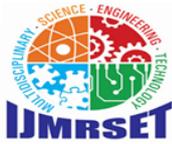# International Journal of Multidisciplinary
Research in Science, Engineering and Technology

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

Impact Factor: 8.206

Volume 9, Issue 1, January 2026

Simultaneous Disease Prediction Using Machine Learning

Prof. Gunasekaran K, K Hemasundhar

Assistant Professor, AMC Engineering College, Bengaluru, India

Department of MCA, AMC Engineering College, Bengaluru, India

ABSTRACT: On this site, you will find a disease prediction tool built with Streamlit and Python. It uses several machine learning techniques, including Naïve Bayes, Random Forest, Decision Trees, and SVMs, to identify diseases such as Heart Disease, Diabetes, and Parkinson's Disease. This tool lets users input basic health metrics like Blood Pressure, Cholesterol, Pulse Rate, and Heart Rate to get real-time predictions about their health status. After collecting and analyzing the health metrics, the best model will be chosen and saved to a file for future use through pickling. The system is designed to adapt for the future, allowing for the addition of other diseases, including chronic and skin diseases. This helps with early diagnosis, prevention, and in some cases, saves lives.

Additionally, this project offers a simple, user-friendly interface that encourages ongoing health monitoring and helps individuals become more aware of their health. The model also supports healthcare professionals by serving as a decision support tool for their practices. Furthermore, this tool demonstrates how AI-driven preventive healthcare tools can make it easier for people in remote and underserved communities to access healthcare services.

KEYWORDS: Diabetes, Heart, Liver, KNN, Random Forest, XG Boost, Streamlit, Lung Cancer, Chronic Kidney Disease.

I. INTRODUCTION

In today's digital age, data plays an important role in many areas, especially in healthcare, which generates large amounts of patient information. This project presents a unified framework for predicting diseases. It aims to fix the problems with current systems that usually focus on one disease at a time, like separate platforms for diabetes, cancer, or skin issues. Our goal is to create an integrated system that can diagnose multiple diseases based on symptoms users enter, providing real-time and accurate results.

The proposed solution uses Django for backend development and Streamlit for deployment. It includes machine learning models to assess and predict diseases such as diabetes, heart disease, and malaria, with plans to add more conditions in the future. We also conduct a focused analysis of liver, diabetes, and heart diseases, looking at how they relate to one another. Users interact with the system by entering symptoms and selecting the disease they are concerned about. The system then uses the appropriate pre-trained model, saved with Python's pickle module, to analyze the data and deliver prediction results. The model not only identifies the disease but also assesses its severity. If a symptom isn't recognized, the user can confirm its addition to the database, which helps the system improve over time.

We implement various machine learning algorithms like Naïve Bayes, K-Nearest Neighbors (KNN), Decision Tree, Random Forest, and Support Vector Machine (SVM). We compare these to find the most accurate and efficient model. The ultimate goal is to create a scalable and smart platform for early diagnosis and proactive healthcare management. This will reduce the need for users to navigate multiple systems for different conditions.

II. RELATED WORK

The use of Machine Learning (ML) in healthcare has gained a lot of attention because it can extract valuable insights from large medical datasets. In recent years, researchers have created many disease prediction systems. However, most of these systems focus on identifying just one disease, such as diabetes, cancer, or heart conditions. This limits scalability and efficiency, especially in situations where quick and thorough health assessments are needed.

Diabetes is one of the most studied health issues in machine learning. Models trained on datasets like the PIMA Indian Diabetes dataset have successfully used algorithms such as Naïve Bayes, Decision Trees, Support Vector Machines (SVM), and Logistic Regression. These models have shown good accuracy in binary classification tasks, mainly predicting whether a patient has diabetes. However, these systems cannot connect with other health models or analyze multiple diseases at the same time.

Heart disease prediction has also been widely researched using various statistical and machine learning methods. The Cleveland Heart Disease dataset is one of the most commonly used in these studies. Researchers have applied Decision Trees, Random Forest, K-Nearest Neighbor (KNN), and Logistic Regression to evaluate patient risk based on factors like cholesterol levels, resting blood pressure, and maximum heart rate. While these models perform well on their own, they are not designed to make integrated predictions with other diseases.

Research on Parkinson's disease mainly focuses on signal processing and speech-related features. Sakar et al. developed a model based on voice measurements that effectively classified Parkinson's patients using SVM and other classifiers. Similar methods have used tremor and motion data as features. However, like other single-disease models, these approaches cannot easily adjust to predict other illnesses.

