Heart Disease Detection by Machine Learning: using Hyperparameter Tuning to Increase Accuracy and the Deployment of Custom Models Online

Akshat Kishore

Student

The Shri Ram School Aravali

Gurgaon, India

Abstract— **In this paper, I discuss a method for detecting cardiac illness utilising artificial intelligence and machine learning algorithms, and making these systems publicly available. We demonstrate how artificial intelligence may be used to forecast if someone would get cardiac disease. A python-based application is created for healthcare research in this study since it is more dependable and aids in tracking and establishing various kinds of health monitoring apps. We demonstrate categorical variable manipulation and the conversion of categorical columns in data processing. We tested a range of machine learning models to achieve the goal of the research and compared the accuracy of each of these models to find the most accurate. We outline the key stages of application development, including the gathering of databases, applying logistic regression, parameter tuning, assessing the features of the dataset, deploying the model and connecting it to the front-end using APIs.**

Keywords— Artificial Intelligence, Machine Learning, Healthcare, AI in Healthcare

I. INTRODUCTION

In England, CVD (Cardio Vascular Disease) causes close to 34% of all fatalities, compared to 40% in the European Union. As risk factors for CVD become more common in formerly low-risk nations, the rate of CVD is expected to climb globally. At the current rate of 80%, cardiovascular disease (CVD) will surpass infectious illness as the leading cause of death in the majority of developing countries by 2020 [1]. Not only is CVD a major cause of death, but it also ranks first in the world for years of life lost due to impairment. The graph below analyses leading death causes globally. [2]

Over 75% of premature CVD, according to the World Health Organization (WHO), is avoidable, and lowering risk factors can help lessen the rising burden of CVD on patients and healthcare providers [3]. Although age is an established risk factor for the development of CVD, postmortem data shows that the process of acquiring CVD in later years is not unavoidable. As a result, risk mitigation is essential.

These illnesses are generally associated with blood vessel blockages or narrowing, which can lead to heart attacks, strokes, and angina. Heart disorders come in a variety of forms, including those that damage the heart's rhythm, valve, or muscle. Machine learning is greatly useful for evaluating whether anybody has had cardiac disease. In either event, if they are anticipated in advance, physicians would find it much simpler to gather vital information for diagnosing and treating patients.

II. THE PROBLEM

Numerous studies and investigations have been conducted on reducing the risk of heart disease. Based on blood pressure, smoking habits, cholesterol and blood pressure levels,
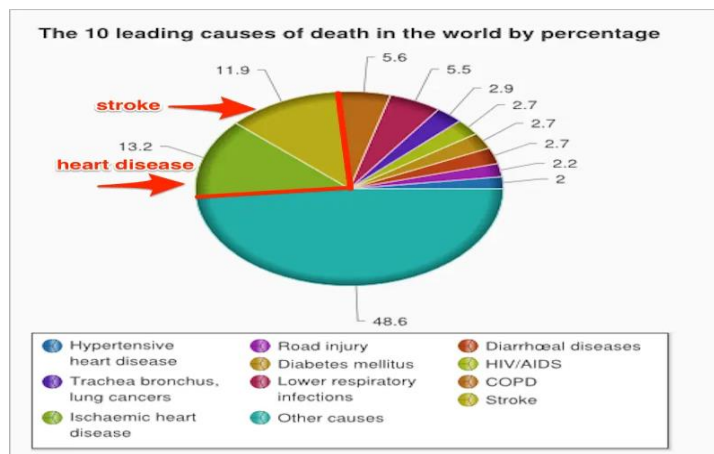

Figure 1: Graph showing global death causes

diabetes, and data from population studies, it is now possible to forecast the development of heart disease [4]. These prediction algorithms have been modified by researchers into simpler score sheets that patients may use to determine their risk of developing heart disease. A common risk prediction criterion used in algorithms for heart disease prediction is the Framingham Risk Score (FRS). The goal of this work was to create a classification algorithm-based intelligent system for heart disease prediction based on risk factor categories.

