Effective Heart Disease Prediction Using IBM Auto AI Service

1. INTRODUCTION

1.1 Overview: Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide. As per the World Health Organization (WHO) report, 17.9 million deaths occurred from cardiovascular diseases (CVDs) in 2019, representing 32% of all global deaths and having an annual mortality rate of greater than 17.7 million.
Heart failure is a common event caused by CVDs and Here the dataset contains 9 features that can be used to predict mortality by heart failure.
In this project, We built a model using Auto AI and build a web application where we can showcase the prediction of heart failure.

1.2 Purpose: After completing this project, following objectives can be achieved:

- Work with Watson Studio
- Create a project in Watson Studio
- Use Auto Ai experiment to create a model
- Deploy the ML model as a webserver
- Integrating Model and Node-RED Service
- Build an Application using Node-RED which takes inputs from the user and showcases the prediction on UI

2. LITERATURE SURVEY

2.1 Existing problem: The primary method used by the physicians was the auscultation method for distinguishing between normal and abnormal cardiac sounds. Every heart disease was identified by the physicians listening to these sounds of the heart using stethoscopes. The auscultation technique used by professional doctors to diagnose a heart disease has some drawbacks. The clarification and classification of distinct sounds in the heart are associated with the abilities and practices of the doctors, which are gained after lengthy examinations.
Apart from the manual method, various machine learning methods have been proposed for CVD detection. Research was conducted to classify the most relevant attributes of heart disease prediction [1].
[2] covered many aspects of cardiac illness, as well as a model based on supervised

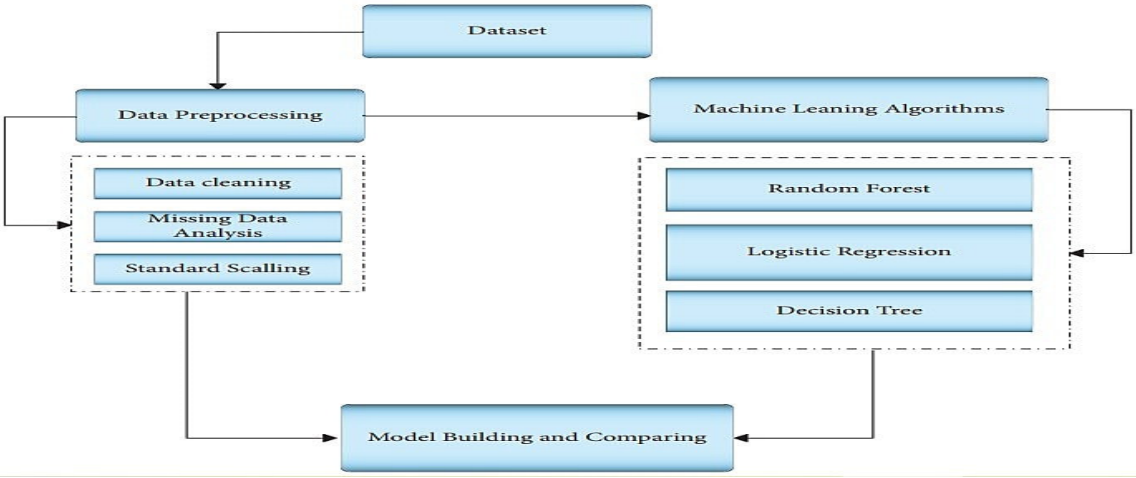
learning techniques such as Random Forest (RF), Decision Tree (DT), and Logistic Regression (LR). It makes use of an existing dataset from the UCI Cleveland database of heart disease patients.

2.2 Proposed solution: Manual approaches for the identification of heart disease are biased and prone to interexaminer variability. In this regard, machine learning algorithms are efficient and reliable sources to detect and categorize persons suffering from heart disease and those who are healthy. The purpose of this project is to forecast the likelihood of individuals getting heart disease.
Machine learning and artificial intelligence based automatic prediction model, which is built by IBM services for heart failure rate prediction where the dataset is trained and a model is built. The auto AI instance is created in the IBM Watson Studio and machine learning services are linked with it.
The NodeRED service is used to deploy the model as a final application. The accuracy along with precision and recall measures and metrics were chosen automatically by the system as best ones.

