|-----------|----------|----------------|-------------|----------|
| SVM | 93.3 | 34.6 | 66.6 | 57.9 |
| RF | 79.3 | 64.0 | 56.6 | 56.7 |
| DT | 61.1 | 61.02 | 48.8 | 52.8 |

The above Table I. shows accuracy and other metrics for each algorithm. For every disease we have trained with different machine learning models, out of that best accuracy model is taken.

Following are the performance matrix for each algorithm:

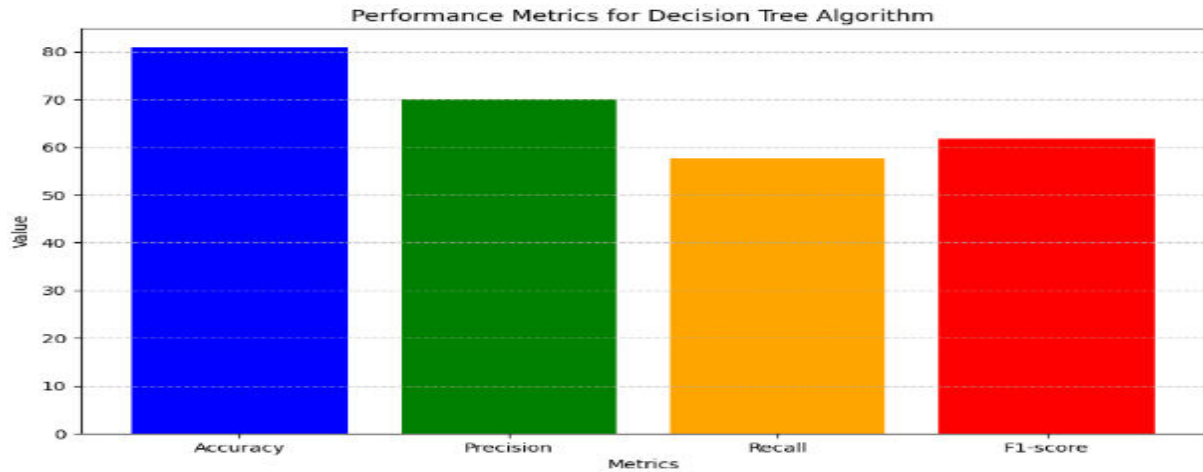


FIG.4 Performance Metrics for Decision Tree

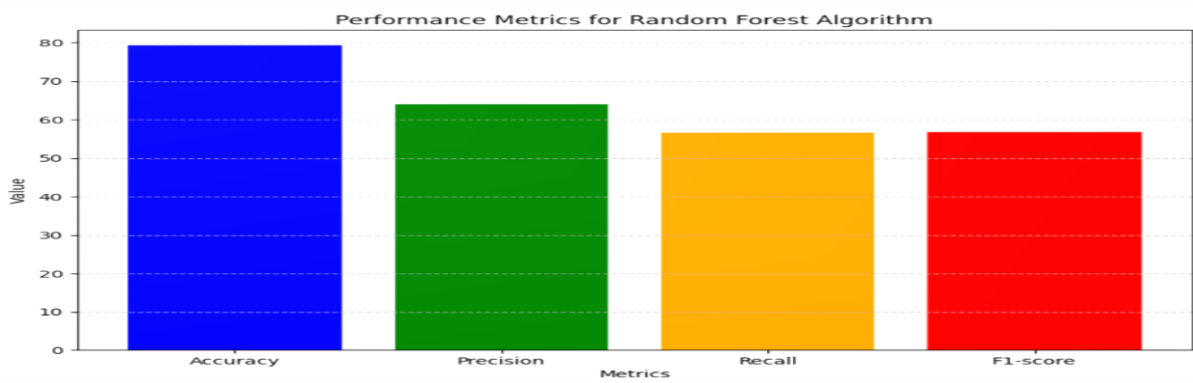


FIG.5 Performance Metrics for Random Forest

