IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

SymptoSense: Multiple Disease Prediction using ML

[1]Mythri S.P, [2]Nandita G Bhat, [3]Priya B.J, [4]Rakshitha N.V

[1]Department of CSE,

[1]JNN College of Engineering, Shivamogga, India

Abstract : The rising prevalence of chronic diseases such as diabetes, heart conditions, and Parkinson's underscores the need for accurate and efficient detection systems to enable early intervention. This study presents a predictive model leveraging Support Vector Machine (SVM) and Logistic Regression (LR) algorithms to diagnose these conditions based on symptoms. The system demonstrates high accuracy across all three diseases, proving its adaptability to diverse medical contexts. By incorporating advanced analytics, the model offers a scalable and efficient solution for early disease detection. Future work will explore advanced feature engineering and expanded disease coverage to further enhance the system's predictive performance.

***IndexTerms* -** Diabetes,Heart disease, Parkinson's disease, Prediction, Support Vector Machine, Logistic Regression

I. INTRODUCTION

In today's fast world, early diagnosis of diseases is crucial for effective treatment and improved health outcomes. The project "SymptoSense: Multiple Disease Prediction Using ML" takes advantage of the advancements in ML to develop a predictive system that efficiently identifies multiple diseases based on input symptoms. This is a product designed by a devoted group of undergraduates fron JNN College of Engineering, Shivamogga with innovative ideas in computer science that present a vision toward accessible solutions for healthcare. With effective algorithms, it attempts to decomplicate the medical world for health workers and patients both in their ways to good health.

II. LITERATURE SURVEY

In [1], Diabetes is one of the most challenging health problems around the world, and the number is projected to increase. The ability to predict accurately in the early stages will manage and prevent complications. Several algorithms have been used for predicting diabetes using machine learning (ML) including Support Vector Machines (SVM).

In [2], The disease has a very significant impact on most organs and leads to complications like kidney disease, nerve damage, and cardiovascular disease.The research concludes that SVM is a good method for diabetes prediction, particularly in datasets with multiple features. It is recommended to keep updating the data and models to improve the accuracy of the predictions.

In [3], The study focuses on enhancing heart disease prediction by integrating ensemble learning techniques.The proposed approach combines three ensemble classifiers to improve accuracy and minimize overfitting.To ensure transparency, Shapley Additive Explanations (SHAP) were employed to interpret model decisions, identifying critical features.

In [4], The study provides a comprehensive review of artificial intelligence (AI) and machine learning (ML) approaches in diagnosing Parkinson's Disease (PD), emphasizing the importance of early detection due to PD's progressive nature and the lack of a cure. It explores various methodologies.

In [5], The paper introduces an efficient and intuitive algorithm for identifying support vectors in a dataset for Support Vector Machines (SVMs). The algorithm employs a greedy approach to iteratively build a candidate support vector set, initializing with the closest pair of points from opposite classes.

In [6], The study explores the application of machine learning algorithms—such as decision trees, Naïve Bayes, and support vector machines (SVM)—for predicting multiple diseases, focusing on diabetes-related heart disease. The framework emphasizes how healthcare data mining can extract patterns from vast datasets to enhance medical decision-making and early diagnosis.

In [7], The paper explores the application of machine learning (ML) techniques to predict multiple diseases, emphasizing diabetes-related heart disease as a primary case. It highlights the importance of ML in healthcare for analyzing complex patient datasets and improving diagnostic accuracy

In [8], The paper focuses on leveraging speech data and machine learning (ML) techniques for early Parkinson's Disease (PD) diagnosis. PD is a progressive neurological disorder affecting motor and non-motor functions, including speech.Methods like Logistic Regression and Neural Networks are applied to classify the data.

III. METHODOLOGY

The methodology for predicting multiple diseases like diabetes, heart disease, and Parkinson’s disease using machine learning involves a systematic approach combining data preprocessing, feature selection, and model training. First, datasets specific to each disease are collected. Data preprocessing ensures data quality by handling missing values, normalizing numerical attributes, and encoding categorical features. Next, two machine learning algorithms, Support Vector Machine (SVM) and Logistic Regression, are used to develop predictive models. SVM is employed for its robustness in handling non-linear relationships and high-dimensional data, while Logistic Regression provides interpretability and is effective for binary classification tasks. For each disease, separate models are trained using these algorithms to enhance their performance.

