ISSN: 2582-7219

# International Journal of Multidisciplinary Research in Science, Engineering and Technology

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

Impact Factor: 8.206

Volume 8, Issue 5, May 2025

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

# Multiple Disease Prediction using Machine Learning

**Vanipenta.Anitha Reddy[1], Podila.Lahari[2], Valapatla.Naga Swetha[3], Shaik.Sirajuddin[4]**

Undergraduate Students, Department of ECE, R.V.R & J.C College of Engineering, Chowdavaram, Guntur A.P., India

**ABSTRACT:** The Multiple Disease Prediction System is an interactive web-based application developed using Python and Stream lit, designed to assist in the early detection of various critical diseases including diabetes, heart disease, parkinson's disease, and breast cancer. The system integrates multiple pre-trained machine learning models like Support Vector Machine (SVM) for Diabetes, Logistic Regression for Heart Disease, Support Vector Machine (SVM) for parkinson 's disease, and Logistic Regression for Breast Cancer to analyze user-inputted medical parameters and return instant diagnostic predictions. Each disease module in the application takes specific input features relevant to that condition and processes them through its respective model to determine the likelihood of the disease being present. This platform serves as a prototype for a multi-disease diagnosis assistant that enhances healthcare accessibility and awareness. By streamlining the prediction process and enabling users to receive quick feedback, the system empowers individuals to take proactive steps toward their health. The user-friendly interface makes the tool approachable even for those with minimal technical background. Overall, this project demonstrates the impactful convergence of machine learning and healthcare, aiming to support early diagnosis and prompt medical consultation, thereby potentially reducing the risk of severe health outcomes.

**KEYWORDS:** Machine Learning, Disease Prediction, Support Vector Machine (SVM), Logistic Regression, Stream lit, Diabetes, Heart Disease, Parkinson's Disease, Breast Cancer, Web-based Application.

## I. INTRODUCTION

In today's rapidly evolving healthcare landscape, early and accurate disease detection plays a pivotal role in improving patient outcomes and reducing the burden on medical infrastructure. With the advancement of artificial intelligence and machine learning, there is growing potential to enhance traditional diagnostic methods by integrating intelligent systems capable of processing and interpreting complex clinical data. This project presents a streamlined, interactive web-based platform developed using Stream lit that enables users to predict the likelihood of multiple critical diseases namely, diabetes, heart disease, parkinson's disease, and breast cancer based on clinical inputs.

The system leverages pre trained machine learning models to analyze patient-specific parameters and deliver near-instant diagnostic predictions. Specifically, it employs a Support Vector Machine (SVM) model for diabetes prediction, Logistic Regression for heart disease detection, a Support Vector Machine (SVM) for identifying parkinson's disease, and Logistic Regression for breast cancer classification. Each disease module within the application is designed to capture the most relevant medical indicators through a user friendly interface, reducing the need for specialized medical knowledge to interact with the system. By incorporating multiple diseases into a single platform, this solution enhances accessibility, efficiency, and scalability making it a practical tool for both patients and healthcare professionals.

Ultimately, this project aims to bridge the gap between data driven healthcare and end user usability. The integration of predictive analytics with a simplified interface not only supports early intervention but also empowers individuals to take proactive steps in managing their health. Through this research, we hope to demonstrate the tangible benefits of multi disease diagnostic tools and encourage broader adoption of AI-powered systems in preventive healthcare.

## II. LITERATURE SURVEY

Machine learning techniques are increasingly effective in predicting various diseases, including diabetes, heart disease, parkinson's, and breast cancer. Studies have demonstrated high accuracy using algorithms such as SVM, Logistic Regression, and Tensor Flow with Keras, often deployed through user-friendly platforms like stream lit. These systems

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

enhance early diagnosis and accessibility but still face challenges, including limited disease coverage, input complexity, lack of machine learning model interpretability, and deployment hurdles. Overall, while current models show strong potential, future research should focus on developing more integrated, scalable, and interpretable solutions for real-world healthcare applications.