There have been emerging efforts toward dual or multi-disease systems. For example, Kumari and Rani introduced a Flask-based web application that predicts both heart disease and diabetes, but it required users to choose the disease model manually. Other researchers have used ensemble learning techniques to improve classification performance across a limited number of disease categories. These systems show promise but lack modular design and the ability to select models dynamically.

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. While deep learning improves accuracy, it requires significant computational power and is harder to interpret compared to traditional ML models. The demand for multi-disease prediction systems is growing, especially for remote healthcare and telemedicine applications. Some research focuses on creating systems where symptoms are matched with multiple models at the same time to predict various diseases. For instance, Sharma and Sharma analyzed multiple classification algorithms across several disease datasets to evaluate their performance and adaptability in integrated systems. 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 capacity to expand into other categories. Each disease model is trained using optimal classifiers like Naïve Bayes, Random Forest, SVM, and Decision Tree. Models are serialized using Python pickling for quick access. The platform also supports symptom validation, where unrecognized symptoms are flagged and can optionally be added to the dataset. Moreover, the system analyzes performance across different classifiers to ensure the best and most efficient model is deployed for each disease. This approach ensures accuracy, adaptability, and ease of use, addressing many limitations of previous systems. Lastly, integrating such a model into healthcare is crucial, especially for rural or under-resourced areas. As Rajesh and Karthik point out, AI systems can significantly enhance disease surveillance and early detection in populations with limited access to specialists. These systems also prevent DoS attacks and protect an SDN controller, thus delivering promising results in two tested scenarios.

III. PROPOSED WORK

This project proposes a Multiple Disease Prediction System that uses the XGBoost algorithm. It will be integrated with Streamlit to offer a user-friendly and interactive interface. The system will predict the likelihood of eight chronic diseases: Diabetes, Heart Disease, Parkinson's Disease, Liver Disease, Hepatitis, Lung Cancer, Chronic Kidney Disease, and Breast Cancer. It will do this by using clinical and lifestyle data as inputs. XGBoost was chosen for its high accuracy, speed, and ability to handle complex medical datasets. Users can enter their health data, and the system will provide immediate risk assessments for each disease. Streamlit makes the system accessible, allowing everyone to engage with the model easily. The main goal is to improve early detection, enable prompt interventions, and support informed medical decisions. This system serves as a reliable tool in preventive healthcare, reducing long-term healthcare costs and enhancing patient outcomes. Therefore, the proposed system is a solution that offers accurate predictions for multiple diseases, is quick, and can scale for early diagnosis.
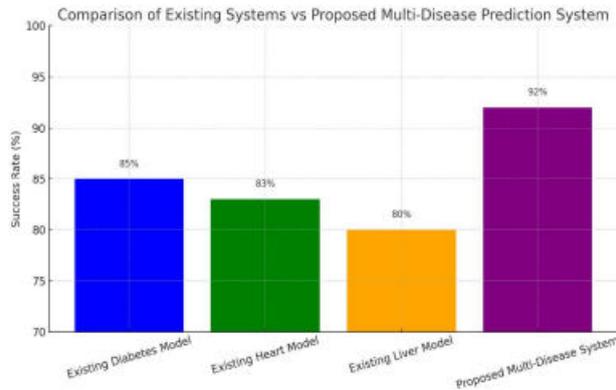
The current system shown in Fig. 1 tends to miss low levels of violence during real-time scenarios. The bar plot reveals that the trained model (class 2) has a limited number of instances. This suggests that it may struggle to accurately detect violence. In contrast, class 1 identifies violence more effectively.

IV. METHODOLOGY

The proposed system uses a machine learning pipeline designed to predict several diseases, specifically Diabetes, Liver Disease, and Heart Disease. The process begins by gathering datasets from reliable hospitals. This is followed by data preprocessing to eliminate missing values, standardize features, and remove outliers. Next, feature selection methods identify the most important health parameters such as glucose level, cholesterol, blood pressure, and age.

Various machine learning algorithms are employed, including Naïve Bayes, Random Forest, Decision Tree, SVM, K-Nearest Neighbors (KNN), and XGBoost. Each algorithm's performance is evaluated based on accuracy, precision, recall, and F1-score. XGBoost receives particular attention for predicting heart disease and optimizing the overall model due to its strong performance and good balance between bias and variance.

The most accurate model for each disease is saved using Python's pickle module, and a web platform is developed with Django and Streamlit for user interaction. When users enter health parameters and select a disease, the system retrieves the corresponding model to provide a disease prediction. The system continually improves as new symptoms are validated, and users are encouraged to add them to the database. This entire process enables quick and accurate disease predictions while remaining user-friendly, as analyses for different diseases are combined in one place.

