International Journal of Computer Techniques – Volume 12 Issue 4, July - August - 2025

MULTIPLE DISEASE PREDICTION USING MACHINE LEARNING

Vinodhini S
Assistant Professor - IT
Velammal Engineering College
Chennai, Tamilnadu.
vinodhini@velammal.edu.in

Vimala Imogen P
Assistant Professor- IT
Velammal Engineering College
Chennai, Tamilnadu.
vimalaimogenp@velammal.edu.in

Aishwarya A S
UG Scholar - IT
Velammal Engineering College
Chennai, Tamilnadu.
aishwarya21ice@gmail.com

Madhu Bharathi S
UG Scholar - IT
Velammal Engineering College
Chennai, Tamilnadu.
madhubhar13@gmail.com

***Abstract:** This project presents a unified disease prediction system using Streamlit and Python, employing machine learning algorithms like Naïve Bayes, Random Forest, Decision Tree, and SVM to identify conditions such as Heart Disease, Diabetes, and Parkinson's Disease. By inputting basic health parameters like blood pressure, cholesterol, pulse rate, and heart rate, users can get real-time predictions on their health status. The most accurate model is selected and saved using pickling for deployment. Designed for scalability, the system can later include other chronic and skin diseases, helping in early diagnosis, preventive care, and potentially reducing mortality rates. It offers a simple, user-friendly interface, promoting regular health monitoring and increasing awareness about personal well-being. This model supports healthcare professionals by acting as a decision-support tool, reducing diagnostic workload and improving patient engagement. Moreover, it emphasizes the importance of accessible, AI-driven tools in preventive healthcare, especially in remote or underserved regions where early detection can be life-saving.*

***Keywords:** Diabetes, Heart, Liver, KNN, Random Forest, XG Boost., Streamlit,lung cancer ,chronic kidney disease.*

1. INTRODUCTION

In today's digital era, data plays a crucial role across various domains, with the healthcare sector generating massive volumes of patient-related information. This project introduces a unified framework for disease prediction, aiming to overcome the limitations of existing systems that typically focus on a single disease, such as separate platforms for diabetes, cancer, or skin conditions. Our objective is to build an integrated system capable of diagnosing multiple diseases based on user-input symptoms, offering real-time, accurate results.The proposed solution leverages Django for backend development and Streamlit for deployment, integrating machine learning models to assess and predict diseases like Diabetes, Heart Disease, and Malaria, with scope for expanding to more conditions in the future. Additionally, a focused analysis is conducted on Liver, Diabetes, and Heart diseases, considering their interrelation. Users interact with the system by entering symptoms and selecting the disease of concern. The system dynamically invokes the appropriate pre-trained model (saved using Python's pickle module) to evaluate the data and return prediction results. The model not only detects the disease but also determines its severity. If a symptom isn't recognized, the user is prompted to confirm its addition to the database, enhancing the system's adaptability over time. Machine learning algorithms such as Naïve Bayes, K-Nearest Neighbors (KNN), Decision Tree, Random Forest, and Support Vector Machine (SVM) are implemented and compared to identify the most accurate and efficient model. The ultimate aim is to provide a comprehensive, scalable, and intelligent platform for early diagnosis and proactive healthcare management, reducing the need for users to access multiple systems for different diseases.