Parshant and Dr. Anu Rathee (2023)[1] focused on using ML algorithms like SVM, Random Forest, and Logistic Regression to predict diseases such as heart disease and diabetes. Their work stands out for its structured comparison of model performance and high accuracy for common illnesses. However, it only addresses three diseases and does not cover aspects like deployment or user interface design.

Prof. Vivek Pandey and his team (2023)[2] combined ML, deep learning, and big data to develop a scalable and adaptable disease prediction system. Their model is designed for real-time predictions using cloud-based platforms, making it suitable for large-scale healthcare applications. Still, the research lacks detailed evaluation of individual disease models and offers no insight into practical deployment.

Kallepalli Reshma, Pasumarthi Niharika, Javvadi Haneesha, Kodithala Rajavardhan, Sana Swaroop (2023)[3] built a web-based disease prediction tool using Stream lit. It targeted diabetes, heart disease, and Parkinson's using interpretable models like SVM and Logistic Regression. The system achieved solid accuracy, particularly for Parkinson's (89%) and heart disease (85%). Despite its practical interface, the study did not incorporate advanced methods like deep learning or cover a broader range of diseases.

Sundari V, Shri Kumar A, Vishwa P (2024)[4] investigated various ML techniques to predict diseases like diabetes and Parkinson's, comparing SVM with other models such as Logistic Regression and Random Forest. They also explored feature selection techniques to enhance model performance, with SVM achieving 87% accuracy. However, the study primarily emphasized model accuracy and paid limited attention to deployment or real-world application.

In the conclusion, the literature highlights significant progress in using machine learning for multi-disease prediction, with strong model accuracy across common conditions. However, most studies lack depth in deployment, scalability, and real-world application integration.

## III. PROPOSED METHODOLOGY

The proposed model architecture for the Multiple Disease Prediction System is built around a user-friendly stream lit web application that integrates four separately trained machine learning models to predict diabetes, heart disease, parkinson's disease, and breast cancer. Each model, saved in `.sav` format using `pickle`, is loaded at the beginning of the application. Through a sidebar menu, users can seamlessly navigate between different disease prediction modules.

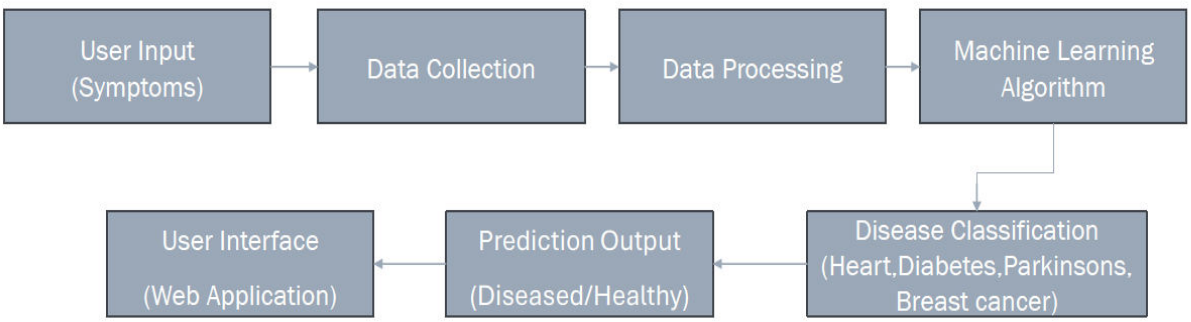



Figure 1: Block diagram for proposed methodology

Each module presents an organized input form where users provide health metrics specific to the disease being assessed. Upon form submission, the corresponding model predicts the likelihood of disease, and the results are

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

instantly displayed on the same page. The app enhances usability with a clean, column-based layout for input fields and a background image for aesthetic appeal, ensuring the entire experience feels interactive, informative, and visually pleasant.

