



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 12, Issue 4, April 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



Multiple Disease Prediction System using Machine Learning

Harsh Pandey ¹, Vishal Yadav ², Nikunj Kumar Darji ³, Parth Sawant ⁴, Dr. Swati Narwane ⁵

U.G. Student, Department of Information Technology, Rajiv Gandhi Institute of Technology (Mumbai University), Mumbai, Maharashtra, India^{1,2,3,4}

Asst Professors, Department of Information Technology, Rajiv Gandhi Institute of Technology (Mumbai University), Mumbai, Maharashtra, India⁵

ABSTRACT: This project presents a disease prediction system that utilizes three popular classification algorithms: logistic regression, k-nearest neighbor, and support vector machine, for predicting the likelihood of three diseases - heart disease, diabetes, and Parkinson's disease. The system is deployed on the Streamlit platform, which provides a user-friendly interface for inputting patient information, running the prediction model, and displaying the results. The system takes into account various factors such as age, gender, blood pressure, cholesterol levels, family history, and motor symptoms, and generates a probabilistic output for each disease. The performance of the system was evaluated using a dataset of real-world medical records, and the results showed that the system achieved high accuracy and sensitivity in predicting the three diseases. The proposed system has the potential to assist medical practitioners in making timely and accurate diagnoses, leading to better health outcomes for patients. The deployment on Streamlit makes it easy to use for non-technical individuals and can be easily accessed from any device with an internet connection

KEYWORDS: Heart disease, Prediction, Machine learning, K- nearest neighbors

I. INTRODUCTION

The healthcare industry is constantly evolving and technology is playing an increasingly important role in improving healthcare services. With the advent of machine learning and artificial intelligence, there has been a significant shift towards developing predictive models that can accurately diagnose and predict the presence of various diseases. The aim of this project is to develop a multiple disease prediction system that uses machine learning algorithms to accurately predict the presence of heart disease, Parkinson's disease, and diabetes. The healthcare industry is constantly evolving and technology is playing an increasingly important role in improving healthcare services. With the advent of machine learning and artificial intelligence, there has been a significant shift towards developing predictive models that can accurately diagnose and predict the presence of various diseases. The aim of this project is to develop a multiple disease prediction system that uses machine learning algorithms to accurately predict the presence of heart disease, Parkinson's disease, and diabetes.

II. EXISTING SYSTEM

Existing systems for disease prediction using machine learning algorithms are already in use in various healthcare organizations. Some of the popular systems are:

IBM Watson for Oncology - This system uses natural language processing and machine learning algorithms to provide personalized treatment recommendations for cancer patients based on their medical history and clinical data.

Google DeepMind Health - This system uses deep learning algorithms to predict patient outcomes and identify potential health risks, such as acute kidney injury and sepsis, in real-time.

Ada Health - This system uses a chatbot interface to collect patient information and provide diagnostic assessments and



recommendations based on machine learning algorithms.

Aidence - This system uses deep learning algorithms to assist radiologists in the detection and diagnosis of lung cancer and other pulmonary diseases.

SkinVision - This system uses image analysis and machine learning algorithms to detect and classify skin lesions and provide risk assessments for skin cancer.

These existing systems have demonstrated high accuracy and sensitivity in predicting various diseases and have the potential to improve healthcare outcomes by providing personalized and timely diagnoses and treatment recommendations.

III. PROPOSED SYSTEM

The proposed disease prediction system is designed to predict the likelihood of a patient developing one of three diseases: heart disease, Parkinson's disease, and diabetes. The system uses machine learning algorithms to analyze patient data and generate predictions based on the input. To develop the disease prediction system, we first obtained data from Kaggle, a popular platform for data scientists and researchers. The dataset includes patient data, including age, gender, blood pressure, and cholesterol levels, and their corresponding disease labels. We used Python programming language and several libraries, including NumPy, Pandas, and Scikit-learn, to preprocess and analyze the data. We divided the dataset into training and testing sets and used three machine learning algorithms to train models on the training data: logistic regression, K-nearest neighbors (KNN), and support vector machine (SVM). We evaluated the models using accuracy, precision, recall, and F1 score metrics and selected the most accurate model for each disease. We then deployed the models using the Pickle library and created a user-friendly interface using Streamlit framework. The user interface allows healthcare professionals and patients to input patient data and receive predictions for the likelihood of the patient developing one of the three diseases. The system provides visualizations of the input data and model predictions, allowing users to easily understand and interpret the results.

