Heart Disease Prediction

A *Final Year Project* submitted

in partial fulfilment of the requirements

for the Degree of

Bachelor of Technology

in

Computer Science and Engineering

Under Guidance Of:

Dr Prashant Kumar

Assistant Professor

Submitted By:

Gyanesh Kumar (17103033)

Sonal Keshri (17103083)

Dr. B R Ambedkar National Institute of Technology

Jalandhar, Punjab

Session: 2017-2021

ACKNOWLEDGEMENT

It's rightly said that there are hundreds of people working behind the curtain to make a Play successful. The final result of our project required a lot of guidance and help from many people and we are extremely fortunate to have got this throughout the journey of our project. Whatever we are today is only due to such supervision and assistance and we would not forget to thank them from the bottom of our heart. We are extremely thankful to Professor A L Sangal, Head, Department of Computer Science & Engineering who has provided all his direct and indirect support.

We are thankful to the In-charge, Major Project Final Year, for providing us mentor and all other support. We are extremely thankful to our mentor Dr Prashant Kumar, who believed in our idea and suggested new ideas wherever required. He has supported us fully in giving solutions to our problems.

We are extremely thankful to have got constant encouragement and guidance from all the Faculties of Department of Computer Science & Engineering(CSE) who gave us their time and suggestions and then helped us in successfully completing our project work. Also, we would like to extend our sincere esteems to all staff in the laboratory for their timely support.

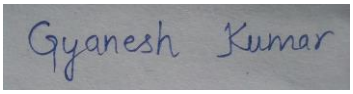
Thank you.

Gyanesh Kumar (17103033)

Sonal Keshri (17103083)

DECLARATION

I hereby certify that the work, which is being presented in the project, entitled **"Heart Disease Prediction"** in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology and submitted in the Department of Computer Science and Engineering, National Institute of Technology Jalandhar, is an authentic record of my own work carried out during a period from August 2020 to May 2021 under the supervision of Dr. Prashant, Assistant Professor, Department of Computer Science & Engineering, National Institute of Technology Jalandhar. The matter presented in this project has not been submitted by us for the award of any other degree of this or any other Institute/University.



Gyanesh Kumar **Sonal Keshri** **Dr. Prashant Kumar**

Assistant Professor

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Table Of Contents

1. Abstract

Among all fatal diseases, heart attacks diseases are considered as the most prevalent. The number of people suffering from heart disease is on the rise (health topics, 2010). The report from the World Health Organization shows us that a large number of people die every year due to heart disease all over the world. Heart disease is also stated as one of the greatest killers in Africa.

One of the most important uses of data analysis is that it helps in keeping human bias away from medical conclusions with the help of proper statistical treatment. By use of data mining for exploratory analysis because of nontrivial information in large volumes of data. The health care industries collect huge amounts of data that contain some hidden information, which is useful for making effective decisions for providing appropriate results and making effective decisions on data, some data mining techniques are used to better the experience and conclusion that have been given.

Heart predictor system will use the data mining knowledge to give a user-oriented approach to new and hidden patterns in the data. The knowledge which is implemented can be used by the healthcare experts to get better quality of service and to reduce the extent of adverse medicine effect.

2. Problem Statement

Heart disease can be managed effectively with a combination of lifestyle changes, medicine and, in some cases, surgery. With the right treatment, the symptoms of heart disease can be reduced and the functioning of the heart improved. The predicted results can be used to prevent and thus reduce cost for surgical treatment and other expenses.

The overall objective of our work will be to predict accurately with few tests and attributes the presence of heart disease. Attributes considered form the primary basis for tests and give accurate results more or less. Many more input attributes can be taken but our goal is to predict with few attributes and faster efficiency the risk of having heart disease.

Decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the data set and databases. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients.
Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to identify inefficiencies and best practices that improve care and reduce costs.

Objective:

The main objective of this research is to develop a heart prediction system. The system can discover and extract hidden knowledge associated with diseases from a historical heart data set. Heart disease prediction system aims to exploit data mining techniques on medical data sets to assist in the prediction of heart diseases.

- Provides a new approach to concealed patterns in the data.
- Helps avoid human biases.
- To research & implement the most accurate classifier that classifies the disease as per the input of the user.
- Reduce the cost of medical tests

The proposed system will integrate clinical decision support with computer-based patient records (Data Sets). This will reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome.

3. Feasibility Study Report

3.1 Title

Heart Disease Prediction System

3.2 Abstract

Heart predictor system will use the data mining knowledge to give a user-oriented approach to new and hidden patterns in the data. The knowledge which is implemented can be used by the healthcare experts to get better quality of service and to reduce the extent of adverse medicine effect.

3.3 Keywords

- **Machine Learning:**

  Machine learning (ML) is a category of algorithm that allows software applications to become more accurate in predicting outcomes without being explicitly programmed.

- **Decision Tree:**

  Decision Trees are a non-parametric supervised learning method used for both classification and regression tasks.

- **Random Forest:**

  Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the

class that is the mode of the classes or mean/average prediction of the individual trees.

- **Support Vector Machines (SVMs):**

  These are supervised machine learning algorithms used to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points.

- **K Nearest Neighbor:**

  The nearest neighbors algorithms are a simple, easy-to-implement set of supervised machine learning algorithms that can be used to solve both classification and regression problems.

3.4 Description of the project

The main objective of this research is to develop a heart prediction system. The system can discover and extract hidden knowledge associated with diseases from a historical heart data set. Heart disease prediction system aims to exploit data mining techniques on medical data sets to assist in the prediction of heart diseases.

- Provides a new approach to concealed patterns in the data.
- Helps avoid human biases.
- To research & implement the most accurate classifier that classifies the disease as per the input of the user.
- Reduce the cost of medical tests

3.5 Project Scope

The scope of the project is that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome.

This suggestion is promising as data modeling and analysis tools, e.g., data mining has the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

Using this model, we will be able to predict the presence/absence of given disease based on the parameters provided. Results obtained from different models are compared side by side based on the speed and accuracy of computation.

The results are not guaranteed to be 100% accurate as these are totally based on the dataset used to train the model.

3.6 Technical Feasibility

- Period of Completion: At least 7-8 months.
- Processor: Intel® Core™ i5 processor or better.
- RAM and Memory: 8-16 GB of DDR3 RAM or better
- Python versions: 3.X
- Software / Tools Required: Jupyter Notebook (IPython), Android Studio, NodeJS
- Programming Languages: PYTHON, KOTLIN/JAVA, TYPESCRIPT

3.7 Risk Analysis

- **Technical:**
  Android & Python Application Supporting Devices as mobiles, tablets, desktops, laptops and Machine Learning

- **Economic Feasibility:**

This project doesn’t require much cost in development. It only requires cost for database management

- Internet connectivity is required for the system

3.8 Deliverables

- Product Research Report
- Android App
- Admin Panel
- ML Model
- Backend

4. PRODUCT RESEARCH

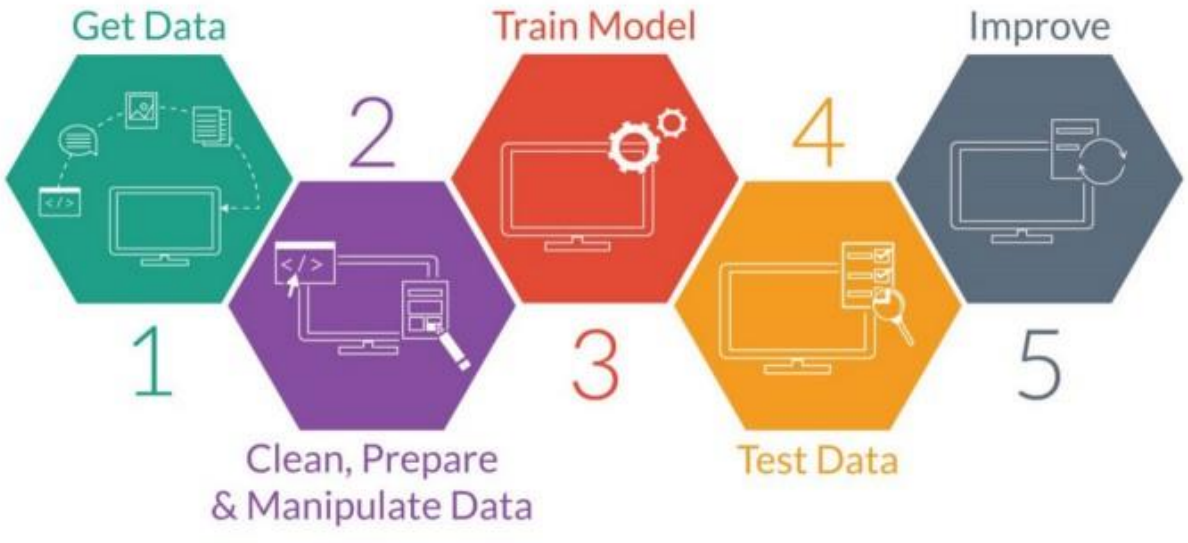


Fig 4.1 ML Model Research Steps

4.1 Gathering Data

Our Predictor (Y, Positive or Negative diagnosis of Heart Disease) is determined by 13 features (X)
The dataset has been taken from UCI Database. There are 14 columns in the dataset, where the patient_id column is a unique and random identifier. The remaining 13 features are described below.