**1. Data Collection**

To begin with, relevant datasets were sourced from Kaggle.com, a well-known platform offering a wide range of high-quality datasets. Separate datasets were selected for each condition-diabetes, heart disease, parkinson's disease, and breast cancer. These datasets contain the necessary clinical features required for training machine learning models to predict each respective disease.

**2. Data Preprocessing**

After collecting the data, a thorough pre processing phase is carried out to prepare it for model training. This involves cleaning the data by handling missing entries, removing duplicate records, and applying feature scaling or normalization techniques. These steps are crucial to ensure consistency and to enhance the performance and accuracy of the machine learning models.

**3. Model Selection**

- **Diabetes – Support Vector Machine (SVM):** For diabetes prediction, the Support Vector Machine was a natural fit due to its ability to handle high-dimensional data and detect complex patterns in relatively small datasets. Since the symptoms and clinical values can vary widely across individuals, SVM helps in drawing a clear boundary between healthy and diabetic cases, even when the differences are subtle.

- **Heart Disease – Logistic Regression:** Logistic Regression was selected for heart disease prediction because of its simplicity and effectiveness in binary classification problems. Medical data related to heart health-like cholesterol levels and blood pressure tends to be well-structured, and logistic regression can model this kind of data well while also offering interpretable results that make sense in a clinical context.

- **Parkinson's disease – Support Vector Machine (SVM):** In the case of Parkinson's disease, the data revolves around vocal measurements, which are often non-linear and complex. SVM is particularly good at handling this kind of data, especially when using non-linear kernels. It can capture the small but meaningful variations in voice patterns that might be missed by simpler models.

- **Breast Cancer – Logistic Regression:** For breast cancer detection, Logistic Regression was the preferred choice because of its speed and high accuracy on structured datasets. The tumor features used for classification such as texture, size, and shape are clearly defined, allowing the model to perform efficiently while keeping the system lightweight and easy to maintain.

**4. Training and Testing**

To ensure reliable and accurate predictions, each machine learning model underwent a structured training and testing process. First, relevant datasets were collected for each disease, containing real patient data with both input features and diagnosis labels. The data was then pre processed this included cleaning, normalization, and selecting the most important features for each condition.

Once the data was ready, it was split into two parts: a training set and a testing set, typically in an 80:20 ratio. The training set was used to teach the model how to identify patterns and relationships between the input features and the disease outcome. During this phase, the model learns to adjust its internal parameters to minimize prediction errors.

After training, the model was evaluated on the testing set, which contains data it has never seen before. This step helps assess how well the model can generalize to new cases. Accuracy, precision, recall, and other metrics were calculated to measure performance. Only models that demonstrated strong results during testing were selected for integration into the final application.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

**5. Model Deployment**

Once the machine learning models were trained and tested, the next step was to make them accessible through a user-friendly interface. To do this, each model was saved using python's pickle module, which allows trained models to be stored and reused without retraining. These saved models were then integrated into a stream lit web application-a lightweight and interactive python framework ideal for deploying data science projects. When a user enters input data into the web app, the corresponding model is loaded in real-time, processes the data, and returns a prediction instantly. This approach ensures fast, responsive interaction while keeping the back-end logic hidden from the user, making the system both powerful and easy to use.

**Key Features:**

- **User Input:** Users can input their medical information, including age, gender, blood pressure, cholesterol levels, and other relevant factors.
- **Disease Prediction:** The application utilizes machine learning models to predict the likelihood of having diabetes, Parkinson's disease, Breast cancer and heart disease based on the inputted medical data.
- **Prediction Results:** The predicted disease outcomes are displayed to the user, providing an indication of the probability of each disease.
- **Visualization:** Visualizations are generated to highlight important features and provide insights into the prediction process.
- **User-Friendly Interface:** The web application offers an intuitive and user-friendly interface, making it easy for individuals without technical knowledge to use the prediction tool.