2. RELATED WORK

The application of Machine Learning (ML) in healthcare has gained considerable attention due to its capacity to derive meaningful insights from vast medical datasets. In recent years, researchers have developed numerous disease prediction systems, but most are restricted to identifying a single disease, such as diabetes, cancer, or heart conditions. The limitations of these one-disease-per-model systems restrict scalability and efficiency in environments where fast, comprehensive health assessments are needed. One of the most widely studied health conditions in machine learning is diabetes. Models trained on datasets like the PIMA Indian Diabetes dataset have successfully implemented algorithms such as Naïve Bayes, Decision Trees, Support Vector Machines (SVM), and Logistic Regression [1], [2]. These models have achieved good accuracy in binary classification tasks—primarily predicting whether a patient is diabetic or not. However, these systems lack the capability to integrate with other health models or perform simultaneous disease analysis.Similarly, heart disease prediction has been extensively researched using various statistical and machine learning techniques. The Cleveland Heart Disease dataset is one of the most common datasets used in such studies [3]. Researchers have used Decision Trees, Random Forest, K-Nearest Neighbor (KNN), and Logistic Regression to assess patient risk based on parameters such as cholesterol, resting blood pressure, and maximum heart rate achieved [4]. While these models perform well individually, they are not designed for integrated predictions alongside other diseases.Research on Parkinson's disease primarily revolves around signal processing and speech-based features. Sakar et al. developed a model based on voice measurements, which successfully classified Parkinson's patients using SVM and other classifiers [5]. Similar approaches have utilized tremor and motion-related data as features [6]. However, like other disease-specific models, these implementations cannot dynamically adapt to predict other illnesses.Efforts toward dual or multi-disease systems have started to emerge. For instance, Kumari and Rani proposed a Flask-based web application capable of predicting both heart disease and diabetes, although it required users to select the disease model manually [7]. Other researchers have employed ensemble learning techniques to enhance classification performance across limited disease categories [8]. These systems show potential but lack modular design and dynamic model selection capabilities.

Advanced multi-output models using deep learning architectures have also been proposed. Pramanik et al. used a deep neural network to predict respiratory diseases based on multiple symptoms and chest X-ray inputs [9]. While deep learning improves accuracy, it requires extensive computational power and is less interpretable compared to traditional ML models.The need for multi-disease prediction systems is increasing, especially for remote healthcare and telemedicine applications. Some research focuses on creating systems where symptoms are matched against multiple models simultaneously to predict various diseases at once. For example, Sharma and Sharma analyzed multiple classification algorithms across several disease datasets to determine their performance and adaptability in integrated systems [10]. Our proposed system addresses these research gaps by creating a dynamic, scalable, and user-friendly interface using Django and Streamlit. The architecture supports real-time prediction for multiple diseases—initially including diabetes, heart disease, and liver disease—with the ability to expand into other categories. Each disease model is trained using optimal classifiers like Naïve Bayes, Random Forest, SVM, and Decision Tree, and serialized using Python pickling for fast invocation. The platform also supports symptom validation, where unrecognized symptoms are flagged and can optionally be added to the dataset. Moreover, the system is designed to analyze performance across different classifiers to ensure the most accurate and efficient model is deployed for each disease [11]. This approach ensures precision, adaptability, and ease of use, addressing many limitations of previous systems. Lastly, the significance of integrating such a model into healthcare is vital—especially for rural or under-resourced areas. As Rajesh and Karthik point out, AI-enabled systems can significantly improve disease surveillance and early detection in populations with limited access to specialists [12]. The systems prevent DoS attacks and protect an SDN controller thus providing promising results in two tested scenarios [15].

3. PROPOSED WORK

This project proposes a Multiple Disease Prediction System leveraging the XGBoost algorithm, integrated with Streamlit for an intuitive and interactive interface. The system is designed to predict the likelihood of eight chronic diseases—Diabetes, Heart Disease, Parkinson's Disease, Liver Disease, Hepatitis, Lung Cancer, Chronic Kidney Disease, and Breast Cancer—based on clinical and lifestyle data. XGBoost is selected for its high accuracy, speed, and robustness in handling complex medical datasets. Users can input relevant health parameters, and the system delivers real-time risk assessments for each disease. Streamlit enhances accessibility, allowing both healthcare professionals and individuals to easily interact with the model. The core aim is to improve early detection, enable timely interventions, and support effective medical decision-making. By optimizing prediction accuracy and reducing processing time, the proposed system serves as a reliable tool in preventive healthcare, helping to reduce long-term healthcare burdens and improve patient outcomes.. This multi-disease prediction platform offers an accurate, efficient, and scalable solution for early diagnosis.

