



# Multiple Disease Prediction using Machine Learning

<sup>1</sup>Ankita Mitra, <sup>2</sup>Kabita Singh, <sup>3</sup>Ranjit Kumar Yadav, <sup>4</sup>Tamal Barman, <sup>5</sup>Shyamapriya Chatterjee, <sup>6</sup>Sujata Kundu

<sup>1,2,3,4</sup> Student, Department of Information Technology, Narula Institute of technology, Kolkata

<sup>5,6</sup> Assistant Professor, Department of Information Technology, Narula Institute of technology, Kolkata

## Abstract

An extensive research project called Multiple Disease Prediction using Machine Learning, Deep Learning, and Streamlit aims to forecast multiple diseases, such as diabetes, heart disease, lung cancer, Parkinson's disease, and breast cancer. Machine learning algorithms, including support vector machines (SVM) and logistic regression, random forest, decision tree are used in this project. Streamlit Cloud and the Streamlit library are used to deploy the models, offering an intuitive interface for disease prediction. There are five disease options available on the application interface: diabetes, Parkinson's disease, breast cancer, lung cancer disease, and heart disease. The user will be asked to provide the relevant information needed for the prediction model after choosing a particular disease. The application quickly produces the disease prediction result after the requirements are provided, showing whether or not the person is impacted by the condition. This research uses machine learning techniques to address the demand for exact disease prediction, enabling early detection and prevention. People may more readily determine their risk for different diseases because of the Streamlit library's and Streamlit Cloud's simple interface, which improves accessibility and usability. The various models' high accuracy levels show how successful the machine learning algorithms used in disease prediction are.

## Keywords

Machine learning, deep learning, datasets, streamlit, heart disease prediction, diabetes detection, lung cancer detection, breast cancer detection.

## Introduction

The Multiple Disease Prediction System using Machine Learning is an advanced healthcare application that utilizes machine learning algorithms to accurately predict the likelihood of a patient having multiple diseases based on their symptoms. The project "Multiple Disease Prediction using Machine Learning, Deep Learning and Streamlit" focuses on predicting five different diseases: diabetes, heart disease, Parkinson's disease, breast cancer and lung cancer. The prediction models are built using machine learning algorithms, including Support Vector Machine.

(SVM) for diabetes and Random Forest for Parkinson's disease and breast cancer, Neural Network for heart disease and lung cancer. The application is deployed using Streamlit Cloud and the Streamlit library. The project begins by collecting relevant data from Kaggle.com, which is then pre-processed to prepare it for training and testing the prediction models. Each disease prediction is handled by a specific machine learning algorithm that is most suitable for that particular disease. The application interface offers five options, each corresponding to a specific disease. When a user selects a particular disease, the application prompts for the necessary parameters required by the corresponding model to predict the disease result. The user provides the required parameters, and the application displays the prediction result based on the input. To deploy the prediction models, Streamlit Cloud and the Streamlit library are utilized. Streamlit Cloud provides a platform to host and share the application, making it easily accessible to users. The Streamlit library simplifies the process of developing interactive and user-friendly model.

## Objective

In an attempt to find an accurate machine learning model for multiple disease prediction, the main objectives of this research work are Early Detection and Diagnosis, Risk Assessment and Prevention, Personalized Medicine, Resource Optimization, Clinical Decision Support, Public Health Surveillance.

## Methodology

The following concisely describes the Multiple Disease Prediction Project's methodology:

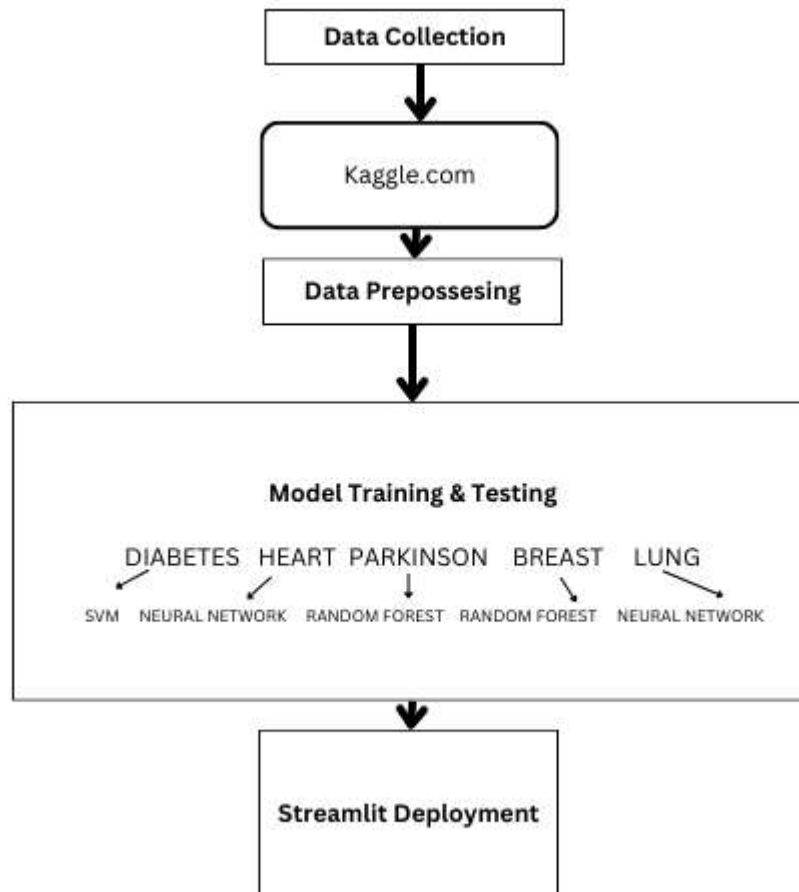
1. Data Collection: Kaggle.com, a well-known website for dataset access, is the source of the data. Particularly, information on diabetes, heart disease, lung disease, Parkinson's disease, and breast cancer is gathered.

2. Data Preprocessing: To ensure the quality and suitability of the data collected for training the machine learning models, preprocessing is applied. This includes handling missing values, eliminating duplicates, and carrying out scaling features or data normalization.

3. Model Selection: For every disease prediction task, a different set of machine learning algorithms is selected. The methods for a variety of diseases are chosen from support vector machines (SVM) and logistic regression, neural network, and random forest based on their effectiveness and fit for the particular prediction tasks.

4. Training and Testing: Two sets of preprocessed data are created: one for training and the other for testing. The testing data is used to assess the models' performance after they have been trained using the training data. The evaluation metric used to determine whether each model's performance is accurate or not.

5. Model Deployment: An interactive web application is made using Streamlit and its cloud deployment features. With five possibilities for predicting diseases—heart disease, lung cancer, diabetes, Parkinson's disease, and breast cancer—the application has an easy-to-use interface. The application asks the user to enter the required information for the prediction when they select a specific disease.



## Proposed System

We have applied different methods, including dimensionality reduction, which minimizes features while removing information from label encoding, which converts textual data to numerical data, and data standardization, which matches the data. To improve the performance of the model, we take simple models and apply the methods that are most suited for the dataset.

With the use of machine learning algorithms such as support vector machines (SVM), logistic regression, neural network, decision tree, and random forest, the proposed system is a comprehensive disease prediction project that can predict a number of diseases, including diabetes, heart disease, parkinson's disease, breast cancer, and lung cancer. With the help of Streamlit, the system's user-friendly interface and input parameters are intended to deliver precise disease predictions that may be implemented on Streamlit Cloud. The Kaggle platform, a well-known data science community, provides the models' data, which is gathered and preprocessed to guarantee its quality and suitability for model training. Following preprocessing, the corresponding machine learning algorithms for each disease are trained using the preprocessed data. To determine how accurate the trained models are at predicting diseases, testing is conducted.

## Input & Output Design

**Input Design:** The user must enter unique data for each condition into the Multiple Disease Prediction System. The system asks for the data it needs when the user chooses a certain disease from the sidebar menu.

Five disease options are available on the application's sidebar menu: diabetes, Parkinson's disease, breast cancer, lung cancer, and heart disease. The application asks for the necessary parameters for a given disease prediction when the user clicks on one of the diseases. The factors that are looked for must be significant and necessary for precise disease prediction, according to the input design. An easy and intuitive way for the user to enter the parameters should be provided.

**Output Design:** The expected result of the system indicates whether or not the individual is impacted by the chosen disease. The output design needs to display the outcome in an easily comprehensible manner. Once the user provides the parameters, the system needs to show the output. One possible presentation format for the output might be:

Prediction result - Positive

"Prediction: The person is affected by [Disease Name]"

Prediction result - Negative

"Prediction: The person is not affected by [Disease Name]."  
The output needs to be shown on the user interface so that the user can quickly understand the outcome of the prediction.

Overall, the output design clearly displays the prediction result on the user interface, and the input design makes sure the user can enter the required parameters for disease prediction.

