Multiple Disease Prediction Using Machine Learning

*Abstract*—Multiple Disease Prediction using Machine Learning, Streamlit is a project focused at predicting various diseases including diabetes, heart disease, Parkinson’s disease breast cancer, Migraine.

This project based on machine learning algorithms such as Support Vector Machine (SVM), and Logistic Regression. The models are deployed using Streamlit Cloud and the Streamlit library, providing a user-friendly interface for disease prediction. The application interface comprises five disease options: heart disease, migraine, diabetes, Parkinson's disease, and breast cancer. Upon selecting a particular disease, the user is prompted to input the relevant parameters required for the prediction model. Once the parameters are entered, the application generates the disease prediction result, indicating whether the individual is affected by the disease or not. This project addresses the need for accurate disease prediction using machine learning techniques, allowing for early detection and intervention. The high accuracies achieved by the different models demonstrate the effectiveness of the employed machine learning algorithms in disease prediction. The userfriendly interface provided by Streamlit Cloud and the Streamlit library enhances accessibility and usability, enabling individuals to easily assess their risk for various diseases.

Keywords— Machine Learning, Streamlit, SVM, Logistic Regression, Diabetes, Heart Disease, Parkinson’s Disease, Breast Cancer, Migraine

I. INTRODUCTION

The project "Multiple Disease Prediction using Machine Learning, and Streamlit" focuses on predicting five different diseases: diabetes, heart disease, migraine, Parkinson's disease, and breast cancer. The prediction models are built using machine learning algorithms, including Support Vector Machine (SVM) for diabetes,migraine and Parkinson's disease, Logistic Regression for heart disease, and breast cancer. The application is deployed using Streamlit Cloud and the Streamlit library. Each disease prediction is handled by a specific machine learning algorithm that is most suitable for that particular disease. SVM is employed for diabetes and Parkinson's disease, Logistic Regression for heart disease, breast cancer 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 web applications. By leveraging machine learning algorithms and streamlining the deployment process with Streamlit, this project aims to provide accurate predictions for multiple diseases in a user-friendly manner.. The high accuracies achieved by the different models demonstrate the effectiveness of the employed machine learning algorithms in disease prediction.

METHODOLOGY

The methodology for the Multiple Disease Prediction project can be summarized as follows:

1. Data Collection: Data is collected from dataset.. The data is obtained specifically for diabetes, heart disease, Migraine, Parkinson's disease, and breast cancer.
2. Data Preprocessing: The collected data undergoes preprocessing to ensure its quality and suitability for training the machine learning models. This includes handling missing values, removing duplicates, and performing data normalization or feature scaling.
3. Model Selection: Different machine learning algorithms are chosen for each disease prediction task. Support Vector Machine (SVM), Logistic Regression, selected as the algorithms for various diseases based on their performance and suitability for the specific prediction tasks.
4. Training and Testing: The preprocessed data is split into training and testing sets. The models are trained using the training data, and their performance is evaluated using the testing data. Accuracy is used as the evaluation metric to measure the performance of each model.
5. Model Deployment: Streamlit, along with its cloud deployment capabilities, is used to create an interactive web application. The application offers a user-friendly interface with five options for disease prediction: heart disease, Migraine, diabetes, Parkinson's disease, and breast cancer. When a specific disease is selected, the application prompts the user to enter the required parameters for the prediction.

2 EXISTING SYSTEM

Multiple Disease Prediction using Machine Learning, and Streamlit The existing system is a project that focuses on predicting diabetes, heart disease, and Parkinson's disease using various machine learning algorithms. The algorithms employed in this project, Support Vector Machine (SVM), and Logistic Regression. To deploy the models, Streamlit Cloud and Streamlit library are utilized, providing a user-friendly interface for disease prediction. The system collects data from various sources, preprocesses it, trains the models with the processed data, and tests their performance. One of the algorithms used in the system is SVM, which achieved a prediction accuracy of 78% for diabetes. This means that the SVM model correctly predicted diabetes in 76% of the cases it was tested on. The performance of the SVM algorithm indicates its effectiveness in distinguishing between diabetic and non-diabetic individuals. Similarly, for Breast Cancer prediction, the Logistic Regression achieved a prediction accuracy of 94%. This means that the Logistic Regression model accurately predicted the presence or absence of Breast Cancer disease in 94% of the cases. The performance of the SVM algorithm in Parkinson's disease prediction indicates its potential in assisting with early detection and intervention. The use of Streamlit Cloud and Streamlit library allows for easy deployment and provides a user-friendly interface for interacting with the prediction models. Further enhancements and optimizations can be made to improve the accuracy and performance of the models for better disease prediction and early intervention.