Figure 2 shows the User Flow Diagram for the multi-disease prediction system. It outlines the interaction from user input to prediction result and symptom database update step-by-step. It forwards the request to the backend, loads the model that has been pickled for the selected disease, preprocesses the input, and runs the prediction using the machine learning model. After that, it shows you the results

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

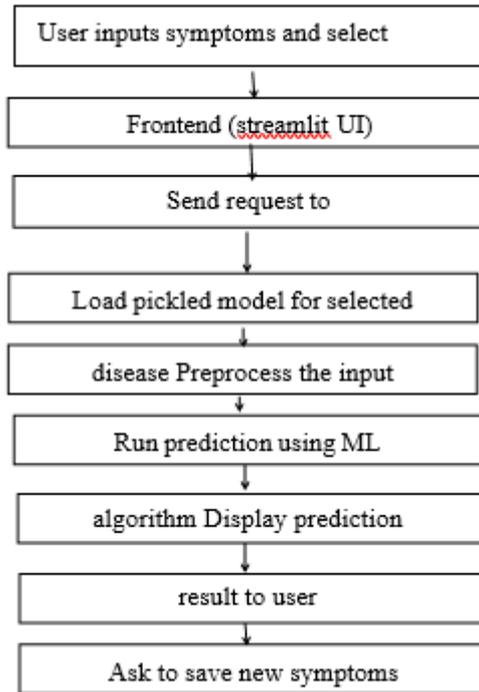(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Fig. 2. Flow Diagram of user

Table 1: Algorithm comparison table

Algorith m	Accur acy	preci sion	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,ov erall

Table 1 shows how different machine learning models compare when it comes to predicting diseases. It looks at accuracy, precision, recall, and F1-score. Naive Bayes (NB) works well for diabetes prediction, especially with smaller datasets. It scores about 84% in accuracy, 0.82 in precision, and 0.81 in recall. Random Forest (RF) performs better, achieving 88% accuracy and around 0.87 for both precision and recall. RF is effective for predicting liver disease because it doesn't easily overfit and can manage various factors in the data. Support Vector Machine (SVM) is a solid choice, reaching 85% accuracy, 0.83 precision, and 0.84 recall. It works well when the data is clearly grouped. K-Nearest Neighbors (KNN) achieves around 82% accuracy, with 0.80 precision and 0.78 recall. It proves useful for diabetes prediction, especially when the situation is a bit complex. However, XGBoost stands out with an impressive 92% accuracy and 0.91 for both precision and recall. It excels at predicting heart disease because it can handle large datasets effectively and minimizes overfitting. In summary, while simpler models like NB and KNN are effective in specific scenarios, XGBoost and RF are the top performers for complex disease predictions involving large amounts of data.

V. RESULTS AND DISCUSSION

The system tries to predict whether someone has one of eight long-term illnesses: diabetes, heart issues, Parkinson's, liver problems, hepatitis, lung cancer, kidney problems, or breast cancer. It does this using the XGBoost algorithm.

The system shows good accuracy for all the illnesses, but it performs especially well in predicting heart issues and diabetes. This indicates that the algorithm can manage challenging medical information, even when the data is not evenly distributed among the illnesses. The ROC curves reveal that the system effectively distinguishes between those with a disease and those without. The scores range from 0.87 to 0.98, which is excellent. It almost perfectly differentiates between people with lung cancer or breast cancer and those who are disease-free.

Since the system relies on real medical information, it should be effective in practical situations. The consistently high scores demonstrate that XGBoost excels at medical predictions, even with varying symptoms. It was less effective in predicting hepatitis, possibly due to insufficient data or symptoms being too similar to those of other illnesses. To improve its accuracy, we could explore deep learning or increase the amount of data. Overall, this XGBoost system appears to be a strong tool for assisting doctors in diagnosing these illnesses early.
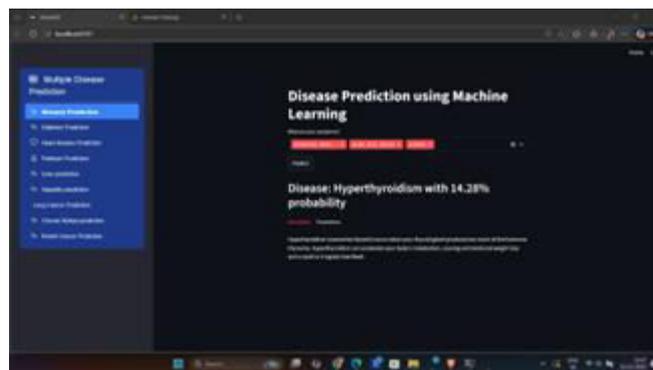

