A SYNOPSIS

ON

"MeDiagnose- Identification of Disease and Suggesting Refernces Based on Symptoms Developed using Machine Learning and Integrating it with Web App "

Submitted in partial fulfillment of the requirements for the award of the Degree of

Bachelor of Technology

of

Poornima University, Jaipur

Session: 2022-2023

Submitted By:

Niharika Jain [2019PUSCEBCEX07635]

Prakhar Saxena [2019PUSCEBCCX07276]

Prikshit Saini [2019PUSCEBCCX07513]

IV Year, Computer Engineering (CE)

Submitted To: Department of Computer Engineering

School of Engineering & Technology,

Poornima University Ramchandrapura, Sitapura Ext., Jaipur, Rajasthan

1.   Abstract

Accurate and on-time analysis of any health-related problem is important for the prevention and treatment of the illness. The traditional way of diagnosis may not be sufficient in the case of a serious ailment. Developing a medical diagnosis system based on machine learning (ML) algorithms for prediction of any disease can help in a more accurate diagnosis than the conventional method. We have designed a disease prediction system using multiple ML algorithms. The data set used had more than 230 diseases for processing. Based on the symptoms, age, and gender of an individual, the diagnosis system gives the output as the disease that the individual might be suffering from. After the disease Identification Health tips and Estimated Cost. is suggested and the medical counsellor is suggested.

2.   Introduction

The intelligent diagnosis problem to be solved in this project includes two aspects. The first one is seeking diagnosis experience according to the relationships between symptoms and diseases. Here, supervised machine learning methods are adopted. The second one is disease prediction based on the symptoms provided by visitors. The first one is the main problem. People often concern the relationships between symptoms and diseases when seeking medical advices. In this project, medical data are divided into three copies, records related to main disease categories, records related to subclass disease types, and records of specific diseases firstly; then two disease recognition methods only based on symptoms for the main disease category identification, subclass disease type identification, and specific disease identification are given. In the method validation part, accuracy of the two diagnosis methods is tested and compared. Results show that automatic disease prediction only based on symptoms is possible for intelligent medical triage and common disease diagnosis. The Features of Me diagnose includes Health tips , Providing Awareness about disease and latest News updates from Medical Field.

3.   Motivation/ Aim/Objective

Disease Prediction using Machine Learning is a system that predicts the disease based on the information provided by the user. It also predicts the disease of the patient or the user based on the information or the symptoms he/she enters into the system and provides accurate results based on that information. If the patient is not very serious and the user just wants to know the type of disease, he/she has been through. It is a system that provides the user the tips and tricks to maintain the health system of the user and it provides a way to find out the disease using this prediction. Now a day's health industry plays a serious role in curing the diseases of the patients so this is often also some quite help for the health industry to tell the user and also it's useful for the user just in case he/she doesn't want to travel to the hospital or the other clinics, so just by entering the symptoms and every one other useful information the user can get to understand the disease he/she is affected by and therefore the health industry also can get enjoy this technique by just asking the symptoms from the user and entering in the system and in just a few seconds they can tell the exact and up to some extent the accurate diseases.

The Features of Me diagnose includes Health tips , Providing Awareness about disease and latest News updates from Medical Field.

4. Technical Details (h/w and s/w requirements) including platform

4.1 Hardware Requirement

Name of Equipment	Quantity
PC with minimum 8GB/1TB Configuration	3

4.2 Software Requirement

Name of Technology	Platform	Used For
Python (Tkinter, Django) ,HTML, CSS, JS	Anaconda	Front End
Python (OpenCV, Machine Learning)	Anaconda	Back End
CSV,SQL	CSV,SQL	Database
Figma	Figma	Designing
Kanban Board	Jira	Project Management

5. Proposed Method /Innovativeness and usefulness of the project

➢ Gathering the Data: Data preparation is the primary step for any machine learning problem. We will be using a dataset from Kaggle for this problem. This dataset consists of two CSV files one for training and one for testing. There is a total of 133 columns in the dataset out of which 132 columns represent the symptoms and the last column is the prognosis.

➢ Cleaning the Data: Cleaning is the most important step in a machine learning project. The quality of our data determines the quality of our machine learning model. So it is always necessary to clean the data before feeding it to the model for training. In our dataset all the columns are numerical, the target column i.e. prognosis is a string type and is encoded to numerical form using a label encoder.

➢ Model Building: After gathering and cleaning the data, the data is ready and can be used to train a machine learning model. We will be using this cleaned data to train the Support Vector Classifier, Naive Bayes Classifier, and Random Forest Classifier. We will be using a confusion matrix to determine the quality of the models.

➢ Inference: After training the three models we will be predicting the disease for the input symptoms by combining the predictions of all three models. This makes our overall prediction more robust and accurate.

At last, we will be defining a function that takes symptoms separated by commas as input, predicts the disease based on the symptoms by using the trained models, and returns the predictions in a JSON format.