3 THEORITICAL ANALYSIS

The proposed study indicated heart disease by examining the classification methods and carrying out performance analysis. The goal of this project is to accurately predict whether or not a patient has heart disease. The input values from the patient's health report are entered by the health professional. The data are incorporated into a model that forecasts the chance of developing heart disease.

3.1 Block diagram:



3.2 Hardware / Software designing:

In this activity, we will be creating the Necessary IBM services. The following are the services that we have to create.

- Watson studio
- Node-RED
- Cloud Object Storage service (COS)
- IBM Watson Machine Learning service (ML)

COS and ML services will be created while creating a Watson Studio Project.

4. EXPERIMENTAL INVESTIGATION

In this activity, we learnt about saving a pipeline as a Watson Machine Learning model, deploy the model, and score it to view a prediction.

This Activity contains the Following Tasks:

- Collect the data set
- Create Watson Studio project
- Add Auto AI experiment
- Run AI Experiment
- Save the model
- Deploy the model

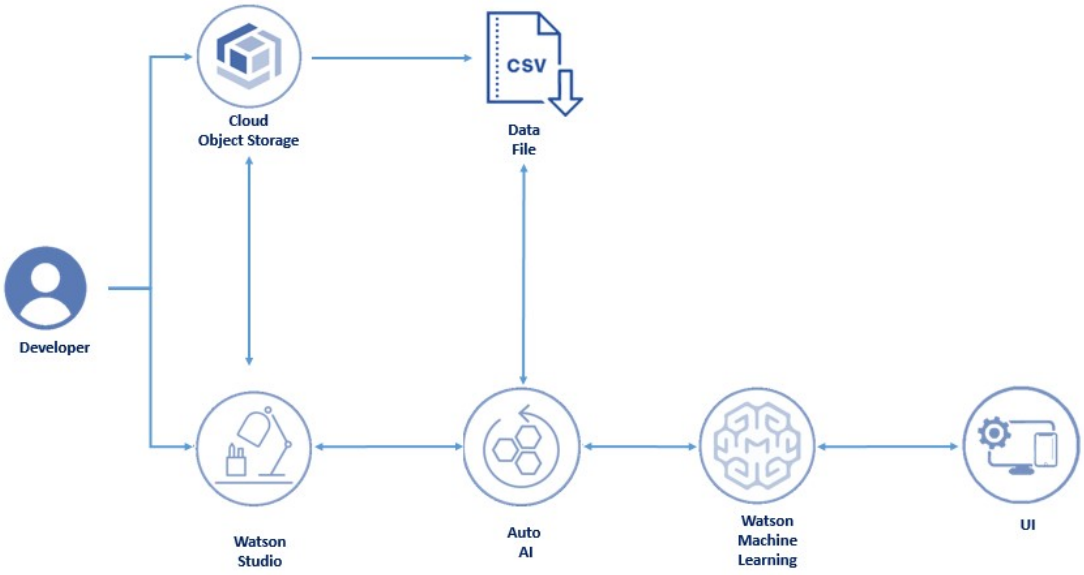
In this Activity, we have collected dataset and then we built a machine learning model that predicts heart failure based on the following parameters in above mentioned dataset.

- AVGHEART BEATS PER MIN
- PALPITATIONS PER DAY
- CHOLESTEROL
- BMI
- AGE
- SEX
- FAMILY HISTORY
- SMOKER LAST 5 YRS
- EXERCISE MIN PER WEEK

Build Node-RED Application: After sucessful deployment of ML mode, we built a User interface which takes inputs from the user. The Model Analyses the Inputs and returns

the Prediction that is showcased on the User interface.

