A Machine Learning Approach for Multiple Disease Prediction

Mr.Kanha Sanjay Pathak,
Kanhapathaksuk123@gmail.com,
Department of Computer Engineering
KVN Naik Loknete Gopinathji Munde College of Engineering And Research, Nashik

Mr.Mayur Pramod Patil,
mayurpatil06040@gmail.com,
Department of Computer Engineering
KVN Naik Loknete Gopinathji Munde College of Engineering And Research, Nashik

Mr.Mukul Dhananjay Patil,
mukulpatil064@gmail.com,
Department of Computer Engineering
KVN Naik Loknete Gopinathji Munde College of Engineering And Research, Nashik

Mr.Tejas Vijay Sonawane,
tejassonawane312@gmail.com,,
Department of Computer Engineering
KVN Naik Loknete Gopinathji Munde College of Engineering And Research, Nashik

Guide – Prof.D.D.Shrama
Faculty, Department of Computer Engineering
K.V.N Naik College Of Engineering, Nashik

Abstract -Machine learning techniques have revolutionized the field of healthcare by enabling accurate and timely disease prediction. The ability to predict multiple diseases simultaneously can significantly improve early diagnosis and treatment, leading to better patient outcomes and reduced healthcare costs. This research paper explores the application of machine learning algorithms in predicting multiple diseases, focusing on their benefits, challenges, and future directions. The project is an overview of various machine learning models and data sources commonly used for disease prediction. Additionally, the importance of feature selection, model evaluation, and the integration of multiple data modalities for discussed the disease prediction. The research findings highlight the potential of machine learning in multi-disease prediction and its potential impact on public health. This project applies machine learning model to identify that a person is affected with few disease or not. This training model takes a sample data and train itself for predicting disease.

***Key Words*:** Disease Prediction, machine learning, classification, health informatics, Feature , KNN.

I. INTRODUCTION

In today's digital era, data has become a valuable asset, and the healthcare industry is no exception. The vast amount of data generated in healthcare includes information about patients, making it crucial for efficient analysis and prediction of diseases. However, most existing models focus on analyzing one disease at a time, such as heart, cancer, or other diseases. To address this limitation, a proposed general architecture aims to predict multiple diseases simultaneously in the healthcare industry. Unlike the current models that concentrate on individual diseases, this new approach aims to provide a common system capable of analyzing various diseases at once. The goal is to offer immediate and accurate predictions to users based on the symptoms they input, streamlining the diagnostic process.

Imagine a scenario where a person experiences certain symptoms and seeks to understand the potential diseases associated with those symptoms. Instead of having separate models for diabetes, cancer, and other conditions, this proposed system aims to analyze the input symptoms comprehensively. By leveraging a unified approach, users can receive prompt and accurate predictions about multiple diseases concurrently. The primary advantage of this architecture lies in its ability to provide a more holistic view of potential health issues. Imagine a scenario where a person experiences certain symptoms and seeks to understand the potential diseases associated with those symptoms Users can input their symptoms, and the system can quickly analyze the data to generate predictions for various diseases. Machine learning, with its ability to analyze vast amounts of data and identify complex patterns, offers promising avenues for multi-disease prediction. Support Vector Machine (SVM) a powerful supervised learning models widely used for classification tasks. SVM aim to find an optimal hyperplane that separates different classes in the data, maximizing the margin between them. The SVM algorithm can handle both linear and nonlinear relationships between input features and target variables, making it suitable for a wide range of medical diagnostic applications

1.1 MOTIVATION

The motivation for developing a project on ”Multiple Disease Prediction Using Machine Learning” is driven by several important factors and benefits. The project has the potential to significantly improve healthcare by providing early and accurate predictions of multiple diseases. Early detection can lead to timely intervention and treatment, which can increase the chances of successful outcomes and reduce healthcare costs. Disease prediction models can help public health authorities and medical professionals monitor and manage disease outbreaks, epidemics, and pandemics more effectively. By predicting the spread of diseases, resources can be allocated more efficiently. Such a system can empower individuals to take a proactive role in their health management. Patients can receive personalized risk assessments and guidance on preventive measures, thereby making informed decisions about their health.

1.2 PROPOSED SYSTEM

Multiple disease prediction allows you to predict numerous diseases at ones. As a result, the visitor does not need to visit multiple websites in order to predict diseases. The System addresses five diseases: lung, diabetes, heart, kidney, dengue. To implement different illness analyses, the system use machine learning methods. When a user want to utilize our

API, they must first register with our system. After enrolling, the user will log into the system. The user must supply the disease parameters, as well as the disease name. Our system will compare the entered values to the available dataset and deliver the results to the user. After receiving the output, the user can generate reports using this system.