1. `age` - age in years
2. `sex` - (1 = male; 0 = female)
3. `cp` - chest pain type
    - 0: Typical angina: chest pain related decrease blood supply to the heart
    - 1: Atypical angina: chest pain not related to heart

- 2: Non-anginal pain: typically esophageal spasms (non heart related)
- 3: Asymptomatic: chest pain not showing signs of disease

4. `trestbps` - resting blood pressure (in mm Hg on admission to the hospital) anything above 130-140 is typically cause for concern
5. `chol` - serum cholesterol in mg/dl
    - serum = LDL + HDL + .2 * triglycerides
    - above 200 is cause for concern
6. `fbs` - (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
    - '>126' mg/dL signals diabetes
7. `restecg` - resting electrocardiographic results
    - 0: Nothing to note
    - 1: ST-T Wave abnormality
        - can range from mild symptoms to severe problems
        - signals non-normal heart beat
    - 2: Possible or definite left ventricular hypertrophy
        - Enlarged heart's main pumping chamber
8. `thalach` - maximum heart rate achieved
9. `exang` - exercise induced angina (1 = yes; 0 = no)
10. `oldpeak` - ST depression induced by exercise relative to rest looks at stress of heart during exercise unhealthy heart will stress more
11. `slope` - the slope of the peak exercise ST segment
    - 0: Upsloping: better heart rate with exercise (uncommon)
    - 1: Flat Sloping: minimal change (typical healthy heart)
    - 2: Downsloping: signs of unhealthy heart
12. `ca` - number of major vessels (0-3) colored by fluoroscopy
    - colored vessel means the doctor can see the blood passing through
    - the more blood movement the better (no clots)
13. `thal` - thallium stress result
    - 1,3: normal
    - 6: fixed defect: used to be defect but ok now
    - 7: reversible defect: no proper blood movement when exercising
14. `target` - have disease or not (1=yes, 0=no) (= the predicted attribute)

4.2 Data Pre-Processing

Datasets in a perfect world is a perfectly curated group of observations with no missing values or anomalies. However, this is not true. Real world data comes in all shapes and sizes. It can be messy, which means it needs to be clean and wrangled. Data cleaning is a necessary part in data science problems.

Machine learning models learn from data. It is crucial, however, that the data you feed them is specifically pre processed and refined for the problem you want to solve. This includes data cleaning, preprocessing, feature engineering, and so on.

Let's say we have a column Gender, with values 1 for Male and 0 for Female. It needs to be converted into two columns with the value 1 where the column would be true and 0 where it will be false.

To get this done, we use the `get_dummies( )` method from pandas. Next, we need to scale the dataset for which we will use the StandardScaler. The `fit_transform( )` method of the scaler scales the data and we update the columns.

4.3 Training & Testing the ML Model

Now we'll Train various Classification Models on the Training set & see which yields the highest accuracy.
In this project, We took 4 algorithms and varied their various parameters and hypertuned them for the final model. We split the dataset into 67% training data and 33% testing data.

4.4 Improvising & Hypertuning the Model

The project involved analysis of the heart disease patient dataset with proper data processing. Then, 4 models were trained and tested with maximum scores as follows:

- K Neighbors Classifier: 87%
- Support Vector Classifier: 83%
- Decision Tree Classifier: 79%
- Random Forest Classifier: 84%

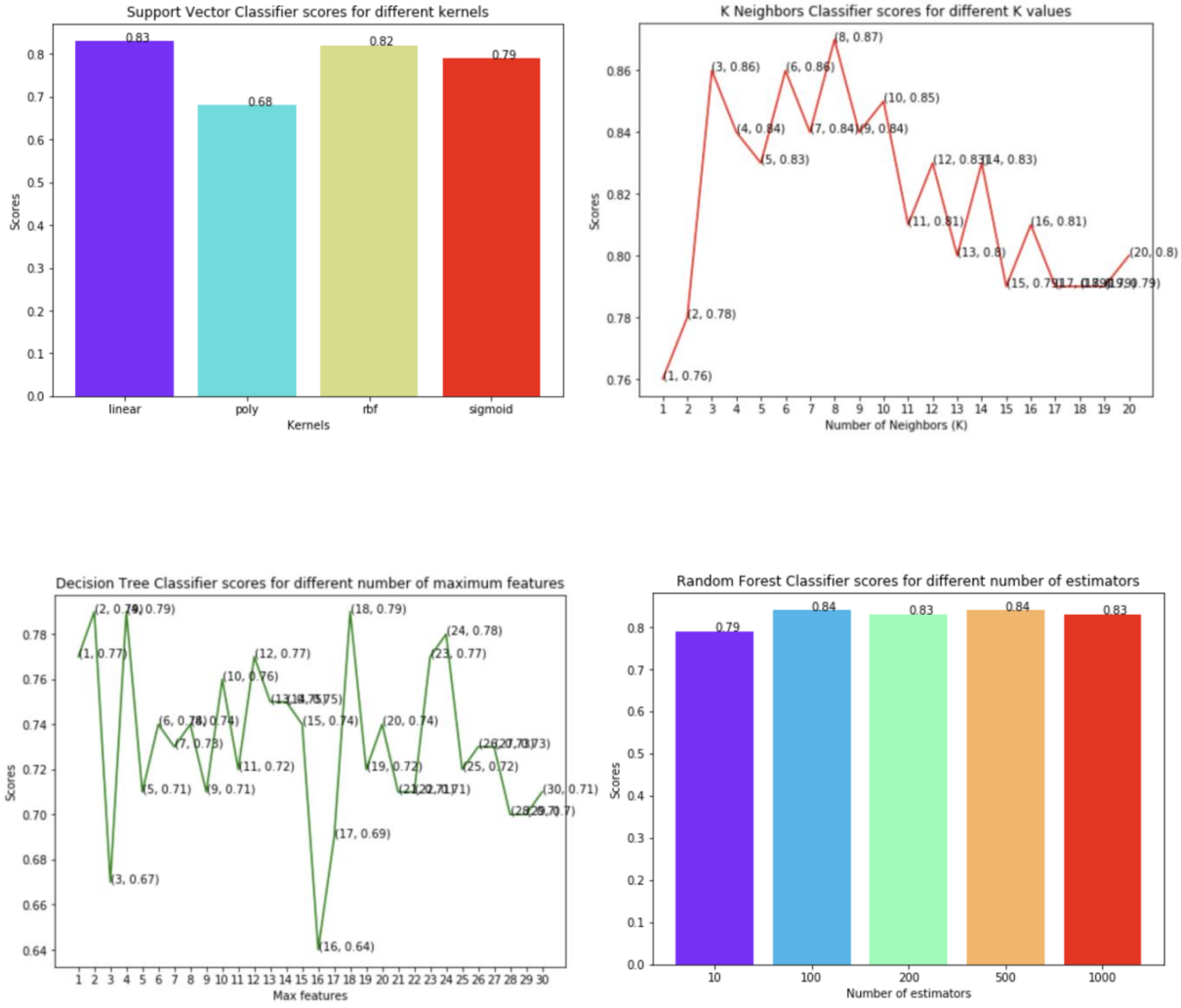


Fig 4.4 Results of Various ML Algorithms

4.5 Result

As we can see in the above figures, the maximum accuracy of the ML Model is achieved using the K Nearest Neighbour (KNN) classification algorithm.
We also hypertuned the model for various values of neighbours ranging from 1 to 20.
After analysing and comparing we can conclude that K Nearest Neighbour Classifier is best suited for our use case and the value for the number of neighbors (hypertuning parameter) should be 8 for best accuracy.

