HEALTHCURE- MACHINE LEARNING APPLICATION FOR MULTIPLE DISEASE PREDICTION

Sanjana Sri V[1], Nithya R[2]
sanjanasri.it20@bitsathy.ac.in, nithyars@bitsathy.ac.in

[1] *Student, Information Technology, Bannari Amman Institute of Technology, Tamil Nadu, India*
[2] *Assistant Professor, Computer Science and Engineering, Bannari Amman Institute of Technology, Tamil Nadu, India*

ABSTRACT

In recent years, the science of machine learning has witnessed amazing advancements and applications in several industries, including healthcare. Because machine learning algorithms can forecast multiple diseases at once, they can potentially improve patient outcomes and alter medical diagnoses. Many machine learning techniques can perform predictive analytics on large amounts of data from many industries. Predictive analytics can be difficult to use in the healthcare industry, but over time, with extensive data analysis, it can assist doctors in making timely decisions regarding the health and treatment plans of their patients. A single ailment at a time is the subject of many machine learning algorithms currently used for health care analysis. For example, the first is used to analyze diabetes, the second to analyze heart disease, and so on. A person must visit several websites to anticipate more than one sickness. A single analysis cannot forecast more than one disease under any common framework. The accuracy of some of the models is worse than others, which can have a major impact on patient care. The test web application uses machine learning to anticipate different diseases, making the process smooth and accessible to the public. Our goal in this study is to create a web application that predicts diseases by utilizing machine learning to make predictions about various illnesses, including heart disease, diabetes, and lung cancer. It could help healthcare professionals make better judgments, improve patient care, and better allocate resources within healthcare systems. It also has the potential for population-level disease surveillance, which would help public health officials quickly identify illness outbreaks and put preventive measures in place.

Keyword: - *Multiple Diseases prediction, machine learning algorithms, common framework, predictive analysis.*

1. INTRODUCTION

In recent years, the science of machine learning has seen incredible developments and applications in several fields recently, including healthcare. Machine learning algorithms have the potential to transform medical diagnostics and enhance patient outcomes by being able to forecast several diseases at once. People today suffer from a wide range of ailments because of their lifestyle choices and the state of the environment. Therefore, early disease prediction becomes a crucial responsibility. However, the physician finds it too challenging to provide an accurate diagnosis based just on symptoms. The hardest thing to do is accurately anticipate an illness. Data mining is crucial in predicting the sickness to solve this issue. Data in medical science is growing at a rapid rate every year. Early patient care has benefited from accurate medical data analysis due to the expansion of data in the medical and healthcare fields. In this project, the Support Vector Machines (SVM) model and the Logistic Regression model are used to predict the existence of three common diseases: Parkinson's disease, diabetes, and heart disease. Diabetes, Parkinson's disease, and cardiovascular illnesses are serious public health issues that have a substantial impact on people's lives and healthcare systems around the globe. The timely identification and precise diagnosis of these illnesses are essential for enhancing patient outcomes, streamlining treatment regimens, and cutting down on

medical expenses. Machine learning presents intriguing paths for multi-disease prediction because of its capacity to evaluate enormous volumes of data and spot intricate patterns. SVMs seek to maximize the margin between distinct classes in the data by identifying the ideal hyperplane to divide them. The SVM technique is appropriate for a variety of medical diagnostic applications since it can handle both linear and nonlinear connections between input data and target variables. This study aimed to create a multi-disease prediction framework with support vector machines (SVMs) and assess its predictive power for heart disease, diabetes, and Parkinson's disease. A thorough dataset was created by utilizing publicly accessible datasets and suitable feature engineering techniques. This dataset included pertinent biomarker, clinical, and demographic data. Using this dataset, the SVM model was trained to discover the complex correlations between the existence of the three diseases and the input features. The results of this study add to the expanding corpus of research on machine learning-based illness prediction, with a particular emphasis on its application to multi-disease prediction. The effectiveness and viability of applying machine learning algorithms to difficult medical diagnoses were clarified by the evaluation and analysis of the model's performance in predicting heart disease, diabetes, and Parkinson's disease. To sum up, this experiment demonstrates the multi-disease prediction domain's potential. By utilizing machine learning, we can attain more precise, quick, and customized results—medical procedures, which result in better patient outcomes and more effective medical systems.