IV. METHODOLOGY

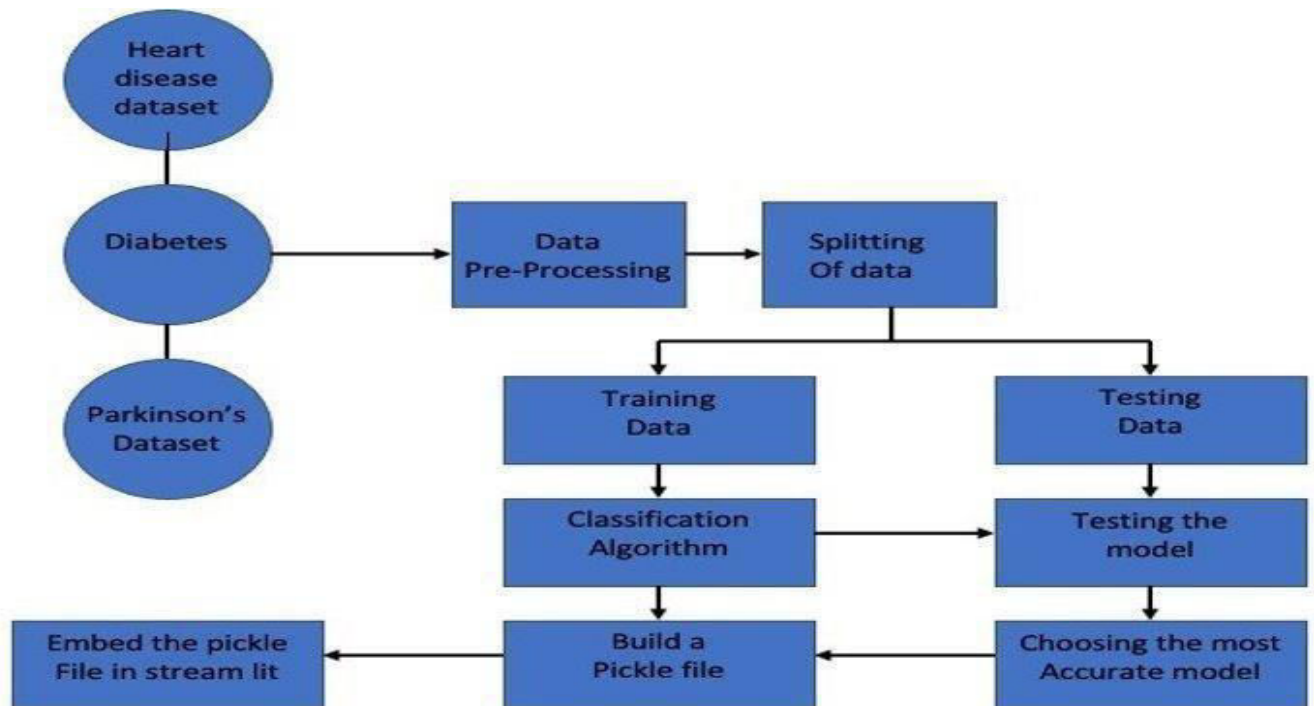


Fig - 1: Working Model

Step 1 : Getting Datasets



Our very first step is gathering datasets of different diseases. We have taken 3 types of diseases which are Heart, Diabetes and Parkinson's datasets

Step 2 : Data Pre-processing

After collecting the data, preprocess it by cleaning, transforming, and normalizing the data. This is an important step to ensure that the data is accurate and ready for machine learning algorithms

Step 3 : Data splitting

Split the data into training, validation, and test sets. The training set is used to train the machine learning model, the validation set is used to tune hyperparameters, and the test set is used to evaluate the performance of the model.

