Survey Paper on 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

---

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. We present an overview of various machine learning models and data sources commonly used for disease prediction. Additionally, we discuss the importance of feature selection, model evaluation, and the integration of multiple data modalities for enhanced disease prediction. The research findings highlight the potential of machine learning in multi-disease prediction and its potential impact on public health. Once more, I am applying 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, Na¨ive Bayesian Networks, Random Forest, Decision Tree, Feature , Selection KNN.

1.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 Machines (SVM) are powerful supervised learning models widely used for classification tasks. SVMs 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 once. As a result, the visitor does not need to visit multiple websites in order to predict diseases. We are addressing three diseases: lung, diabetes, and heart. To implement different illness analyses, we will use machine learning methods. When a user wants 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 our 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. Gather relevant data from various sources, including electronic health records (EHRs), medical imaging, patient surveys, and lifestyle data. This data should cover a comprehensive set of variables such as demographics, medical history, family history, lifestyle choices (e.g., diet, exercise), and more. Clean, normalize, and preprocess the data to handle missing values, outliers, and inconsistencies. This step may also involve feature engineering to extract meaningful features from the raw data. Determine the set of diseases to be predicted. This can range from common chronic conditions like diabetes, heart disease, and hypertension to infectious diseases, cancer, and rare genetic disorders. 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. RELATED WORK

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.[7]G Naveen Kishore and few other authors proposed the work named Prediction Of Diabetes Using Machine Learning Classification Techniques proposed. In this work, various classification algorithms like SVM, Logistic Regression, Decision Tree, KNN, Random Forest are utilized on the 769 instances of the Pima dataset which contain features like Pregnancies, Blood pressure, body mass index, etc. They have reported the highest accuracy as 74.4 %for the classification algorithm Random Forest and the lowest accuracy in this work is attained by the KNN reported as 71.3%. [8]The work "Understanding the lifestyle of people to identify the reasons for Diabetes using data mining" proposed by Gavin Pinto, Radhika Desai, and Sunil Jangid discussed reducing the risk of diabetes disease using data mining techniques and also discussed diabetes sub-classification. The authors used Naïve Bayes and SVM classification algorithms on the dataset collected by a survey using google forms and reported the accuracy of 64.92 for SVM and 60.44 for Naïve Bayes. [9]In the work presented by M .Marimuthu , S . Deiva Rani , Gayatri. R described the cardio diseases in a detailed manner and also applied the classification algorithms like SVM, Decision Tree, Naïve Bayes, K-Nearest Neighbors on the Framingham dataset from Kaggle. The authors compared various machine learning algorithms for the forecast of the risk of heart disease. The highest reported accuracy in this work is 83.60% for the KNN classification algorithm. [10]In the work proposed by Purushottam, Richa Sharma and Dr. Kanak Saxena discuss cardiovascular sickness by using the implementation of Knowledge Extraction based on Evolutionary Learning (java programming technique for making the development model for data mining issues). The highest reported accuracy in this work is 86.7%. [11][12]M. Chinna Rao, K. Ramesh, and G. Subbalakshmi presented a decision support system for heart disease prediction utilizing the Nave Bayes classification method, which discussed the extraction of hidden information heart disease dataset that can address complex queries.[13]Amandeep Kaur and Jyothi Arora presented a study that covered the examination of algorithms such as KNN, SVM, ANN, and Decision Tree on the heart disease dataset and plotted the accuracies graph.[14]Noreen Fatima proposed work on the Cancer forecast 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. [15]Ch. Shravya, K. Pravallika, Shaik Subhani presented the work on Breast cancer prediction using Supervised machine learning techniques on the dataset and also analyzed the results with (PCA)principal component analysis and also used the dimensionality reduction and explained in a well-mannered way.[16]Nikitha Rane, Jean Sunny presented work on the classification of Cancer 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.[17]Dilip Singh Sisodia ,Deepti Sisodia predicted diabetes using classification techniques and reported an accuracy of around 76% on the Pima dataset.

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.

4.1 System Architecture

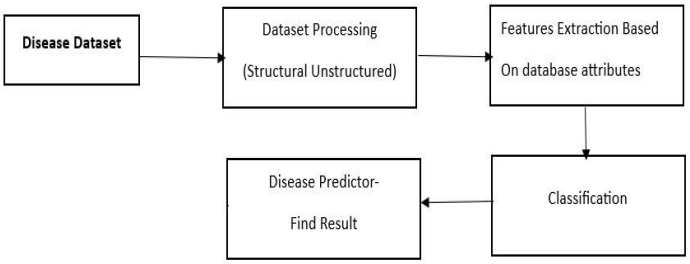

Figure 1: System Architecture

The process begins with a "Disease Dataset," which is raw data related to various diseases. This data undergoes "Dataset Processing," where it is organized and prepared for analysis. The processing can handle both structured and unstructured data . The next step is "Features Extraction Based On Database

Attributes." Here, specific characteristics or attributes from the dataset are identified and extracted for further analysis. These features are crucial as they provide insights and patterns that are instrumental in making predictions . After feature extraction, the data is used in two parallel processes: "Disease Predictor - Find Result" and "Classification." The Disease Predictor uses the extracted features to predict potential diseases based on patterns recognized in the dataset. It determines correlations and associations with various diseases, leading to a result that indicates potential health risks or conditions.

On the other hand, classification involves categorizing the extracted features into different classes or categories. This could mean sorting data based on severity, type of disease, or any other relevant classification metric . In summary, this flowchart represents a comprehensive approach to disease prediction using datasets. Each stage plays a crucial role in ensuring that the data is not only well-processed and analyzed but also effectively utilized to yield accurate predictions that can be instrumental in healthcare planning and management. This process is a testament to how data-driven approaches are revolutionizing healthcare.

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. Through extensive testing and evaluation, the model showcased high accuracy and efficiency in simultaneously predicting various diseases based on input symptoms. 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 cancer, 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.

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. It can assist. Healthcare providers in making informed decisions, enhance patient care, and improve resource allocation within healthcare systems. Furthermore, it holds promise for population-level disease surveillance, enabling prompt detection of disease outbreaks and implementation of preventive measures.

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] 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.

[3] 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.

[4] Arundhuti Chowdary, "Revolution in authentication process by using biometrics," International Conference on Recent Trends in Information Systems, pp. 36-41, 2011.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

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

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