

Early Life Style Disease Prediction Using Ensemble Techniques

¹G Pravalika ²V Sai Harshith ³V Alekhya

B.Tech Final Year Students, Dept. of CSE-Data Science, Sphoorthy Engineering College

⁴M Sreelaxmi, *Asst. Professor, CSE-DS, Sphoorthy Engineering College*

ABSTRACT

In order to reduce health risks, there is a need for an effective application, especially in medical services. Currently, patients must contact a doctor and obtain medications to diagnose early lifestyle diseases, which requires multiple services and is time-consuming. There is scope for developing an end-to-end medical service system that covers all the services required by patients. This includes predicting early lifestyle diseases from given symptoms using machine learning, implementing a chatbot application with a question-and-answer system, and integrating a medicine booking system into a single platform. The entire process is designed as a single web application offering multiple services. In this project, we are developing a web application for early lifestyle detection using the Flask framework, Python programming language, and MySQL database.

Date of Submission: 27-04-2024

Date of Acceptance: 05-05-2024

I. INTRODUCTION

The early life style diseases are increasing exponentially every year. These viruses and bacteria are air transmitted. People are mostly attacked by these when they go to some clinic for a minor checkup or doctor suggestion. To counter this problem, we are trying to design an efficient machine learning algorithm to predict the problem you might have in turn decreasing the need for a common man to visit a hospital. We are taking our project to next level by adding modules to book doctors' slot online and buy medicines online which will also help in minimizing the need to go near a hospital. Our idea has a lot of future scope as we can make this data into logs and use them to find the areas which are affected with the viruses and can alert people to keep their distance from that area. Machine learning is an arising approach that helps in determination of diseases using model information or past data. The Machine learning algorithm has two passes: Training and Testing. A Forecast of a disease by using a patient's symptoms and history machine learning technology is battling for past many years. Machine Learning technology gives a decent stage in the clinical field, so a medical care issue can be solved proficiently. We are applying machine learning to maintain complete hospital data Machine learning technology which allows building models to get rapidly examine information and convey results quicker. With the use of machine learning technology doctors can make good decision for patient diagnoses and treatment options, which prompts improvement of patient medical care administrations. Machine Learning technology gives a decent platform in medical field, so that a healthcare issue can be tackled proficiently. The fundamental target of our project is to make the user Interface as apart of front end and connect it with ml models using Django python web framework. We will predict the disease in 2 variants i.e., in general and for specific disease. There is a need to study and make a framework which will make it simple for an end user to anticipate the constant sicknesses without visiting a specialist or a doctor for the treatment and also predicts the type of doctors they should visit. Machine Learning has great power to analyse and cope up with different diseases so that prediction is more accurate and it is cost effective in the treatment Computers give us information; they engage us and help us in a lot of manners. A chatbot is a program intended to counterfeit smart communication on a text or speech. Yet, this paper concentrates only on text. These systems can learn themselves and restore their knowledge using 8 human assistance or using web resources. This application is incredibly fundamental since knowledge is stored in advance. The system application uses the question-and-answer protocol in the form of a chatbot to answer user queries. This system is developed to reduce the healthcare cost and time of the users, as it is not possible for the users to visit the doctors or experts when immediately needed. The response to the question will be replied based on the user query and knowledge base. The significant keywords are fetched from the sentence and answer to those sentences. If the match is discovered or the significant, answer will be given or similar answers will be displayed. With big data growth in biomedical and healthcare communities, accurate analysis of medical data benefits early disease detection, patient care, and community services. However, the analysis accuracy is reduced when the quality of medical data is incomplete. Moreover, different regions exhibit unique characteristics of certain regional diseases, which may weaken the prediction of disease outbreaks. In

this project, we streamline machine learning algorithms for effective prediction of chronic disease outbreak in disease-frequent communities. We experiment the modified prediction models over real-life hospital data collected from different parts of county in 2013-2015. In this project we use three type so medical services to users using our web application which is developed in flask framework using python programming language and MySQL database.

Objective:

Develop a web application with integration of machine learning techniques with a patient friendly process to predict early life style diseases based on user symptoms. Patients can view predicted results and use chat bot to get information about disease related queries which works on NLP technique. Patients can view available doctors and book appointment and get medical aid. Objective is to design this entire process in a single application to help patients to track disease, book appointment and get information from chatbot.

