# Review Paper Multiple Disease Prediction using Machine Learning

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

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

Dr. Saurabh Saoji
*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*

***Abstract*—Machine literacy ways have revolutionized the field of healthcare by enabling accurate and timely complaint vaticination. The capability to prognosticate multiple conditions contemporaneously can significantly ameliorate early opinion and treatment, leading to better case issues and reduced healthcare costs. This exploration paper explores the operation of machine literacy algorithms in prognosticating multiple conditions, fastening on their benefits, challenges, and unborn directions. We present an overview of colorful machine literacy models and data sources generally used for complaint vaticination. Also, we bandy the significance of point selection, model evaluation, and the integration of multiple data modalities for enhanced complaint vaticination. The exploration findings punctuate the eventuality of machine literacy inmulti-disease vaticination and its implicit impact on public health. Once further, I'm applying machine literacy model to identify that a person is affected with many complaint or not. This training model takes a sample data and train itself for prognosticating complaint.**

***Index Terms*—Machine Learning Disease Prediction, Disease data, Machine Learning.**

## I. Introduction

In recent times, the field of machine literacy has witnessed remarkable advancements and operations in colorful disciplines, including healthcare. The capability to prognosticate multiple conditions contemporaneously using machine literacy models has the implicit to revise medical diagnostics and ameliorate patient issues. This exploration paper explores the application of the Support Vector Machines (SVM) model to prognosticate the presence of three current conditions heart complaint, diabetes, and Parkinson's complaint. Cardiovascular conditions, diabetes, and Parkinson's complaint are significant public health enterprises that put a considerable burden on individualities and healthcare systems worldwide. Beforehand discovery and accurate opinion of these conditions play a vital part in perfecting patient prognostic, optimizing treatment plans, and reducing healthcare costs. Machine literacy, with its capability to dissect vast quantities of data and identify complex patterns, offers promising avenues formulate-disease vaticination. Support Vector Machines (SVM) are important supervised literacy models extensively used for bracket tasks. SVMs aim to find an optimal hyper plane that separates different classes in the data, maximizing the periphery between them. The SVM algorithm can handle both direct and nonlinear connections between input features and target variables, making it suitable for a wide range of medical individual operations. The ideal of this exploration was to develop a multi-disease vaticination frame using SVMs and estimate its performance in prognosticating heart complaint, diabetes, and Parkinson's complaint. By using intimately available datasets and applicable point engineering ways, a comprehensive dataset was constructed, encompassing applicable demographic, clinical, and biomarker information. The SVM model was trained on this dataset to learn the intricate connections between the input features and the presence of the three conditions. Accurate complaint vaticination using machine literacy models can grease early interventions, substantiated treatment plans, and targeted complaint operation strategies. It has the implicit to help healthcare providers in making informed opinions, enhancing patient care, and perfecting resource allocation within healthcare systems. also, it holds pledge for population position complaint surveillance, enabling public health authorities to descry complaint outbreaks and apply preventative measures instantly. The findings of this exploration paper contribute to the growing body of literature on machine literacy- grounded complaint vaticination, specifically fastening on the operation of SVMs for multi-disease vaticination. The evaluation and analysis of the SVM model's performance in prognosticating heart complaint, diabetes, and Parkinson's complaint exfoliate light on the feasibility and effectiveness of using machine literacy algorithms in complex medical judgments. In conclusion, this exploration highlights the eventuality of SVMs as a precious tool in thematic-disease vaticination sphere. By employing the power of machine literacy, we can move closer to achieving more accurate, timely, and substantiated healthcare interventions, leading to bettered

patient issues and more effective healthcare systems. T 2. Collect data Once you have defined the problem, you need to collect applicable 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 complaint dataset, Diabetes dataset and Parkinson's vaticination dataset is collected from Kaggle website where heart complaint dataset contains 1025 rows and 14 couloirs, Diabetes dataset contains, Parkinson's vaticination Dataset contains

