CardioVigilant: Cardiovascular Decompensation Forecasting

PROJECT PHASE #3

Aishwarya Chand, Prajakta Jhade, Shreya Thakur

PROJECT OVERVIEW

*One person dies every **33 seconds** in the United States from cardiovascular disease. About 695,000 people in the United States died from heart disease (according to the data collected in 2021)—that's 1 in every 5 deaths. Globally, the scale of mortality due to this disease is even more surprising with a record of 20.5 million.*

Cardiovascular diseases are one of the major healthcare concerns and leading causes of mortality globally. Early intervention plays a crucial role here and enables healthcare providers to tailor treatment plans and adjust medications accordingly. Despite advancement in cardiovascular care, predicting and preventing cardiovascular decompensation remains a significant challenge.

Our Web Application: CardioVigilant aims to transform cardiovascular healthcare, equipping healthcare providers with an unparalleled tool for precision forecasting and redefine standards in patient care, ultimately saving lives and enhancing the quality of cardiovascular health globally.

PHASE-1 OVERVIEW

In the project phase one, we have performed data cleaning and exploratory data analysis (EDA) to prepare the data for the development of a Cardiovascular Decompensation Forecasting model. Initially, null values were removed from critical columns, including Sex, RestingBP, Cholesterol, MaxHR, and heart disease, to maintain the integrity of the data.

Categorical columns are expected as binary but containing more than the counted on the categories were corrected. Outliers on Age, especially the records with ages less than 20, were detected as deviations and removed. Gender variance was maintained by removing records with null values and correcting inconsistencies in naming. ChestPainType and ExerciseAngina columns were cleaned for consistency, also by performing renaming and grouping values. RestingBP and MaxHR Columns, we have imputed null values with mean to maintain dataset completeness. Label encoding was applied to convert categorical data into numerical which is suitable for modelling.

During EDA, the dataset disclosed insights as: males has higher incidence of heart disease as compared to females; some types of chest pain, like Asymptomatic (ASY), were more closely prone to heart disease and other factors like RestingECG and FastingBS showed higher susceptibility to get the heart issues. In Addition, a correlation matrix was used to identify the trends between various features and t he targets.

PHASE 2

In Phase 2 of our project, we delved deeper into the application of various machine learning algorithms for predicting heart disease. We explored a range of algorithms including Logistic Regression, Random Forest, Artificial Neural Network (ANN), Naive Bayes, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree, and XGBoost. Each algorithm was carefully selected based on its suitability for binary classification tasks and its ability to handle the complexities of medical datasets.

Throughout Phase 2, we conducted extensive experimentation to train and fine-tune each model. This involved parameter tuning, cross-validation, and evaluation using a variety of metrics such as accuracy, precision, recall, F1-score, ROC-AUC score, and misclassification rate. We also leveraged techniques like RepeatedStratifiedKFold and GridSearchCV to ensure robustness and optimize model performance.

Overall, Phase 2 of our project was characterized by rigorous experimentation, evaluation, and analysis of machine learning algorithms for heart disease prediction. By leveraging a diverse range of techniques and models, we aimed to develop a robust and accurate predictive model with practical utility in healthcare settings.

UTILIZATION OF MODELS FROM PHASE 2

In Phase 2 of our project, we carefully considered the selection and utilization of machine learning models, with a particular focus on effectively predicting heart disease risk. Among the various algorithms explored, we ultimately chose to implement the Artificial Neural Network (ANN) due to its ability to capture intricate patterns and nonlinear relationships inherent in medical datasets, particularly those related to cardiovascular health.
While K-Nearest Neighbor (KNN) may have demonstrated marginally better output values in some aspects, the decision to prioritize the Artificial Neural Network (ANN) for heart disease prediction was driven by several key factors.

The ANN model was structured with three dense layers, each utilizing the Rectified Linear Unit (ReLU) activation function to introduce nonlinearity. Additionally, we incorporated Dropout layers after the first two dense layers to mitigate overfitting by randomly dropping a fraction of neurons during training.