1.1 Problem Statement

Many machine learning models now in use for health care analysis focus on a single disease at a time. As an illustration, the first is for the analysis of diabetes, the second is for the analysis of heart disease, and so on. A person must visit several websites to anticipate more than one sickness. A single analysis cannot forecast more than one disease under any common framework. The accuracy of some of the models is worse than others, which can have a major impact on patient care.

Fig -1: Disease Prediction

1.2 Background of The Work

Healthcare systems around the world are faced with the task of accommodating a growing patient population with a variety of medical ailments. Not only do early diagnoses lead to better patient outcomes, but they also lessen the overall demand for healthcare resources. Treatment effectiveness is hampered by traditional diagnostic techniques, which frequently rely on patient-reported symptoms or identify diseases at an advanced stage. By using patterns and insights from large datasets, machine learning provides a promising way to predict diseases early on. A more dependable prediction model is made possible by the system's increased accuracy and robustness due to the integration of several algorithms. With the rapid advancements in technology and the widespread usage of mobile phones as user-friendly devices, an application that forecasts the prevalence of lifestyle diseases such as diabetes, heart disease, and other conditions is highly beneficial. With a few pieces of information, such as symptoms and diagnostic reports, a disease predictor can help you make critical predictions about a condition that is still present but is unknown. For future reference, it also helps to have a thorough understanding of the signs, causes, and other crucial elements.

2. RELATED WORK

To better understand the body of knowledge surrounding the use of machine learning techniques specifically, Support Vector Machines (SVM) and logistic regression for the prediction of various diseases, such as Parkinson's disease, diabetes, and cardiovascular disease, this research project conducted a literature review. The survey includes studies that have looked at comparable research goals, approaches, and results, offering insightful information and laying the groundwork for the current undertaking. When it comes to diagnosing common childhood disorders, this model performs comparably to skilled pediatricians, exhibiting good diagnostic accuracy across several organ systems. This work offers a proof of concept for putting an AI-based system into practice to help doctors handle massive volumes of data, enhance diagnostic assessments, and offer clinical decision assistance when there is uncertainty or complexity in the diagnosis.[1] The advantages of an AI system like this are probably going to be widespread, even though the impact would be more noticeable in places where there is a relative shortage of healthcare practitioners.[2] This work highlights the potential of SVM models in multi-disease prediction and advances the field of disease prediction using machine learning. We can get closer to providing more precise, prompt, and individualized healthcare interventions by utilizing machine learning, which will ultimately enhance patient outcomes and create more effective healthcare systems. [3] With the use of machine learning techniques and clinical approaches, it is possible to demonstrate how frameworks and systems have emerged to enable the fundamental and direct application of machine learning (ML) models to sophisticated information inquiry. This study presents a thorough near-investigation of three models that execute clinical records, with each model achieving an accuracy score of up to 98%.[4] Lastly, a disordered lattice and precision value investigation of the document is conducted. Comparative studies of a similar nature have been carried out in the context of diabetes and Parkinson's disease prediction, demonstrating the advantages and disadvantages of various models and their suitability for use in scenarios including several diseases. Optimization and Feature Selection Methods: The performance of illness prediction models has been enhanced by the widespread use of feature selection and optimization approaches. The review of the literature highlights the expanding corpus of work on machine learning-based illness prediction, with a particular emphasis on the use of SVM models for multi-disease prediction. It draws attention to how well SVM and logistic predict Parkinson's disease, diabetes, and heart disease and stresses the significance of feature selection, model optimization, and comparison analysis. The survey offers a thorough grasp of the body of literature, enabling the current research effort to have a strong foundation and pointing out possible directions for future study and advancements in the use of SVM models for multi-disease prediction.

2.1 Scope of Study

- Medical professionals and patients can all benefit from the disease prediction that explains the importance of early prevention.
- Technological breakthroughs, especially in the field of machine learning, are driving a revolutionary revolution in the healthcare industry.
- As the burden of various health conditions increases, there is a rising demand for comprehensive, early, and accurate disease prediction systems to meet the difficulties.
- The goal of the current project is to create a strong and useful application that can be easily integrated into healthcare processes by promoting interdisciplinary collaboration.