5.Tools & Technologies

5.1 Framework & Libraries

- **NestJs -** Nest (NestJS) is a framework for building efficient, scalable NodeJS server-side applications. It uses progressive JavaScript, is built with and fully supports TypeScript(yet still enables developers to code in pure JavaScript) and combines elements of OOPd (Object Oriented Programming), FP (Functional Programming), and FRP (Functional Reactive Programming).

- **Flask -** Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions

- **ML Libraries -** sklearn, pandas

5.2 Programming Languages

- Python

- Javascript
- Typescript
- Kotlin
- XML/JSON

5.3 Architecture Patterns Used

- MVVM (Model-View-ViewModel) for Android App
- MVC(Model-View-Controller) for web backend

5.4 DataBase Used

- Sqlite - Development Environment
- PostgreSql - Production Environment

5.5 Tools

- Jupyter Notebook
- VsCode
- Android Studio
- WebStorm

6. Software Workflow

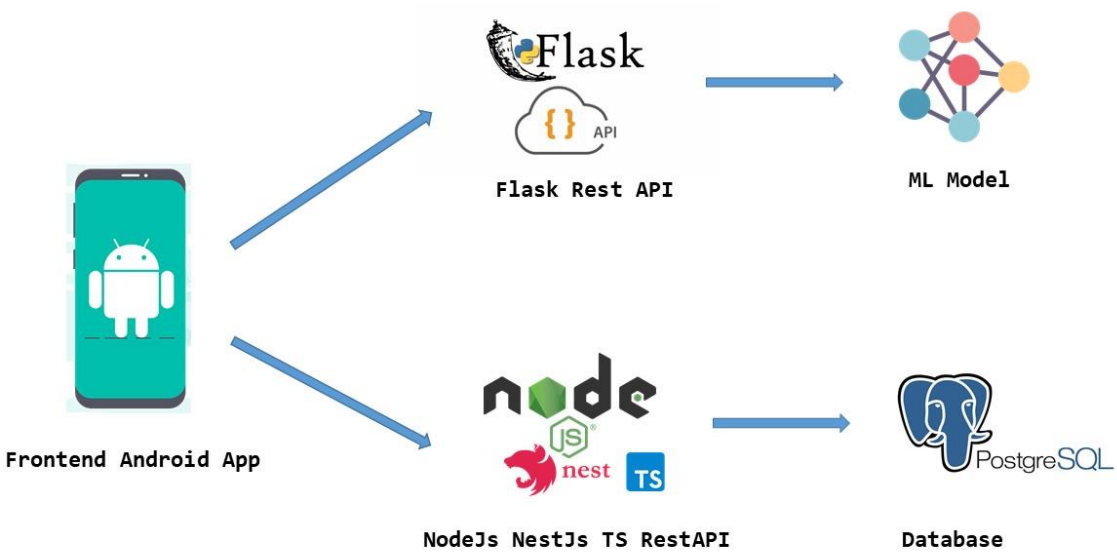


Fig 6.1 Software Workflow

- Android App Calls APIs of Python Flask Rest API for predicting the results
- Flask API internally uses our ML model to predict result
- For other Rest APIs Android App calls the NodeJs Backend
- Data on App can be added by accessing the admin panel of NodeJs Admin
- NodeJs Backend is connected to PostgreSQL Database

7. Working Project Description

7.1 Android App

7.1.1 Supported Languages

- To increase the accessibility of our app we have given the support for English as well as Hindi.

- During the first usage of the app or from clicking the globe icon we can launch the **“Choose Language”** screen and change the preferred language.

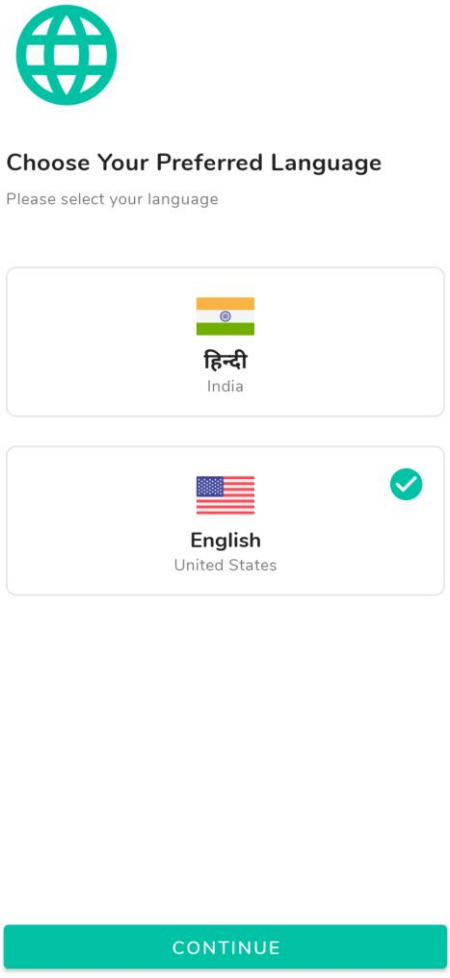


Fig 7.1.1 Change Language

7.1.2 Home Screen

- On the home screen, a brief is shown about the app and the motivation behind it.

- Click the **“TAKE THE TEST NOW”** button launches the form screen for heart disease prediction

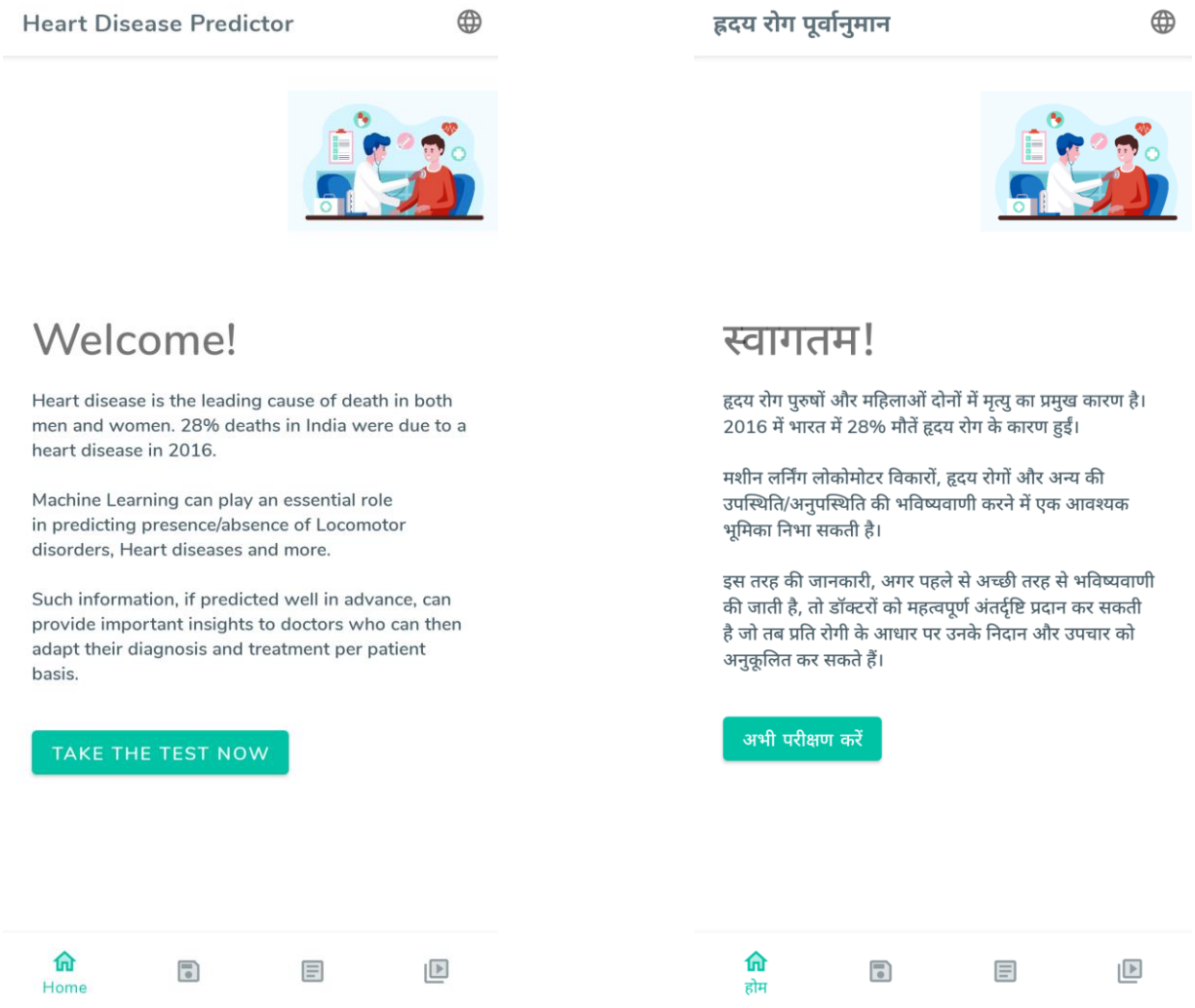


Fig 7.1.2 Home Screen

7.1.3 Prediction Test Screen

- Clicking the **“TAKE THE TEST NOW”** button launches the form screen for heart disease prediction
- After filling the various details of the person, we can click on the predict button to get the prediction result