For this project we wanted to gain more insight and hence we went though a couple of research and review paper (added in reference) which helped us to choose Artificial Neural Network as the model .

Heart disease prediction often involves analyzing complex, nonlinear relationships between various risk factors and the likelihood of disease occurrence. ANNs, with their deep learning architecture and ability to learn hierarchical representations of features, are well-suited to capture these intricate patterns present in raw medical data.

We opted for the Adam optimizer and binary cross-entropy loss function to compile the model, aiming to minimize loss and maximize accuracy. Training was conducted over 100 epochs with a batch size of 32, and validation was performed on unseen data to assess generalization performance.

In evaluating the effectiveness of the ANN model, we focused on key metrics such as binary cross-entropy loss, accuracy, and validation accuracy. The model exhibited promising results, achieving an accuracy of 90.60% on the training data and 85.33% on the validation data. The relatively high validation accuracy suggests that the model can generalize well to new patient data, which is critical for practical utility in real-world healthcare applications.

By prioritizing the ANN model, our goal was to provide healthcare professionals with a robust and reliable tool for accurately assessing heart disease risk. The ANN's ability to effectively learn from data and generalize to unseen instances enables it to generate accurate predictions, thereby empowering clinicians to make informed decisions and initiate timely interventions. Ultimately, our aim is to leverage the capabilities of the ANN to improve patient outcomes by facilitating early detection and proactive management of heart disease, ultimately contributing to better overall healthcare delivery.

WHAT CAN YOU AS A USER LEARN FROM OUR PRODUCT ?

By analyzing their health data, users can understand the personalised risk for cardiovascular diseases/decomposition. This includes factors like tendency to behave, lifestyle factors, and environmental factors that may contribute to their cardiovascular health status.

Through the predictive analytics capabilities of CardioVigilant, users can learn to plan for their long term cardiovascular health by understanding the future risk and taking preventive measures accordingly. This proactive approach can help users mitigate future cardiovascular risks and improve their overall quality of life.

PROJECT EXTENSION AND PRACTICAL APPLICATION

CardioVigilant aims to transform cardiovascular healthcare, equipping healthcare providers with an unparalleled tool for precision forecasting and redefine standards in patient care, ultimately saving lives and enhancing the quality of cardiovascular health globally.

For now we have extended our project by creating the user Interface where users can input their required medical details and get the prediction However, we can further extend this project by Incorporating wearable devices that can provide real-time data on patients' vital signs, activities, and other relevant metrics. The continuous stream of data can significantly enhance the accuracy of predictive analytics and early detection of cardiovascular decompensation.

Analysing historical data of patient’s, genetic propensities, lifestyle factors, and environmental factors can be used to personalise risk assessment for cardiovascular diseases. This approach allows providers to take preventive measures based on individual risk profiles, leading to more targeted and effective care.

Expanding CardioVigilant to include remote patient monitoring capabilities enables healthcare providers to monitor patient’s heart health outside clinical settings. By collecting data according to the symptoms of a patient , healthcare providers can work promptly in case of any deviations from the normal health status, thus preventing cardiovascular decompensation and reducing hospitalizations.

RUNNING THE HEART DISEASE PREDICTION MODEL:

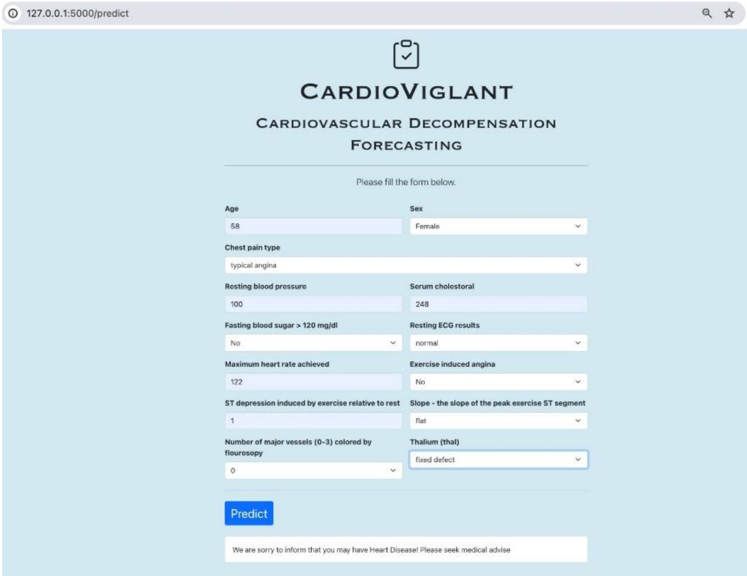
Tech Stack Used:
1. Python and Data Science Libraries
2. File Handling techniques
3. HTML/CSS
4. Flask

By following these steps, you can run our heart disease prediction model locally and obtain predictions using the Flask AP

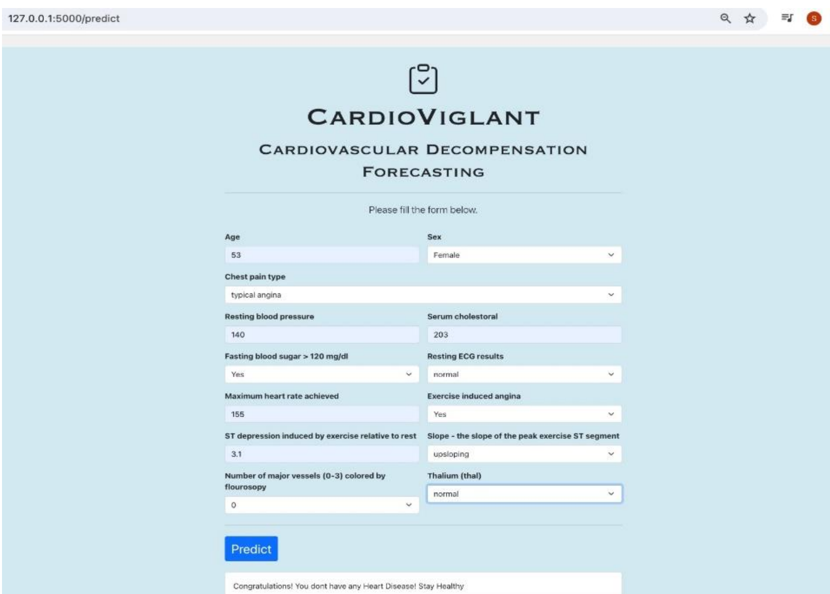
1. Run the command `python app.py` to run the app.py file. Which will take you to our web page on your local host.
   - The app.py code uses the pretrained model that we have created in previous phase and run the model for our user. We have created the pickle file for our model.

2. Input Data and Get Predictions
   - On the home page, you'll find input fields to enter information for heart disease prediction, such as age, cholesterol level, blood pressure, etc.
   - After entering the required information, submit the form.
   - The Flask API processes the input data using the trained machine learning model.
   - You will receive predictions regarding the likelihood of heart disease based on the input data