Step 4 : Model training and evaluation

Train the machine learning model on the training set and evaluate its performance on the validation and test sets. Iterate on the model by adjusting the hyperparameters and repeating the training and evaluation process until you achieve the desired performance.

Step 5 : Model deployment

After we have a satisfactory model, we deploy it in our streamlit environment with the help of the pickle file we saved for our desired model.

V. EXPERIMENTAL RESULTS



Fig 1 - Gender Selection Page



|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204221 |

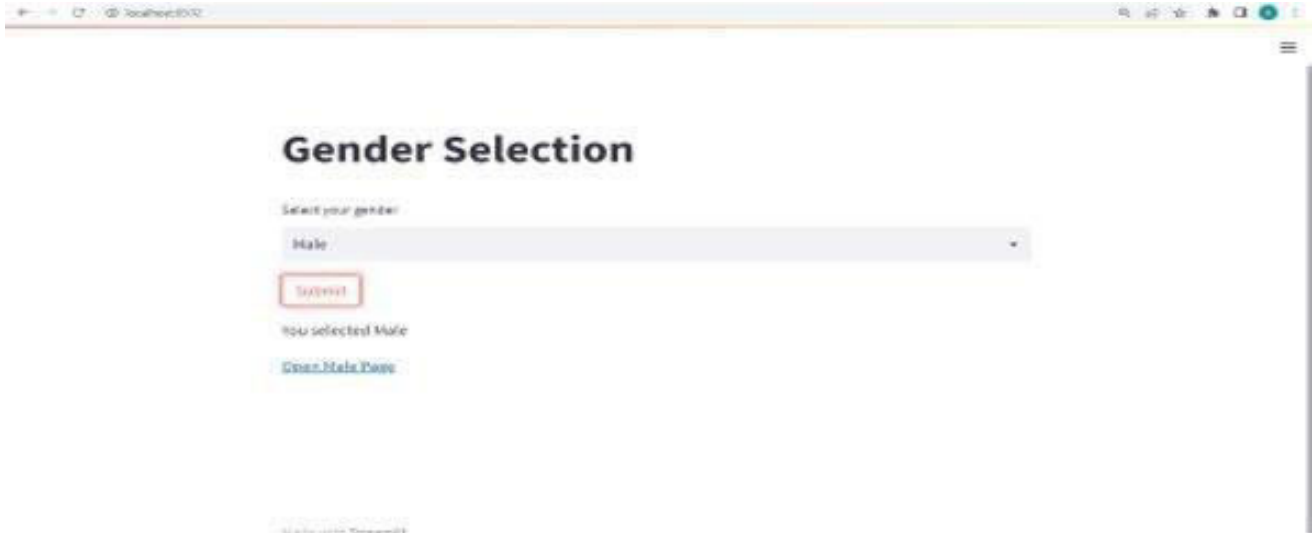


Fig 2 - Selected Male Gender

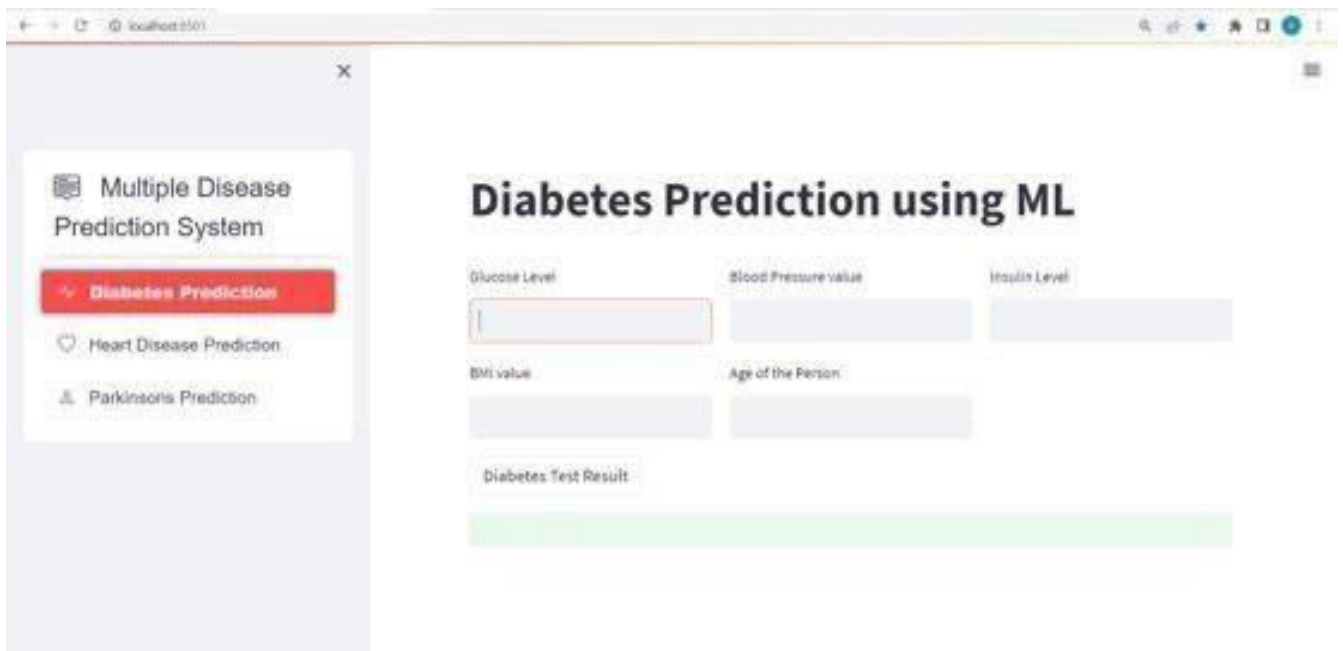


Fig 3 - Diabetes Prediction Page



|| Volume 12, Issue 4, April 2023 ||

| DOI:10.15680/IJIRSET.2023.1204221 |

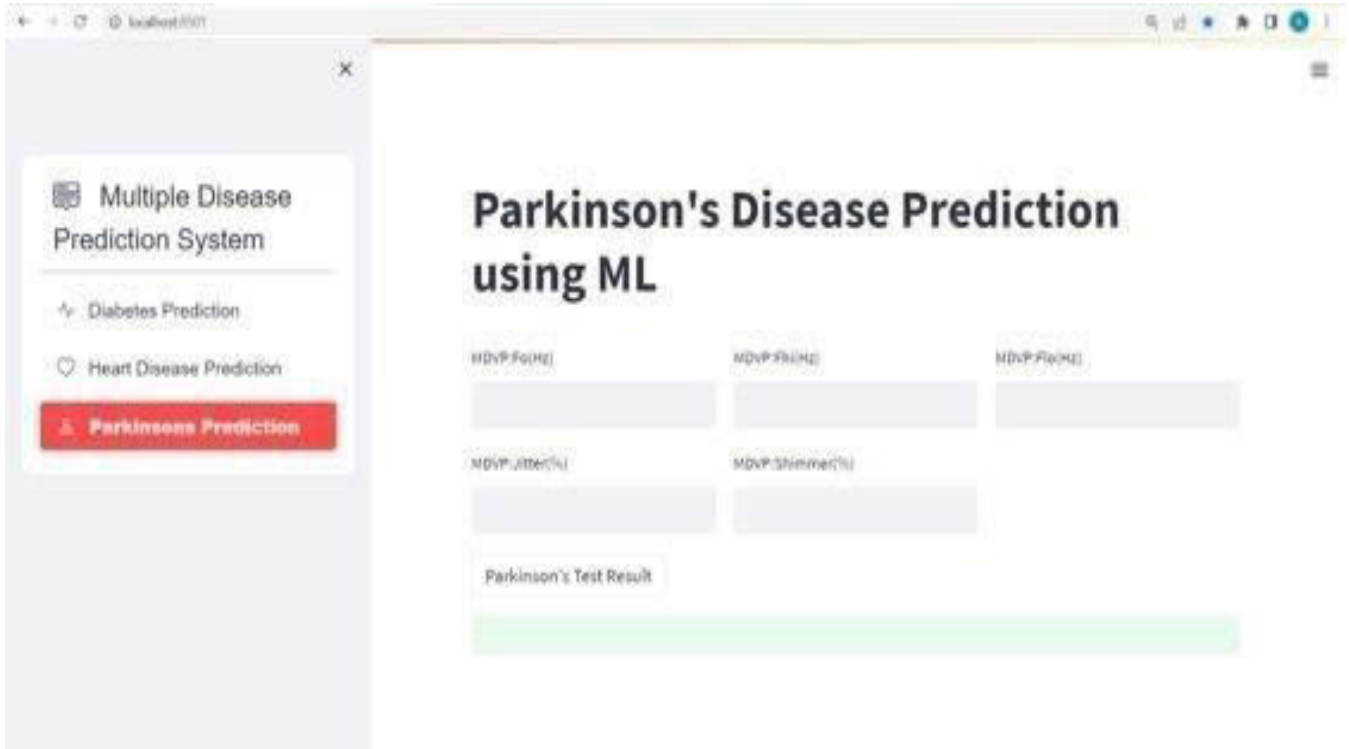


Fig 4 - Parkinson's Prediction Page