2.2 Results imply

- Early Detection of Diseases: Large datasets can be analyzed using machine learning models, Early detection greatly improves patient outcomes by enabling prompt intervention and therapy.
- Healthcare Prevention: Disease prediction models identify people at risk before symptoms appear, allowing healthcare to change from reactive to proactive.
- By enabling healthcare practitioners to carry out lifestyle changes and preventive measures, this proactive approach lessens the total strain on the healthcare system.
- Personalized Medicines: Individual patients can have disease prediction models customized to their specific genetic composition, way of life, and medical background.

3. PROPOSED METHODOLOGY

Healthcare has seen a paradigm shift in recent years because of the application of cutting-edge technology, most notably machine learning, to diagnosis and research. Disease prediction is one of the most innovative uses of machine learning in healthcare. The goal of this project is to create an accurate and early disease detection system by applying multiple machine-learning algorithms completely and effectively. Create a disease prediction system that can manage a variety of medical datasets. Apply and enhance a variety of machine learning techniques to the prediction of disease. Analyze the system's sensitivity, specificity, and accuracy in terms of performance main objectives are to improve various aspects of disease prediction, prevention, and management.

3.1 Main Objectives

1.Early Detection: Large datasets can be analyzed using machine learning models, which can then be used to find patterns and minute correlations that conventional diagnostic techniques would miss. Early detection greatly improves patient outcomes by enabling prompt intervention and therapy. Disease prediction models identify people at risk before symptoms appear, allowing healthcare to change from reactive to proactive...

2. Accurate Diagnosis: Various types of medical data, such as genetic information, electronic health records, lifestyle factors, and imaging data, can be analyzed using machine learning algorithms. Predictions are more accurate when a person's health situation is understood holistically, which is made possible by this all-encompassing method. Individual patients can have disease prediction models customized to their specific genetic composition, way of life, and medical background. Treatment regimens can be tailored with the use of personalized predictions, which maximizes therapeutic results.

3. Reduces Healthcare Costs: Predicting diseases early on will help allocate healthcare resources more effectively, which will reduce the burden on hospitals, clinics, and patients. Resource efficiency is essential, particularly considering global health concerns including pandemics and the rising incidence of chronic illnesses.

3.2 Methodology

Data Collection from Kaggle.com, a well-known website for dataset access, is the source of the data. Particularly, information on diabetes, heart disease, and Parkinson's disease is gathered. Data Preprocessing is to guarantee the quality and appropriateness of the acquired data for training the machine learning models, preprocessing is applied. This cover managing missing values, eliminating duplicates, and carrying out feature scaling or data normalization. For every illness prediction task, a different set of machine learning algorithms is selected. The algorithms for different diseases are chosen from Support Vector Machine (SVM), and Logistic Regression based on their effectiveness and fit for the prediction tasks. Two sets of pre-processed data are created: training and testing. The testing data is used to assess the models' performance after they have been trained using the training data. Each model's performance is evaluated using accuracy as the evaluation metric. The interactive application is made with Streamlit where the user needs to enter the necessary parameters for the prediction.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction
count	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000	768.000000
mean	3.845052	120.894531	69.105469	20.536458	79.799479	31.992578	0.471876
std	3.369578	31.972618	19.355807	15.952218	115.244002	7.884160	0.331329
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.078000
25%	1.000000	99.000000	62.000000	0.000000	0.000000	27.300000	0.243750
50%	3.000000	117.000000	72.000000	23.000000	30.500000	32.000000	0.372500
75%	6.000000	140.250000	80.000000	32.000000	127.250000	36.600000	0.626250
max	17.000000	199.000000	122.000000	99.000000	846.000000	67.100000	2.420000

Fig-2: Data collection

• Import the necessary libraries and the datasets.

• Split the data into features and target variables.

• Split the data into training and testing sets.

• Train the model on the training data.

• Save the model using the pickle module.

• Use Streamlit to create a web application that allows users to input their own data and make predictions using the saved model.

