Disease Prediction Using Machine Learning

1st: V.JAshwin

(M.Sc. Data Science and Business Analytics) Department of Data science and Business Analytics Vels Institute of science, Technology & Advanced studies, Chennai, india

2nd mrs m.s nithya priya

Department of data science and business analytics vels institute of science, technology & advanced studies, chennai, india

1. Abstract

Accurate and early diagnosis of diseases is critical in modern health care. However, rising number of diseases challenge the timely diagnosis, it has become very necessary to develop system

The project explains about the ML based system, it takes patient symptoms and predicts the diseases. The system preprocesses the input data to make predictions. By assisting system healthcare professionals in decision making, this helps to reduce diagnostic errors provide early warnings and to provide guidance for the investigation On our results the machine learning models can achieve high accuracy, making this approach a practical tool for clinical support and to detect the disease easily.

,

2. INTRODUCTION

Health care plays an important role in human life, it early detects the disease easily and can save lives. With the growth of medical data. machine learning offers a new approach to support diagnosis. The main purpose of the project is to analyse the symptoms provided by a patient and to predict the disease . It also helps to understand the possible risks and encourages medical consultation.Objectivesofthe

Project:

1. To create the data set of common disease and their symptoms
2. Implement machine learning algorithms for predicting diseases
3. Evaluate the accuracy of different models
4. Develop a user -friendly system to provide predictions

3. Literature Survey

Previous research highlights of machine learning of disease prediction:

>Machine learning Approach: random trees and decision trees perform well on system based data set which can handle both categorical and numerical data

>Deep learning approaches: useful for complex datasets, medical imaging and health records

>Mobile health applications: several applications provide guidance to patients using ML From these studies, machine learning methods are highly effective for predicting diseases

PROPOSED METHODOLOGY:

The methodology includes following steps:

>Data collection

>data preprocessing

>model selection

>about training

>model testing

>Prediction

1. data collection:gather data sets about the diseases from the reliable medical things
2. data processor:to handle the unknown things and to divide the data
3. model selection:To apply algorithms
4. about training:to train the models on 80%of the dataset
5. model testing:Evaluate the model using the remaining 20% of the data set
6. prediction:to analyze the disease based on the inputs

5. SYSTEM ARCHITECTURE WORK FLOW DIAGRAM:

[Input symptoms]-->[Data processing]- ->[Feature Encoding]-->. [ML Model training]-->[predicted disease]

DESCRIPTION:

Input symptoms: the symptoms which is provided by the user Data pre processing: to clean the data and handle any missing values Feature encoding: to convert symptoms into numerical inputs ML model training: train random or forest decision tree algorithm using the data set Predict symptoms: To know about the disease based on the inputs.

Output: To display probable disease

6. PHASESOFTHEPROJECT

PHASE1-data collection:

>to collect the datasets of symptoms and diseases

>and ensure common diseases are covered

PHASE2-Datapreprocessing:

>to handle the inconsistent data

>convert the categorical data into symptom into numerical format >split the dataset into training and sets(80-20 ratio)

PHASE3-Model training:

>TO train the multiple models. > helps to reduce over fitting PHASE 4- model evaluation: >Evaluate the model based Bon matrices like accuracy, precision PHASE 5-Prediction:>Develop a user-friendly interface like GUI or web application > to allow users to input symptoms and receive disease predictions

7. INPUT

Symptom	value
Fever	1
Cough	1
Headache	0
Fatigue	1

Musclepain0

>1—>symptom present
>0—->symptom absent
And this vector is processed by the machine learning model to generate a prediction

8. PSEUDO CODE AND MPLEMENTATION

Pseudo code Start Load dataset Preprocess data (handle missing values, encode symptoms) model Evaluate model on test data Accept user symptom input Predict disease Display predicted disease End

Python implementation example: Import pandas as pd from sklearn.model_selection import train_test_split

from sklearn.ensemble import RandoForestClassifier
from sklearn.metrics import accuracy_score

#Load dataset data=pd.read_csv("disease_dataset.csv")

#split features and target
x=data.drop("Disease",axis=1)
y=data["disease"]

#split x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2)

model=RandomForest classifier()
model.fit(x_train,y_train)

#predict y_pred=model.predict(x_test)
print("predicted disease:",prediction[0])

9. Output Example:

Input:fever=1, cough=1, Headach e=0, fatigue=1 Output:predicted disease→flu The system predicts the most likely disease based on the symptom input.

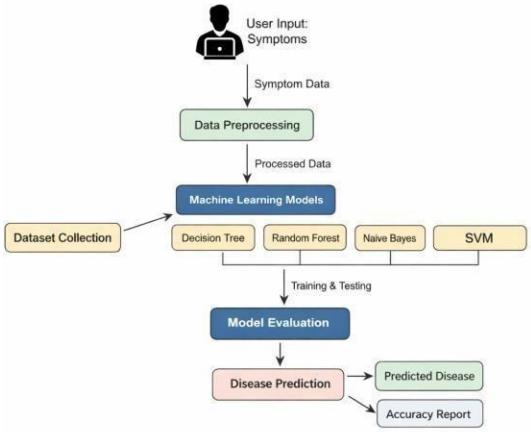
11.RESULTANDDISCUSSION

>Random Forest performs best due to the ensemble learning

>model accuracy depends on the dataset quality

>diseases and symptoms improves prediction reliability

10.SCREEN SHOTS / CHARTS / GRAPHS



12. **CONCLUSION**:

Machine learning enables highly accurate disease prediction and patient history to facilitate early diagnosis