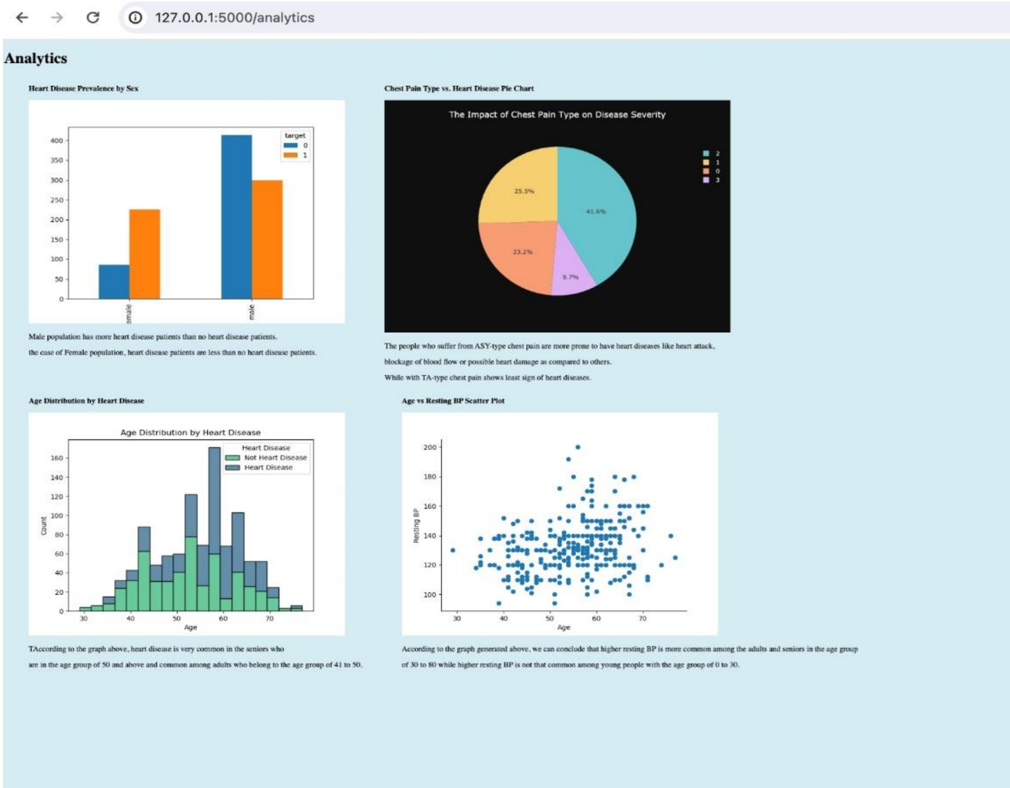
PREDICTION THAT USER HAVE LIKELIHOOD TO GET HEART ATTACK



PREDICTION THAT USER HAVE LIKELIHOOD TO NOT GET HEART ATTACK



Analytics Page:-



DATASET DETAILS

• Age: of the patient in years.
• Sex: of the patient, categorized as Male (M)-[0] or Female (F)-[1].
• ChestPainType: Describes the chest pain categorized as Typical Angina (TA)[0], Atypical Angina (ATA)[1], Non-Anginal Pain (NAP)[2], or Asymptomatic (ASY)[3]. Different types of chest pain may indicate different heart conditions.
• RestingBP: Represents the resting blood pressure of the patient measured in mmHg. Blood pressure is a common risk factor for heart disease.
• Cholesterol: Serum cholesterol level measured in mm/dl. High cholesterol levels are linked with an increased risk of developing heart disease.
• FastingBS: Fasting Blood Sugar Indicates whether the patient has fasting blood sugar or not. Levels > 120 mg/dl (1) or not (0).
• RestingECG: resting electrocardiogram as Normal[0], showing ST-T wave abnormalities (ST)[1], or indicating left ventricular hypertrophy (LVH)[2].
• MaxHR: Represents the maximum heart rate achieved. This attribute's numeric value lies between 60 and 202.
• ExerciseAngina: Indicates whether the patient experiences exercise-induced angina [Y: Yes, N: No].
• Oldpeak: Represents the ST depression measured during exercise. ST depression can indicate myocardial ischemia, which is a lack of blood flow to the heart muscle.
• ST_Slope: Describes the slope of the peak exercise ST segment as Upsloping[0], Flat[1], or Downsloping. [2][Up: upsloping, Flat: flat, Down: downsloping]

• Heart Disease: Output of predicting heart disease based on the input features mentioned above. [1: heart disease, 0: Normal]

REFERENCES

ANN Model –
https://www.tensorflow.org/tutorials/quickstart/beginner

https://scikit-learn.org/stable/

Decision Tree – https://developers.google.com/machine-learning/decision-forests/decision-trees

Random Forest - https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

https://core.ac.uk/download/pdf/234686628.pdf