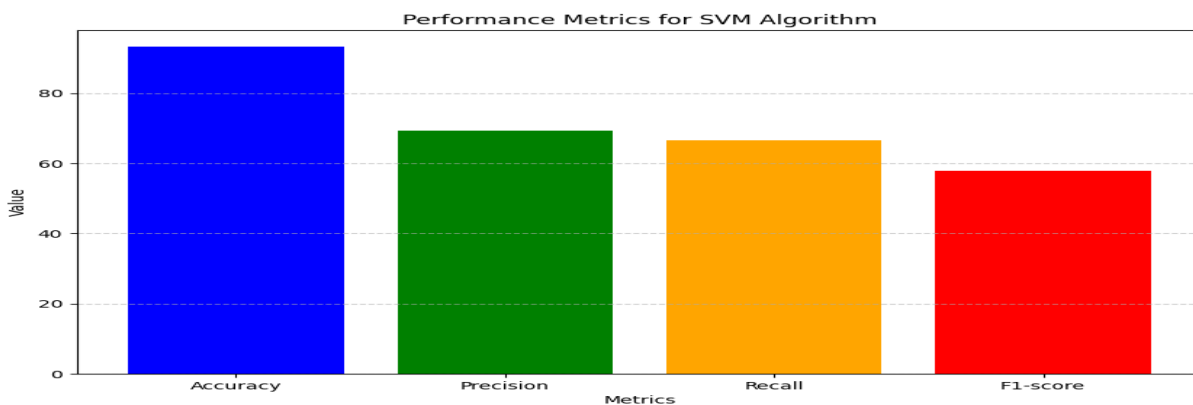


FIG.6 Performance Metrics for SVM

Figure 4, Figure 5 and Figure 6 show the comparisons of Accuracy, Precision, Recall and F1-score for each of the models implemented for the diseases respectively.

A User Interface is developed through which the end user can enter its respective attributes for prediction.

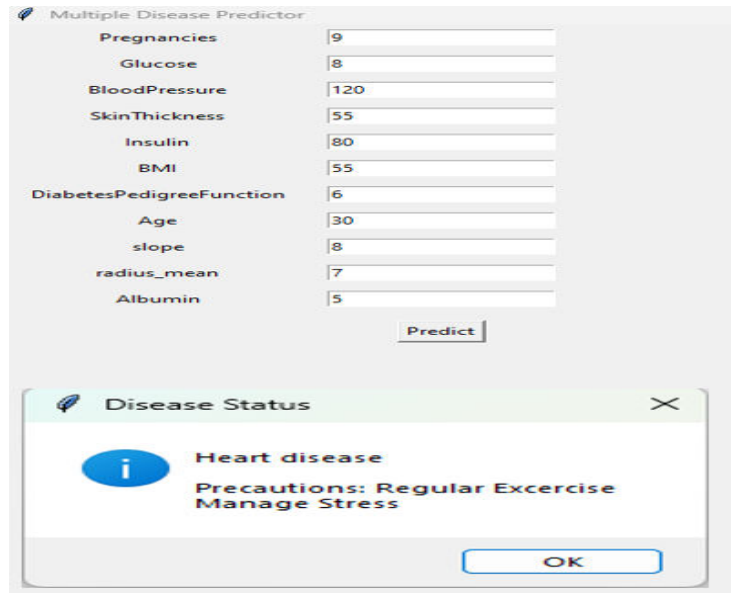


FIG 7 Output for heart disease for positive case

Above Fig 7. displays the positive case for heart disease i.e. (The Person is having heart disease)

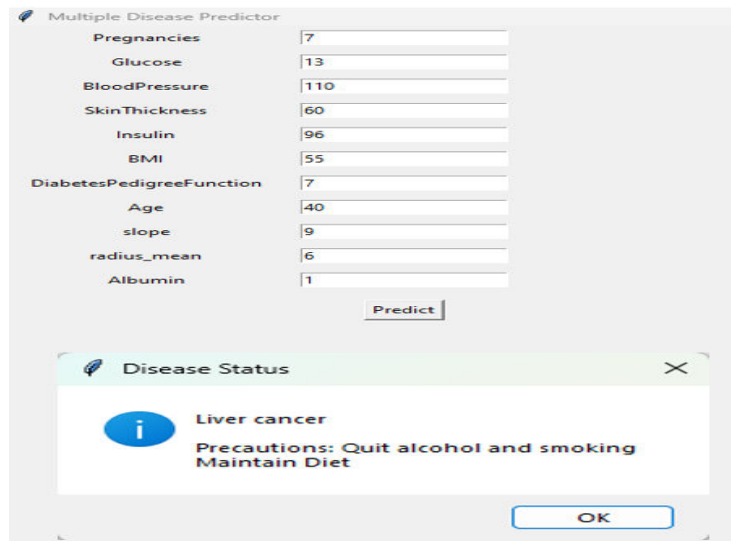
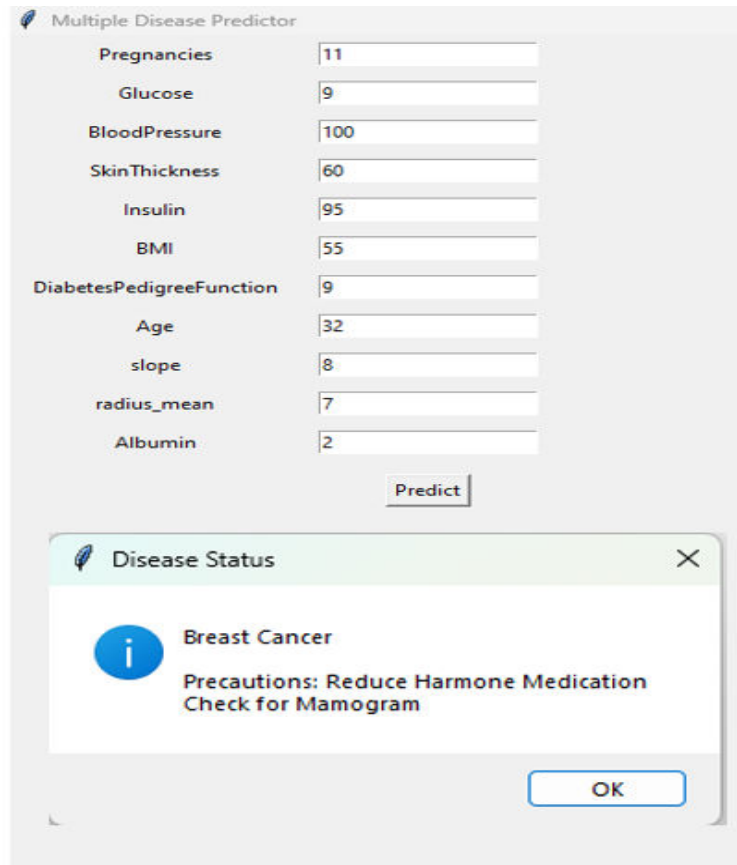