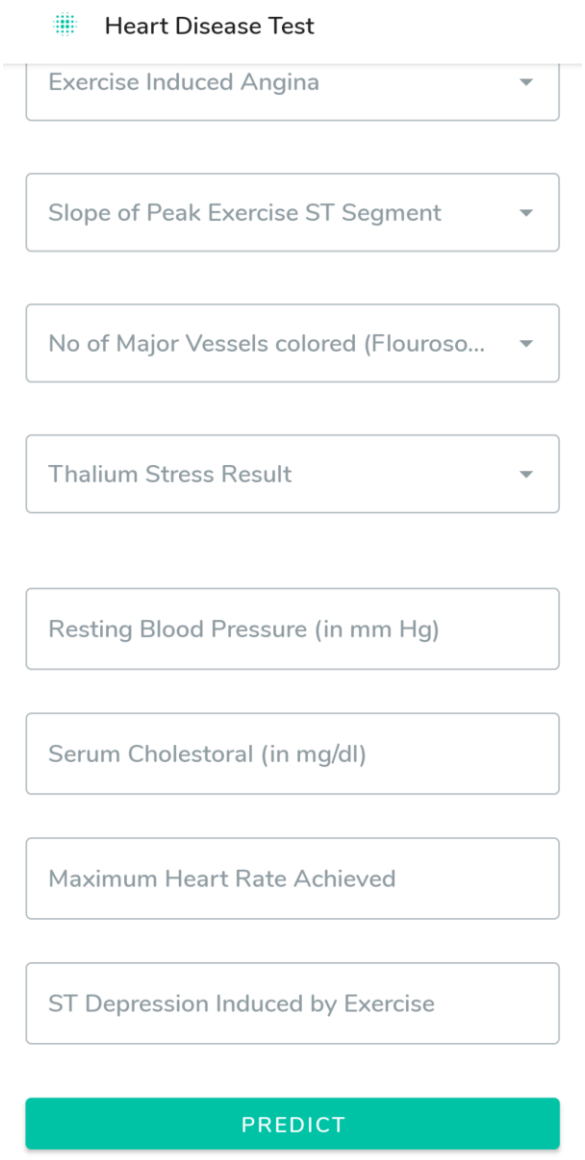


Fig 7.1.3 Home Screen

7.1.4 Prediction Response Screen

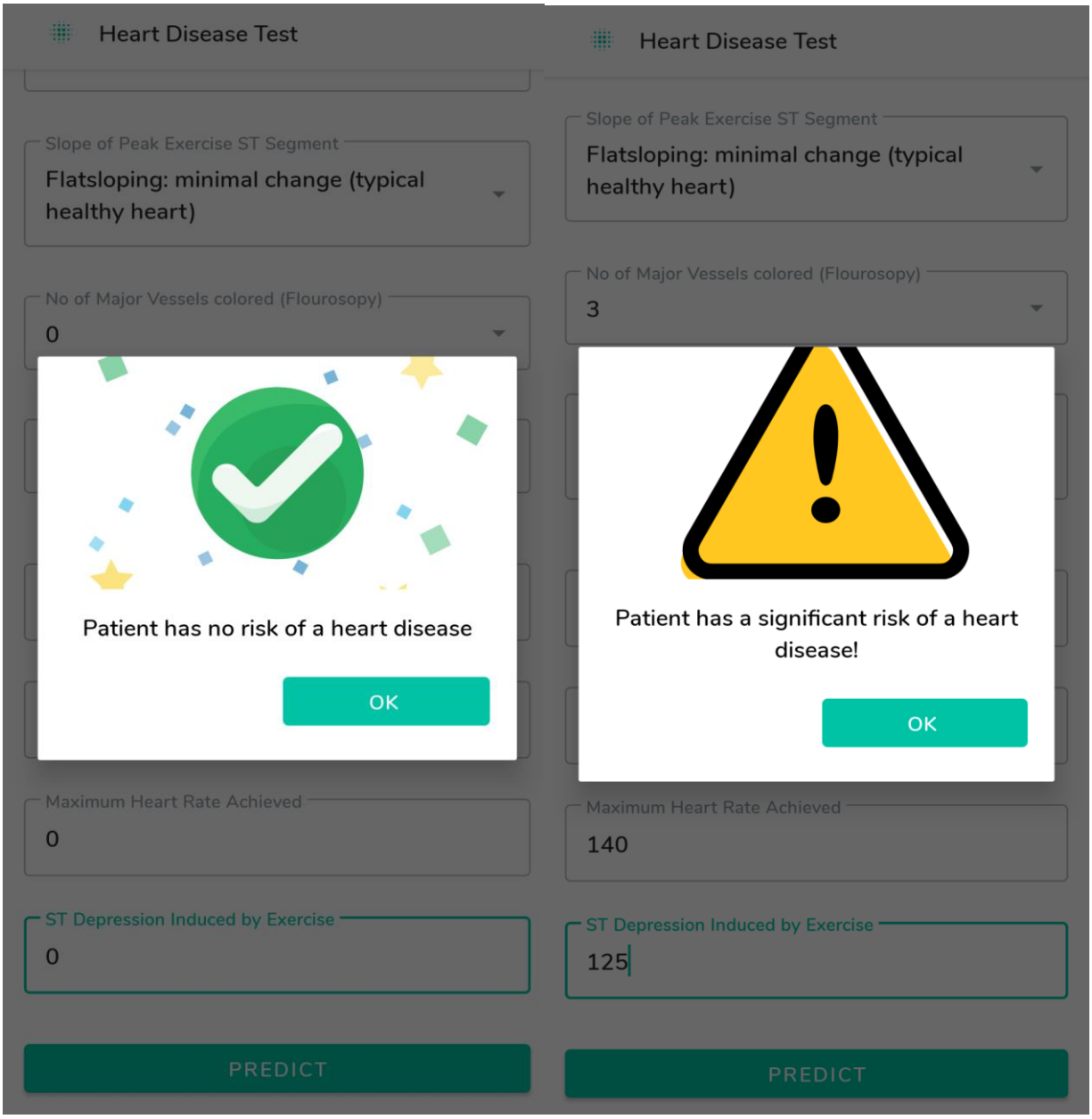


Fig 7.1.4 Prediction Response Screen

7.1.5 Report History Screen

We can see the report of previously generated reports by going to the reports section by clicking on the reports icon on the bottom navigation bar.