## Result

Here the result of applying different types of algorithms is shown below:

| DISEASES  | SVM | RANDOM FOREST | DECISION TREE | LOGISTIC REGRESSION | NEURAL NETWORK |
|-----------|-----|---------------|---------------|---------------------|----------------|
| DIABETES  | 78% | 72%           | 64%           | -                   | 77%            |
| HEART     | -   | 82%           | 88%           | 83%                 | 91%            |
| PARKINSON | 87% | 94%           | 74%           | -                   | 89%            |
| BREAST    | -   | 96%           | -             | 60%                 | 94%            |
| LUNG      | -   | -             | -             | 94%                 | 96%            |

In diabetes detection we have applied 4 algorithms those are support vector machine, random forest, decision tree, logistic regression, neural network and got highest accuracy in svm. In heart disease prediction, we have applied 4 algorithms those are random forest, decision tree, logistic regression, and neural network and got highest accuracy in neural network. In Parkinson disease prediction we have applied 4 algorithms those are svm, random forest, decision tree, neural network and got highest accuracy in random forest algorithm. In breast cancer detection we have applied 3 algorithms those are random forest, logistic regression, neural network and got highest accuracy in random forest algorithm. In lung cancer detection we have applied 2 algorithms those are logistic regression and neural network and we got highest accuracy in neural network.

## Conclusion

This report presents the findings of several studies conducted in the area of improving patient-doctor interactions through technology. Our proposed system aims to bridge the gap between patients and doctors, helping both groups achieve their goals more effectively.

### 1. Machine Learning for Disease Prediction

- The system employs various machine learning methods to support the prediction of multiple diseases. This approach ensures accurate and timely diagnosis, benefiting patients by providing early detection and treatment options.

### 2. Trust Development

- Unlike many current systems that merely automate processes, our system focuses on building trust with users. This is achieved through transparent and reliable predictions, making users more confident in the system's capabilities.

### 3. Incorporating Physician Recommendations

- To further ensure user trust and maintain the integrity of doctors' practices, our system integrates physician recommendations. This dual approach ensures that while patients

receive accurate predictions and advice, doctors' businesses are not adversely affected. Instead, it enhances their practice by providing a tool that supports their decision-making process.

By addressing these critical areas, our system aims to improve the overall healthcare experience for both patients and doctors, fostering a more collaborative and trustworthy environment.

## Future Scope

To further enhance the capabilities and impact of our proposed system, several additional features are envisioned for future development. These features aim to provide comprehensive support and convenience to users, making the healthcare experience even more accessible and efficient.

### 1. Near Hospital Locator

- **Feature:** The system will include a module that identifies the nearest hospitals to the user's location.
- **Benefit:** This will help patients quickly find and access emergency medical services, improving response times and potentially saving lives.

### 2. Available Doctors

- **Feature:** Users will be able to view a list of available doctors in their vicinity, including their specialties, consultation hours, and availability status.
- **Benefit:** This ensures that patients can find and book appointments with the right doctors conveniently, reducing waiting times and improving patient satisfaction.

### 3. Nearest Medicine Shop with Map Directions

- **Feature:** Integration with mapping services to provide directions to the nearest pharmacies.
- **Benefit:** Patients can easily locate and navigate to nearby medicine shops, ensuring they can promptly obtain prescribed medications and other medical supplies.

### 4. Live Consultation with Doctors

- **Feature:** The system will support live video consultations with doctors.
- **Benefit:** This feature allows patients to receive medical advice and consultations from the comfort of their homes, enhancing accessibility and convenience, especially for those unable to visit a clinic in person.

By incorporating these future features, our system aims to provide a comprehensive, user-friendly platform that supports patients in managing their health more effectively and efficiently. These enhancements will not only improve the user experience but also contribute to better health outcomes through timely and accessible medical support.

## Bibliography

1) Machine Learning for Healthcare: Review, Opportunities, and Challenges Authors: Chen M, Hao Y, Hwang K, Wang L, Wang L. Publication: IEEE Transactions on Cybernetics, 2017.

2) Title: Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records Authors: Miotto R, Li L, Kidd BA, Dudley JT. Publication: Scientific Reports, 2016.

3) Title: Predictive Modelling of Hospital Readmission Rates using Electronic Medical Record-wide Machine Learning: A Case-Study using Mount Sinai Heart Failure Cohort Authors: Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Publication: Pacific Symposium on Biocomputing, 2017.

4) <https://www.geeksforgeeks.org/disease-prediction-using-machine-learning/>

5) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8896926/>

6) <https://ieeexplore.ieee.org/document/9791739>