IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, November 2023

Health-Guard: Disease Prediction and Recommendation System

Kunal Suresh Pattanshetti, Rohan Khare and Prof. Suvarna Karankal
Department of Artificial Intelligence & Data Science
ISBM College of Engineering, Nande, Pune, India
kunalp0201@gmail.comandrohankumarkhare@gmail.com
rohankumarkhare@gmail.com and karankal.suvarna@gmail.com

Abstract: *One sector that has seen a significant shift as a result of machine learning is the medical field. Machine learning has brought a drastic change in the healthcare sector by predicting the diseases very accurately, fast and within short time. So, if we go see the traditional way of diagnosing a disease is through a specialist doctor where they go through certain procedure like studying symptoms and asking for tests from the patients, however this method is less accurate and takes a lot of time. However, on basis of study, different machine learning algorithms like SVM, KNN, Logistics regression, Decision tree have proved highly fit for prediction of diseases related to Kidney, Heart, Lungs, Parkinsons, etc. with high accuracy. Therefore, the machine learning based software would prove beneficial to the doctor's, medical staff and even the normal people for fast diagnose of the disease which would be more accurate and within a short period of time. So, we have proposed a ML based disease prediction software which will consist of two parts one for the normal people's use and other for the medical faculties use, these parts will be consisting of different diseases which will be predicted on the basis of different medical test results in case of medical faculties part and based on symptoms in case of normal people's part.*

Keywords: Machine Learning, Disease Prediction, SVM, Logistic Regression, KNN

I. INTRODUCTION

Disease diagnosis is a complex process that requires extensive medical knowledge and careful analysis of patient symptoms, medical history, and test results. Doctors must be able to differentiate between different diseases that have overlapping symptoms and identify the most likely diagnosis for a patient. However, the number of known diseases keeps increasing, making accurate diagnosis challenging even for experienced physicians. Recent advances in machine learning and artificial intelligence have opened new possibilities for developing automated and intelligent systems for medical diagnosis and disease prediction. Machine learning methods have the potential to uncover hidden patterns in patient data that can assist doctors in making more accurate diagnoses and treatment decisions.

In this project, we aim to develop a machine learning-based system for predicting multiple diseases based on input patient symptoms and health parameters. The system utilizes supervised learning algorithms including support vector machines (SVM), logistic regression, k-nearest neighbours (KNN), and Naive Bayes to predict disease risk based on patient data, where we are going to use a dataset of electronic health records of thousands of patients to train and evaluate the models. It will provide doctors with probable disease diagnoses for a patient along with confidence scores based on the predictive models. The system will also provide recommended actions and next steps tailored to the predicted diseases such as additional tests, specialist referral, or medication. The ability to predict probable diseases and give tailored recommendations will assist doctors in quickly narrowing down on the most likely diagnosis. This can potentially improve diagnosis accuracy, reduce delays and errors, and improve treatment outcomes for patients. The system can also help patients in rural and remote areas get reliable disease screening without access to skilled doctors. Overall, the multiple disease prediction and recommendation system will demonstrate the usefulness of ML in healthcare and medical decision support.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, November 2023

II. LITERATURE SURVEY

Using machine learning algorithms, a number of studies have been conducted to predict diseases based on symptoms and different medical tests displayed by a person.

The paper [1] "Disease Symptoms Prediction Application" present a disease prediction system using machine learning algorithms to predict diseases based on patient symptoms. Using two disease datasets - a diabetes dataset and a cancer dataset, they evaluate multiple algorithms including GLM Binomial, Random Forest, Naive Bayes, and Principal Component Analysis. For the diabetes dataset, Random Forest achieves is used and for the cancer dataset, they have applied the PCA algorithm for feature reduction.

The paper [2] "Prediction of Disease Using Machine Learning" present a disease prediction system using machine learning algorithms including KNN, Naive Bayes, Logistic Regression and Decision Tree. The system takes symptoms from the user as input and outputs the predicted disease. Using a structured disease dataset containing patient symptoms, the algorithms are trained. Naive Bayes is used for predicting the disease, KNN for classification, Logistic Regression for feature extraction, and Decision Tree for dividing the dataset.