**6. Input and Output Design**

- **Input Design**

  The input design of the Multiple Disease Prediction system is centred around simplicity, relevance, and user experience. When a user accesses the web application, they are presented with a clear, menu-based interface offering five disease prediction options: heart disease, diabetes, parkinson's disease, and breast cancer. Upon selecting a disease, the application dynamically displays the corresponding input fields required for that specific prediction.

  Each disease requires different clinical parameters-for example, glucose level and BMI for diabetes, or voice frequency features for parkinson's. The interface is designed to only prompt for the most essential and disease-specific inputs, minimizing confusion and making the process intuitive. Input fields are labelled clearly, with appropriate formatting and helpful placeholders to guide users. This design ensures that even individuals without a medical or technical background can provide the necessary information accurately and with ease.

- **Output Design**

  After the user submits the required information, the system processes the data using the corresponding machine learning model and quickly displays the prediction result. The output is designed to be simple, direct, and easy to understand.

  The result is presented in one of two clear messages:
  - "Prediction: The person is affected by [Disease Name]."
  - "Prediction: The person is not affected by [Disease Name]."

  These results are shown immediately on the same page, eliminating any need for the user to navigate elsewhere. The output section is styled to draw attention without overwhelming the interface, ensuring the user can quickly interpret the result. This streamlined design makes it easier for users to take informed next steps based on the prediction whether that means seeking medical advice or simply staying reassured.

  Overall, the input and output designs work hand-in-hand to create an efficient and user-centric experience, making advanced health prediction accessible to everyone.

© 2025 IJMRSET | Volume 8, Issue 5, May 2025| DOI:10.15680/IJMRSET.2025.0805115

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

## IV. RESULTS

The machine learning models in the "Multiple Disease Prediction System" were thoroughly evaluated to assess how well they could predict conditions like Diabetes, Heart Disease, Breast Cancer and Parkinson's Disease. Throughout the training process, the models showed steady improvement, with their performance consistently increasing in both training and validation, suggesting they were learning effectively from the data.

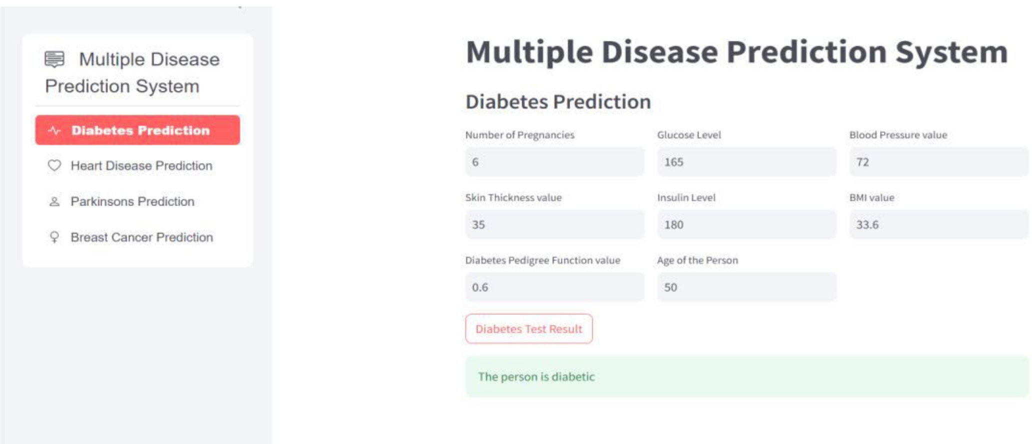
As the training progressed, the models' accuracy steadily went up, while the validation loss decreased, showing they were able to generalize well to new, unseen data. To prevent over fitting, early stopping was used, which automatically halted the training when the validation loss plateaued, ensuring the models didn't simply memorize the training data but learned to make accurate predictions on new cases.

