HEART CONDITION PREDICTION APPLICATION (MyHeartPal)

ABSTRACT

As the saying goes "Charity begins at home" so are the early symptoms of heart malfunctioning, which can be detected at home itself and treated as early as possible before the complications increase. Nearly 17.9 million people are losing their lives due to cardiovascular complaint, which is 32 percent of total death throughout the world. It is a global concern nowadays. The mortality fee because of heart ailment may be decreased through early treatment, for which early-degree detection is an important issue. This is made possible with our application for detecting early symptoms of heart problems. As soon as a person experiences any symptom such as frequent palpitations, shortness of breath, heartache or a combination of all, he can clarify his doubts and check his heart condition through our application. As the first step he/she can chat with our chatbot which is specially designed by us to clear heart condition related queries which after collecting the symptoms from the user informs whether the person does not have CVD symptoms or he should immediately get the tests done for further examination, if the assistant is unsure or the user wants some extra clarification, he/she will be required to play a quiz, for confirmation. When a person has done the necessary medical examinations, after being told to do so through the assistant or quiz or both, he/she needs to enter some of the data obtained from the tests. The AIML model will be trained according to an existing dataset based on data obtained from the patients of the Indian sub-continent and logistic regression will be used to obtain the best result. This will confirm whether he/she has a heart condition that needs medication and cure. Our ultimate objective is to make people aware of their heart condition from their home itself as people tend to ignore the small discomforts and problems related to heart because of their busy schedule and doubt, before it's too late.

Keywords: Cardiovascular disease, Cardiac arrest symptoms, Artificial Intelligence, Machine Learning , Healthcare, Early heart-health detection, Heart condition analysis at home, Heartache, Breathlessness, Chest pain

INTRODUCTION

This application is easy to use and will be easily reachable to the common people. This application is built for every age group and can help detect a heart misfunctioning when a person is least expecting it. We have built a a simple GUI application using the Tkinter library for creating a registration and login screen with options for user authentication. The Home page comprises of options for chat with AI, play quiz to know your heart better, Analysis of heart by ML and logout. The chatbot which specializes in predicting heart condition through symptoms, is built using TensorFlow/Keras and natural language processing techniques for training data. The quiz application is built using the Tkinter library. The quiz allows users to answer a series of questions. Users can answer questions by clicking the "Yes" or "No" buttons, and their responses are stored. Once all questions are answered, the quiz calculates the percentage of "Yes" responses. Based on the percentage, it displays a result health-related feedback using Tkinter's `messagebox.showinfo()` function. If the feedback from the chatbot or the quiz or both is that, the person should get the tests done, he/she must do it. This is when the "Analysis of heart by ML" comes into play. The medical examination data is entered into the desired places. The model is trained on 307 datasets already contributed by various organizations. Based on trained and test data, graphs are plotted and a prediction of heart disease is made.

The modules and libraries used in our project are:
Chatbot-
1.tkinter: This is the standard GUI (Graphical User Interface) library for Python.
2.random: This module provides functions for generating random numbers, choosing random elements from sequences, and shuffling sequences.
3.json: This module provides functions for encoding and decoding JSON data.
4.pickle: It is a module used to implement binary protocols for serializing and de-serializing Python objects.
5.numpy: This library helps support for large, multidimensional arrays and matrices, at the side of a group of mathematical features to perform on those arrays.

6.nltk: It is a natural-language-processing library in Python. It presents an interface for lexical assets which provides a collection of textual content-processing libraries for various tasks such as stemming, tagging, parsing, classification, tokenization, etc..

7.tensorflow. keras: This is a high-level neural networks API written in Python, running on top of TensorFlow. It provides an easy way to build and train deep learning models.

8.WordNetLemmatizer (from nltk.stem): This class is used for lemmatization, which is the process of reducing words to their base or root form.

Quiz-

1.tkinter: This is the standard GUI (Graphical User Interface) library for Python.

2.messagebox: This module provides a set of functions to create message boxes for displaying information or asking for user confirmation.

3.json: This module provides functions for encoding and decoding JSON data.

 Analysis by ML-

1.tkinter: This is the standard GUI (Graphical User Interface) library for Python.

2.datetime: This module supplies classes for manipulating dates and times.

3.numpy: This library provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays.

4.matplotlib: This is a plotting library for Python and provides functions for creating static, animated, and interactive visualizations in Python.

5.backend: This appears to be a custom module in your project, likely containing machine learning models or functions for data processing and prediction.