5. FLOWCHART

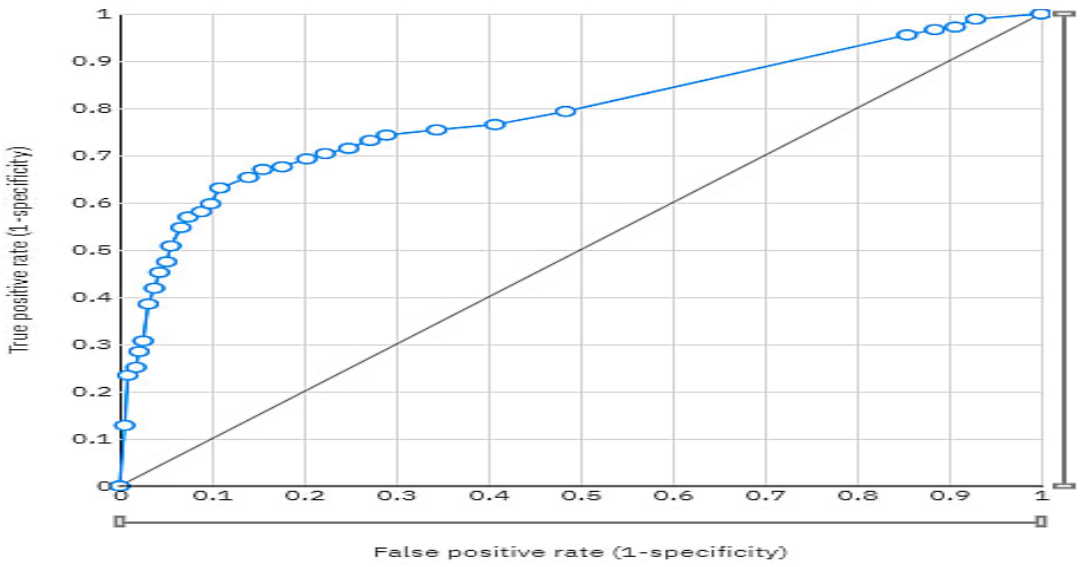


6. RESULT

The auto AI service determines the best algorithm as the Snap Random Forest Classifier algorithm for the given dataset here and automatically classifies it as a binary classification problem with values as Y/N for heart failure. Several algorithms can be chosen and deployed. In our experiment Snap Random Forest Classifier performs better than other ML algorithms. It gives an accurancy of 86.9%. This class implements a random forest classifier using the IBM Snap ML library. It can be used for binary and multi-class classification problems. Random Forest RF is a supervised machine learning algorithm that is widely used for both classification and regression problems. It creates decision trees on randomly selected training datasets, gets the prediction from the collection of trees, and casts a unit vote for the most popular class to classify an input vector.

Performance is evaluated based on following metrics:

1.ROCcurve:



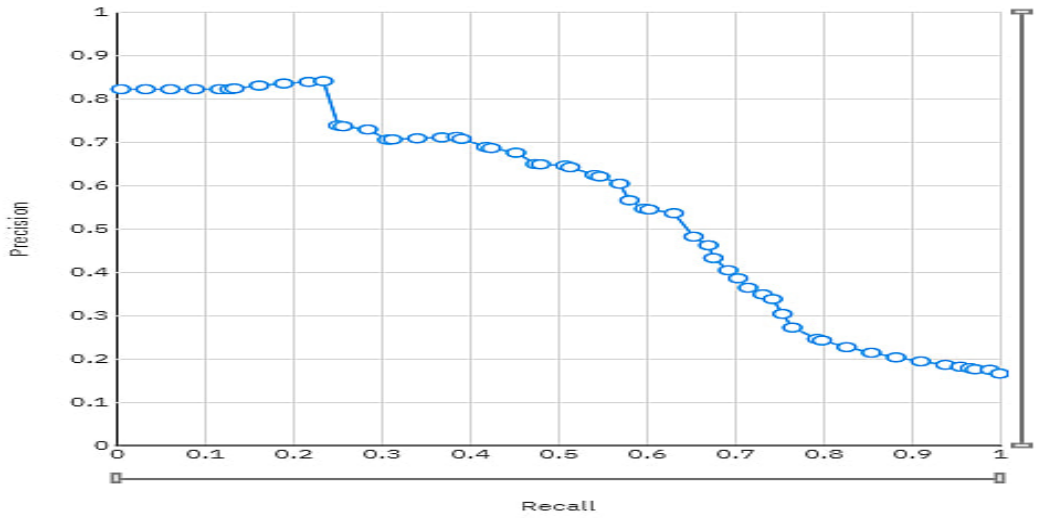
2. Confusion matrix:

Confusion matrix ⓘ

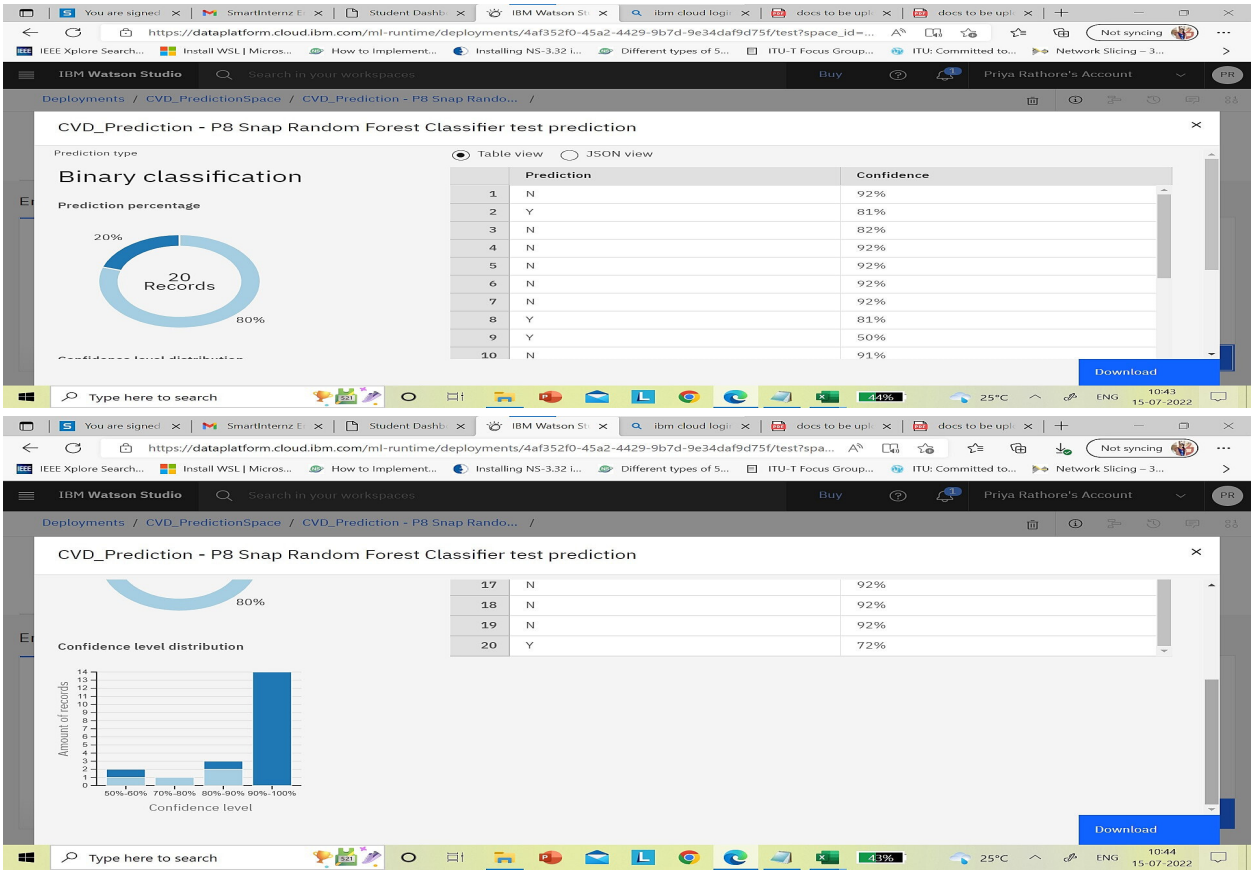
Observed	Predicted		
	Y	N	Percent correct
Y	63	116	35.2%
N	26	875	97.1%
Percent correct	70.8%	88.3%	86.9%