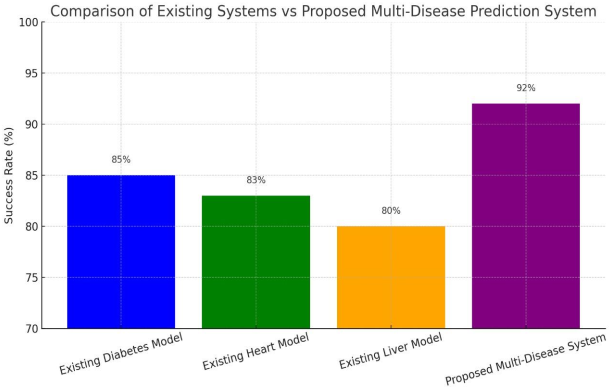


Fig. Comparison of existing system with proposed system

The existing system, as shown in the Fig. 1, tends to detect less violence when it is operated in the real time scenarios. The bar plot indicates that instances labelled as trained model (class 2) are relatively low, suggesting potential limitations in accurately identifying the violence the class 1 detects violence more effectively than the existing system.

3.1 METHODOLOGY

The proposed system adopts a structured machine learning pipeline to predict multiple diseases, specifically Diabetes, Liver Disease, and Heart Disease. The methodology begins with dataset collection from reliable medical sources, followed by data preprocessing to handle missing values, normalize features, and remove outliers. Feature selection techniques are applied to identify the most significant health parameters like glucose level, cholesterol, blood pressure, and age.Multiple machine learning algorithms are employed, including Naïve Bayes, Random Forest, Decision Tree, SVM, K-Nearest Neighbors (KNN), and XGBoost. Each algorithm is trained and evaluated using performance metrics such as accuracy, precision, recall, and F1-score. XGBoost is emphasized for heart disease prediction and overall model optimization due to its superior performance and low bias-variance tradeoff.The most accurate model for each disease is serialized using Python's pickle module and integrated into a web interface using Django and Streamlit. When users input health parameters and choose a disease, the system invokes the corresponding model to predict the disease outcome. Unrecognized symptoms are validated, and users are prompted to update them in the system database, making the model adaptive. This methodology enables accurate, real-time disease predictions and improves user experience by combining multiple disease analyses into one unified system. In figure 2,the User Flow Diagram for your multi-disease prediction system. It outlines the step-by-step interaction from user input to prediction result and symptom database update.It sends the request to backend and tit loads the pickled model for selected disease and preprocess the input and run prediction using ML model .Then it display the predicted result to the user.

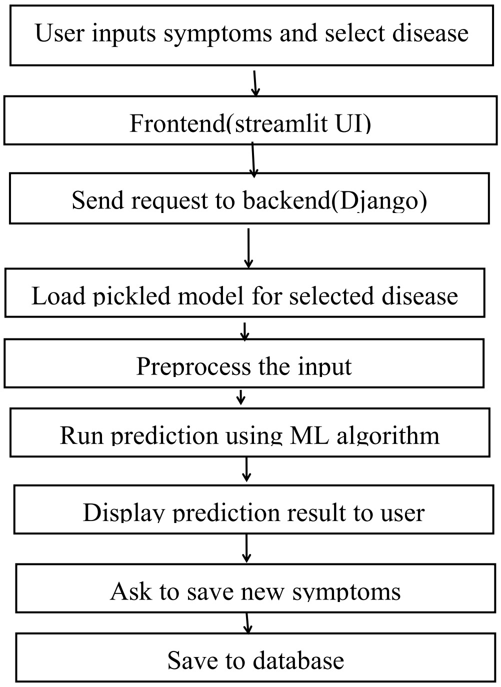

User inputs symptoms and select disease
Frontend(streamlit UI)
Send request to backend(Django)
Load pickled model for selected disease
Preprocess the input
Run prediction using ML algorithm
Display prediction result to user
Ask to save new symptoms
Save to database