Fig. 3. Snapshots of the result

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
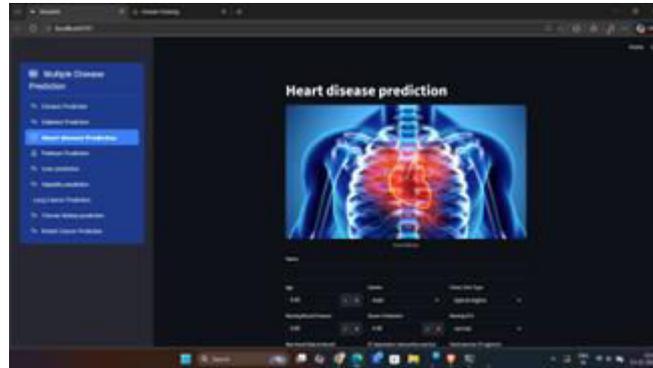

Fig. 4.diabetes disease prediction
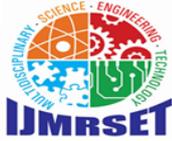
Fig.5.Heart disease prediction

In figure 5, this page predicts heart disease based on user parameters. It indicates positive for the presence of the disease and negative for its absence.
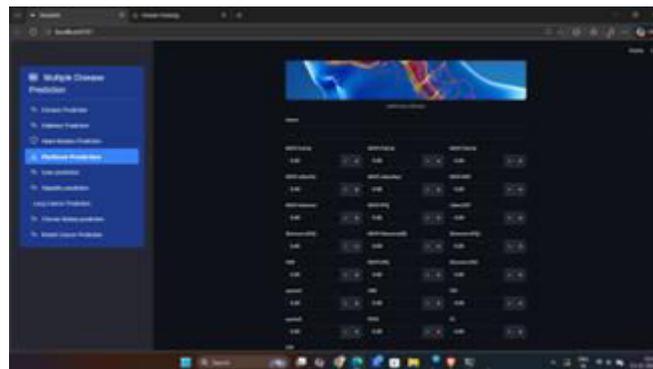

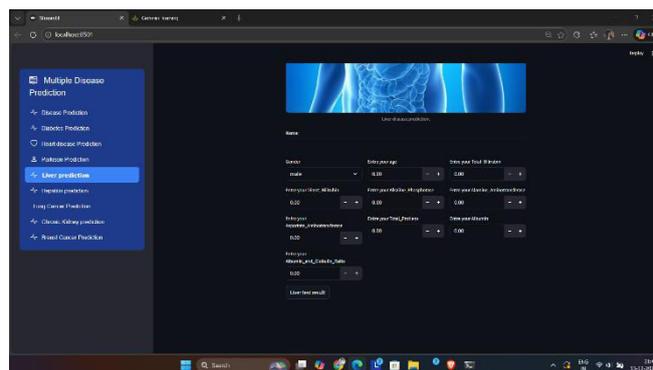Fig.6. Parkinson disease prediction

Fig.7.liver disease prediction

The figures 6 and 7 predict Parkinson's disease and liver disease based on the user's input and provide the output result.
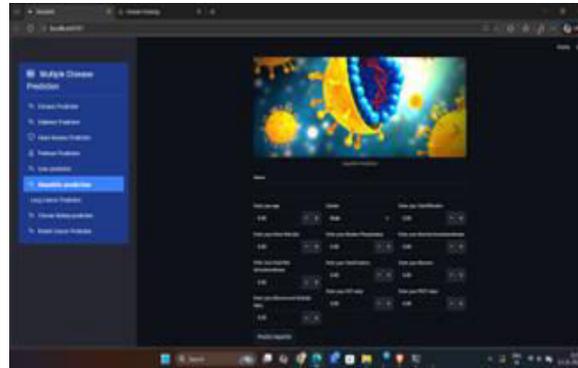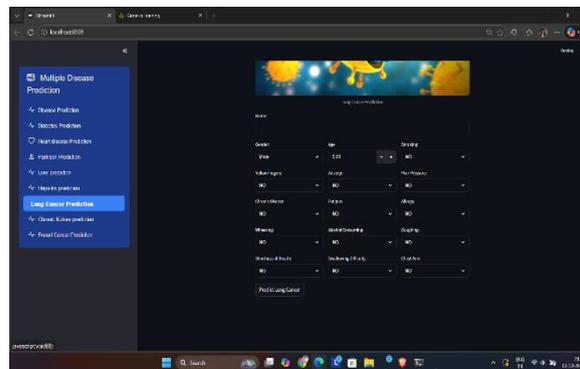
Fig.8.Hepatitis prediction