The paper [3] "Assistance Tool for Prediction and Monitoring Various Diseases based on Machine Learning" The system allows users to input symptoms and parameters related to their health condition. For heart disease prediction, parameters like blood pressure, cholesterol, etc. are taken and Random Forest algorithm is applied because of its high accuracy as compared to Naive Bayes and KNN.

The paper [4] "Multiple Disease Prediction System" the system allows prediction of diabetes, heart disease, and liver disease based on user input parameters. For diabetes prediction, XGBoost, for heart disease KNN, and for liver disease Random Forest was used. After data preprocessing, the models are trained on disease datasets and optimized for accuracy. The trained models are integrated into a Django framework that takes user inputs through a graphical interface and displays disease predictions. The system architecture enables convenient prediction of multiple diseases through a single interface instead of separate systems.

The paper [5] "Multiple Disease Prediction Using Machine Learning" present a machine learning based system for predicting multiple diseases including heart disease, diabetes, and Parkinson's disease. They utilize the Support Vector Machine (SVM) algorithm compared to other models like decision trees and neural networks due to its higher accuracy.

The paper [6] "A Machine Learning Model for Early Prediction of Multiple Diseases to Cure Lives" presents a machine learning model for early prediction of multiple diseases. The main objective of the study is to predict diseases based on symptoms provided by patients using various machine learning algorithms such as Decision Tree, Random Forest, Naïve Bayes, and KNN.

III. PROPOSED MODEL

The proposed model will utilize supervised learning algorithms including Logistic Regression, K-Nearest Neighbours (KNN), and Support Vector Machines (SVM) to predict disease risk based on patient symptoms and health parameters.

Logistic Regression will predict probability of a binary outcome using a logistic function and will be simple and fast to train. K-Nearest Neighbours will classify data points based on closest training examples in feature space and will be effective for multi-class prediction. Support Vector Machines will work by finding an optimal hyperplane that separates classes of data and will be effective for high-dimensional data.

The model will be trained on a large dataset of electronic health records containing patient symptoms, history, diagnoses, and test results. For a new patient, the model will compute probability scores for different diseases using the trained algorithms. It will provide the top predicted diseases with confidence scores to assist diagnosis by the doctor.

The model will also give tailored recommendations for each predicted disease such as suggested medical tests, referral to a specialist, applicable treatments, and recommended lifestyle changes. This will assist doctors in determining targeted next steps for the patient.

Model performance will be evaluated using metrics like accuracy, precision, recall, F1-score on a held-out test set, along with doctor reviews of predictions. In conclusion, this machine learning model utilizing Logistic Regression, KNN, and SVM can potentially improve diagnosis and treatment by predicting likely diseases and giving tailored recommendations. More patient data can further enhance model performance in future work.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, November 2023

The first step will be collecting raw medical data including symptoms, diagnoses, test results, etc. from sources like electronic records. Relevant features will be extracted to create a dataset. This data will undergo pre-processing to handle inconsistencies and prepare it for the models.

Supervised learning models including Logistic Regression, K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) will be trained on the dataset to predict diseases from patient information. The models will be evaluated on test data and the most accurate will be selected.