III.   ETHICS AND FEATURES

We have aimed to incorporate Microsoft's Responsible AI and Ethics Code in our models as described below: [5]

1. Inclusivity- We have made sure that NoWa is available on both iOS and Android platforms, available to be downloaded by almost any user across the globe.
2. Fairness- Our product aims to provide an easy-to-use platform for classification of wastes seen in daily life, giving people access to a tool to help easier segregation of wastes and inversely reduce the waste which ends up in landfills.
3. Privacy- both versions of the app ensure full privacy of the user by using the native permissions feature on both Android and iOS. That is, the app asks for permission whenever camera access is required. The user can select from a range of standard options including "while using app" and "only for once".
4. Reliability and Safety- our machine learning models have been trained on a big dataset and thus has a high accuracy. Thus, the model predicts correct output almost all of the time it runs.

IV.   RESEARCH AIMS

The aims of the research are as follows:

1) To evaluate the methods for detecting heart disease using Artificial intelligence in Python.

2) To critically analyse the earlier actions and use a proper methodological strategy for superscribing the found issue

3) To critically utilise Python language data interpretation techniques for identifying health issues.

4) To evaluate the artefact or product critically utilising cybersecurity strategies, employing the right techniques, and determining the work's weaknesses and strengths.

V.   THE PYTHON MODEL

A. Data

We used publicly available heart disease data ay UCI, obtained through the following creators:

1. Hungarian Institute of Cardiology. Budapest: Andras Janosi,   M.D.
2. University Hospital, Zurich, Switzerland: William Steinbrunn,   M.D.
3. University Hospital, Basel, Switzerland: Matthias Pfisterer, M.D.
4. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation Robert Detrano, M.D., Ph.D. [6]

This data contains 13 parameters which are frequently used to determine the presence of heart disease.
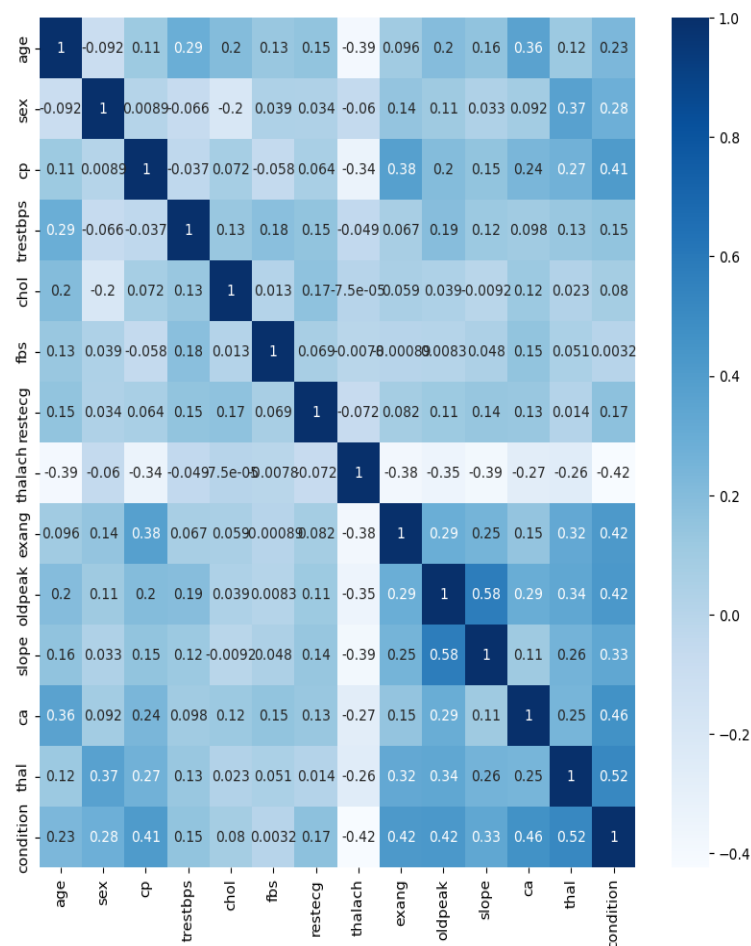