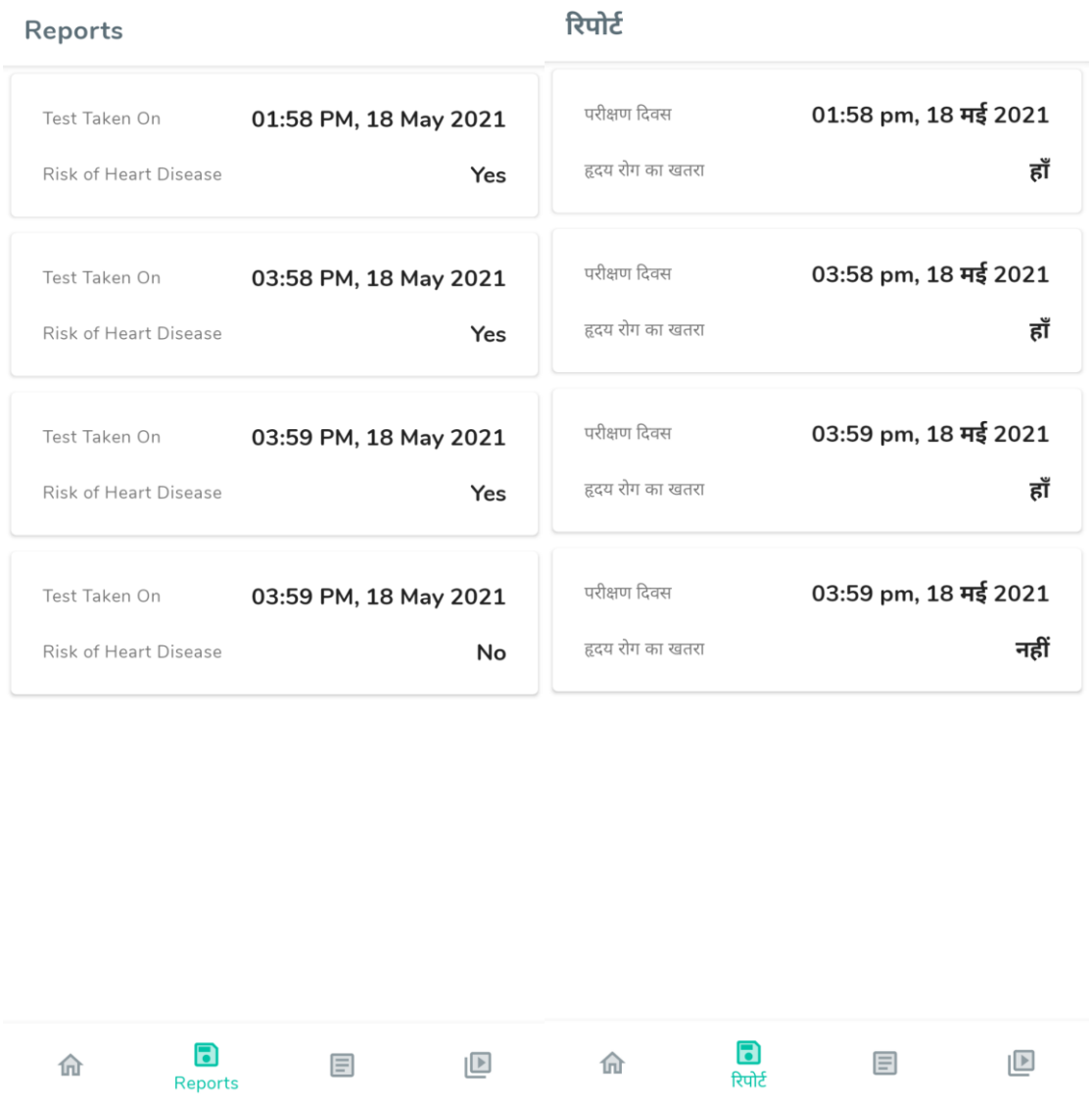


Fig 7.1.5 Report History Screen

7.1.6 Helpful Articles

If the admin wants to show users some important information / blog / article like symptoms of Heart Disease, it can be added from the admin panel and is shown on this screen.

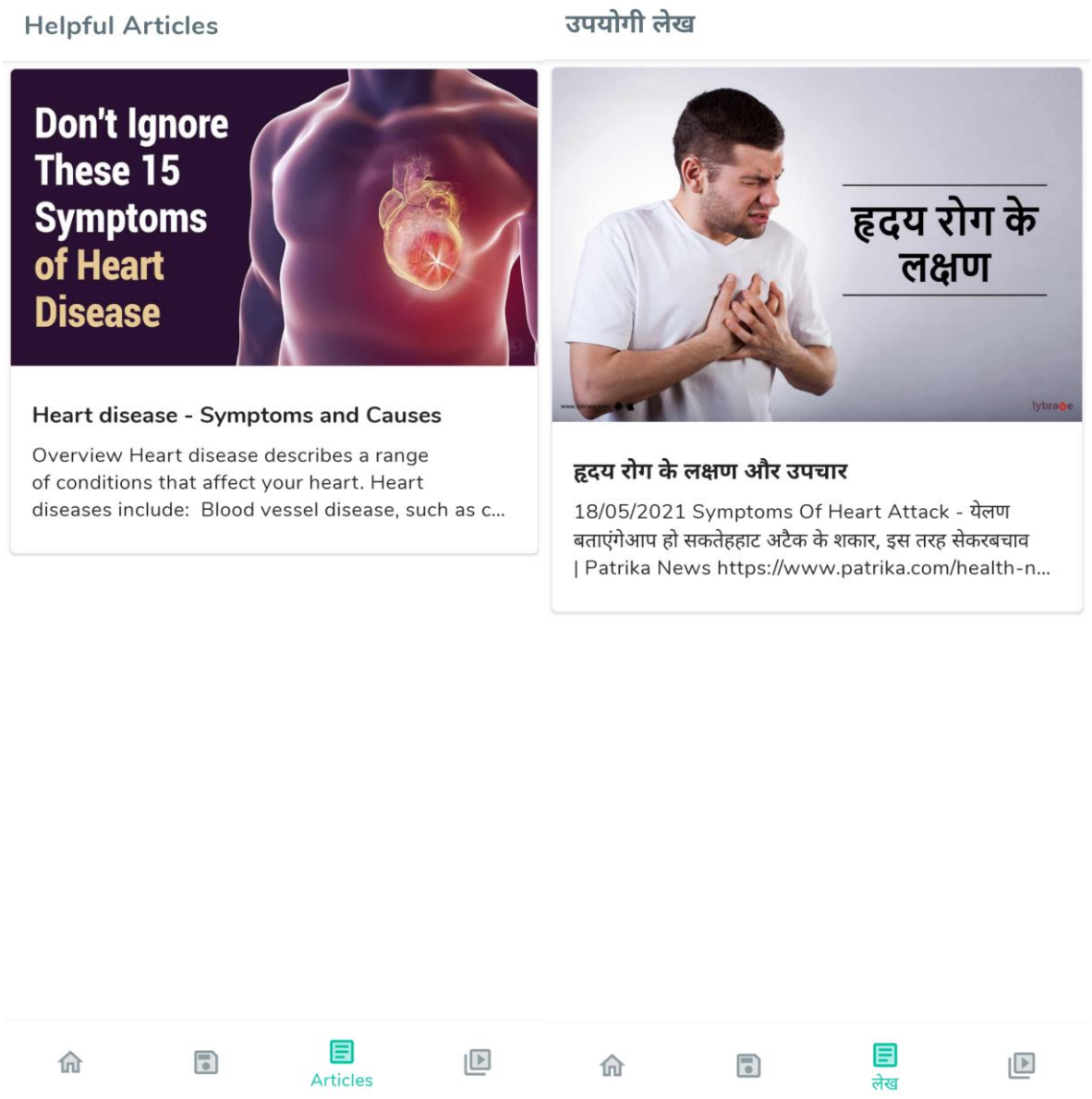


Fig 7.1.6 Helpful Articles Screen

7.1.7 Helpful Videos

If the admin wants to show some YouTube video like symptoms of Heart Disease, It can be added from the admin panel and is shown on this screen.