| S. No | Disease Name | Algorithm Name | Existing Accuracy | Proposed Accuracy |
|---|---|---|---|---|
| 1. | Diabetes Disease | SVM Classifier | 0.85 | 0.88 |
| 2. | Heart Disease | Logistic Regression | 0.79 | 0.86 |
| 3. | Parkinson's Disease | SVM Classifier | 0.85 | 0.87 |
| 4. | Breast Cancer Disease | Logistic Regression | 0.91 | 0.98 |

Figure 2: Comparison of Accuracy of all 4 models

**Visual Output Samples**

- **Diabetes Prediction Results:** In the case of diabetes prediction, the model exhibited a strong capability in correctly classifying individuals with and without the condition. The majority of diabetic cases were accurately detected, reflecting the model's sensitivity, while non-diabetic cases were predominantly identified correctly, reducing the rate of false alerts. The limited number of misclassifications highlights the model's reliability and its applicability as a supportive tool in the clinical diagnosis and management of diabetes.

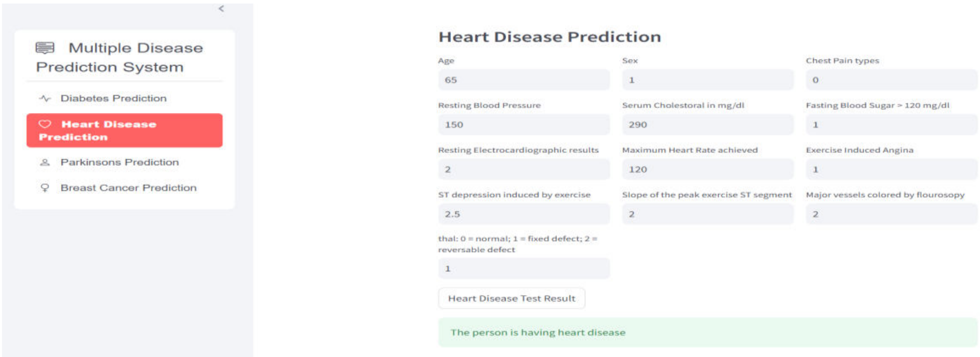



- **Heart Disease Prediction Results:** The prediction results for heart disease reveal that the model successfully achieved a high level of accuracy in distinguishing between affected and unaffected individuals. A significant number of heart disease cases were correctly identified, promoting timely intervention and treatment.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

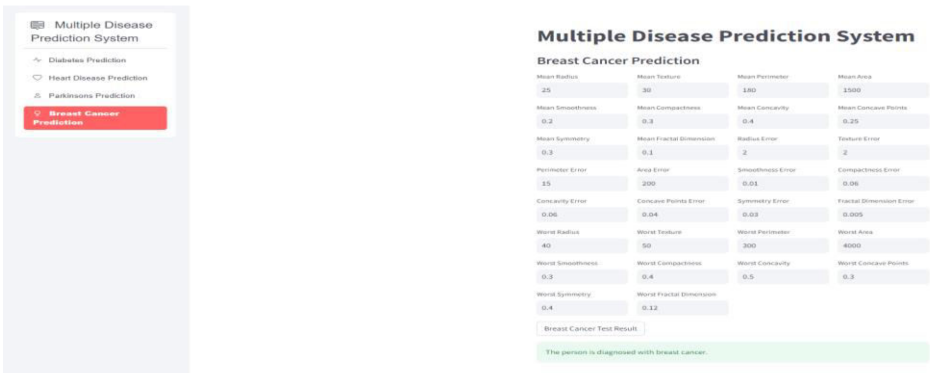
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)




Furthermore, the correct classification of healthy individuals supports the model's effectiveness in minimizing false positives. These outcomes collectively demonstrate the model's potential in contributing to accurate and early heart disease detection.

- **Parkinson's Disease Prediction Results:** For Parkinson's disease prediction, the model demonstrated high precision in identifying individuals affected by the condition. Most patients were correctly classified, ensuring prompt and appropriate clinical attention. Additionally, healthy individuals were largely recognized accurately, reducing the occurrence of erroneous diagnoses. The minimal number of misclassifications indicates the model's strength and highlights its suitability for assisting in the early detection of parkinson's disease.