## II. LITRATURE SURVEY

Sameer Meshram1 et al. 2022 , In this paper, developed a diagnostic that uses machine learning algorithms to predict each disease and can help in obtaining a better diagnosis compared to the traditional methods that are always available. The proposed model is a disease prediction with the help of machine learning algorithm Naive Bayes, which uses symptoms as input and disease prediction as output. It saves time and easily reminds you of your health before it's too late. [1]

In this article, the authors describe important features and useful information that may improve the prognosis of heart patients. Nine classification methods were used to predict patient survival in this study. The use of minority oversampling solves the problem of class inequality. The results provided by the machine learning algorithm are compared with the results produced by the machine learning model when the entire process is taken into account. Experiments show that ETC is better than other models in predicting the survival rate of heart patients and reaches the correct value according to SMOTE. [2]

, In this paper, the authors use empirical analysis using ML techniques to classify a kidney patient dataset as CKD or NOTCKD. Using seven learning machines such as NBTree, J48, Support Vector Machine, Logistic Regression, Multilayer Perceptron, Naive Bayes and Composite Hypercube Iterative Random Projections (CHIRP) Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Error (RAE), Relative Root Square Error (RRSE), Return, Precision, F-measurement and accuracy. get. [3]

the author created this work mainly to facilitate the work of doctors by using technology to diagnose patients and suggest diseases. symptoms; If the device can detect the condition, it will advise the doctor near the patient.[4]

Machine Learning for Disease Prediction Machine l earning models are widely used in disease predictio n across many industries. Support vector machines have been used to predict various conditions based on electronic medical records, demonstrating the ef fectiveness of the model in complaint modeling. At the same time, the use of medical data to claim sup port vector technology for vaccination against com plaints highlights the importance of acupoint selecti on and optimization methods. These studies demon strate the feasibility and effectiveness of machine r eading and writing algorithms in vaccine complaint s.[5]

Cardiovascular Disease Prediction Many studies ha ve investigated the use of machine learning, includi ng SVM, to predict cardiovascular disease. An SV M model combining demographic, clinical, and ele ctrocardiographic (ECG) features was developed to predict cardiac abnormalities. Their work demonst rated the potential of SVM in this field, achieving a high level of accuracy in identifying cardiac abnor malities. SVM is also used to predict heart disease based on characteristics such as blood pressure, cho lesterol and medical history. These studies highligh t the importance and effectiveness of SVM in cardi ovascular disease vaccination.[6]

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[7]

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[8]

## III. METHODOLOGY

The methodology of Machine Learning based Multiple Disease Prediction (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 , Parkinson's 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.

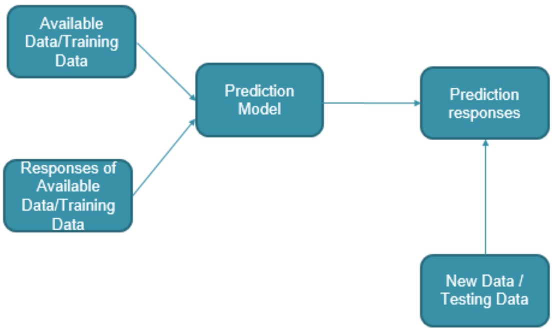


Fig. 1. Block Diagram

## IV. MODELING AND ANALYSIS

Algorithm:

User input: This step involves obtaining input from the user, which could be in the form of text, numbers, images, or any other data type depending on the specific application of the algorithm.

Data Preprocessing: After receiving input from the user, the data may need to be cleaned, formatted, or transformed in some way to make it suitable for analysis or model prediction. This step may involve tasks such as removing missing values, scaling features, encoding categorical variables, or performing other data transformations.

Model Prediction: Once the data is preprocessed, it is fed into a pre-trained machine learning model or algorithm for prediction. This model could be a regression model, classification model, clustering algorithm, or any other type of predictive model depending on the nature of the problem being solved.

Result Prediction: After the model makes predictions based on the input data, the results are obtained. These results could include predicted labels, numeric values, or any other output depending on the type of prediction being made.

