Multiple Disease Prediction Using Machine Learning

Project report submitted for

4th Semester Minor Project-1

in

Department of _____

By,

Umesh Sinha (211010252)

Sanjiv Kushwaha(211010246)

Paras (211010240)

Department of _____

Dr. Shyama Prasad Mukherjee

International Institute of Information Technology, Naya Raipur

(A Joint Initiative of Govt. of Chhattisgarh and NTPC)

Email: iiitnr@iiitnr.ac.in, Tel: (0771) 2474040, Web: www.iiitnr.ac.in

CERTIFICATE

This is to certify that the project titled “MULTIPLE DISEASE PREDICTION USING MACHINE LEARNING” by “UMESH SINHA , SANJIV KUSHWAHA, PARAS” has been carried out under my/our supervision and that this work has not been submitted elsewhere for a degree.

(Signature of Guide)

_____

Guide Name

Position of Guide

Department of _____

Dr. SPM IIIT-NR

Month, Year

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Signature of Author)

_____

Author Name

(Roll Number)

Date : _____

Approval Sheet

This project report entitled “PROJECT TITLE” by “UMESH SINHA, SANJIV KUSHWAHA, PARAS” is approved for Xth Semester X Project.

(Signature of Examiner - I)

MITHLESH DUBEY

Name of Examiner -I

(Signature of Examiner - II)

SANTOSH KUMAR

Name of Examiner -II

(Signature of Chair)

DR. SHRINIVAS NAYAK

Name of Chair

Date: _____ Place: _____

ABSTRACT

Our project's aims to develop a machine learning model capable of predicting the likelihood of an individual developing one of three diseases: diabetes, heart disease, and Parkinson's disease. The proposed project will use the Python programming language along with the NumPy, Pandas, Matplotlib and scikit-learn libraries to analyze large datasets of patient information and identify patterns and correlations that can be used to predict outcomes. of the disease.

The project is to create a web application. Programs, healthcare professionals and patients can use the app to assess their risk for these diseases and make informed decisions about their health.The relevance of this project lies in the growing need for precise and efficient methods to predict the evolution of the disease.

As the prevalence of chronic diseases increases and healthcare costs rise, there is a growing need for tools that help healthcare professionals and patients make informed decisions about their health. The proposed project would create a web-based application that healthcare professionals and patients could use to determine their risk for diabetes, heart disease and Parkinson's disease, making healthcare more accessible and affordable.Researchers working in this field around the world have developed a variety of machine learning models to predict chronic disease risk. Some popular models support vector machines, logistic regression, and support vector machine classifiers. These models have shown promising results in predicting disease outcomes.

List of All Tables

List of All Figures

INTRODUCTION

Researchers working in this field around the world have developed a variety of machine learning models to predict chronic disease risk. Some popular models are support vector machines, logistic regression, and support vector machine classifiers. These models have shown promising results in predicting disease outcomes. The Python programming language is also a popular choice for developing machine learning models due to its ease of use and the availability of various libraries such as NumPy, Pandas, Matplotlib, and scikit-learn.

In India, there is a growing need for accurate and effective methods to predict disease outcomes. As the prevalence of chronic diseases increases and healthcare costs rise, there is a growing need for tools that help healthcare professionals and patients make informed decisions about their health. However, there are user-friendly tools available to healthcare professionals and patients to determine their risk of developing these diseases.

The relevance of this project lies in the growing need for accurate and efficient methods for predicting disease outcomes. With the increasing prevalence of chronic diseases and the rising costs of healthcare, there is a growing need for tools that can help healthcare professionals and patients make informed decisions about their health. The proposed project will create a web application that can be used by healthcare professionals and patients to assess their risk of developing diabetes, heart disease, and Parkinson's disease, thus making healthcare more accessible and affordable.

LITERATURE REVIEW

Section 2.1:

Some Research Paper Review : Based on the literature review, it is observed that machine learning techniques have been widely used for disease prediction in recent years. Several studies have used machine learning algorithms to predict the likelihood of various diseases, such as diabetes, heart disease, and Parkinson's disease.

Section 2.2 :

Some models and their accuracy : "Parkinson's Disease Diagnosis via Joint Learning From Multiple Modalities and Relations" published on 2 september , 2018 .The proposed method is evaluated on the public Parkinson's progression markers initiative dataset. The extensive experimental results show that the multitask framework can effectively boost the performance of regression and classification and outperforms other state-of-the-art methods. The computerised predictions of clinical scores and label for PD diagnosis may offer quantitative reference for decision support as well. gains an ACC of 91.57%, an F-Score of 85.63% and an AUC of 94.78%. Meanwhile, in PD vs. SWEDD, this method obtains an ACC of 92.06%, an F-Score of 95.39% and an AUC of 94.63%. As for NC vs. SWEDD, this method with CSF data compared with MD data has a slightly high ACC value (e.g., 93.89% vs. 91.89%).