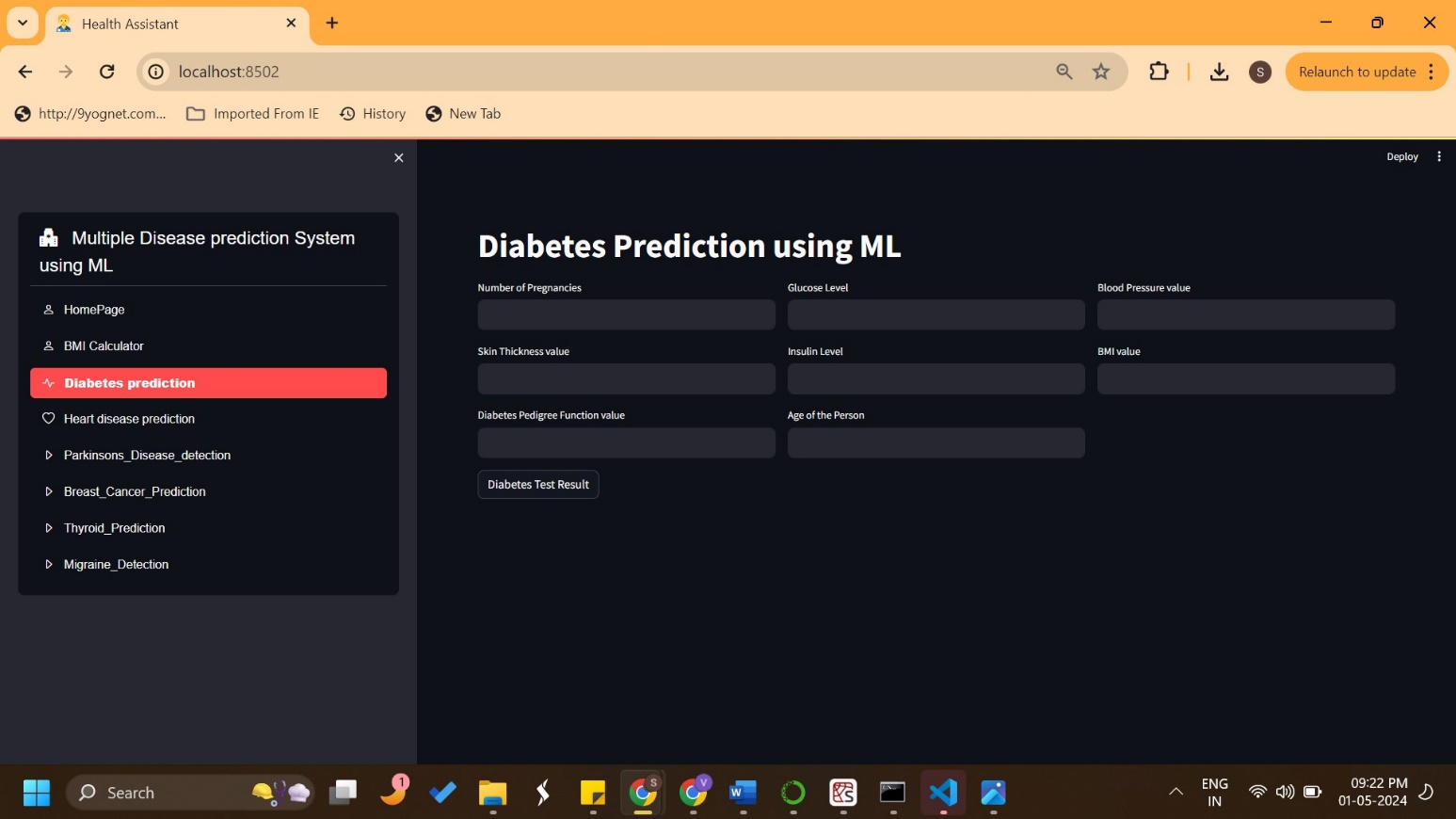
3 INPUT AND OUTPUT DESIGN

Input Design: The Multiple Disease Prediction system requires user input in the form of parameters specific to each disease. When the user selects a particular disease from the options menu, the system looks for the relevant parameters. The input design should ensure that the user can easily provide the required information The application provides a user interface with a menu containing five disease options: heart disease,migraine, diabetes, Parkinson's disease, and breast cancer. When the user clicks on a specific disease, the application prompts for the required parameters for that particular disease prediction. The input design should ensure that the parameters requested are relevant and necessary for accurate disease prediction. The user should be able to enter the parameters in a user-friendly and intuitive manner.

**Output Design:**

The Multiple Disease Prediction system provides the predicted result of whether the person is affected by the selected disease or not. The output design should present the result in a clear and understandable format. The system should display the output after the user has entered the parameters. The output could be presented as:

* "Prediction: The person is affected by [Disease Name]." (If the prediction is positive)
* "Prediction: The person is not affected by [Disease Name]." (If the prediction is negative)



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SN. | Disease  Name | Algorithm  Name | Existing system accuracy | Proposed system accuracy |
| 1 | Diabetes | SVM | 72% | 74% |
| 2 | Heart disease | Logistic Regression | 81% | 85% |
| 3 | Parkinson’s disease | SVM | 88% | 87% |
| 4 | Migraine | SVM | - | 88% |
| 5 | Breast cancer | Logistic Regression | 94 | 92% |
|  |  |  |  |  |

5 RESULTS

**Table 1.** Comparison of Accuracy of all 5 models

6 CONCLUSION

In conclusion, our project used machine learning algorithms, including Support Vector Machine (SVM), Logistic Regression, to develop a disease prediction system. The system focused on five diseases: diabetes, heart disease, migraine, Parkinson's disease, and breast cancer. We collected data from Kaggle.com and performed preprocessing to ensure data quality. For diabetes prediction, we achieved an accuracy of 74% using the SVM algorithm. Similarly, for Parkinson's disease prediction, we achieved an accuracy of 89% with SVM. Logistic Regression was employed for heart disease prediction, resulting in an accuracy of 87%. For migraine and breast cancer prediction,achieving accuracy