- **Breast Cancer Prediction Results:** The performance analysis for breast cancer prediction indicates that the model effectively identified the majority of positive cases, demonstrating a strong ability to detect the presence of the disease. Additionally, a high proportion of negative cases were accurately recognized, thereby reducing the likelihood of unnecessary medical interventions. The occurrence of incorrect predictions remained minimal, underscoring the model's robustness and its potential utility in supporting early and accurate breast cancer diagnosis.



ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

## V. CONCLUSION

This research highlights the potential of machine learning techniques in automating the prediction of multiple diseases, including diabetes, parkinson's disease, breast cancer and heart disease. By employing specialized models for each condition and integrating them into an accessible stream lit web application, the system enables users to obtain fast and reliable predictions based on clinical input parameters. The use of disease-specific datasets, along with appropriate pre processing, contributed to the improved accuracy and sensitivity of the models, demonstrating their applicability for real-world healthcare support.

Additionally, the development of a web-based interface enhances the system's accessibility for a wide range of users, including those without technical backgrounds, thereby promoting proactive health management. Evaluation through performance metrics such as accuracy, F1-score, and confusion matrices confirmed the robustness of the models, while the modular design offers scalability for future expansion to additional diseases. Overall, this work demonstrates the promising role of machine learning in facilitating early diagnosis, personalized healthcare, and preventive interventions.

## REFERENCES

1. PARSHANT , DR. ANU RATHEE," Multiple Disease Prediction Using Machine Learning", 2023.
2. Mallula Venkatesh" MULTIPLE DISEASE PREDICTION USING MACHINE LEARNING, DEEP LEARNING AND STREAM-LIT", 2023.
3. Prof. Vivek Pandey, Tanay Tambe , Vipin Vishwakarma , Raheem Ansari "Multiple Diseases Prediction System using ML, Deep Learning and Big Data ",2023.
4. Kallepalli Reshma, Pasumarthi Niharika, Javvadi Haneesha, Kodithala Rajavardhan, Sana Swaroop" MULTI DISEASE PREDICTION SYSTEM USING MACHINE LEARNING ",2023.
5. Rajendra Acharya U, Fujita H, Oh SL, et al. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals, 2017.
6. Sundari, Shri Kumar, Vishwa, "Disease Prediction using Machine Learning", 2024.
7. Sumit Sharma, Mahesh Parmar," Heart Diseases Prediction using Deep Learning Neural Network Model", 2020.
8. Arora S, Aggarwal P, Sivaswamy J, Automated diagnosis of Parkinson's disease using ensemble machine learning, 2017.
9. Al-Mallah MH, Aljizeeri A, Ahmed AM, et al. Prediction of diabetes mellitus type-II using machine learning techniques. Int J Med Inform, 2014.
10. MULTIPLE DISEASE PREDICTION SYSTEM USING MACHINE LEARNING, Mohammed Azeez*1, Mohammed Adnan*2, Muhammed Mehboob*3, Dr. Smitha Patil*4
11. E. Char, et al., "Ethical Considerations in AI-Driven Healthcare," 2020. M. Wild, et al., "Streamlit for Machine Learning: Creating Interactive Web Apps in Python," 2022.
12. Arora S, Aggarwal P, Sivaswamy J. Automated diagnosis of Parkinson's disease using ensemble machine learning. IEEE Trans Inf Technol Biomed. 2017; 21(1):289-299.
13. Arora, P., Chaudhary, S., \& Rana, M. (2020). Prediction of diabetes using machine learning algorithms: A review. Journal of Ambient Intelligence and Humanized Computing, 11(6), 2575-2589.
14. Al-Mallah MH, Aljizeeri A, Ahmed AM, et al. Prediction of diabetes mellitus type-II using machine learning techniques. Int J Med Inform, 2014.