Figure 2: Correlation plot

We performed data exploration to analyse the dependence of variables with high correspondence to the presence of the disease condition. This was done using Matplotlib in Python to visualise the obtained data and its correlation with the presence of a heart disease.
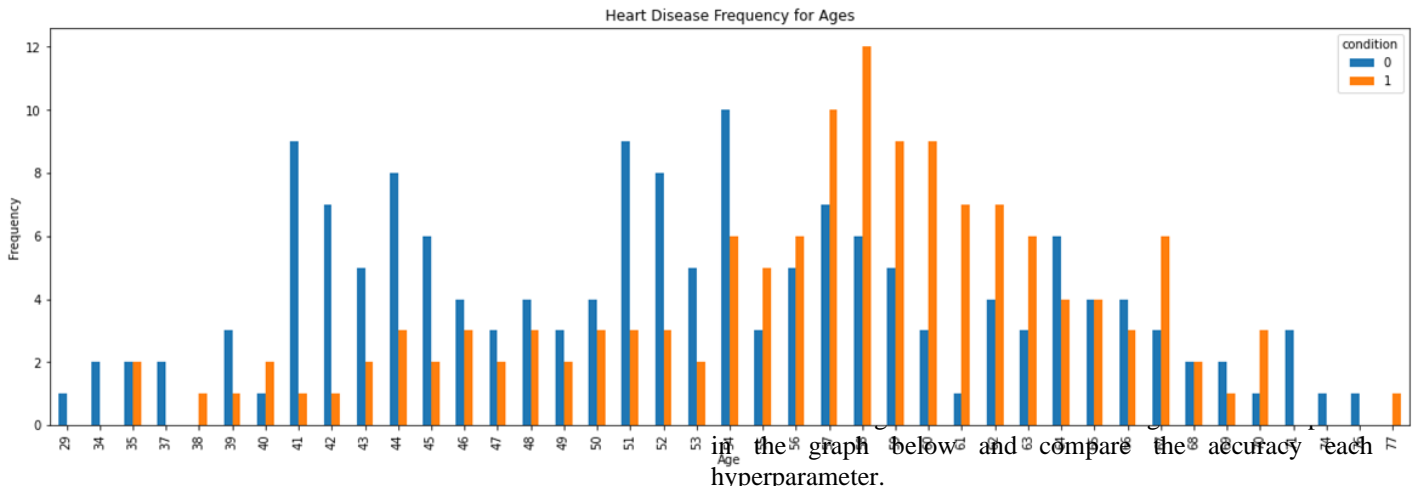
Figure 3: Analysis of heart disease presence based on age

B. The models

As a part of model creation, we tested 7 of the most common models used for use in such classification problems, as summarised in table 1. A summary of the models tested and their basic metrics can be seen in the table below.

The Grid Search Model gives the maximum accuracy of all models due to its hyperparameter tuning algorithm described in the next section.

For example, say there is a model which has 2 hyperparameters, each having exactly 3 possible values. The Grid Search Algorithm will iterate through each of the points in the graph below and compare the accuracy each hyperparameter.

Model type	Precision	Recall	F1-Score
KNN	0.635	0.595	0.595
SVC	0.745	0.725	0.735
GridSearch on SVC	**0.85**	**0.85**	**0.85**
RandomForest	0.82	0.835	0.825
DecisionTree	0.745	0.755	0.74
GaussianNB	0.82	0.835	0.825
Logistic Regression	0.835	0.85	0.84

C. Grid Search: How it Works

Grid Search is a method through which Hyperparameters can be chosen for any particular model. We chose the SVM model considering that it was found to give maximum accuracy to hyper parameter tuning.

Hyperparameters are simply variables that the programmer specifies while building a Machine Learning Model. Thus, tuning these parameters can improve the accuracy of the model. the combination with the best performance, and takes that combination as its starting point.
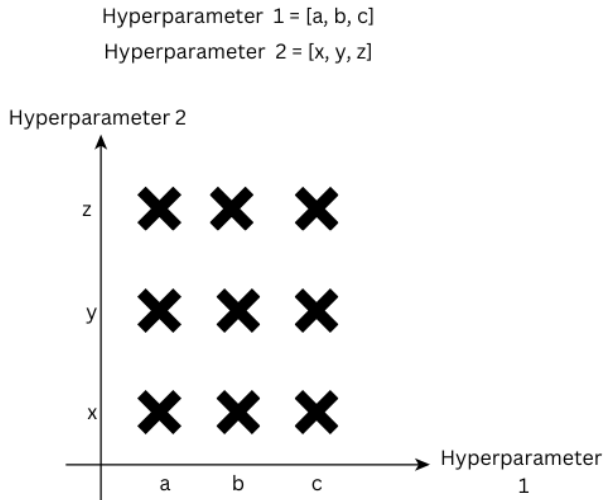
VI. DEPLOYMENT