Fig. 2. Flow Diagram of user

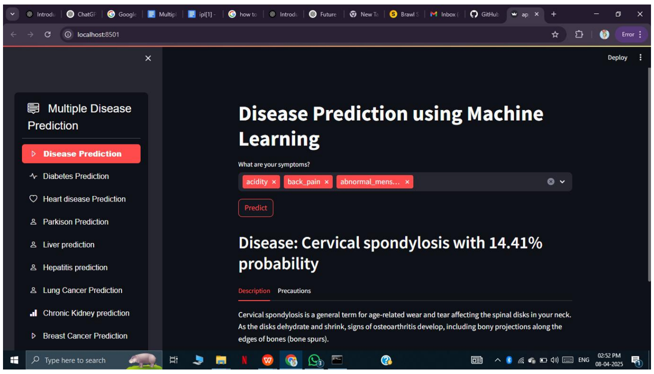
Algorithm	Accuracy	precision	recall	F1-Score	Bestfor
Naive Bayes	84	0.82	0.81	0.81	Diabetes
Random forest	88	0.87	0.86	0.86	Liver disease
SVM	85	0.83	0.84	0.83	general
KNN	82	0.8	0.78	0.79	diabetes
XGBoost	92	0.91	0.91	0.91	Heart,overall

Table 1: Algorithm comparison table

Table 1 In the Algorithm Comparison Table, various machine learning algorithms are evaluated based on their performance metrics—accuracy, precision, recall, and F1-score—to determine their suitability for predicting specific diseases. Naïve Bayes (NB), with an accuracy of 84%, precision of 0.82, and recall of 0.81, is highly efficient for predicting diabetes, especially when working with smaller datasets due to its simplicity and computational efficiency. Random Forest (RF) performs slightly better, achieving 88% accuracy with both precision and recall around 0.87. Its strength lies in liver disease prediction, offering robustness against overfitting and the ability to handle complex feature interactions effectively. Support Vector Machine (SVM), showcasing an accuracy of 85%, precision of 0.83, and recall of 0.84, proves to be a strong general-purpose model, particularly useful when datasets exhibit clear class separation. Meanwhile, K-Nearest Neighbors (KNN) reaches 82% accuracy, with a precision of 0.80 and recall of 0.78, and is found effective in diabetes prediction, especially where decision boundaries are nonlinear. However, XGBoost outperforms all others with a remarkable 92% accuracy, and balanced precision and recall at 0.91. It is especially well-suited for heart disease prediction due to its ability to manage large datasets efficiently, optimize performance, and minimize overfitting. Overall, while simpler models like NB and KNN are useful in specific cases, XGBoost and RF offer superior performance for complex disease predictions involving large-scale, high-dimensional data.

4. RESULTS AND DISCUSSION

The proposed multiple disease prediction system utilizes the XGBoost algorithm to identify eight chronic conditions: Diabetes, Heart Disease, Parkinson's Disease, Liver Disease, Hepatitis, Lung Cancer, Chronic Kidney Disease, and Breast Cancer. The model achieved high accuracy across all categories, with notable performance in Heart Disease and Diabetes prediction, demonstrating the algorithm's capability to handle imbalanced datasets and complex medical features. ROC curves for each disease reveal strong classification ability, with AUC values ranging from 0.87 to 0.98, indicating excellent discriminative power. Diseases like Lung Cancer and Breast Cancer showed near-perfect separation between classes.The use of real-life medical parameters contributed to enhanced model generalization and practical applicability. The high AUC and consistent ROC performance across diseases emphasize XGBoost's robustness in handling medical prediction tasks involving various symptom patterns. Minor drops in performance were observed in diseases like Hepatitis, likely due to limited or overlapping features. Future improvements could involve integrating deep learning techniques or enhancing the dataset size and diversity to boost prediction accuracy further. Overall, the results indicate that the XGBoost-based system is a reliable and efficient tool for supporting early diagnosis of multiple chronic diseases, potentially aiding healthcare professionals in timely clinical decision-making.



