Multiple Disease Prediction Using Machine Learning Algorithm

Farzana Jawale
Student of Department of Computer Engineering
Nutan Maharashtra Institute of Engineering and Technology
Pune, India

Ritik Singh
Student of Department of Computer Engineering
Nutan Maharashtra Institute of Engineering and Technology
Pune, India

Dr. Naveenkumar Jayakumar
Department of Computational Intelligence
Vellore Institute of Technology Vellore

Dr. Saurabh Saoji
Nutan Maharashtra Institute of Engineering and Technology
Pune, India

***Abstract*—** Due to the large quantum of information, it's delicate for croakers to directly descry the symptoms of the complaint and make an early opinion of the complaint. There are two ways to determine whether a specific complaint is present in the case's body. It takes a lot of time manually. Using machine literacy algorithms makes our job easier. thus, each contagion has its own specific uses. There's no universal system or practice for prognosticating colorful conditions. thus, the proposed system has an operation that can prognosticate numerous conditions with the help of stoner input. Algorithms used in the proposed system; support vector machines, logistic retrogression, decision trees and KNN algorithms. The results presented by the system are in double format similar as" yes" or" no"

Index Terms:- Heart Disease, Diabetes, Parkinsons Prediction Machine Learning, accuracy

I. INTRODUCTION

Due to the large quantum of information, it's delicate for medicalprofessionals to directly diagnose symptoms and descry conditions at an earlystage.There are now numerous form operations that can prognosticate health conditions. Machine literacy algorithms and some data mining Technologies are designed to dissect current training patterns in medical treatment, fastening on prognosticating specific crimes in an operation. For illustration, an app for liver webbing, an app for cancer webbing, an app for lung webbing, etc. There's no single machine learning operation that will bring together multiple complaint vaticinationmodels.The plan is grounded on machinelearning.Machine literacy is the use of literal data to make prognostications. The purpose of using this model is to give a stoner-friendly interface to directly prognosticate colorful conditions from a single operation

II. LITERATURE SURVEY

This literature check conducted for this exploration design explores the being body of knowledge regarding the operation of machine literacy ways, specifically Support Vector Machines(SVM), for the vaticination of multiple conditions, including cardiovascular complaint, diabetes, and Parkinson's complaint. The check encompasses studies that have addressed analogous exploration objects, methodologies, and issues, furnishing precious perceptivity and establishing the foundation for the current design.[1]

Machine Learning for Disease Prediction Machine literacy models have been considerably employed for complaint vaticination in colorful disciplines. Employed SVM to prognosticate multiple conditions grounded on electronic health records, demonstrating the model's efficacity in relating complaint patterns. also, employed SVM for complaint vaticination using clinical data, emphasizing the significance of point selection and model optimization ways. These studies establish the applicability and effectiveness of machine literacy algorithms in complaint vaticination.[2]

Heart Disease Prediction Several studies have explored the use of machine literacy, including SVM, for heart complaint vaticination. Developed a SVMbased model to prognosticate heart complaint using a combination of demographic, clinical, and electrocardiogram(ECG) features. Their study achieved high delicacy in detecting heart complaint, emphasizing the eventuality of SVM in this sphere. also, employed SVM to prognosticate heart complaint grounded on features similar as blood pressure, cholesterol situations, and medical history. These studies punctuate the connection and effectiveness of SVM in heart complaint vaticination.[3]

- Diabetes vaticination The vaticination of diabetes using machine literacy models, including SVM, has garnered significant attention employed SVM to prognosticate diabetes grounded on clinical and inheritable features, demonstrating the model's eventuality for accurate diabetes threat assessment. also, employed SVM to prognosticate diabetes using features similar as glucose situations, body mass indicator, and blood pressure. These studies emphasize the effectiveness of SVM in diabetes vaticination and emphasize the significance of incorporating applicable features.[4]

- Parkinson's complaint vaticination Machine literacy ways, including SVM, have been explored for the vaticination of Parkinson's complaint. Employed SVM to prognosticate the inflexibility of Parkinson's complaint grounded on voice features, achieving promising results. also, employed SVM to prognosticate Parkinson's complaint using voice recordings, pressing the eventuality of SVM innon-invasive and accessible vaticination styles. These studies demonstrate the feasibility of SVM in Parkinson's complaint vaticination and its implicit for early discovery.[5]

III. METHODOLOGY

The methodology of Machine Learning based Multiple Disease Predictiosn (MLMDP) is implemented in the following steps:

1. Define the problem: First, you need to clearly define the problem you are trying to solve. This includes understanding the domain, the data you have, and the expected output.