Hyperparameter 1 = [a, b, c]
Hyperparameter 2 = [x, y, z]

Figure 4: Working of Grid Search

Figure 5: Complete analysis of the workflow of creating and deploying the model

Therefore, the grid search algorithm iterates through the following number of cases:

$$N = n_1 . n_2 . n_3 \ldots n_n$$

Where n_1 is the number of possibilities in the first hyperparameter, n_2 is the number of possibilities in the second hyperparameter and so on till n hyperparameters.
Grid search can also be represented mathematically as a cartesian product of the sets of hyperparameters. For example, consider two hyperparameters as below:

$$HP_1 = [a, b, c]$$
$$HP_2 = [x, y, z]$$

$$N = [(a,x),(a,y),(a,z),(b,x),(b,y),(b,z),(c,x),(c,y),(c,z)]$$

Where N is the set of all iterations that the grid search will perform. Such as set is called a Cartesian product of HP_1 and HP_2.

Now, as we can see above, grid search checks for each of the parameter's combinations possible. On the other hand, an algorithm like random search would look through a random set of these parameter combinations for better computation times.

Therefore, GridSearch results in maximum accuracy but can be slightly slow to compute. However, for out of 7000 patients, the GridSearch code ran in less than second. [8]
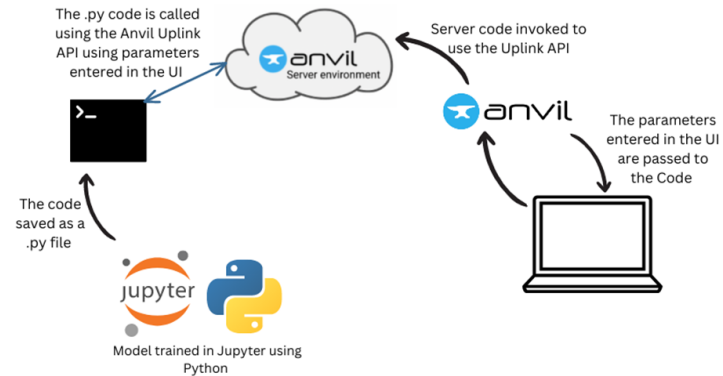
VII. RESULTS

The website has been published on the internet. It has a list of parameters which the user must enter and click on the "Calculate Possibilities" button.

Following this, the server code is called and the entered parameter are indirectly passed to the model.predict() function which gives the result as 0 or 1.

These results are then simply converted to String outputs and displayed to the user on the website in less than 3 seconds.

VIII. CONCLUSION

The usage of artificial intelligence can be used to predict heart diseases and can have the following advantages:
1. Prevention of early risks such as embolic strokes.
2. Patients can immediately contact doctors to confirm any diagnoses.
3. General purpose doctors who are not specialised in cardiovascular disease can use this tool to help in diagnosis.
4. Doctors in rural areas with minimal medical study can use the tool as an aid.

REFERENCES

[1] Jack Stewart, Gavin Manmathan and Peter Wilkinson "Primary prevention of cardiovascular disease: A review of contemporary guidance and literature," National Library of Medicine

[2] Lauren F Friedman "Just 2 things cause a quarter of all deaths in the world," Business Insider

[3] "Cardiovascular diseases (CVDs)," WHO

[4] Zaibunnisa L. H. Malik, Momin Fatema, Nikam Pooja and Gawandar Ankita "Heart disease prediction using artificial intelligence," IJERT

[5] "Heart disease dataset," UCI Machine Learning Repository

[6] Petro Liashchynskyi and Pavlo Liashchynskyi "Grid search, random search, genetic algorithm: a big comparison for NAS," ArXiv.org

[7] Tong Yu and Hong Zhu "Hyper-Parameter optimization: A review of algorithms and applications," ArXiv.org