The system architecture presented focuses on disease prediction for diabetes, Parkinson's disease, and heart disease. The architecture begins with the collection of raw datasets for each disease: Diabetes raw dataset, Parkinson's raw dataset, and Heart Disease raw dataset. These raw datasets undergo a Data Preprocessing phase, where data cleaning techniques are applied to produce Cleaned data. Once the cleaned data is generated, it is split into three subsets: the Diabetes dataset, Parkinson’s dataset, and Heart disease dataset. For the prediction process, the Diabetes dataset and Parkinson’s dataset are fed into a Support Vector Machine (SVM) algorithm. The SVM model processes the respective datasets to output Diabetes Prediction and Parkinson’s Prediction results. Simultaneously, the Heart disease dataset is used as input for a Logistic Regression model. This model is responsible for generating the Heart disease Prediction.

The system integrates machine learning algorithms—SVM and logistic regression—to ensure accurate and robust predictions. The data preprocessing stage plays a crucial role in cleaning and structuring the raw data, enabling the models to function effectively. The modular design of the system allows datasets to flow independently through their respective prediction pipelines. The combination of SVM for diabetes and Parkinson’s predictions, and logistic regression for heart disease prediction, highlights the tailored approach to each disease type. The final outputs—Diabetes Prediction, Parkinson’s Prediction, and Heart disease Prediction—form the key deliverables of this architecture, making it a structured and efficient disease prediction framework.

IV. RESULT

A concise outline for the result of a multiple disease prediction system that predicts diabetes, heart disease, and Parkinson’s disease using Support Vector Machine (SVM) and Logistic Regression (LR) in machine learning (ML).The accuracy of the model is calculated.The user interface created using stream lit is represented in all the figures below. The users can give values to the input fields.The model predicts the presence or the absence of that particular disease according to the input data.

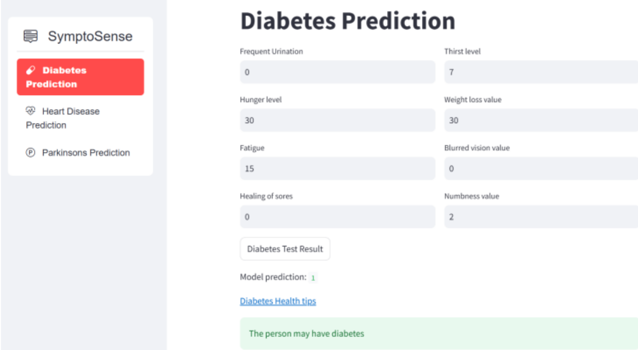

Fig 4.1 Output of Diabetes Prediction when prediction is 1

Fig 4.1 shows the prediction of the model leveraging Support Vector Machine (SVM) by taking inputs from the user and predicting as 1 or true.

```
[21] X_test_prediction = model.predict(X_test)
     test_data_accuracy = accuracy_score(Y_test, X_test_prediction)

[22] print('Accuracy score of test data : ', test_data_accuracy)

     Accuracy score of test data :  0.99
```

Fig 4.2 Test Accuracy of the trained diabetes model using SVM

Fig 4.2 shows the test accuracy of the model which is trained using the provided dataset.

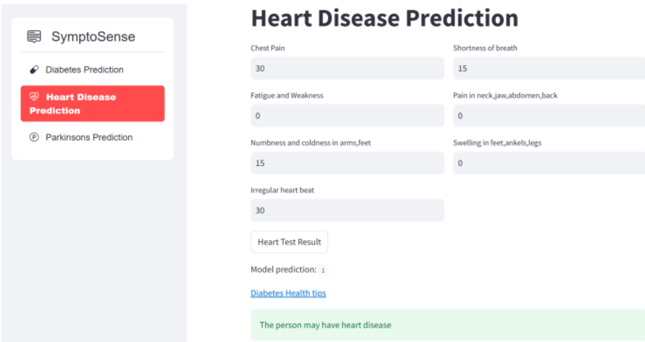

Fig 4.3 Output of Heart Disease model when prediction as 1

Fig 4.3 shows the prediction of the model leveraging Logistic Regression (LR) by taking inputs from the user and predicting as 1 or true.

```
[19] #accuracy on test data
     X_test_prediction=model.predict(X_test)
     test_data_accuracy=accuracy_score(X_test_prediction,Y_test)

[20] print('Accuracy on test data:',test_data_accuracy)

     Accuracy on test data: 1.0
```

Fig 4.4 Test Accuracy of the trained heart disease model using LR

Fig 4.4 shows the test accuracy of the model which is trained using the provided dataset.