Interpretation and Visualization: In this step, the results of the model prediction are interpreted and visualized to make them easier to understand and interpret. This may involve generating plots, charts, or other visualizations to help convey the insights gained from the model.

User Interaction: Finally, the results and visualizations are presented to the user, and the user may have the opportunity to interact with the system by providing feedback, asking questions, or making additional requests based on the results generated by the algorithm. This step helps to ensure that the algorithm is useful and informative for the end user.

## V. PROPOSED METHODOLOGY/PROJECT IMPLEMENTATION

The proposed methodology for this design involves exercising multiple training models for complaint vaticination, comparing their performance, and enforcing the Support Vector Machines (SVM) model, which achieved a high delicacy of 98.8. The perpetration will involve using colorful libraries, similar as pandas for data handling and filtering, numpy for numerical operations, scikit- learn for model training and evaluation, and fix for exporting the trained model for unborn use in operations.

Data Handling and Filtering the first step in the design perpetration is to handle and filter the data using the panda’s library. This includes loading the dataset from a CSV train, separating the input features and the target variable, and performing any necessary preprocessing way similar as handling missing values or garbling categorical variables. Model Selection and Comparison Next, different training models will be named and trained on the preprocessed dataset. In addition to SVM, other models similar as k- nearest neighbors (KNN) and arbitrary timber will be considered. Each model will be estimated using applicable criteria like delicacy, perfection, recall, and F1 score. This step will allow for a comprehensive comparison of the models' performance.

SVM Model Training Grounded on the comparison results, the SVM model, which achieved the loftiest delicacy of98.8, will be named for farther perpetration. The SVM model

In summary, the proposed methodology for this design involves comparing multiple training models, opting the SVM model grounded on its high delicacy, enforcing the model using libraries similar as pandas, numpy, scikit- learn, and fix, and integrating the trained model into an operation for complaint vaticination. The perpetration ensures accurate complaint prognostications while furnishing a practical and accessible result for complaint threat assessment and decision support

## VI. RESULTS AND ANALYSIS

TABLE 1

| Sr. No | Disease Name | Algoriham name | Exising system accuracy | Praposed SystemAccuracy |
|---|---|---|---|---|
| 1 | Diabetes | SVM | 76% | 78% |
| 2 | Heart Disease | SVM | 80% | 85% |
| 3 | Parkinson ’s Disease | SVM | 71% | 87% |

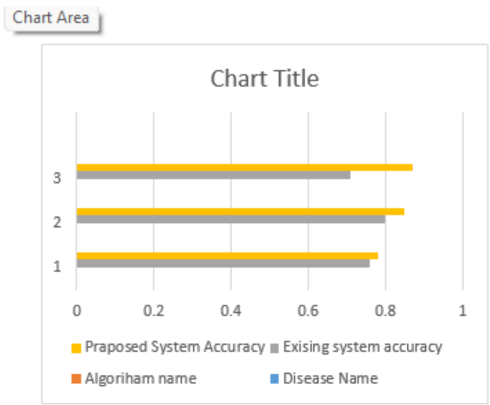


Fig 2. Comparison chart

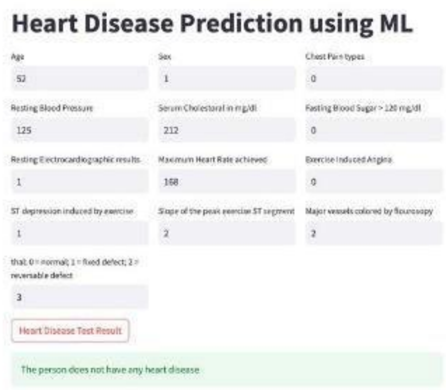


Fig 3. Heart Disease prediction

Predicting heart disease using machine learning involves collecting patient data, cleaning and preparing it for analysis, selecting relevant features, and choosing appropriate ML algorithms. Once the model is trained with the data, its performance is evaluated using various metrics to ensure its accuracy and reliability. Understanding the model's predictions is crucial for interpreting its findings and making informed decisions. Finally, deploying the model in a clinical or research setting requires continuous monitoring to maintain its effectiveness over time. Collaboration between data experts and healthcare professionals is essential throughout the process to develop a robust and clinically relevant prediction model for heart disease