"Machine Learning-Based Heart Disease Prediction: A Study for Home Personalized Care", published on 17 November 2022.This study developed an individual care framework to address heart disease risk using a home system. Machine learning models for predicting heart disease include Logistic Regression, K Nearest Neighbors, Support Vector Machines, Naive Bayes, Decision Trees, Random Forests, and XG Boost. Early and effective detection of heart disease plays an important role in healthcare.The performance of the proposed model was evaluated using the Cleveland Heart Disease dataset from the UCI Machine Learning Repository. Compared to all machine learning algorithms, the Random Forest algorithm showed a better performance accuracy score of 90.16%. The best models assess a patient's health rather than a routine hospital visit.

"Implementing a Web Application to Predict Diabetes Disease: An Approach by Machine Learning Algorithm" ,was published on August,2018 and the main goal of this exploration is to create a web application with higher predictive accuracy based on powerful algorithms. machine learning. It uses a benchmark dataset, Pima Indian, which predicts the onset of diabetes based on diagnostic measures. With a prediction accuracy of 82.35%, artificial neural networks (ANN) showed significant accuracy gains, which prompted us to develop an interactive web application for diabetes prediction.

“Improved Prediction of Diabetes Mellitus using Machine Learning Based Approach” ,published on 5 November 2021. This paper proposes a machine learning-based model to predict diabetic disease. Three supervised machine learning algorithms, namely K-NN, linear SVM and random forest, were selected to predict diabetes for early diagnosis. The area under the curve and the accuracy of each model were obtained using the PIMA India diabetes dataset from the UCI repository. The results of the comparison show that among the three algorithms, random forest is the best model with an accuracy rate of 78.57 and an AUC of 95.08 for the prediction of diabetes risk.

Reference paper name	Published date	Author	Method used
Parkinson's Disease Diagnosis via Joint Learning From Multiple Modalities and Relations	July 2019	Haijun Lei; Zhongwei Huang; Feng Zhou; Ahmed Elazab; Ee-Leng Tan; Hancong Li; Jing Qin	joint regression and classification framework
Machine Learning-Based Heart Disease Prediction: A Study for Home Personalized Care	17 November 2022	Goutam Kumar Sahoo; Keerthana Kanike; Santos Kumar Das; Poonam Singh	Logistic Regression, K - Nearest Neighbor, Support Vector Machine, Naive Bayes, Decision Tree, Random Forest and XG Boost.
Implementing a Web Application to Predict Diabetes Disease An Approach by Machine Learning Algorithm	August 2018	Samrat Kumar Dey; Ashraf Hossain; Md. Mahbubur Rahman	Artificial Neural Network (ANN)

Table 1 Reference Paper Review

CHAPTER 3 PROPOSED SOLUTION

Section 3.1 :

The proposed solution for "Multiple Disease Prediction Using Machine Learning and Python Libraries" includes the following steps:

Data Collection and Preprocessing: A large patient information dataset will be collected and preprocessed to remove any missing or inconsistent data . DELETE. This dataset will include various input data such as age, blood pressure, BMI, blood sugar, number of pregnancies, diabetes pedigree function values, chest pain, maximum heart rate and various other parameters depending on the predicted disease.Using various input data such as number of pregnancies, blood sugar level, age, blood pressure, BMI value, diabetes pedigree function value for diabetes prediction , age, sex, chest pain, maximum heart rate for heart disease prediction and Parkinson's disease prediction, we use inputs values of data distribution 1, distribution 2, PPE, RPDE respectively and D2 will make the model more accurate and efficient.

Feature Selection: Relevant features will be selected from the dataset based on their importance in predicting disease outcome. This will help reduce the dimensionality of the dataset and improve model performance.

Model Development: The dataset will be split into training and testing sets.The model will be developed for diabetes, heart disease and Parkinson's disease prediction algorithms using support vector machines, logistic regression and support vector machine classifiers, respectively, which are known for their high accuracy in classification problems.

Section 3.2 :

Model Optimization: Model hyperparameters will be optimized using techniques such as grid search and random search to improve model accuracy.

Model Evaluation: Model performance will be evaluated using various parameters such as accuracy, precision, recall and F1 score. The model will be compared to other existing models to ensure that it performs better.

Web Application Development: User-friendly web applications will be developed using the Streamlit framework.

The app will allow users to enter their personal information and receive predictions about their risk of diabetes, heart disease and Parkinson's disease.