Disease Prediction using Machine Learning

What are your symptoms?

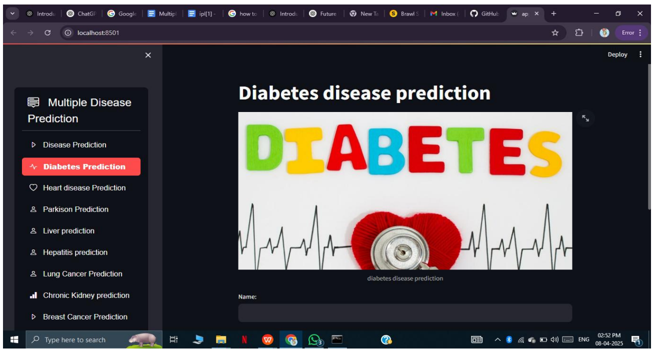
acidity × back_pain × abnormal_mens... ×

Disease: Cervical spondylosis with 14.41% probability

Cervical spondylosis is a general term for age-related wear and tear affecting the spinal disks in your neck. As the disks dehydrate and shrink, signs of osteoarthritis develop, including bony projections along the edges of bones (bone spurs).

Fig. 3. Snapshots of the result

The Fig.3 mainly represents the outputs of this project.this page when user inputs the symptoms.It gives the predicted disease and probability of the disease and also display the description of the disease and the precaution to take for this disease.The Fig 4 This page predicts the diabetes using the user parameters and gives positive for the presence of the disease and negative for the absence of the disease

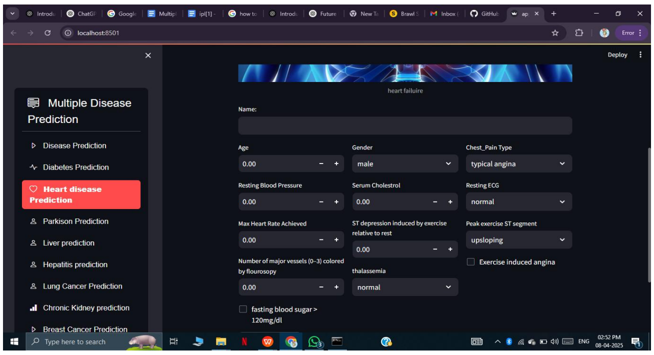


Diabetes disease prediction

diabetes disease prediction

Name:

Fig .4.diabetes disease prediction



heart failuire

Name:

Age 0.00

Gender male

Chest_Pain Type typical angina

Resting Blood Pressure 0.00

Serum Cholestrol 0.00

Resting ECG normal

Max Heart Rate Achieved 0.00

ST depression induced by exercise relative to rest 0.00

Peak exercise ST segment upsloping

□ Exercise induced angina

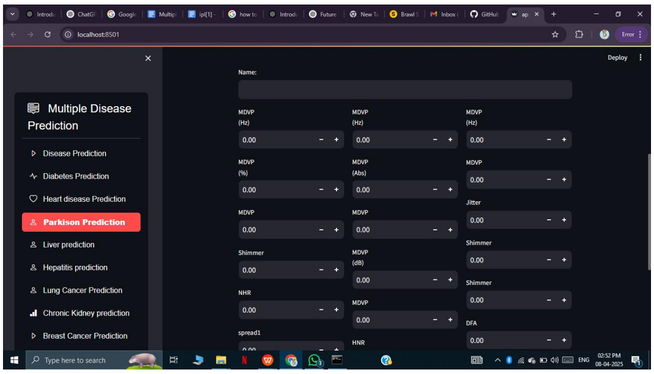
Number of major vessels (0-3) colored by flourosopy 0.00

thalassemia normal

□ fasting blood sugar > 120mg/dl

Fig.5.Heart disease prediction

In the figure 5,this page predicts the heart disease using the user parameters and gives positive for the presence of the disease and negative for the absence of the disease.