FIG 8 Output for Liver disease for positive case

Above Fig 8. displays the positive case for Liver disease i.e. (The Person is having Liver disease)



| Parameter | Value |
|--------------------------|-------|
| Pregnancies | 11 |
| Glucose | 9 |
| BloodPressure | 100 |
| SkinThickness | 60 |
| Insulin | 95 |
| BMI | 55 |
| DiabetesPedigreeFunction | 9 |
| Age | 32 |
| slope | 8 |
| radius_mean | 7 |
| Albumin | 2 |

Disease Status

Breast Cancer

Precautions: Reduce Hormone Medication
Check for Mamogram

OK

FIG 9 Output for Brest cancer disease for positive case

Above Fig 9. displays the positive case for Brest Cancer disease i.e. (The Person is having Brest cancer disease)

The model uses 11 input attributes to predict the presence of diabetes, heart disease, liver disease, or breast cancer and upon clicking the predict button, it displays the predicted disease (diabetes, heart disease, liver disease, breast cancer) or "No disease" if none are detected.

VI. CONCLUSION

The Multiple Disease Prediction System using Machine Learning offers a transformative approach to healthcare, leveraging advanced algorithms to accurately predict multiple diseases from patient medical history and symptoms. Utilizing decision trees, support vector machines, and random forests, the system achieves high predictive accuracy, aiding medical professionals in informed decision-making and personalized treatment planning. Early disease prediction and prevention facilitated by this system can significantly improve healthcare outcomes and reduce costs. Continued refinement of these algorithms promises further enhancement in prediction accuracy for diseases like breast cancer, diabetes, liver disease, and heart diseases. This comprehensive machine learning approach optimizes healthcare delivery, resource utilization, and the implementation of personalized healthcare strategies.

REFERENCES

1. T. Subba and T. Chingtham, "Comparative Analysis of Machine Learning Algorithms with Advanced Feature Extraction for ECG Signal Classification," in *IEEE Access*, vol. 12, pp. 57727-57740, 2024, doi: 10.1109/ACCESS.2024.3387041.

2. M. Kalaivani and R. Shalini, "Multi-Disease Prediction Techniques using Machine and Deep Learning," 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, 2023, pp. 1-5, doi: 10.1109/INCET57972.2023.10170201.
3. P. Prasad, "The Application of Machine Learning Techniques to the Diagnosis of Breast Cancer," 2023 3rd International conference on Artificial Intelligence and Signal Processing (AISP), VIJAYAWADA, India, 2023, pp. 1-7, doi: 10.1109/AISP57993.2023.10134824.
4. A. Sharma, J. Pathak and P. Rajakumar, "Disease Prediction using machine learning algorithms," 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2022, pp. 995-999, doi: 10.1109/ICACITE53722.2022.9823744.
5. S. Narayanan, N. M. Balamurugan, M. K and P. B. Palas, "Leveraging Machine Learning Methods for Multiple Disease Prediction using Python ML Libraries and Flask API," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2022, pp. 694-701, doi: 10.1109/ICAAIC53929.2022.9792807.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details