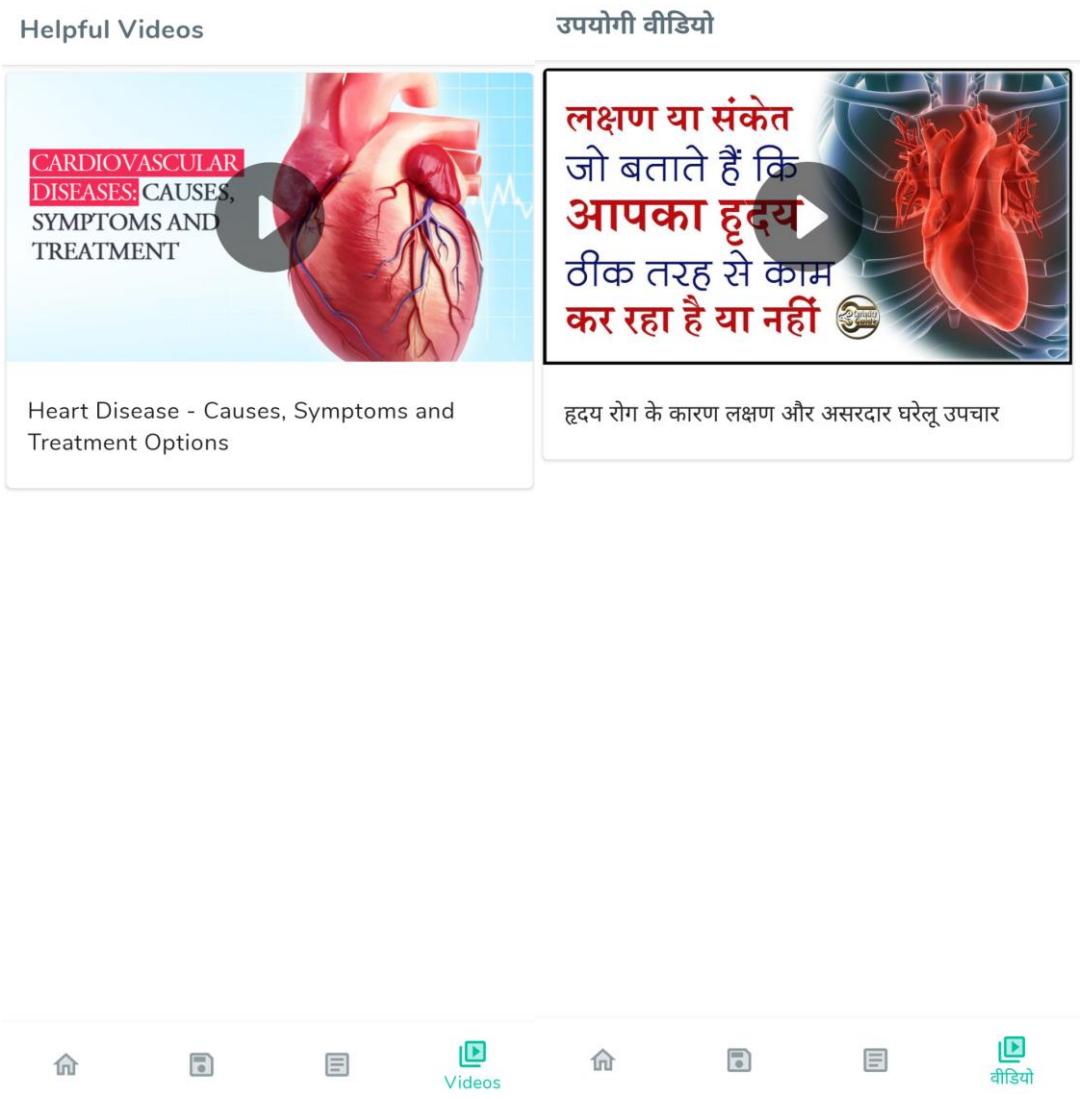


Fig 7.1.7 Helpful Videos Screen

7.1.8 Article Detail Screen

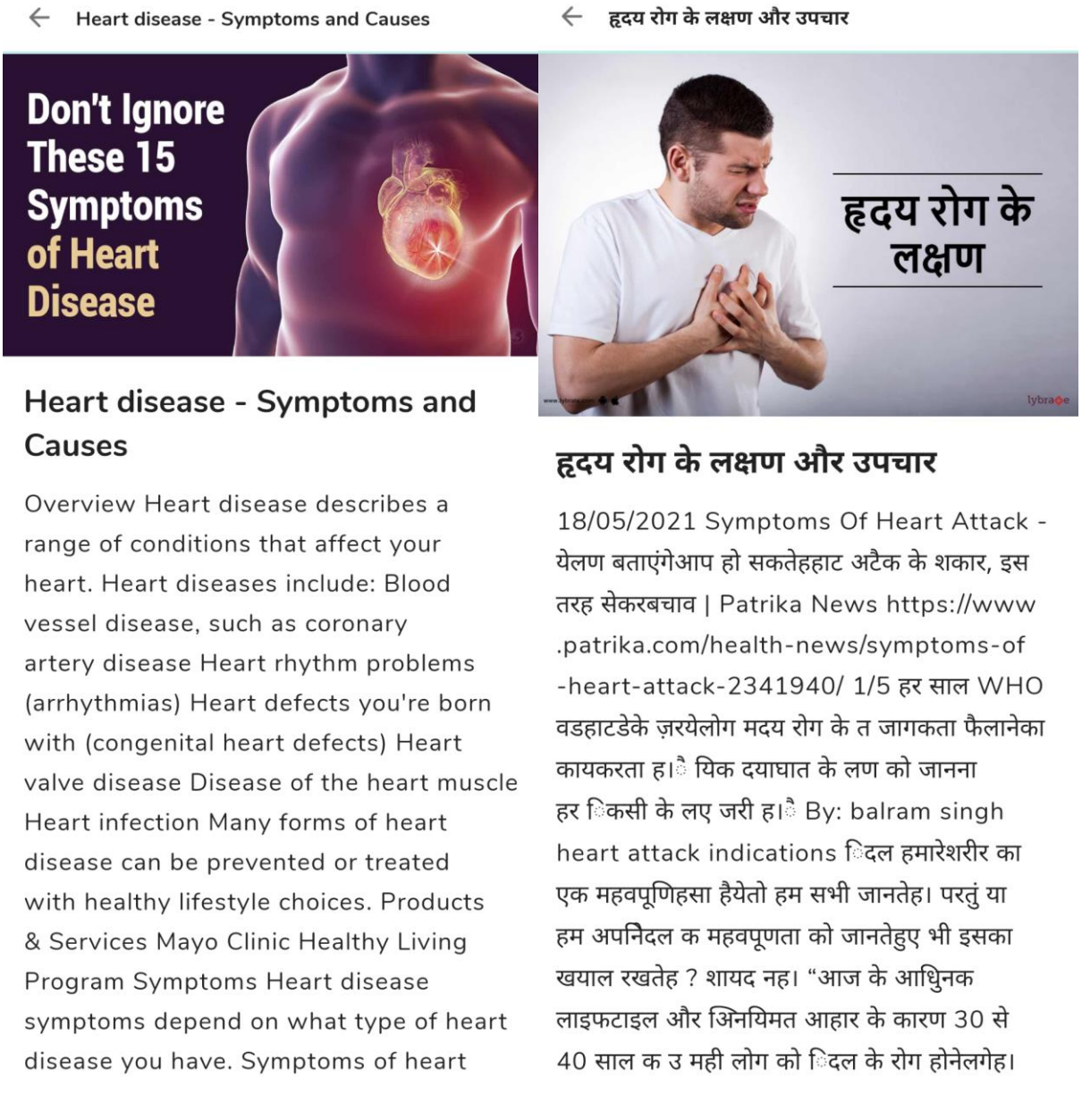


Fig 7.1.8 Article Detail Screen

7.1.9 Video Detail Screen

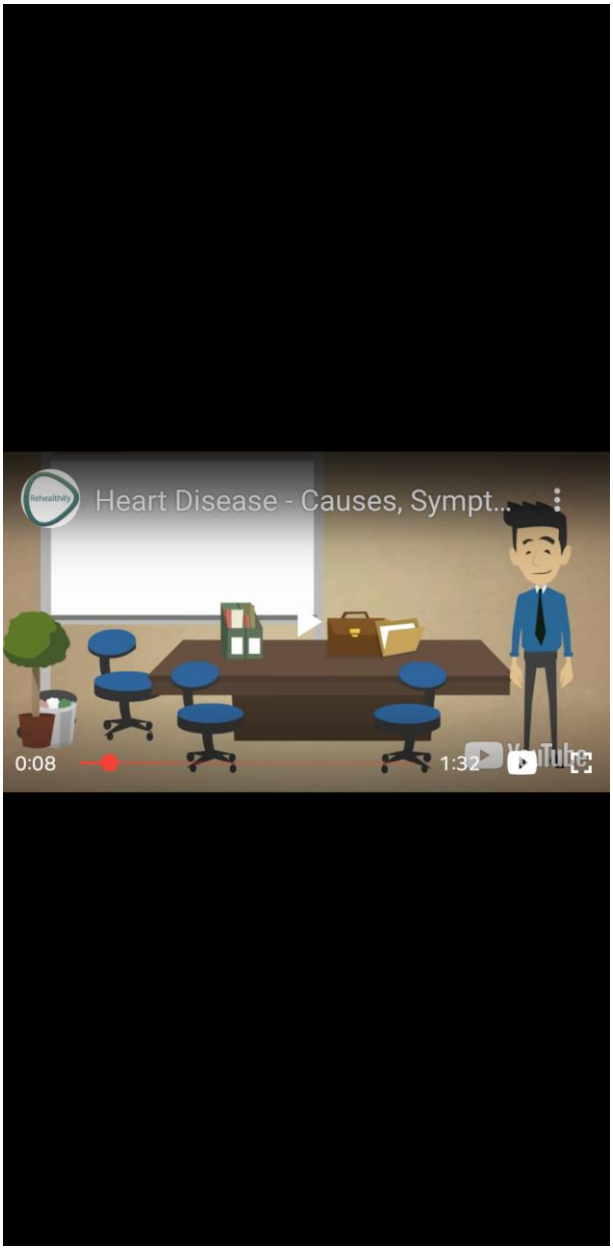


Fig 7.1.9 Video Detail Screen

7.2 Admin Panel

7.2.1 Login Screen

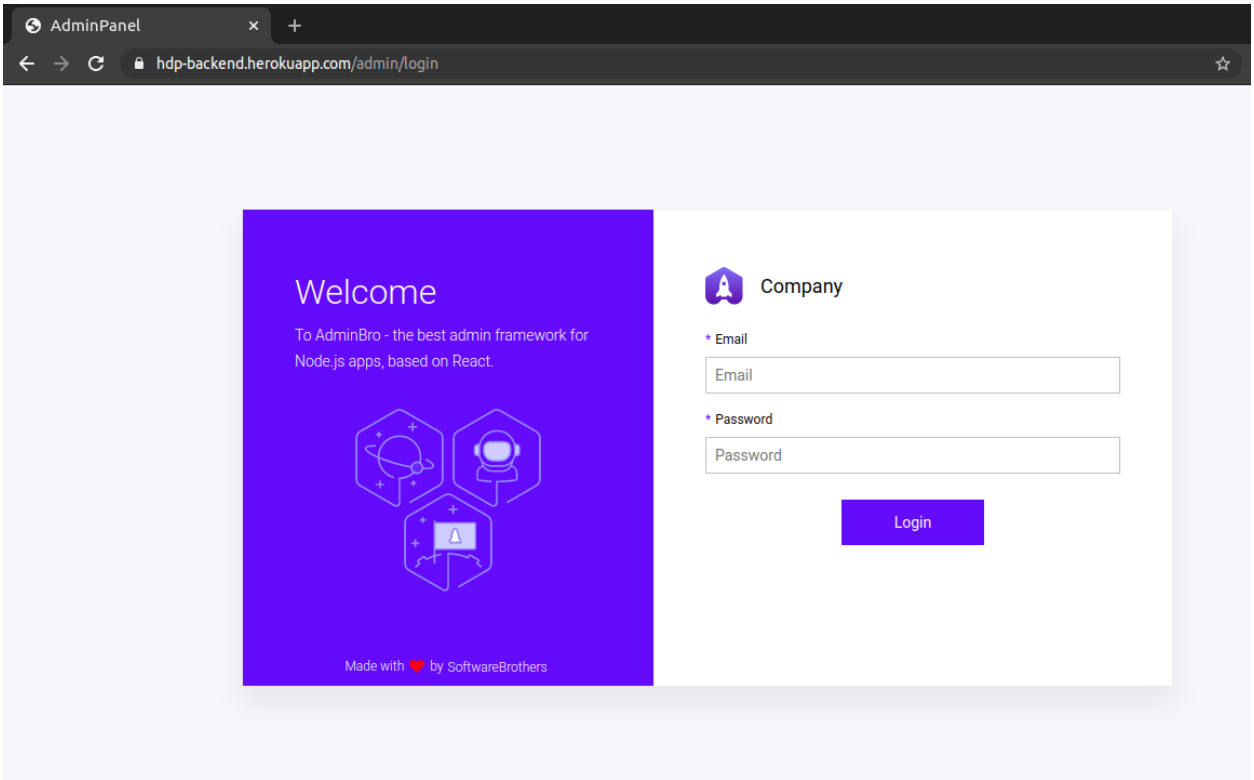


Fig 7.2.1 Admin Panel Login Screen

7.2.2 Add Article

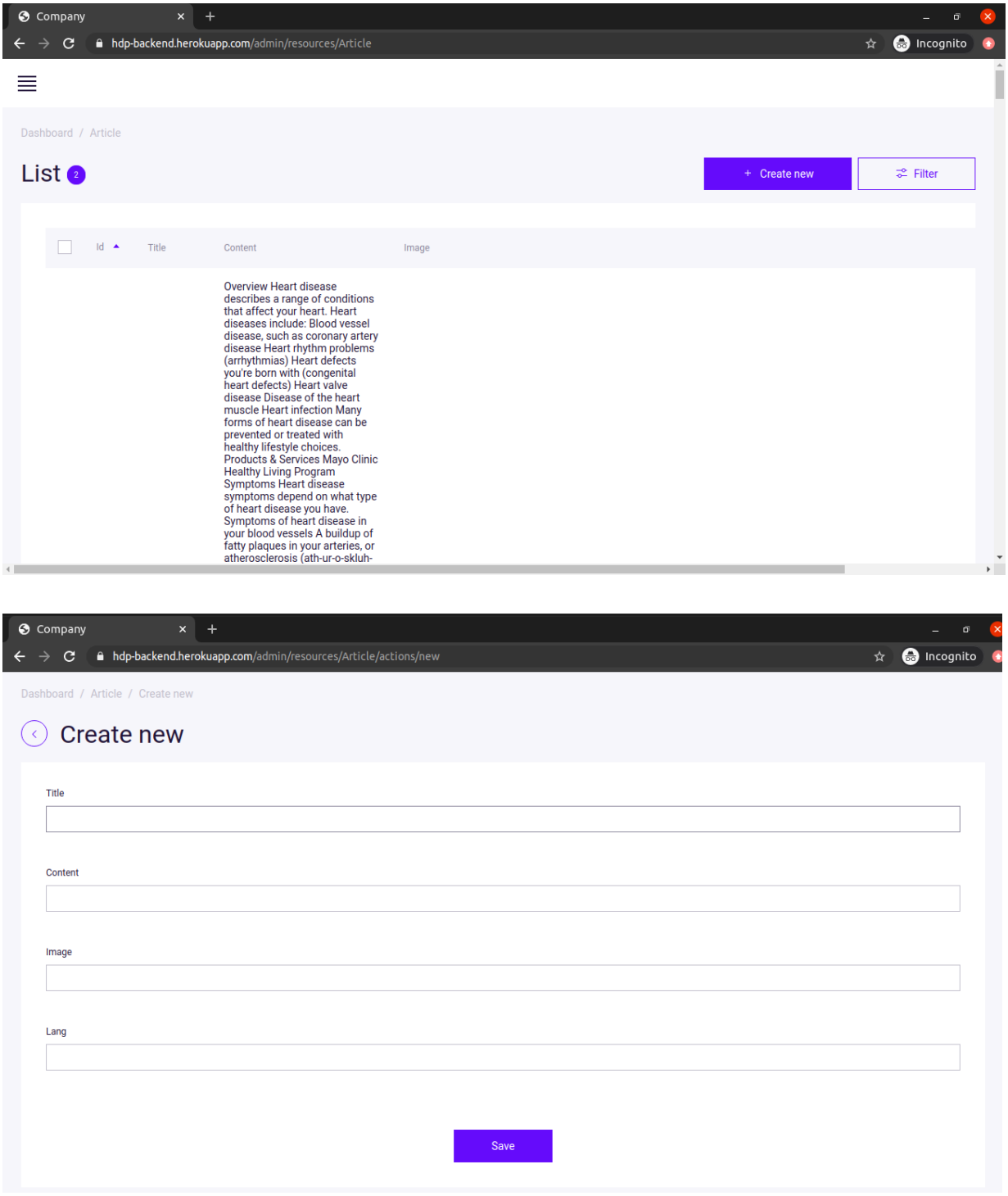


Fig 7.2.2 Admin Panel Article Screens

7.2.3 Add Video

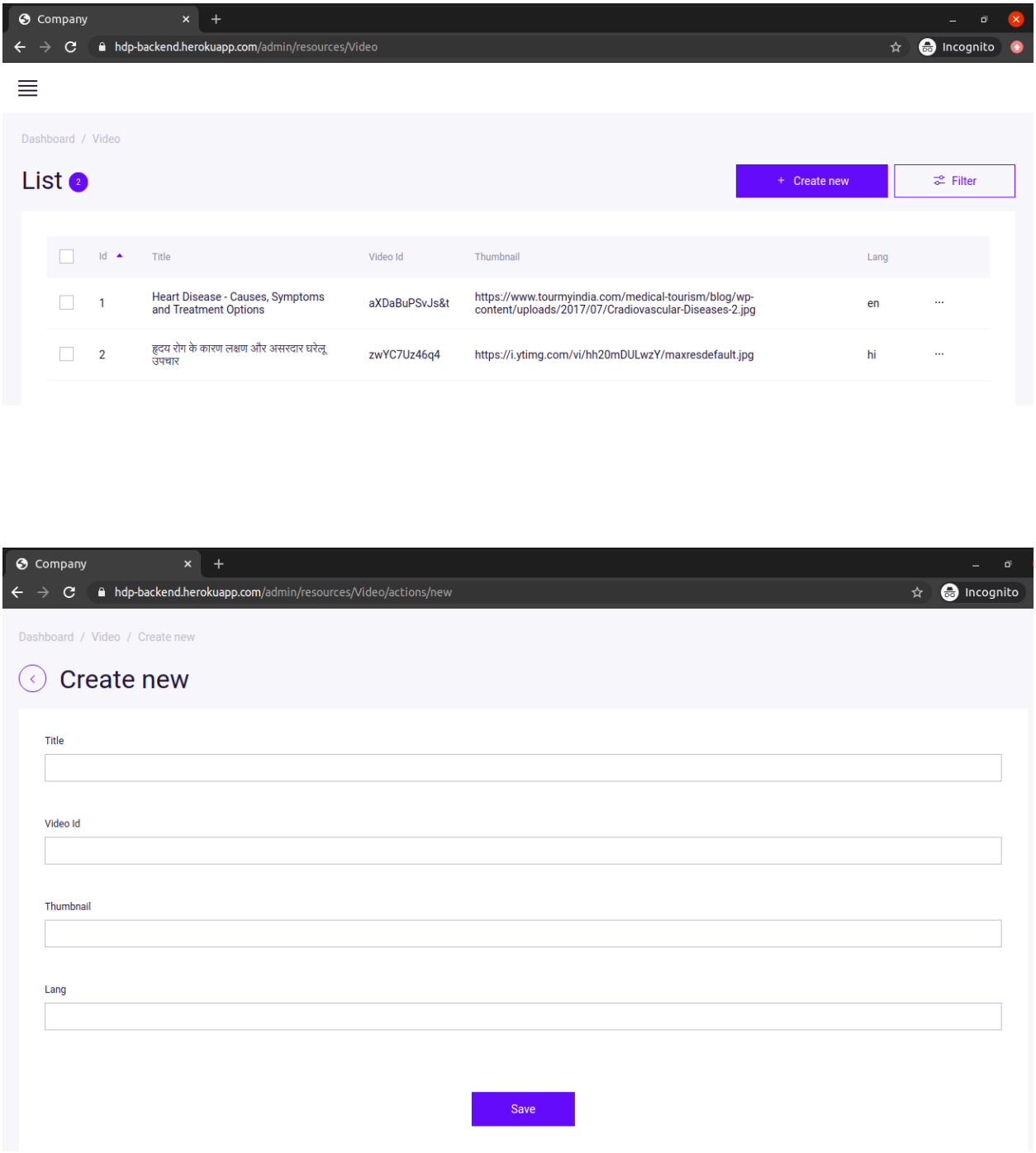


Fig 7.2.3 Admin Panel Video Screens

7.2.4 Swagger API Docs

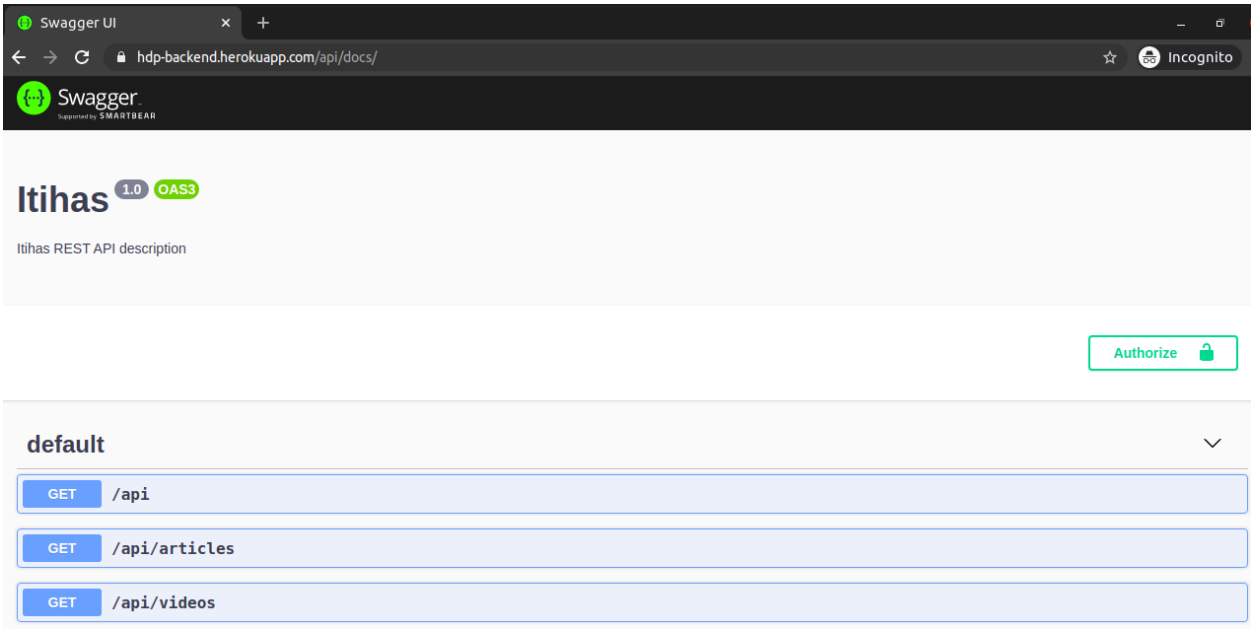
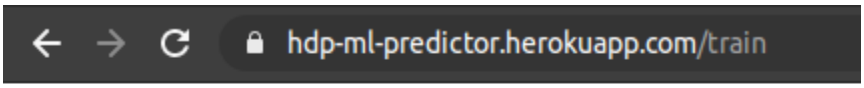


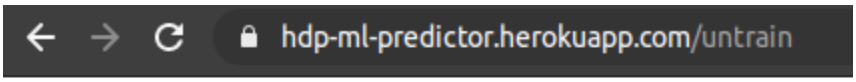
Fig 7.2.4 Swagger API Docs

7.3 ML Model

We added two APIs to train and untrain the model on the fly as well