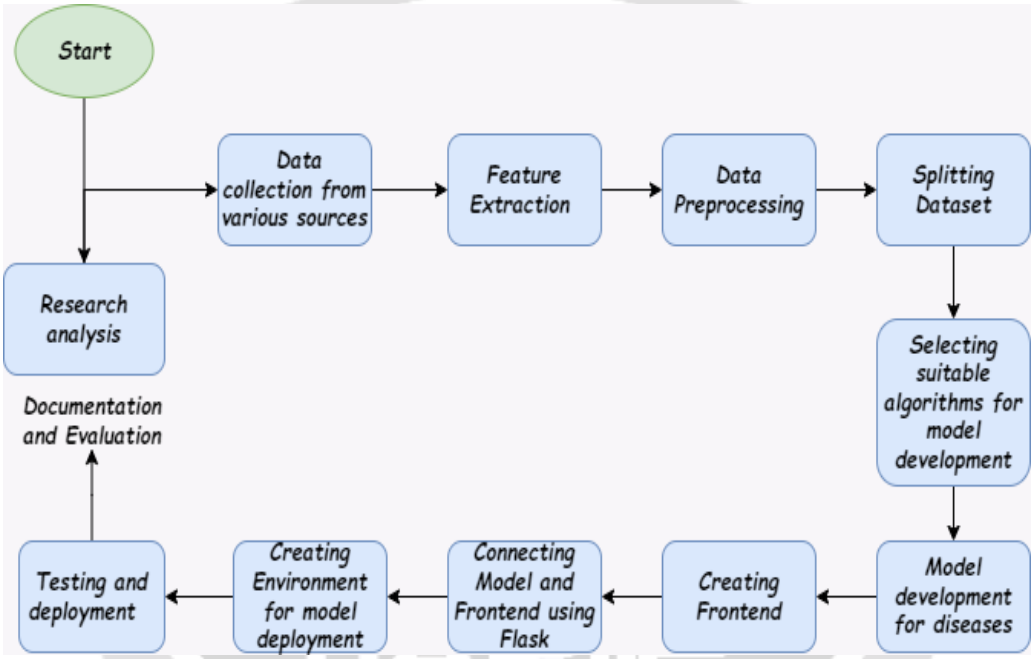

Fig -3: Workflow

4. PROPOSED MODULES

Data Collection- This process involves the collection of various and large data from the source. Based on the required attributes the data will be collected.
Data Preprocessing - Data Preprocessing is Processing the data which is already been gathered, This process involves processing data values.
Algorithm Selection- In this process, the algorithm selection will be performed based on the accuracy in prediction values such as SVM and Logistic Regression.
Model Training- This module involves training the selected object detection algorithm on the preprocessed data. The training process involves optimizing the algorithm's parameters to make it more accurate.

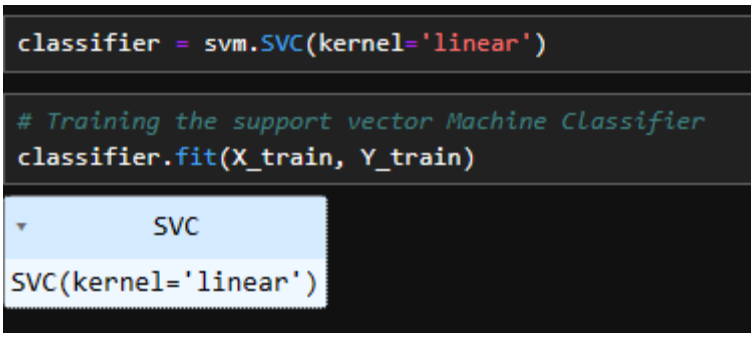

Fig –4: Model Training

Model Testing and Evaluation- This Module involves testing the trained object detection model on new set of data's or values to evaluate its accuracy and performance.