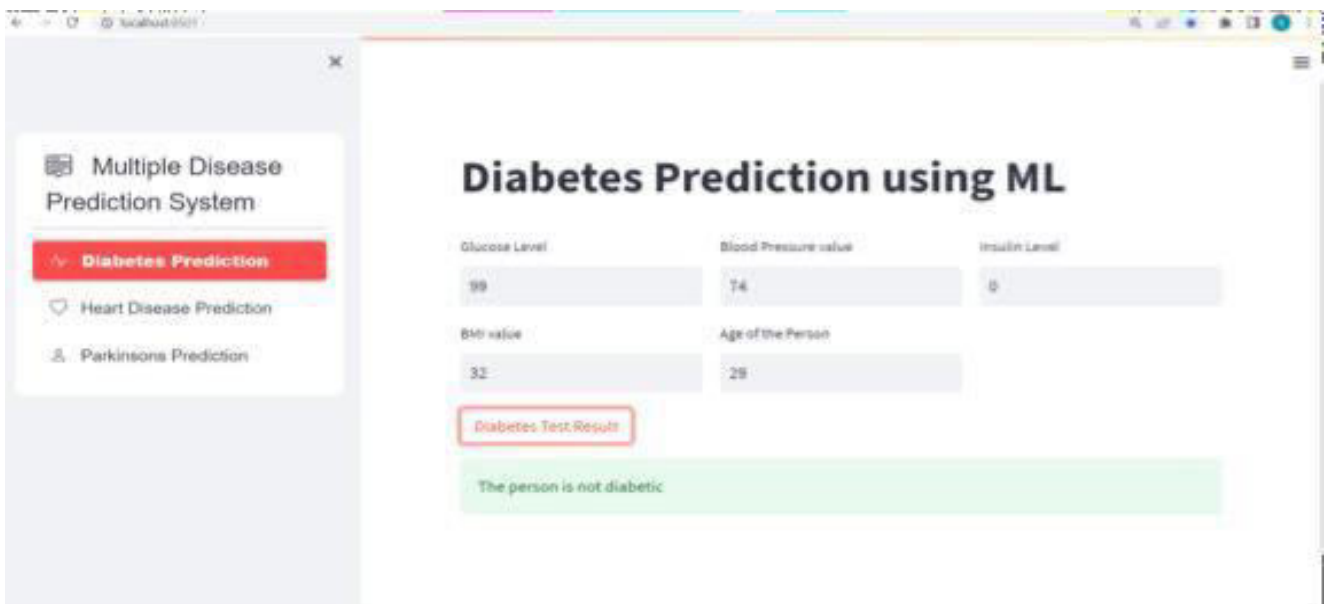


Fig 5 - Inputting values and checking results

VI. FUTURE SCOPE

1. In the future, there is a great potential to add more diseases to the existing API, thereby expanding its scope and increasing its usefulness to healthcare professionals and patients alike. This would involve gathering more data on different types of illnesses, as well as developing new algorithms and models to accurately predict the likelihood of