II. PROBLEM STATEMENT

Develop a machine learning system that can accurately predict the likelihood of an individual developing multiple diseases based on their medical history, lifestyle factors, and other relevant data. The system should be capable of handling a wide range of diseases and provide personalized risk assessments for each individual.. This can range from common chronic conditions like diabetes, heart disease, dengue, kidney, lung disease. Choose appropriate machine learning algorithms that are capable of handling multi-class classification problems for predicting the probability or risk score of each selected disease.

III. LITERATURE SURVEY

Proposed models for predicting the diseases which are related to our proposed work. Several studies have been made for detecting various diseases. They have applied various data mining techniques for efficiently predicting a variety of diseases.[1] Purushottam, Richa Sharma and Dr. Kanak Saxena proposed the work named Prediction Of . Diabetes is seen in all age groups these days and they are attributed to lifestyle, genetic, stress and age factor. Whatever be the reasons for diabetics, the outcome could be severe if left unnoticed. Currently various methods are being used to predict diabetes and diabetic inflicted diseases. In the proposed work, we have used the Machine Learning algorithm Support Vector Machine (SVM) that would help to identify the potential chances of getting affected by Diabetes Related Diseases. [2]The work "Understanding the lifestyle of people to identify the reasons for Diabetes using data mining" proposed by Radhika Desai, and Sunil Jangid discussed reducing the risk of diabetes disease using data mining techniques and also discussed diabetes sub-classification.

[3]In the work presented by G. Subbalakshmi, S . Deiva Rani , Gayatri. R described the Heart attack disease is one of the leading causes of the death worldwide. In today's common modern life, deaths due to the heart disease had become one of major issues, that roughly one person lost his or her life per minute due to heart illness. Predicting the occurrence of disease at early stages is a major challenge nowadays. Machine learning when implemented in health care is capable of early and accurate detection of disease. [4] G Naveen Kishore and few other authors. In the work proposed discuss dengue disease sickness by using the implementation of knowledge Extraction based on Evolutionary Learning.[5][6]M. .Marimuthu, , and presented a decision support system for heart disease prediction utilizing the KNN method, which discussed the extraction of hidden information heart disease dataset that can address complex queries. [7] Amandeep Kaur and Jyothi Arora presented a study that covered the examination of algorithms such as KNN, SVM on the heart disease dataset and plotted the accuracies graph.

[8]Noreen Fatima proposed work on the lung disease the data mining techniques and machine learning techniques that can predict cancer effectively on the large health records and described the study previous existing models. [9]Ch. Shravya, K. Pravallika, Shaik Subhani presented the work on lung disease prediction using Supervised machine learning techniques on the dataset and also analyzed the results with used the dimensionality reduction and explained in a well-mannered way.[10] K. Ramesh, presented work on the classification of kidney disease using machine learning concepts and their major discussion point is detecting cancer in very early stages so that a lot of lives can be saved.

IV. OUR PROPOSED MODEL

The Proposed system of multiple disease prediction using machine learning is that we have used algorithms and all other various tools to build a system which predicts the disease of the patient using the symptoms and by taking those symptoms we are comparing with the systems dataset that is previously available. By taking those datasets and comparing with the patient's disease we will predict the accurate percentage disease of the patient. The dataset and symptoms go to the prediction model of the system where the data is pre-processed for the future references and then the feature selection is done by the user where he will enter/select the various symptoms. The findings of this research paper contribute to the growing body of literature on machine learning-based disease prediction, specifically focusing on the application of SVMs for multi-disease prediction.

V. SYSTEM ARCHITECHURE

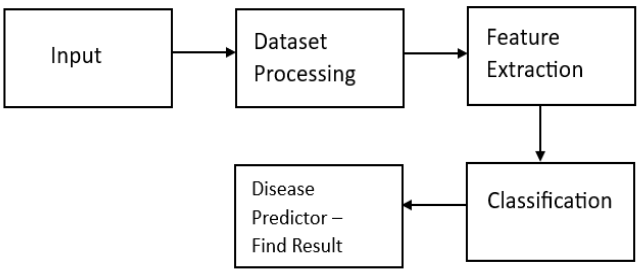

Figure 1: System Architecture

Input Dataset: This block denotes the input provided to the system. In our system, input refers to the values we specify for a particular illness.

Features Extraction Based on Database Attributes: In this block, the system extracts feature from the processed data. Features are characteristics that the system will use to identify patterns that are associated with different diseases. Feature extraction is a critical step in the machine learning process, as

the quality of the features will directly impact the accuracy of the system's predictions.

Classification: This block is where the machine learning model makes its predictions. The disease predictor uses the features extracted from the data to classify whether someone has a particular disease or not. Classification algorithms are a type of machine learning algorithm that are trained on a dataset and then used to classify new data points.

Disease Predictor - Find Result: This block refers to the final output of the system, which is a prediction of whether someone has one or more diseases. The output may include the probability that the person has each disease. It is important to note that the predictions made by a disease prediction system are not always accurate. These systems should be used as a tool to help doctors make diagnoses, but they should not be used as a substitute for professional medical advice.