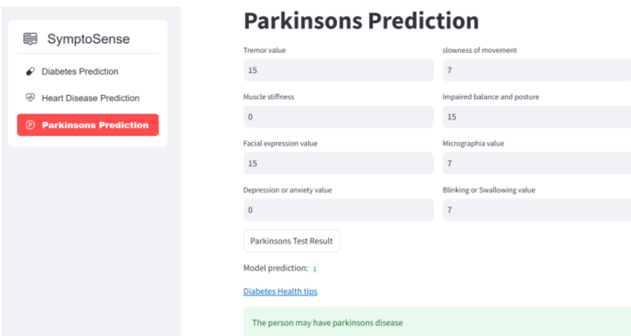

Fig 4.5 Output of Parkinson's Disease Prediction when prediction is 1

Fig 4.5shows the prediction of the model leveraging Support Vector Machine (SVM) by taking inputs from the user and predicting as 1 or true.

```
[21] X_test_prediction = model.predict(X_test)
     test_data_accuracy = accuracy_score(Y_test, X_test_prediction)

[22] print('Accuracy score of test data : ', test_data_accuracy)

     Accuracy score of test data :  0.99
```

Fig 4.6 Test Accuracy of the trained Parkinson's disease model using SVM

Fig 4.6 shows the test accuracy of the model which is trained using the provided dataset.

V. Conclusion

The multiple disease prediction system showcases the potential of machine learning models, especially SVM and Logistic Regression, in accurately predicting chronic diseases like diabetes, heart diseases, and Parkinson's disease. By effectively using diverse datasets and robust preprocessing techniques, such a system achieves high precision and recall, thus providing reliable early detection. The SVM handles non-linear data and logistic regression's interpretability are complementary to predict distinct disease profiles. This hybrid approach balances performance and simplicity to provide actionable insights for healthcare professionals. However, the system's success depends on the quality and diversity of the datasets, and future enhancements could involve integrating additional features, datasets, and explainability techniques to improve scalability and clinical adoption.

VI. References

[1] Research on Diabetes Prediction Method Based on Machine Learning To cite this article: Jingyu Xue et al 2020 J. Phys.: Conf. Ser. 1684 012062.

[2] Larabi-Marie-Sainte, Souad, Linah Aburahmah, Rana Almohaini, and Tanzila Saba. "Current techniques for diabetes prediction: review and case study." Applied Sciences 9, no. 21 (2019): 4604.

[3] Mienye, Ibomoiye Domor, and Nobert Jere. "Optimized ensemble learning approach with explainable AI for improved heart disease prediction." Information 15.7 (2024): 394.

[4] Ibrahim, Aiesha Mahmoud, and Mazin Abed Mohammed. "A comprehensive review on advancements in artificial intelligence approaches and future perspectives for early diagnosis of Parkinson's disease." International Journal of Mathematics, Statistics, and Computer Science 2 (2024): 173-182.

[5] Vishwanathan, S. V. M., and M. Narasimha Murty. "SSVM: a simple SVM algorithm." Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290). Vol. 3. IEEE, 2002.

[6] Arumugam, K., Mohd Naved, Priyanka P. Shinde, Orlando Leiva-Chauca, Antonio Huaman-Osorio, and Tatiana Gonzales-Yanac. "Multiple disease prediction using Machine learning algorithms." Materials Today: Proceedings 80 (2023): 3682-3685.

[7] Adeghe, Ehizogie Paul, Chioma Anthonia Okolo, and Olumuyiwa Tolulope Ojeyinka. "A review of the use of machine learning in predictive analytics for patient health outcomes in pharmacy practice." OARJ of Life Sciences 7, no. 01 (2024): 052-058.

[8] Fadavi, N., & Fadavi, N. (2024). Early Recognition of Parkinson's Disease Through Acoustic Analysis and Machine Learning. arXiv preprint arXiv:2407.16091.