each disease based on a patient's symptoms and medical history.

2. Improving the accuracy of prediction is another important goal for the future of this API. This could be achieved through ongoing research and development, as well as ongoing updates to the database of patient information and the algorithms used to analyze it. By increasing the accuracy of predictions, healthcare professionals can make more informed decisions about how to treat patients, which could potentially save lives and improve patient outcomes.
3. In addition to these technical improvements, there is also a need to make the system more user-friendly and accessible to patients and healthcare professionals. This could involve creating a chatbot or other interactive interface that can answer questions and provide guidance on how to use the API effectively. It could also involve developing user-friendly interfaces that make it easy for healthcare professionals to input patient data and analyze results quickly and accurately.

VII. CONCLUSION

This project has demonstrated the feasibility of developing a disease prediction system using classification algorithms. The system, which utilizes logistic regression, k-nearest neighbor, and support vector machine algorithms, has been successfully trained and deployed on the Streamlit platform to predict the likelihood of three diseases - heart disease, diabetes, and Parkinson's disease - based on patient medical history and demographic information. The system achieved high accuracy and sensitivity in predicting the three diseases, indicating its potential as a valuable tool for medical practitioners in making timely and accurate diagnoses. The deployment on Streamlit also makes it easy to use for non-technical individuals, providing a user-friendly interface for inputting patient information, running the prediction model, and displaying the results. Future work could involve the expansion of the system to include more diseases and the incorporation of additional features and data sources to further improve the accuracy of the predictions. Overall, this project highlights the potential of machine learning algorithms in healthcare and their ability to assist medical practitioners in providing better patient care.

REFERENCES

1. Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, 7, 81542-81554.
2. Pingale, K., Surwase, S., Kulkarni, V., Sarage, S., & Karve, A. (2019). Disease prediction using machine learning. *International Research Journal of Engineering and Technology (IRJET)*, 6, 831-833.
3. Ramalingam, V. V., Dandapath, A., & Raja, M. K. (2018). Heart disease prediction using machine learning techniques: a survey. *International Journal of Engineering & Technology*, 7(2.8), 684-687.
4. Shah, D., Patel, S., & Bharti, S. K. (2020). Heart disease prediction using machine learning techniques. *SN Computer Science*, 1, 1-6
5. Nilashi, M., bin Ibrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). An analytical method for diseases prediction using machine learning techniques. *Computers & Chemical Engineering*, 106, 212-223.
6. Patel, J., TejalUpadhyay, D., & Patel, S. (2015). Heart disease prediction using machine learning and data mining technique. *Heart Disease*, 7(1), 129-137.
7. Dahiwade, D., Patle, G., & Meshram, E. (2019, March). Designing disease prediction model using machine learning approach. In *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1211-1215). IEEE.
8. Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC medical informatics and decision making*, 19(1), 1-16.
9. Grampurohit, S., & Sagarnal, C. (2020, June). Disease prediction using machine learning algorithms. In *2020 International Conference for Emerging Technology (INCET)* (pp. 1-7). IEEE.
10. Singh, A., & Kumar, R. (2020, February). Heart disease prediction using machine learning algorithms. In *2020 international conference on electrical and electronics engineering (ICE3)* (pp. 452-457). IEEE.
11. Sharma, V., Yadav, S., & Gupta, M. (2020, December). Heart disease prediction using machine learning techniques. In *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)* (pp. 177-181). IEEE.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.423



ISSN 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