2. Collect data: Once you have defined the problem, you need to collect relevant data that can be used to train the machine learning algorithm. This could involve gathering data from different sources or generating synthetic data.

3. For this MLMDP system Heart Disease dataset, Diabetes dataset and Parkinson's prediction dataset is collected from Kaggle website where heart disease dataset contains 1025 rows and 14 cols, Diabetes dataset contains , Parkinsons prediction Dataset contains .

4. Data Pre-processing: Data pre-processing involves cleaning, transforming, and normalizing the data so that it can be used for training the machine learning algorithm. This step is crucial as the quality of the data can significantly impact the performance of the model.

5. Dividing the data: Two sets of data must be created: a training set and a testing set. The model is trained on the training set, and its performance is assessed on the testing set.

6. Select a model: Once the data is pre-processed and split, you need to select an appropriate machine learning model for the problem. This involves understanding the strengths and weaknesses of different models and selecting the one that is best suited for the problem.

7. You can now train the machine learning algorithm using the training set after choosing the model and the data. Finding the ideal set of parameters that reduces the discrepancy between the model's predictions and the actual output entails applying an optimization technique.

8. A model needs to be evaluated using the testing set after it has been trained. To evaluate how effectively the model is doing, metrics like accuracy, precision, and recall must be calculated.

9. Tune the model: Depending on the evaluation's findings, you might need to make changes to the model's architecture or its parameters.

10. Deploy the model: Once the performance of the model is good, than it can deployed, and used for make predictions on new data. This could involve integrating the model into a larger system or creating an API that allows others to use the model.

11. Monitor and maintain the model: Finally, it is important to monitor and maintain the model over time to ensure that it continues to perform well as new data becomes available. This could involve retraining the model with new data or updating its parameters as needed.

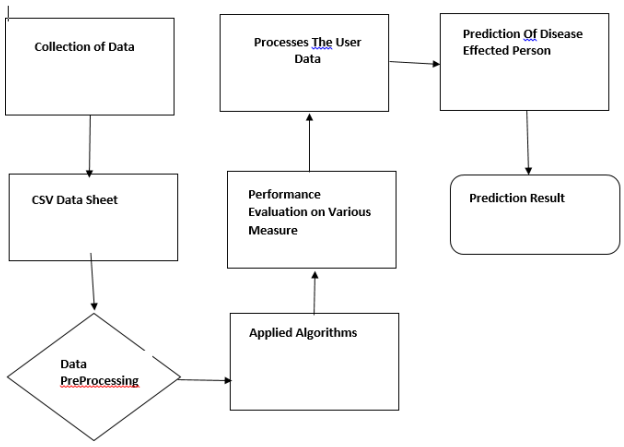

Fig 1 System Architecture

IV. MODELING AND ANALYSIS

❖ Algorithm:

- Step 1: User input.
- Step 2: Data Preprocessing.
- Step 3: Model Prediction.
- Step 4: Result Prediction.
- Step 5: Interpretation and Visualization.
- Step 6: User Interaction.

V. PROPOSED SYSTEM

This project is mainly concentrated on building an application which incorporates more than one disease with a user friendly interface. One potential solution to this problem is to use machine learning algorithms to build a single, unified model that can predict the presence or absence of multiple diseases simultaneously.