Fig.9.lung disease prediction

The figures 8 and 9 predict hepatitis and lung cancer based on user parameter input and provide the output results.
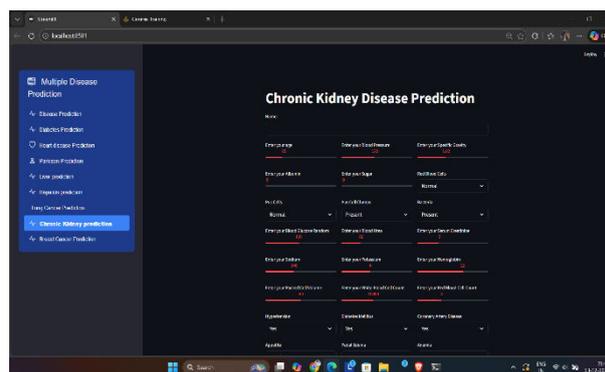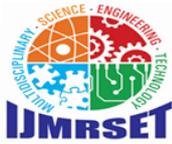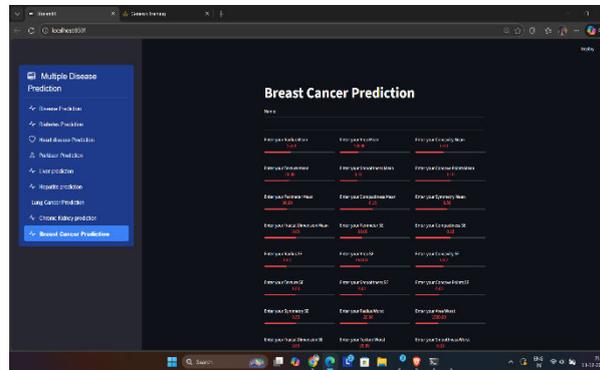

Fig.10.chronic kidney prediction

Fig.11.breast cancer prediction

The figure 10 and figure 11 show predictions for kidney and breast cancer based on user input parameters. They provide the output results.

VI. CONCLUSION

So, this project shows how machine learning, especially with models like XGBoost, can be great at predicting different diseases. It looks at symptoms and health info and gets pretty accurate results. That means it could really help doctors diagnose stuff early and make better decisions. Because it can automatically make predictions and handle tons of data, it lightens the load on healthcare folks and can help patients get better care. Right now, it looks like it has potential, but if we can make it even better and hook it up to real-time healthcare systems, it could become a super useful clinical tool. It's adaptable, accurate, and can handle a lot, so it could be a game-changer for digital healthcare down the road.

REFERENCES

[1] **S. Rajaraman, J. Siegelman, P. O. Alderson, L. Folio and S. Antani**, "**Deep learning for multi-class disease classification in chest radiographs**," IEEE Access, vol. 8, pp. 24312–24323, 2020.

[2] **A.Sharma and S. Kumar**, "**Machine learning techniques for disease prediction: A survey**," International Conference on Computing, Power and Communication Technologies (GUCON)2019, pp. 887–892.

[3] **M. A. Khan, M. Alhaisoni, A. Hussain, et al.**, "**Human disease diagnosis using machine learning algorithms**," IEEE Access, vol. 8, pp. 42090–42105, 2020.

[4] **N. Patel and N. Shah**, "**Predicting chronic diseases using machine learning algorithms**," International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2019, pp. 55–60.

[5] **R. Ahmed, A. Jalal and K. Kim**, "**A patient diagnosis system using machine learning classification algorithms**," IEEE International Conference on Applied Sciences and Technology (ICAST), 2020, pp. 145–150.

[6] **P. K. Singh and A. Gupta**, "**Heart disease prediction using machine learning**," IEEE International Conference on Intelligent Engineering and Management (ICIEM), 2020, pp. 348–353.

[7] **M. Chen, Y. Hao, K. Hwang, L. Wang and L. Wang**, "**Disease prediction by machine learning over big healthcare data**," IEEE Access, vol. 5, pp. 8869–8879, 2017.

[8] **G. M. Khan, S. Islam and R. U. Islam**, "**A robust multi-disease prediction model using ensemble learning**," IEEE International Conference on Information and Communication Technology (ICICT), 2021, pp. 1–6.

[9] **V. Chaurasia and S. Pal**, "**A novel approach for heart disease prediction using decision tree and Naive Bayes**," International Journal of Advanced Computer Science and Information Technology, 2018.

[10] **S. Tuli, N. Basumatary and S. Tuli**, "**Multi-disease classification using hybrid deep learning models**," IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), 2021.

INTERNATIONAL JOURNAL OF

MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING AND TECHNOLOGY