Motivation:

In present scenario machine learning is a fast-growing method to help users in short time period. Medical filed is the most researched area in machine learning for predicting various disease and help patients to know about disease based in symptoms. Taking this as main motivation web application is developed to help patients to get all information about disease tracking, chat bot suggestion and appointment booking from single website.

II. LITERATURE SURVEY

BASE PAPER-1

TITLE:

Disease prediction by machine learning over big data from healthcare communities.(2017) AUTHORS: M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang,“”, ” IEEE Access, vol. 5, no. 1, pp. 8869– 8879, 2017. ABSTRACT: With big data growth in biomedical and healthcare communities, accurate analysis of medical data benefits early disease detection, patient care, and community services. However, the analysis accuracy is reduced when the quality of medical data is incomplete. Moreover, different regions exhibit unique characteristics of certain regional diseases, which may weaken the prediction of disease outbreaks. In this paper, we streamline machine learning algorithms for effective prediction of chronic disease outbreak in disease-frequent communities. We experiment the modified prediction models over real-life hospital data collected from central China in 2013-2015. To overcome the difficulty of incomplete data, we use a latent factor model to reconstruct the missing data. We experiment on a regional chronic disease of cerebral infarction. We propose a new convolutional neural network (CNN)-based multimodal disease risk prediction algorithm using structured and unstructured data from hospital. To the best of our knowledge, none of the existing work focused on both data types in the area of medical big data analytics. Compared with several typical prediction algorithms, the prediction accuracy of our proposed algorithm reaches 94.8% with a convergence speed, which is faster than that of the CNN-based unimodal disease risk prediction algorithm.

III. EXISTING SYSTEM

Recent studies have investigated the patient risk prediction problem within the domain of active learning with relative similarities, particularly in medical applications utilizing Electronic Health Records (EHR). This approach involves an interactive patient risk prediction method that actively solicits feedback from medical experts based on the relative similarity of patients. In this proposed system, machine learning algorithms such as Random Forest and Naive Bayes are employed for both training and prediction tasks. These algorithms leverage EHR data to predict the risk of patients for specific diseases.

Disadvantages:

The existing methods for disease prediction are not automated processes; instead, they rely on manual matching of patients' records. As a result, the time taken for disease prediction is prolonged, as data comparison is based on stored health records. Additionally, the accuracy of these existing methods is typically below 60 percent.

IV. PROPOSED SYSTEM

In the proposed system, we leverage a lifestyle disease dataset sourced from Kaggle. This dataset comprises features representing symptoms, while the labels denote the type of disease. To enable 11 disease prediction, we employ multiple machine learning algorithms, namely Decision Trees (DT), Logistic Regression (LR), Naive Bayes (NB), and Support Vector Machines (SVM). Subsequently, we evaluate the accuracy of each model and select the most efficient algorithm for prediction purposes. In addition, we develop a Flask web application designed around a hospital scenario, featuring a chatbot interface. This interface enables users to

input symptoms, upon which the selected machine learning algorithm generates predictions regarding potential lifestyle diseases. By integrating various machine learning techniques with a user-friendly web interface and chatbot functionality, our system aims to deliver accessible and accurate early predictions of lifestyle diseases, thereby facilitating proactive healthcare management.

Advantages:

Automation streamlines disease prediction by eliminating the need for manual record matching. Improved accuracy is achieved through the utilization of diverse datasets and the implementation of multiple algorithms. Incorporating comprehensive data sources and employing various machine learning algorithms further enhances the robustness of predictions. A user-friendly interface provided by a chatbot-based web application facilitates proactive healthcare management by enabling early prediction of diseases.

V. FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential. Three key considerations involved in the feasibility analysis are

- ECONOMICAL EASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY:

This study evaluates the economic impact of implementing machine learning and chatbot technology in healthcare services. Initial investment in hardware, software, and skilled personnel may be required, but long-term benefits include improved healthcare services, cost savings through reduced doctor visits, and time efficiency for users.

TECHNICAL FEASIBILITY:

The project reliance on machine learning for disease prediction, integration with Django web framework, chatbot development, and data logging and analysis are all technically feasible given advancements in technology and available resources.