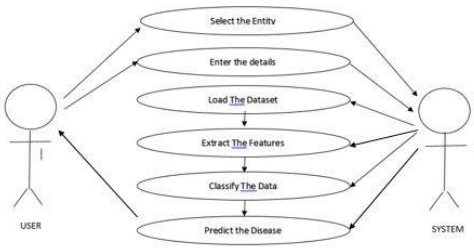

Fig. 2. Use case Diagram

VI. FUTURE SCOPE

Incorporating more data sources: Currently, multiple disease prediction systems typically rely on electronic health records and medical imaging data. In the future, other data sources such as wearable devices, social media, and environmental data could be integrated into these systems to provide a more comprehensive picture of a patient's health. Addressing data bias: As with all machine learning algorithms, bias in the training data can lead to inaccurate predictions and perpetuate health disparities. Future work should focus on developing methods to address and mitigate data bias, such as using more diverse and representative datasets, and incorporating fairness and equity considerations into the algorithm development process. Advancing personalized medicine: Multiple disease prediction using machine learning has the potential to enable more personalized and precise medicine, by predicting an individual's risk of developing specific diseases based on their unique medical history and other factors. Future work should focus on developing personalized treatment plans based on these predictions, including targeted prevention strategies and personalized treatment options.

VII. CONCLUSION

Multiple disease prediction using machine learning is a promising approach to healthcare that has the potential to revolutionize the way we diagnose and treat diseases. By using machine learning algorithms to analyze large amounts of patient data, we can identify patterns and correlations that may not be immediately apparent to human clinicians. This approach has the potential to enable earlier diagnosis, better treatment, and improved patient outcomes. While there are challenges and limitations to the use of machine learning in healthcare, such as the risk of bias and the need for diverse and representative data, ongoing research and development in this field is helping to address these challenges and unlock the full potential of multiple disease prediction using machine learning. As technology continues to evolve and more data becomes available, it is likely that machine learning algorithms will become increasingly sophisticated and accurate, leading to even better patient outcomes and more personalized medicine. Multiple disease prediction using machine learning has the potential to transform healthcare, and it is an exciting area of research that holds great promise for the future.

VIII. REFERENCES

[1] Liang H, Tsui BY, Ni H, et al. Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. Nat Med. 2019;25(3):433- 438.

[2] [2] Deo RC. Machine learning in medicine. Circulation. 2015;132(20):1920-1930.

[3] Rajendra Acharya U, Fujita H, Oh SL, et al. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. Inf Sci (Ny). 2017;415-416:190-198.

[4] Paniagua JA, Molina-Antonio JD, LopezMartinez F, et al. Heart disease prediction using random forests. J Med Syst. 2019;43(10):329.

[5] Poudel RP, Lamichhane S, Kumar A, et al. Predicting the risk of type 2 diabetes mellitus using data mining techniques. J Diabetes Res. 2018;2018:1686023.

[6] Al-Mallah MH, Aljizeeri A, Ahmed AM, et al. Prediction of diabetes mellitus type-II using machine learning techniques. Int J Med Inform. 2014;83(8):596-604.

[7] Tsanas A, Little MA, McSharry PE, Ramig LO. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. J R Soc Interface. 2012;9(65):2756-2764.

[8] Arora S, Aggarwal P, Sivaswamy J. Automated diagnosis of Parkinson's disease using ensemble machine learning. IEEE Trans Inf Technol Biomed. 2017;21(1):289-299.

[9] Ahmad F, Hussain M, Khan MK, et al. Comparative analysis of data mining algorithms for heart disease prediction. J Med Syst. 2019;43(4):101.

[10] Parashar A, Gupta A, Gupta A. Machine learning techniques for diabetes prediction. Int J Emerg Technol Adv Eng. 2014;4(3):672-675.

[11] Breiman L. Random forests. Mach Learn. 2001;45(1):5-32.

[12] [12] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer; 2009.

[13] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011;12:2825-2830.

[14] McKinney W, van der Walt S, Lamoureux C, et al. Data structures for statistical computing in Python. In: Proceedings of the 9th Python in Science Conference; 2010.

[15] Pickle Python object serialization. Python documentation. https://docs.python.org/3/library/pickle.html. Accessed May 26, 2023.

[16] Huang ML, Hung CC, Hsu CY, et al. Predicting ischemic stroke using the Framingham Stroke Risk Score and a simple decision rule in Chinese patients with type 2 diabetes. Diabetes Care. 2010;33(2):427-429.

[17] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and

Regression Trees. Wadsworth and Brooks; 1984