```
{"message":"Model Trained"}
```




```
{"message":"Model Untrained"}
```

Fig 7.3 ML Model Train/Untrain APIs

8. Conclusion & Future Scope

8.1 Conclusion:

- We successfully delivered the Android App which predicts the risk of heart disease using predefined attributes of the patient which is predicted with the help of K Nearest Neighbour(KNN) Algorithm.
- Along with the prediction system system we also included the support for vernacular Hindi language as well to increase the accessibility of our App.
- We also have implemented the features of report history, helpful articles. helpful videos in our app
- We have also provided the support of Admin Panel for easy manipulation of data of our app.
- To train and untrain the ML Model on the fly, we have included to HTTP APIs as well

8.2 Future Scope:

- The accuracy of the ML Model can be improved by using a better dataset.
- The accuracy can also be improved by implementing a feedback mechanism by asking the user themselves about their results
- Support for other vernacular languages can be provided as well
- Support for emergency services like Ambulance can also be incorporated in future releases.

9. References

- Heart Disease Prediction using Machine Learning and Data Mining
  https://www.researchgate.net/publication/342210798_Heart_Disease_Prediction_using_Machine_Learning_and_Data_Mining
- Heart Disease UCI Database - https://archive.ics.uci.edu/ml/datasets/Heart+Disease
- Prediction of Heart Disease Using Machine Learning Algorithms
  https://www.researchgate.net/publication/326733163_Prediction_of_Heart_Disease_Using_Machine_Learning_Algorithms
- NestJs - https://docs.nestjs.com/
- Flask - https://flask.palletsprojects.com/en/2.0.x/