SOCIAL FEASIBILITY:

The project reducing healthcare costs and time, providing an easy-to-use interface, addressing privacy concerns, and promoting public health awareness align with social needs and are likely to be well-received by users.

VI. SYSTEM ARCHITECTURE

A system architecture or systems architecture is the conceptual model that defines the structure, behavior, and more views of a system. an architecture description is a formal description and representation of a system. organized in a way that supports reasoning about the structures and behaviors of the system.

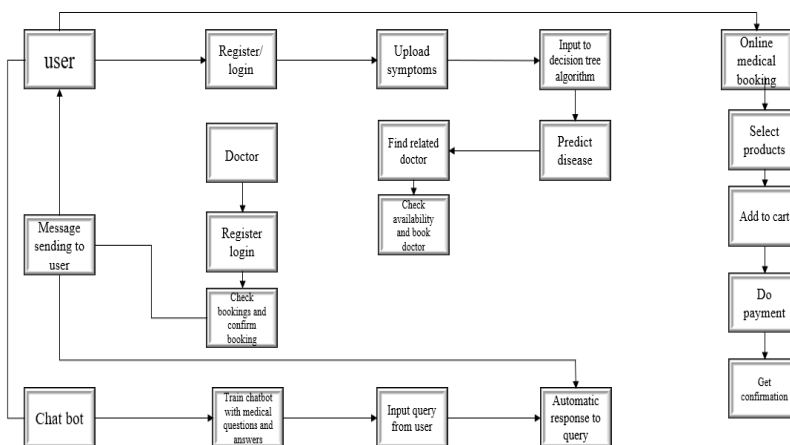


Fig1: System Architecture

UML DIAGRAMS
USE CASE DIAGRAMS

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

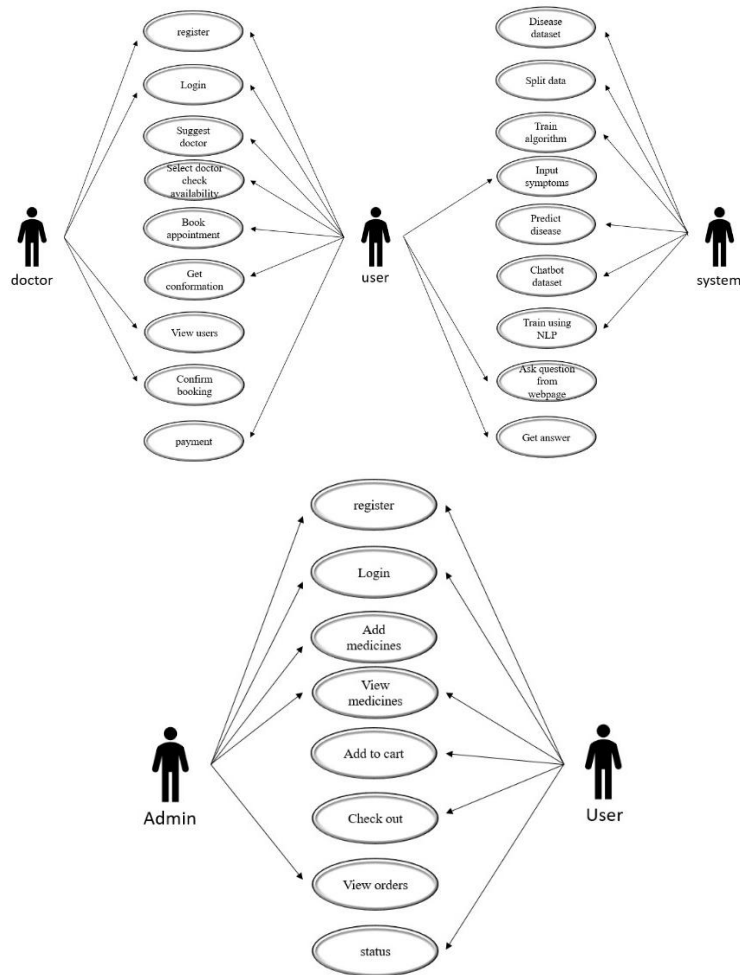


Fig2: use case diagram

SEQUENCE DIAGRAMS

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

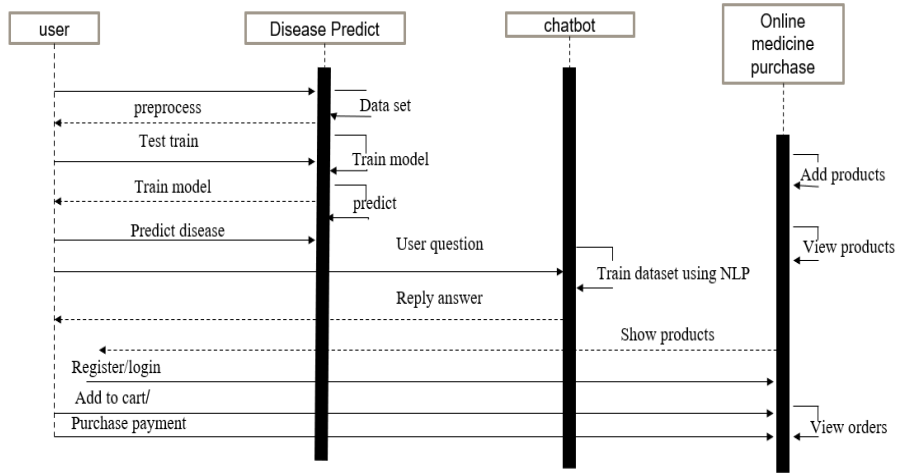


Fig3: Sequence diagram

CLASS DIAGRAM

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.



Fig4:class diagram

ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

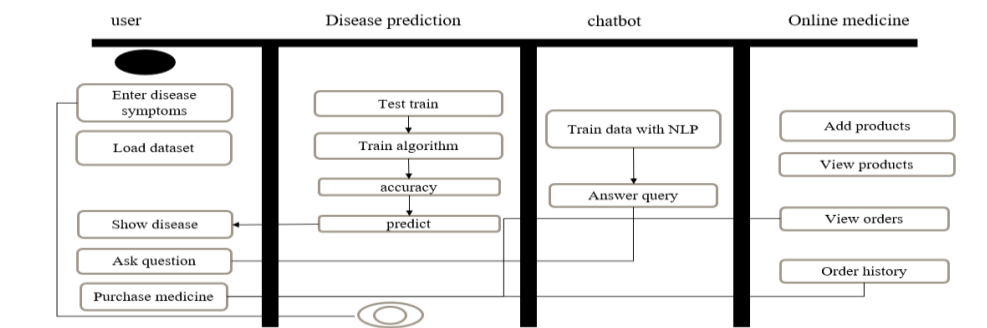


Fig5:activity diagram

VII. MODULES

1. Automatic disease prediction using machine learning (random forest, decision tree, Logistic regression). All three algorithms are used to train medical data and best performed algorithm is used as main model for final prediction.