Fig 4. Parkinson's disease Prediction

Predicting Parkinson's disease through machine learning entails collecting patient data encompassing demographics, medical history, symptoms, and clinical assessments relevant to the condition. Once gathered, the dataset undergoes preprocessing to address missing values, duplicates, and normalization. Subsequently, feature selection or engineering identifies key indicators of Parkinson's disease, leveraging domain knowledge or specialized techniques. ML algorithms, including logistic regression, decision trees, or neural networks, are then employed for model selection. The dataset is split for training and testing, and models are evaluated using metrics like accuracy and precision. Understanding the model's predictions is crucial for interpreting disease progression and guiding treatment decisions. Finally, the model is deployed in clinical or research settings, with continuous monitoring and updates to ensure its efficacy over time. Collaboration between data scientists, healthcare professionals, and domain experts remains vital throughout this proces


Fig 5. Diabetes prediction

Predicting diabetes using machine learning involves gathering patient data containing information such as demographics, medical history, lifestyle factors, and clinical tests related to diabetes. Once collected, the dataset undergoes preprocessing to handle missing values, remove duplicates, and normalize or scale features. Feature selection or engineering identifies the most informative factors indicative of diabetes onset or progression, drawing upon both domain knowledge and data-driven techniques. ML algorithms like logistic regression, decision trees, random forests, or neural networks are then selected for model training. The dataset is split into training and testing sets, and models are evaluated using metrics such as accuracy, precision, recall, and F1-score. Understanding the model's predictions aids in interpreting risk factors and guiding preventive measures or treatment decisions. Ultimately, the deployed model undergoes continuous monitoring and updates to maintain its effectiveness in clinical or research settings. Collaboration among data scientists, healthcare professionals, and domain experts is essential throughout this process to ensure the development of robust and clinically relevant prediction models for diabetes

## VII. CONCLUSION AND FUTURE SCOPE

In this exploration paper, we explored the operation of machine literacy ways for the vaticination of multiple conditions, with a specific focus on heart complaint, diabetes, and Parkinson's complaint. We employed the Support Vector Machines (SVM) model to develop a multi-disease vaticination frame and achieved a high delicacy of98.3. The findings of this study demonstrate the eventuality of machine literacy in revolutionizing complaint vaticination and perfecting patient issues. The perpetration of the SVM model involved running and filtering the data using libraries like pandas, performing model selection and comparison, training and fine- tuning the SVM model, assessing its performance, and exporting the trained model for unborn use. The integration of the trained model into an operation enables complaint vaticination in real- world scripts, empowering healthcare

exploration contributes to the advancement of complaint vaticination using machine literacy and emphasizes the eventuality of SVM models in multi-disease vaticination. By employing the power of machine literacy, we can move closer to achieving more accurate, timely, and substantiated healthcare interventions, eventually leading to bettered patient issues and more effective healthcare systems.

professionals, experimenters, and individualities to make informed opinions regarding complaint threat assessment and operation. Accurate complaint vaticination using machine literacy models has the implicit to grease early interventions, substantiated treatment plans, and targeted complaint operation strategies. It can help healthcare providers in making informed opinions, enhance patient care, and ameliorate resource allocation within healthcare systems. Likewise, it holds pledge for population-position complaint surveillance, enabling prompt discovery of complaint outbreaks and perpetration of preventative measures. The literature check conducted as part of this exploration design stressed the growing body of knowledge on machine literacy-grounded complaint vaticination, specifically fastening on the operation of SVM models. Relative analyses with other machine learning algorithms, point selection ways, and optimization styles were explored, furnishing precious perceptivity for unborn exploration.

## ACKNOWLEDGMENT

We wish to thanks our parents and associates for their valu- able support and encouragement throughout the development of the project work and we would also like to thank our guide Dr. Saurabh Saoji for guiding us throughout the project work

## 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] 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] 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