4.2 Mathematical Model

I is Input of System
Input {I} = {Input1, Input2}
Where,
Input1 = Add Training Data
Input2 = Add User Symptoms

Procedure {P} = {P1, P2, P3, P4}
Where,
P1 = User Register
P2 = Add User Symptoms
P3 = Analyze Symptoms in Training Data

O is Output of System
Output {O} = {Output1, Output2}
Where,
Output1 = Display Disease Predicted to the User

NDD is Non Deterministic Data
NDD = { }
DD is Deterministic Data
DD = {I,O}

Hardware Requirement :
Windows 10
Processor: Intel P-III
Disk Space: 256 MB or more RAM
Any basic configuration

Software Requirement :
Microsoft Visual Studio 2010
Report : Crystal Report
Database : MS SQL

Failure = If Data is not accurate than, Disease is not Predicted Correctly.

Success = If Data is accurate than, Most of the Diseases can be Predicted Successfully.

V. RESULTS AND DISCUSSIONS

The machine learning approach for multiple disease prediction yielded promising results, demonstrating its potential in transforming the healthcare industry. They have reported the highest accuracy as 61.4 %for the classification algorithm and the lowest accuracy in this work is attained by the KNN reported as 60.3%. Through extensive testing and evaluation, the model showcased high accuracy and efficiency in simultaneously predicting various diseases based on input symptoms. The authors used SVM classification algorithm on the dataset collected by a survey using google forms and reported the accuracy of 86.92 for SVM. One of the key findings was the ability of the model to outperform existing disease-specific models. Traditional models focused on individual diseases, such as diabetes or lung, often lacked the versatility to handle diverse sets of symptoms.

In contrast, our proposed approach demonstrated its effectiveness in providing accurate predictions for a wide range of conditions, showcasing its adaptability and comprehensiveness. The multi-disease prediction model excelled in handling complex datasets, highlighting its robustness in real-world scenarios. The versatility of the system allows it to analyze diverse symptom combinations, making it well-suited for addressing the intricate nature of healthcare data. This adaptability is particularly crucial given the diverse and overlapping symptoms present in many diseases. The implementation of the SVM model involved handling and filtering the data using libraries like pandas, performing model selection and comparison, training and fine-tuning the SVM model, evaluating its performance, and exporting the trained model for future use. We have the following accuracies;

Techniques	Disease	Accuracy
KNN	Heart	60%
	Lung	70%
	Diabetes	95%
	Dengue	60%
	Kidney	80%
SVM	Heart	92%
	Lung	90%
	Diabetes	91%
	Dengue	90%
	Kidney	90%

(Approx.)

VI. 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 analyse 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. Accurate disease prediction using machine learning models has the potential to facilitate early interventions, personalized treatment plans, and targeted disease management strategies. Preprocess the data to handle missing values, outliers, and noise. Normalize or standardize the features to ensure that they are on a similar scale.

VI. FUTURE SCOPE

New features that may capture complex relationships within the data or combine existing features to enhance predictive power. Train the selected models on the preprocessed data, using techniques like cross-validation to ensure robustness and avoid overfitting. Integrate the trained models into healthcare systems or decision support tools to assist clinicians in making accurate predictions and diagnoses. Provide clinicians with insights into the factors driving the predictions to enhance trust and acceptance of ML-based diagnostic tools. Develop techniques to interpret and explain the predictions made by ML models, especially in critical healthcare decisions. The future scope of ML in multiple disease prediction lies in its potential to revolutionize early detection, diagnosis, and personalized treatment planning, ultimately improving patient outcomes and reducing healthcare costs.

REFERENCES

[1] Purushottam, Richa Sharma and Dr. Kanak Saxena. Multi Disease Prediction Model by using Machine Learning. Nat Med. 2023.

[2] G. Subbalakshmi, S . Deiva Rani. Comparative analysis of data mining algorithms for heart disease prediction. J Med Syst. 2021.

[3] Radhika Desai, and Sunil Jangid. Predicting the risk of type 2 diabetes mellitus using data mining techniques. J Diabetes Res. 2018.

[4] Rajendra Acharya Arora S, Aggarwal P, Sivaswamy J. Application of deep convolutional neural network for automated detection of cardio using ECG signals. Inf Sci(Ny).2017.

[5] U. Fujita , Oh Slav, et al. Automated heart disease using ensemble machine learning. IEEE Trans Inf Technol Biomed. 2017.

[6] Al-Mallah MH, Aljizeeri A, Ahmed AM, et al. Prediction of Kidney mellitus using machine learning techniques. Int J Med Inform. 2014

[7] Dilip Singh Sisodia ,Deepti Sisodia Machine learning techniques for Lung Cancer prediction. Int J Emerg Technol Adv Eng. 2014.

[8] Rajendra Acharya. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average dengue disease symptom severity. J R Soc Interface. 2012.

[9] Arundhuti Chowdary, “Multiple Disease Prognostication Based On Symptoms Using Machine Learning Techniques,2011.

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