2. Chat bot service:
Using this service user will register with web application and have option to use chat bot to get automatic response from trained question and answers data which is done using NLP technique.
3. Online Doctor service:

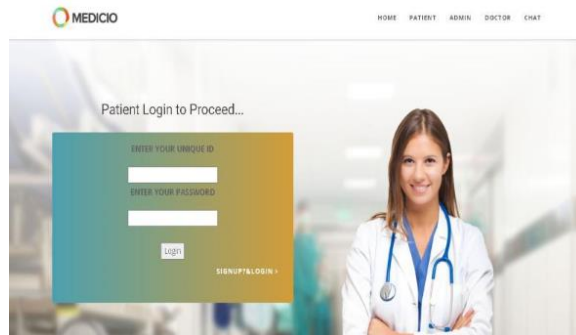
Using this service user can book doctors based on predicted disease using machine learning algorithm based on given input symptoms. Users can book doctors slots based on timings and get conformation from doctor.

4. Online medicine booking Service:

Using this service user will have option to purchase medicines from online store and add products to cart and do payment.

This application helps users to get all type of medical services in emergency situation and get response based on trained data without any doctor consolation.

VIII. RESULTS OUTPUT



This picture shows digital interface displays fields for patient login details with username and password.



This picture shows a user interface presents options for adding symptoms for users to input their specific medical concerns.



This picture shows results screen emerges after inputting symptoms, displaying tailored information or suggestions based on the provided medical concerns.