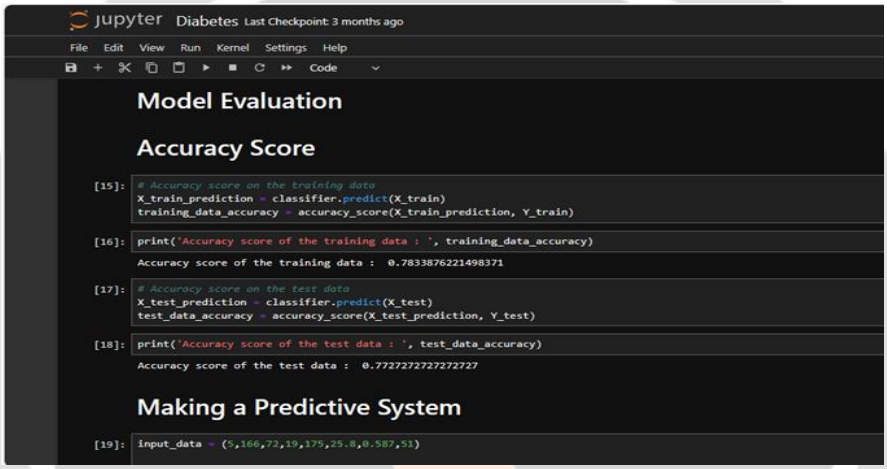

Fig -4: Model Evaluation

4.1 System Requirements

Hardware Requirements

- Processor: core i3/15
- RAM: 2-4GB

Software Requirements

- Platform: Windows 7/8/10
- Coding Language: Python

4.2 Technology and Tools

Technology:

- Machine learning algorithms
    1. SVM
    2. Logistic Regression
- Streamlit library
- Python
- Flask frameworks

Tools:

- Python 3.12
- Anaconda Navigator
- Spyder
- Jupyter notebook

4.3 Software Architecture

- Data Collection and Feature Processing:

  Gather real-time or historical data from various sources, such as medical sensors, electronic health records (EHRs), wearable devices, and other healthcare systems. Clean and preprocess the raw data to handle missing values, outliers, and noise. Normalize or standardize the data to ensure consistent scales and improve model performance. Feature engineering: Extract relevant features from the raw data or generate new features to improve prediction accuracy.

- Prediction Model and Model Training:

  Implement a machine learning model that predicts the likelihood of diseases based on the preprocessed data. The model can be based on traditional machine learning algorithms or more advanced deep learning models, depending on the complexity of the data and prediction requirements. The model has been trained and tested by using different machine learning algorithms to predict the most accurate prediction level for the dataset.

- Deployment of the model:

  Save the trained model using the pickle module. Use Streamlit to create a web application that allows users to input their data and make predictions using the saved model. Deployment and evaluation of the results.

5. CONCLUSION

This prediction system reduces the gap between doctors and patients where both can easily access and get results in clear and understandable form. The application provides a user-friendly interface that contains a menu with three diseases. This prediction system requires the input parameters specific to each disease. Users can enter the input parameters of disease to be predicted and they will get the results whether the person gets affected or not. The prediction system provides the predicted result for the person by entering the relevant and necessary features. The effectiveness of the model and algorithm has been verified through simulation, with prediction accuracy. This system provides support for multiple disease prediction using different Machine Learning algorithms. Multiple disease prediction is a powerful approach that holds the potential to revolutionize the way we detect, understand, and respond to events. Our Proposed System aims at bridging the gap between Doctors and Patients which will help both classes of users in achieving their goals. The present approach of many systems focuses only on automating this process for a single disease which lacks in building for multiple disease system.

6. REFERENCES

[1]. Liang H, Tsui BY, Ni H, et al. (2019) “Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence.” Nat Med. 2019;25(3):433- 438.
[2]. Parshant, Dr. Anu Rathee (2023) “Multiple Disease Prediction Using Machine Learning” ISSN: 2456-8880.
[3]. Talasila Bhanuteja, Kilaru Venkata Narendra Kumar, et al.,(2021) “Symptoms Based Multiple Disease Prediction Model using Machine Learning Approach” ISSN: 2278-3075
[4]. Paniagua JA, Molina-Antonio JD, Lopez Martinez F, et al. (2019) “Heart disease prediction using random forests.” J Med Syst. 2019;43(10):329
[5]. Aishwarya Mujumdara, Dr. Vaidehi V (2019)” Diabetes Prediction using Machine Learning Algorithms”
[6]. A. K. Shrivas and S. Kumar Sahu,(2018) "Classification of Chronic Kidney Disease using Feature Selection Techniques", IJCSE, vol. 6, no. 5, pp. 649- 653,