6.os: It is a module that provides a portable way of using operating system-dependent functionality.

7.messagebox: This module provides a set of functions to create message boxes for various purposes like displaying warnings, errors, or asking for user input.

8.PIL (Python Imaging Library): Although not explicitly imported, it seems like you're using image files (PhotoImage) for buttons and labels, which are part of the PIL package.

9.ttk: This submodule within tkinter provides access to the Tk themed widget set, which includes more modernlooking widgets compared to the standard tkinter widgets.

LITERATURE SURVEY

Subhajit Roy Subhojit Malik et al. [1] published their research named RSVM: A Promising Approach for Early Heart Disease Prediction, on the cardiovascular complications that have rapidly increased after COVID-19 pandemic. An approach has been adopted using supervised machine learning technique known as Rule-based Support Vector Machine (RSVM). It had been introduced in this paper for early detection of this disease. After K-Means clustering, the proposed rule engine was formed . Before this computational model was applied, the data was first cleaned, preprocessed and then outliers have been removed. Ekta Maini,Bondu Venkateswarlu et al. [2] published their research work named Machine learning–based heart disease prediction system for Indian population. The study presents the development and evaluation of a machine learning (ML)-based prediction system for cardiovascular diseases (CVDs) tailored specifically for the Indian population, particularly focusing on rural areas. They applied Traditional ML algorithms to anonymized medical records collected from a tertiary hospital in South India, considering a range of risk factors including lifestyle attributes associated with heart diseases. The study employed five ML algorithms: k-Nearest Neighbours, Naïve Bayes, Logistic Regression, AdaBoost, and Random Forest, to predict the risk of CVDs. The Random Forest algorithm exhibited the highest performance, achieving a diagnostic accuracy of 93.8% on the validation dataset. Dhai, Abdelkamel et al. [3] published their research named Using Machine Learning for Heart Disease Prediction. In this paper they have researched on heart disease from data analytics point of view. They have used data analytics to detect and predict disease's patients. They Started with a preprocessing phase, in which they selected the most relevant features by the correlation matrix, after that they applied three data analytics techniques (neural networks, SVM and KNN) on data sets of different sizes, in order to study the accuracy and stability of each of them and found out that neural networks are easier to configure and obtain good results (93% accuracy). Chayakrit Krittanawong, Hafeez Ul Hassan Virk et al. [5] Their work is named Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques. The main aim of the model was building a potential machine learning model to predict heart disease in the early stages and employing several feature selection techniques in order to identify

significant features. Three unique approaches had been applied for feature selection such as chi-square, ANOVA, and mutual information, and the selected feature subsets had been denoted as SF1, SF2, and SF3, respectively. Six different machine learning models - logistic regression (C1), support vector machine (C2), K-nearest neighbor (C3), random forest (C4), Naive

Bayes (C5), and decision tree (C6) were applied to find an optimistic model with the best-fit feature subset. It turned out that random forest was successful in providing the most optimistic performance for SF3 feature subsets which provided 94.51% accuracy, 94.87% sensitivity, 94.23% specificity, 94.95 area under ROC curve (AURC), and 0.31 log loss. The overall performance of the applied model with the selected features indicated that the model was highly potential for clinical use to predict heart disease in the initial stages with low cost and less time.

METHODOLOGY

Chatbot: To respond to heart related queries, we have set up a simple GUI using Tkinter library. The queries entered by the users are tokenized and lemmatized, the sentences are then converted into bag of words representation. The intent of the user query has been predicted using a trained neural network model. We have predefined the intents and the response is retrieved based on it.

Quiz: It is a simple quiz application where users answer questions with "yes" or "no" responses. Based on their answers, it provides a result indicating their health status.

Analysis by ML: It involved creating a comprehensive GUI for collecting patient data, performing heart health analysis, and displaying the results using graphs and analysis reports. It collects input data from various GUI elements, including entry fields and comboboxes. Then, it plots the input data on four different graphs using Matplotlib. Finally, it predicts the likelihood of a heart disease based on the input data using a machine learning model.

Backend: This Python backend is used to build a typical machine learning workflow for evaluating a logistic regression model for heart disease prediction. It loads the heart disease dataset from a CSV file using Pandas. It provides statistical measures (count, mean, standard deviation, min, quartiles, max) about the dataset and displays the distribution of the target variable. It splits the dataset into features (X) and the target variable (Y) and then scales the features to standardize the dataset's features to have a mean of 0 and a standard deviation of 1. Finally, it initializes and trains a logistic regression model using the training data, makes predictions on both the training and test datasets using the trained model, calculates the accuracy of the model's predictions on both the training and test datasets and prints it.