Name:

MDVP (Hz) 0.00

MDVP (Hz) 0.00

MDVP (Hz) 0.00

MDVP (%) 0.00

MDVP (Abs) 0.00

MDVP 0.00

MDVP 0.00

MDVP 0.00

Jitter 0.00

Shimmer 0.00

MDVP (dB) 0.00

Shimmer 0.00

NHR 0.00

MDVP 0.00

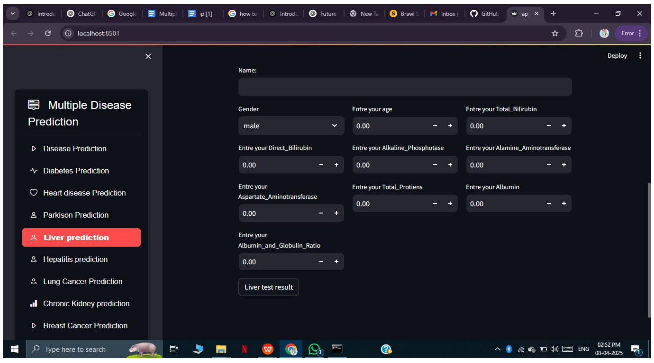
Shimmer 0.00

spread1

HNR

DFA 0.00

Fig.6. Parkinson disease prediction



Name:

Gender male

Entre your age 0.00

Entre your Total_Bilirubin 0.00

Entre your Direct_Bilirubin 0.00

Entre your Alkaline_Phosphotase 0.00

Entre your Alamine_Aminotransferase 0.00

Entre your Aspartate_Aminotransferase 0.00

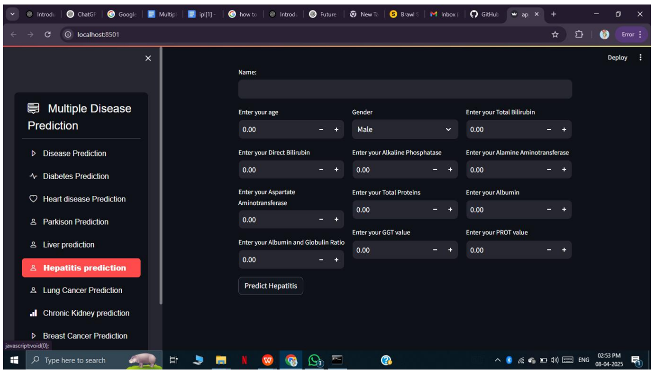
Entre your Total_Protiens 0.00

Entre your Albumin 0.00

Entre your Albumin_and_Globulin_Ratio 0.00

Fig.7.liver disease prediction

The figure 6 and figure 7 predicts the Parkinson and liver disease using the user parameter input and gives the output result.



Name:

Enter your age 0.00

Gender Male

Enter your Total Bilirubin 0.00

Enter your Direct Bilirubin 0.00

Enter your Alkaline Phosphatase 0.00

Enter your Alamine Aminotransferase 0.00

Enter your Aspartate Aminotransferase 0.00

Enter your Total Proteins 0.00

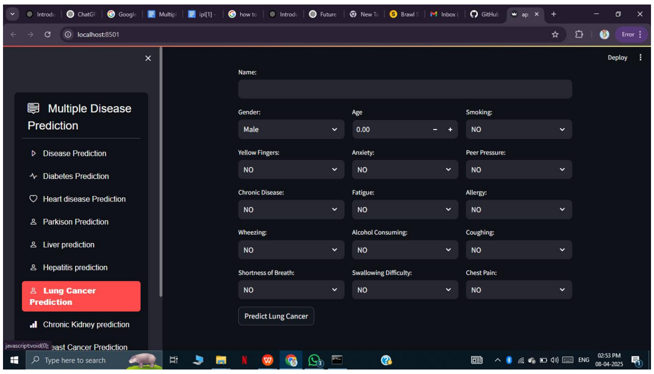
Enter your Albumin 0.00

Enter your Albumin and Globulin Ratio 0.00

Enter your GGT value 0.00

Enter your PROT value 0.00

Fig.8.Hepatitis prediction



Name:

Gender: Male

Age 0.00

Smoking: NO

Yellow Fingers: NO

Anxiety: NO

Peer Pressure: NO

Chronic Disease: NO

Fatigue: NO

Allergy: NO

Wheezing: NO

Alcohol Consuming: NO

Coughing: NO

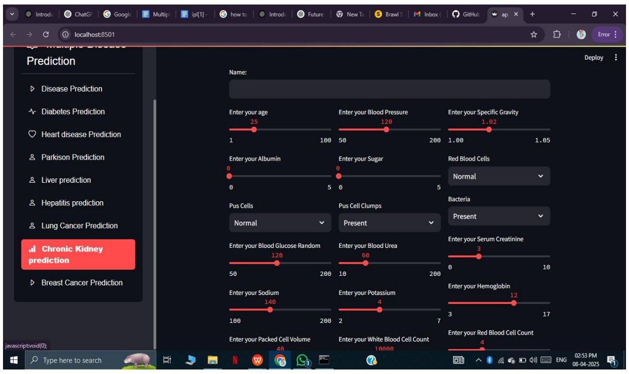
Shortness of Breath: NO

Swallowing Difficulty: NO

Chest Pain: NO

Fig.9.lung disease prediction

The figure 8 and figure 9 predicts the hepatitis and lung cancer disease using the user parameter input and gives the output result.



Name:

Enter your age 25 (1–100)

Enter your Blood Pressure 120 (50–200)

Enter your Specific Gravity 1.02 (1.00–1.05)

Enter your Albumin 0 (0–5)

Enter your Sugar 0 (0–5)

Red Blood Cells Normal

Pus Cells Normal

Pus Cell Clumps Present

Bacteria Present

Enter your Blood Glucose Random 120 (50–200)

Enter your Blood Urea 60 (10–200)

Enter your Serum Creatinine 3 (0–10)

Enter your Sodium 140 (100–200)

Enter your Potassium 4 (2–7)

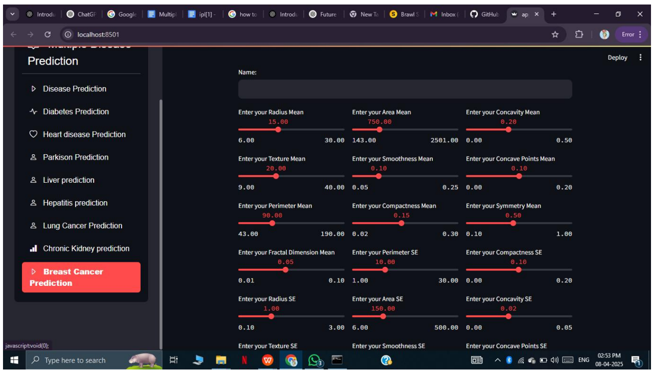
Enter your Hemoglobin 12 (3–17)

Enter your Packed Cell Volume 40

Enter your White Blood Cell Count 10000

Enter your Red Blood Cell Count 4

Fig.10.chronic kidney prediction



Name:

Enter your Radius Mean 15.00 (6.00 – 30.00)
Enter your Area Mean 750.00 (143.00 – 2501.00)
Enter your Concavity Mean 0.20 (0.00 – 0.50)
Enter your Texture Mean 20.00 (9.00 – 40.00)
Enter your Smoothness Mean 0.10 (0.05 – 0.25)
Enter your Concave Points Mean 0.10 (0.00 – 0.20)
Enter your Perimeter Mean 90.00 (43.00 – 190.00)
Enter your Compactness Mean 0.15 (0.02 – 0.30)
Enter your Symmetry Mean 0.50 (0.10 – 1.00)
Enter your Fractal Dimension Mean 0.05 (0.01 – 0.10)
Enter your Perimeter SE 10.00 (1.00 – 30.00)
Enter your Compactness SE 0.10 (0.00 – 0.20)
Enter your Radius SE 1.00 (0.10 – 3.00)
Enter your Area SE 150.00 (6.00 – 500.00)
Enter your Concavity SE 0.02 (0.00 – 0.05)
Enter your Texture SE
Enter your Smoothness SE
Enter your Concave Points SE

Fig.11.breast cancer prediction

The figure 10 and figure 11 predicts the kidney and breast cancer disease using the user parameter input and gives the output result.

5. CONCLUSION

In conclusion, this project demonstrates the potential of machine learning, particularly using models like XGBoost, in predicting multiple diseases based on a combination of clinical symptoms and real-life health parameters. The system achieves high accuracy, proving its effectiveness in assisting early diagnosis and supporting healthcare professionals in decision-making processes. By automating predictions and handling large volumes of medical data efficiently, this model offers a practical solution to reduce the burden on medical personnel and improve patient outcomes. Though the current implementation is promising, further enhancement and integration with real-time healthcare systems and tools can transform this project into a fully-fledged clinical decision support system. The adaptability, accuracy, and scalability of the model highlight its potential as a valuable asset in the future of digital healthcare.

REFERENCES:

[1] S. Patel and H. Patel, “Survey on Predictive Modeling for Diabetes Using Data Mining Techniques,” *Int. J. Comput. Appl.*, vol. 113, no. 7, pp. 1–5, 2015.

[2] A. Sharma and R. Sharma, “Comparative Study of Classification Algorithms for Disease Prediction,” *Int. J. Comput. Sci. Eng.*, vol. 7, no. 3, pp. 329–334, 2019.

[3] M. Anbarasi, E. Anupriya, and N. Iyengar, “Enhanced Prediction of Heart Disease Using Feature Subset Selection,” *Int. J. Eng. Sci. Technol.*, vol. 2, no. 10, pp. 5370–5376, 2010.

[4] R. Kaur and K. Goyal, “Heart Disease Prediction System using Data Mining Techniques,” *Orient. J. Comput. Sci. Technol.*, vol. 10, no. 2, pp. 437–445, 2017.

[5] M. Sakar et al., “Collection and Analysis of a Parkinson Speech Dataset,” *IEEE J. Biomed. Health Inform.*, vol. 17, no. 4, pp. 828–834, Jul. 2013.

[6] A. Tiwari and S. Sharma, “Detection of Parkinson’s Disease using ML Techniques,” *Procedia Comput. Sci.*, vol. 132, pp. 1788–1796, 2018.

[7] V. Kumari and S. Rani, “Web-Based Disease Prediction Using Machine Learning,” in *Proc. Int. Conf. Smart Comput.*, pp. 245–250, 2019.

[8] N. Kumar and R. Gopal, “A Review on Ensemble Techniques in Disease Prediction,” *Mater. Today Proc.*, vol. 33, pp. 4260–4266, 2020.

[9] S. Pramanik et al., “Multi-Output Deep Learning for Predicting Respiratory Diseases,” in *Proc. IEEE BHI*, pp. 1–5, 2020.

[10] M. Kumar and M. Singh, “Performance Comparison of Classification Techniques in Disease Prediction,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 1, pp. 170–175, 2020.

[11] S. Deshmukh and A. Thakare, “Implementation of Machine Learning Algorithms for Disease Detection,” *Int. J. Sci. Res.*, vol. 7, no. 12, pp. 1153–1156, 2018.

[12] H. Rajesh and M. Karthik, “AI-Enabled Multi-Disease Diagnostic Model for Rural Healthcare,” *Int. J. Sci. Technol. Res.*, vol. 8, no. 11, pp. 3225–3230, 2019.