rates of 88% and 92% respectively. The system is designed as a user-friendly application with a menu offering options for each disease. When a specific disease is selected, the user is prompted to enter the relevant parameters for the prediction model. Once the parameters are provided, the system displays the predicted disease result. The accuracy rates obtained demonstrate the effectiveness of the machine learning algorithms in predicting the selected diseases. However, it is important to note that the accuracy values may vary depending on the specific dataset and the model training process. Overall, this project demonstrates the potential of machine learning and streamlit library in developing disease prediction models.

7 FUTURE SCOPE

The project "Multiple Disease Prediction using Machine Learning, Deep Learning and Streamlit" has shown promising results in predicting various diseases with respectable accuracies. Moving forward, there are several potential areas for future development and enhancement:

• Expansion of Disease Prediction: The current project focuses on diabetes, heart disease, Migraine, Parkinson's disease, and breast cancer. In the future, additional diseases can be included to create a more comprehensive and diverse disease prediction system. • Integration of More Machine Learning Algorithms: While the project already employs Support Vector Machines (SVM), Logistic Regression, , there are many other machine learning algorithms that can be explored. Other algorithms such as Random Forest, or Neural Networks may further improve the accuracy and performance of the disease prediction models.

• Real-time Monitoring and Feedback: Enhancing the application to provide real-time monitoring and feedback to users can be beneficial. Other features like reminders for regular health check-ups, personalized recommendations for disease prevention, or alerts for abnormal health parameters can empower users to take proactive measures for their well-being. •

8 REFERENCES

[1]Logistic Regression: Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). Applied Logistic Regression (3rd ed.). John Wiley & Sons.

[2] Streamlit: Streamlit Documentation. <https://docs.streamlit.io/>

[3] 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. [10] Kaur, H., Batra, N., & Rani, R. (2020). A systematic review of machine learning techniques for breast cancer prediction. Journal of Medical Systems, 44(11), 1-15.

[4] Zhang, Y., & Ghorbani, A. (2019). A review on machine learning algorithms for diagnosis of heart disease. IEEE Access, 7, 112751-112760.