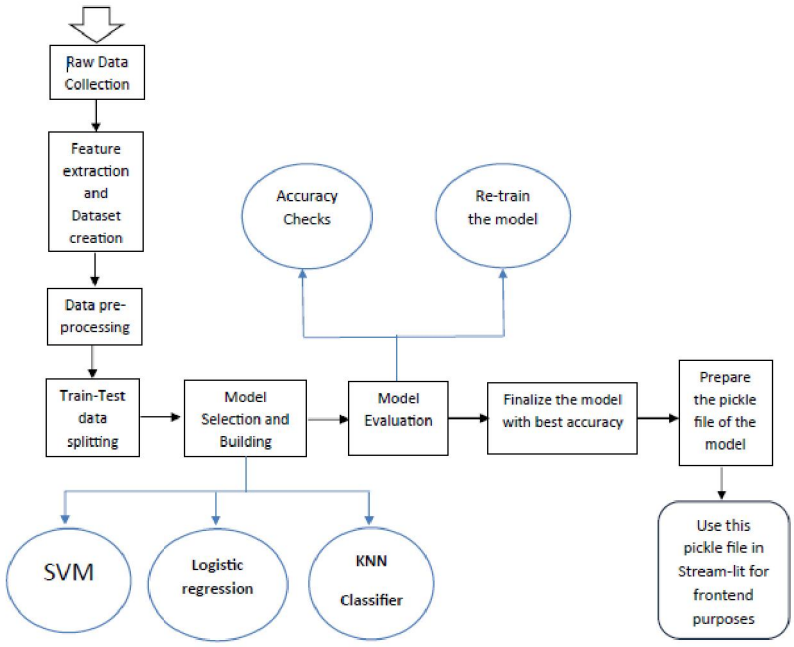

Figure 1: System Architecture for The Proposed Model

The optimized model will be finalized, retrained on all data, and tested further to ensure robust predictions. The final model parameters will be saved to a pickle file for integrating into the frontend app.

In summary, the system will collect and prepare medical data, develop disease prediction models using Logistic Regression, KNN and SVM, evaluate and select the best model, and save it to a file for use in real-time prediction based on user inputs. This will assist doctors by providing reliable and accurate disease risk assessments.

IV. CONCLUSION

In conclusion, this review paper provided a comprehensive overview of the potential for machine learning models to transform medical diagnosis and disease prediction and how all of them can be used together in order to make a platform feasible for all the sufferers to diagnose a disease and find appropriate solution for the same. The detailed analysis of various supervised learning algorithms such as SVM, logistic regression, KNN, and Naive Bayes demonstrated their ability to uncover patterns in patient data that can aid physicians in making timely and accurate diagnoses. The examination of existing research revealed the capabilities of machine learning in improving diagnosis for diseases such as cancer, diabetes, neurological disorders, and more. The promise of these technologies to enhance disease screening and detection, even in remote areas without access to doctors, was also highlighted.

Building on these findings, this report proposed a machine learning model that utilizes these supervised learning techniques to predict multiple diseases from patient symptoms and health parameters. By providing doctors with probable diagnoses complete with confidence scores and tailored recommendations for next steps, the model can assist in expediting and enhancing the diagnosis process. With further training on large datasets of electronic health records, the performance of the model can be improved, as evaluated through metrics like accuracy, precision, and recall. Overall, this review paper provided the basis for understanding how the predictive capabilities of machine learning can

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, November 2023

be harnessed to augment physicians' expertise in tackling the increasing complexity of disease diagnosis, leading to better healthcare outcomes. The proposed model outlined in this report is a step toward realizing that vision.

REFERENCES

[1]. Sethi, R., Thumar, A., Jain, V., & Chavan, S. (2019). Disease prediction application based on symptoms. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 641–646. https://doi.org/10.32628/cseit1952189

[2]. Kulkarni, V., Surwase, S., Pingale, K., Sarage, S., & Karve, A. (2020). Prediction of Disease Using Machine Learning. International Research Journal of Engineering and Technology (IRJET), 7(5), 2458-2463.

[3]. Kumar, A., & Pathak, A. (2021). Assistance Tool for Prediction and Monitoring Various Diseases based on Machine Learning. Turkish Journal of Computer and Mathematics Education, 12(6), 4013-4023.

[4]. Singh, A., Yadav, A., Shah, S., & Nagpure, R. (2022). Multiple Disease Prediction System. International Research Journal of Engineering and Technology (IRJET), 9(3), 2395-0072.

[5]. PARSHANT, & RATHEE, A. (2021). Multiple Disease Prediction Using Machine Learning. Iconic Research and Engineering Journals, 6(12). ISSN: 2456-8880.

[6]. Kumar, A., & Pathak, A. (2021). A Machine Learning Model for Early Prediction of Multiple Diseases to Cure Lives. Turkish Journal of Computer and Mathematics Education, 12(6), 4013-4023.