PROPOSED APPROACH

Firstly, we learned about training and test data, neural networks and various python libraries. Instead of using an API for building the chatbot, we have customized the training model to enable it to recognize certain health related queries. Then we have built a simple quiz to give better confirmation about the possibility of an underlying heart issue. The quiz comprises of yes/no type questions. We went through the various studies conducted earlier especially the [4] Ekta Maini,Bondu Venkateswarlu et al. " A heart disease prediction system based on machine leaning , for Indian population: An exploratory study performed in South India". It presented the development and evaluation of a machine learning (ML)-based prediction system for cardiovascular diseases (CVDs) tailored specifically for the Indian population, particularly focusing on rural areas. They considering a range of risk factors including lifestyle attributes associated with heart diseases. The study employed five ML algorithms: k-Nearest Neighbors, Naïve Bayes, Logistic Regression, AdaBoost, and Random Forest, to predict the risk of CVDs. Highest performance was exhibited by the Random Forest algorithm. We decided to go with the Logistic regression, which we applied on the dataset available on public cloud. The accuracy on training data is 0.847107 and accuracy on test data obtained is 0.786885. The output consisted of the four graphs and the report that printed whether the user has a heart disease or not.
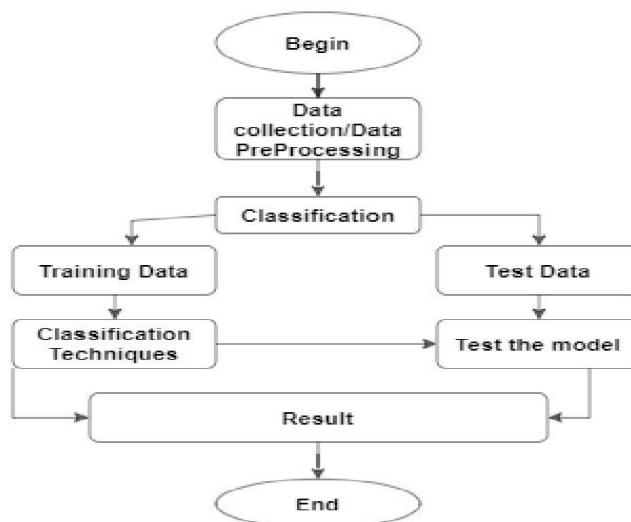
Fig. 1: Generic Model Predicting Heart Disease

CONCLUSION

MyHeartPal is an easily operated heart disease predictor which cautions people of an underlying heart issue in the early stages before it gets fatal. It's usefulness lies in identifying and addressing the early signs of heart malfunctioning in the comfort of one's home amidst one's busy schedule when one is most likely to ignore it.

The integration of a machine learning algorithm, specifically a logistic regression model trained on a dataset of various health parameters, adds a layer of sophistication to the application. The practical utility of the application lies in its capacity to predict the severity of a potential heart issue and guide users on the urgency of seeking medical attention. In essence, this console-based application represents a holistic approach to early heart issue detection, combining userinitiated conversations, targeted questioning, and machine learning analysis. By empowering individuals to take control of their heart health and facilitating informed decisions about medical consultation, the application contributes significantly to preventive healthcare and overall well-being.

Scope for improvement-

1.We can modify our quiz set by asking questions which are relevant for the KNN specification to get the desired inputs.

2.When we have achieved enough accurateness, we can extend our model to make more critical decisions like what should be the next steps is a heart related issue is detected like proposing methods like echo cardiography, magnetic resonance imaging, etc..

ACKNOWLEDGMENT

REFERENCES

[1] *Echo Research and Practice.* 2019 by Seetharam K., Kagiyama N., Sengupta P. (Jan,2020)

[2] Burke L. E., Ma J., Azar K. M. J., et al. Current science on consumer use of mobile health for cardiovascular disease prevention. *Circulation.* 2015,

[3] Jiang F., Jiang Y., Zhi H., et al. Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology.* 2017.

[4]Ekta Maini,Bondu Venkateswarlu et al. Machine learning–based heart disease prediction system for Indian population: An exploratory study done in South India.

[5] Chayakrit Krittanawong, Hafeez Ul Hassan Virk, Sripal Bangalore el al. Machine learning prediction in cardiovascular diseases: a meta-analysis