Symptom1	Symptom2	Symptom3	Symptom4	Symptom5	Disease
123blurred_and_distorted_vision	sinus_pressure	redness_of_eyes	sinus_pressure	redness_of_eyes	Common Cold
123back_pain	back_pain	back_pain	back_pain	back_pain	Drug Reaction
146swelling_of_stomach	throat_irritation	runny_nose	neck_pain	mild_fever	Drug Reaction
123acute_liver_failure	throat_irritation	sinus_pressure	phlegm	malaise	Common cold
123runny_nose	sinus_pressure	weakness_in_limbs	fluid_overload	congestion	Drug Reaction
123runny_nose	sinus_pressure	weakness_in_limbs	fluid_overload	congestion	Common cold
178sinus_pressure	fluid_overload	congestion	diarrhoea	weakness_in_limbs	Common Cold
123fluid_overload	belly_pain	continuous_fuel_of_urineback_pain		spinning_movements	Urinary Infection
123back_pain	loss_of_smell	painful_walking	acute_liver_failure	depression	Drug Reaction

IX. MODULES

We were able to build a ML algorithm that accurately predict the diseases with six symptoms input and also help the customer in booking and buying medicine and help him get in contact with doctor. In this application machine learning based medical helper application is developed by testing disease dataset with multiple machine learning algorithms and most accurate algorithm is used to predict disease which is used in flask web framework. Using this framework health website is designed which has doctor appointment booking, chat bot helper, medicine booking, disease prediction all health-related services are integrated in single application.

Future Work

In future disease data for different disease are collected and trained using Deep learning methods to get more effective results and accuracy. Segmentation of MRI scans can be applied to dataset can be integrated to website. We can collect the disease data in logs and try marking areas with large virus spreads and alert users. This tech will also help government and make their jobs easier.

REFERENCES

- [1]. Talha Mahboob Alam a, Muhammad Atif Iqbal a, Yasir Ali a, Abdul Wahab b, Safdar Ijaz b, Talha Imtiaz Baig b, Ayaz Hussain c, Muhammad Awais Malik b, Muhammad Mehdi Raza b, Salman Ibrar b, Zunish Abbasd, "A model for early prediction of diabetes", vol. 5, no. 1, pp. 8869–8879, 2019
- [2]. C. Beulah Christalin Latha, S. Carolin Jeeva "Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques," Springer Data Mining Knowl. Discovery, vol. 16, no. 4, pp. 1070–1093, 2019.
- [3]. IM. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C. Youn, "Wearable 2.0: Enable human-cloud integration in next generation healthcare system," IEEE Commun., vol. 55, no. 1, pp. 54–61, Jan. 2017.
- [4]. Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "HealthCPS: Healthcare cyberphysical system assisted by cloud and big data," IEEE Syst. J., vol. 11, no. 1, pp. 88–95, Mar. 2017.
- [5]. L. Qiu, K. Gai, and M. Qiu, "Optimal big data sharing approach for telehealth in cloud computing," in Proc. IEEE Int. Conf. Smart Cloud (Smart Cloud), Nov. 2016, pp. 184– 189.
- [6]. Disease and symptoms Dataset –www.github.com.
- [7]. Heart disease Dataset-WWW.UCI Repository. com 37
- [8]. Ajinkya Kunjir, Harshal Sawant, NuzhatF.Shaikh, "Data Mining and Visualization for prediction of Multiple Diseases in Healthcare," in IEEE big data analytics and computational intelligence, Oct 2017 pp.2325.
- [9]. Shanthi Mendis, PekkaPuska, Bo Norrving, World Health Organization (2011), Global Atlas on Cardiovascular Disease Prevention and Control, PP. 3– 18. ISBN 978-92-4-156437-3.
- [10]. Amin, S.U.; Agarwal, K.; Beg, R., "Genetic neural network based data mining in prediction of heart disease using risk factors", IEEE Conference on Information & Communication Technologies (ICT), vol., no.,pp.1227-31,11- 12 April 2013. Chat Bot: 1.K. Oh, D. Lee, B. Ko and H. Choi, "A Chatbot for Psychiatric Counseling in Mental Healthcare Service Based on Emotional Dialogue Analysis and Sentence Generation", 2017 18th IEEE International Conference on Mobile Data Management (MDM), pp. 371-375, 2017.