Overall, the proposed solution aims to create a highly accurate and efficient disease prediction model that healthcare professionals and patients can use to make informed decisions about their health. The use of advanced machine learning algorithms and a user-friendly web application will make the model easily accessible and usable by a variety of users.

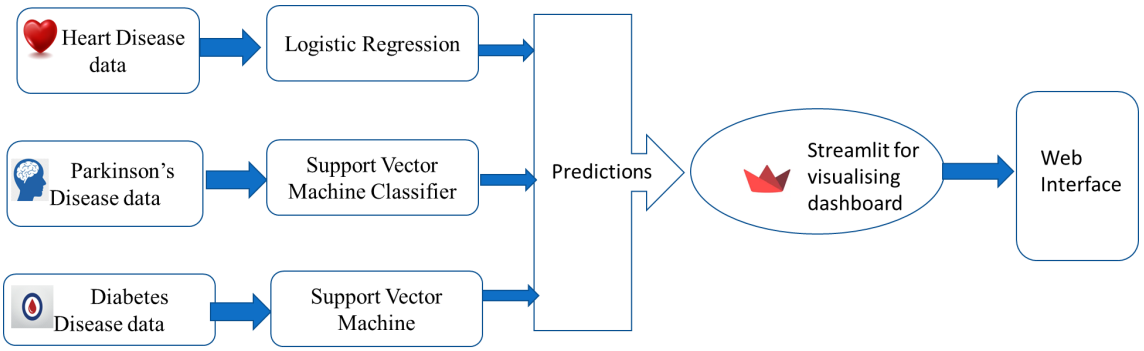

Figure1. Block Diagram of multiple disease prediction using Machine Learning and Web Deployment

CHAPTER 4 RESULTS

Section 4.1 :

The success of the project will be judged on the accuracy and reliability of the machine learning models as well as the usability and efficiency of the web applications developed.

Overall, the proposed project has great potential to contribute to the field of disease prediction research and provide healthcare professionals and patients with valuable tools to make informed decisions about their health.

Section 4.2 :

Support vector machines, support vector classification, and logistic regression are algorithms used to train machine learning models. However, the specific characteristics or variables used to train each disease model are as follows:

For the prediction of diabetes: age, BMI (body mass index), blood sugar level, number of pregnancies, genealogical function of the diabetes, blood pressure. For cardiac prediction: age, sex, type of chest pain, resting blood pressure, cholesterol level, fasting blood glucose, resting ECG results, peak heart rate reached, exercise-induced angina, ST depression exercise-induced.

For the prediction of Parkinson's disease: age, gender, Unified Parkinson's Disease Rating Scale (UPDRS), time of diagnosis, number of previous medication adjustments, and five treatment methods for voice signal (spread1, spread2, PPE, RPDE , and DFA).The model is using **accuracy_score()** to find the accuracy score for the test data.

These characteristics, or variables, were selected based on their relevance and importance in predicting the likelihood of developing each disease.It will be pre-processed, cleaned and converted into a format suitable for machine learning before being used to train models.

CHAPTER 5 CONCLUSIONS

Section 5.1 :

Finally, our project on multi-disease prediction using machine learning has shown promising results in predicting patients with multiple diseases. Using machine learning algorithms, you can analyze large data sets and identify key features indicative of certain diseases. This has the potential to help healthcare professionals make more accurate diagnoses and provide timely treatment.

However, it is important to note that machine learning models are not foolproof and should be used in conjunction with clinical expertise and judgment. Further research and development is needed to ensure the accuracy and reliability of these models.

Overall, our project demonstrates the potential of machine learning in healthcare and the importance of further research in this area. Overall, the proposed project has great potential to contribute to the field of disease prediction research and provide healthcare professionals and patients with valuable tools to make informed decisions about their health.

REFERENCES

Journal :

1. Haijun Lei; Zhongwei Huang; Feng Zhou; Ahmed Elazab; Ee-Leng Tan; Hancong Li; Jing Qin “Parkinson's Disease Diagnosis via Joint Learning From Multiple Modalities and Relations”,July 2019.

2. Goutam Kumar Sahoo; Keerthana Kanike; Santos Kumar Das; Poonam Singh ”Machine Learning-Based Heart Disease Prediction: A Study for Home Personalized Care”, August 2022.

3. Samrat Kumar Dey; Ashraf Hossain; Md. Mahbubur Rahman ”Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm”, August 2018.

4.Madhumita Pal; Smita Parija; Ganapati Panda “Improved Prediction of Diabetes Mellitus using Machine Learning Based Approach”,August 2021.

5.Vijeta Sharma; Shrinkhala Yadav; Manjari Gupta ”Heart Disease Prediction using Machine Learning Techniques”, March 2021.