6. Market Potential and Competitive Advantage

The Global Healthcare Predictive Analytics Market Size was valued at USD 9.5 Billion in 2021 and is estimated to reach the market value of USD 87.5 Billion by 2030, growing at a CAGR of 28.2% from 2022 to 2030. The growing prevalence of chronic disorders and the increasing expenditures on these diseases has been troublesome for people with low and middle income. This factor has exploded the healthcare predictive analytics market revenue throughout the forecast timeframe from 2022 to 2030. According to the Centers for Disease Control and Prevention (CDC), people with chronic and mental health conditions account for 90% of the United States' USD 3.8 trillion in yearly healthcare expenditure. As a result, the potential cost savings from lowering chronic disorder treatment are substantial. For instance, more than 877,500 Americans die each year from heart disease or stroke, accounting for one-third of all deaths. These diseases also have an economic impact, costing the US healthcare system USD 216 billion per year.

7. Plan of work / Work flow (including SRS, DFD, ERD Diagrams)

The waterfall model is a classical model used in system development life cycle to create a system with a linear and sequential approach. It is termed as waterfall because the model develops systematically from one phase to another in downward fashion. The waterfall approach does not define the process to go back to the previous phase to handle changes in requirement. The waterfall approach is the earliest approach that was used for software development.The work Methodology used is Waterfall along with scrum sprints of agile methods using the Jira tool The Kanban Board.

Step 1: Analysis of all the requirements.

Step 2: Gathering of data for the project.

Step 3: Analyzing the Data for training.

Step 4: Training the data and calculating efficiency.

Step 5: Designing of the layout of different modules and front end.

Step 6: Making every module functional.

Step 7: Make a single unit by combining all modules.

Step 8: Proceed for the testing of the designed system.

Step 9: Check whether the system is compatible with every focused environment or not.

Step 10: Make changes if required.

Table 1: Workflow of Project

S.No.	Work to be Done
1.	Mission Planning
2.	Concept Development
3.	Documentation
4.	Detailed Design
5.	Project System Design
6.	Implementing functionality into modules
7.	Combining all functionality into single unit

8.	Testing Phase I
9.	Testing Phase II
10.	Final Presentation to Faculty

8. Conclusion / Expected outcome and benefits to society

In this project, neural network and SVM machine learning methods are given to solve the automatic disease diagnosis problem only based on symptoms. In our methods, each symptom is a feature. The methods work in three layers, which are main disease category identification, subclass disease type identification, and specific disease identification. The methods are suitable for the diagnosis of common diseases and disease triage for specialized diseases. The availability in practice is proved and analyzed in the experiments of this paper. In addition, future research is also required to investigate automatic symptom extraction and discuss the maximum number size of symptoms.

9. Team Members

S.No.	Name	Semester	Role	Branch
1	Niharika Jain	VII	Back-end Developer and ML Engineer	B.Tech CE
2	Prakhar Saxena	VII	Front End Devlopment and Security Administrator	B.Tech CTIS
3	Prikshit Saini	VII	Database Administrator and Content Developer	B.Tech CTIS

10. Guide Name & Designation

Name	Designation	Institutional Address with Phone No.
Dr. A.ANUSHYA	Associate. Professor, FCE, Poornima University	Poornima University IS-2027 to 2031, Ramchandrapura, P.O. Vidhani Vatika, Sitapura Extension, Jaipur-303905 Phone:0141-6500250

11. References

1) S. Muchhala, H. Sodhani and S Geedh, "Disease Prediction using ML", JETIR, vol. 8, no. 11, 2021.
2) S.M.K. Dourjoy, A.M.G.R. Rafi, Z.N. Tumpa and M. Saifuzzaman, "A comparative study on prediction of dengue fever using ML algorithm", Advances in Distributed Computing and ML, pp. 501-510, 2021.
3) N. Nezu, Y. Usui, A. Saito, H. Shimizu, M. Asakage, N. Yamakawa, et al., "ML approach for intraocular disease prediction based on aqueous humor immune mediator profiles", Ophthalmology, vol. 128, no. 8, pp. 1197-1208, 2021.
4) A. Saravanan, L. Chitra, S. S. Chandran, B. S. Aravind, J. N. Kumar, S. Jayaprakash, et al., "Distinguished DC-DC Converter for an Electric Vehicle", 2022 6th International Conference on Computing Methodologies and Communication (ICCMC), pp. 578-582, 2022, March.

5) A. Ramakrishnan and M. S. Ramkumar, "Review on Backup Storage System in Wind Energy Conversion System", 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), pp. 861-869, 2022, April.

6) M. S. Ramkumar, C. Reddy, A. Ramakrishnan, K. Raja, S. Pushpa, S. Jose, et al., "Review on Li-Ion Battery with Battery Management System in Electrical Vehicle", Advances in Materials Science and Engineering, 2022.