Less correct — More correct

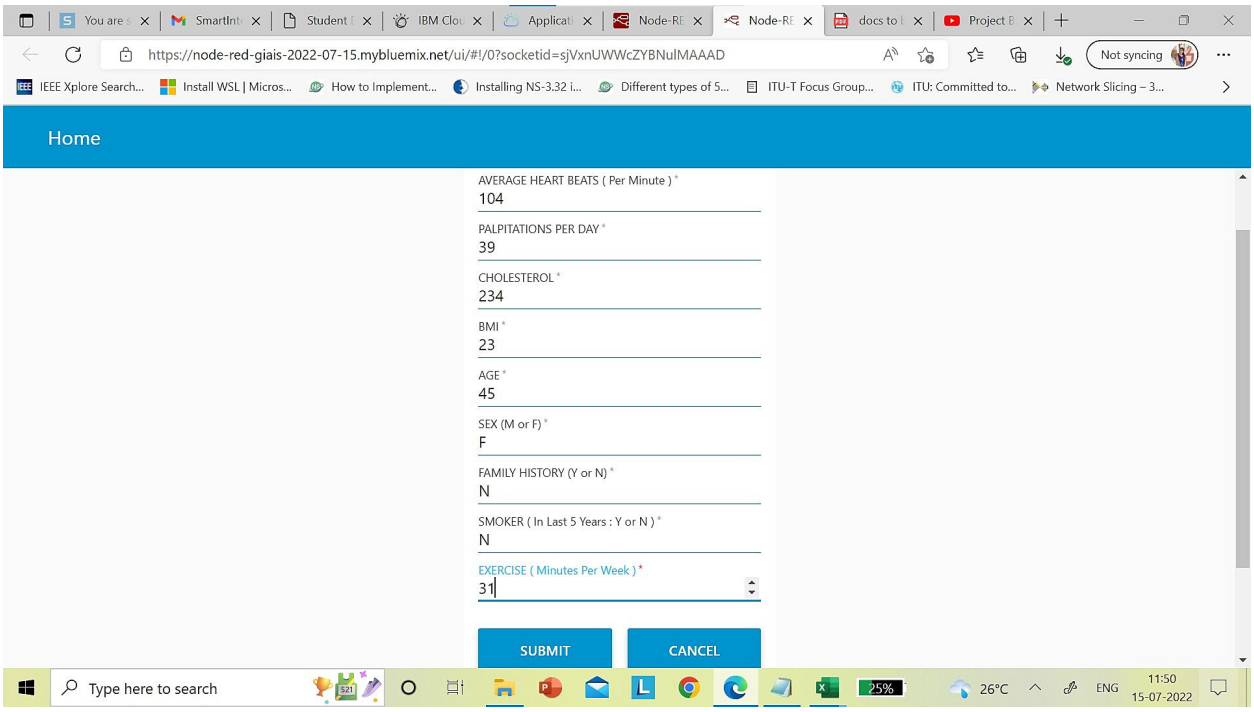
3. Precision recall Curve:

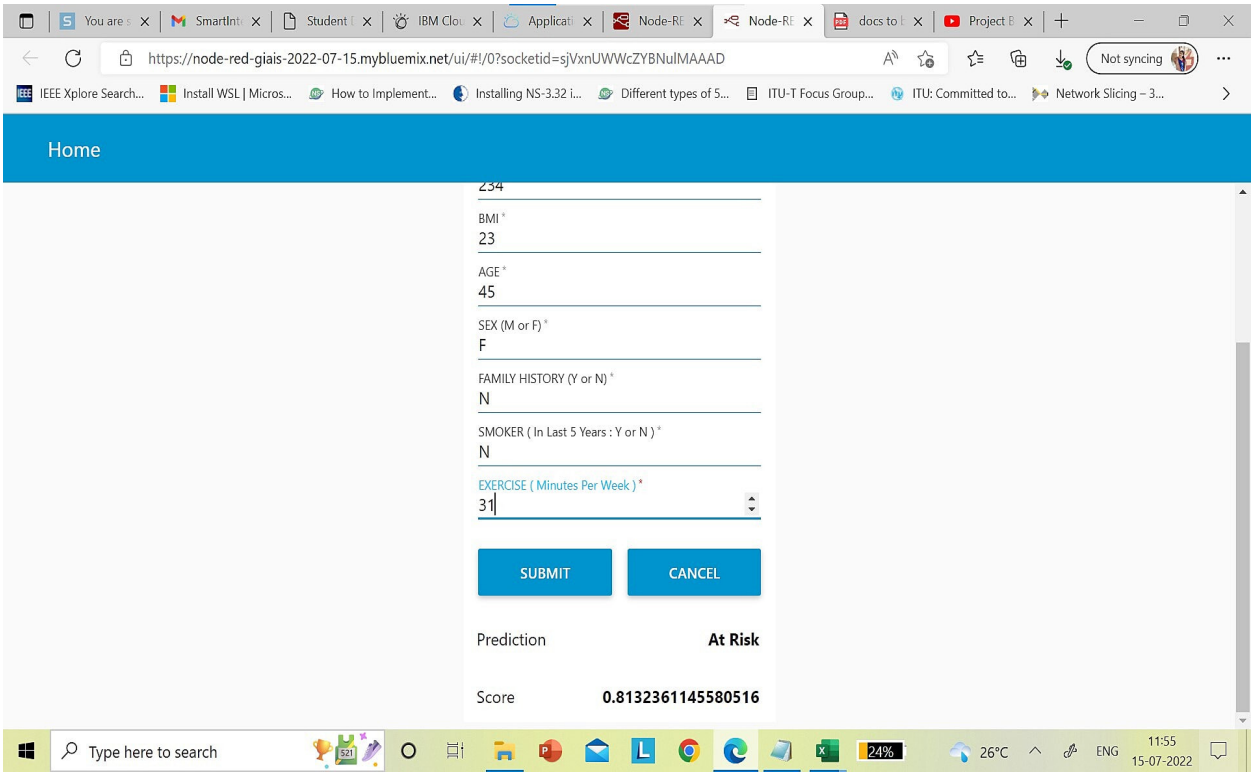


6.1 Output: Screenshots of prediction is attachted below:



6.2 Dashbord using Node Red: Screenshots of dashborad is attachted below:





7. ADVANTAGES & DISADVANTAGES

7.1 Advantages: We can clearly see the advantages of this analysis in term of comparing different classifiers to classify the CVD dataset, and the benefit of having a reliable feature selection method for CVD prediction with using minimal number of attributes instead of having to consider all available ones.

7.2 Disadvantages: This project is conducted with limited data. In the future, large-size datasets with a greater number of attributes or fused datasets that may be obtained by combining two different datasets can be used to enhance the diagnosis procedure.

8. APPLICATIONS

A physician's evaluation of the patient's medical history, physical examination report, and analysis of concerning symptoms are used to diagnose heart diseases. However, the findings of this method of diagnosis are insufficient in detecting heart disease patients. Furthermore, it is both costly and computationally challenging to examine. We build a noninvasive prediction system to handle these issues using machine learning classifiers. Heart diseases are efficiently diagnosed using an expert decision system relying on machine learning classifiers.

9. CONCLUSIONS

Different classifiers were used and compared to classify the CVD dataset, and we concluded the benefit of having a reliable feature selection method for CVD disease prediction with using minimal number of attributes instead of having to consider all available ones.

10. FUTURE SCOPE

As an extension to this work, and some sort of limitation to the work performed here, different types of classifiers can be included in the analysis and more in depth sensitivity analysis can be performed on these classifiers, also an extension can be made by applying same analysis to other bioinformatics diseases' datasets, and see the performance of these classifiers to classify and predict these diseases.

11. BIBILOGRAPHY

[1] Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., & Saboor, A. (2020). Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare. IEEE Access,Vol. 8, 107562- 107582.

[2] Antovani, R. G., Horváth, T., Cerri, R., Vanschoren, J., & de Carvalho, A. C. (2016, October). Hyperparameter tuning of a decision tree induction algorithm. In 2016 5th Brazilian Conference on Intelligent Systems (BRACIS) IEEE, 37-42.

[3] Chandrasegar, T., & Choudhary, A. (2019, March). Heart Disease Diagnosis using a Machine Learning Algorithm. In 2019 Innovations in Power and Advanced Computing Technologies (i-PACT), IEEE, Vol. 1, 1-4.

[4] H. Khdair and N. Dasari, “Exploring machine learning techniques for coronary heart disease prediction,” International Journal of Advanced Computer Science and Applications, vol. 12, no. 5, 2021.

[5] J. Minou, J. Mantas, F. Malamateniou, and D. Kaitelidou, “Classification techniques for cardio- vascular diseases using supervised machine learning,” Medical Archives, vol. 74, no. 1, p. 39, 2020.

[6] NCD (2015). www.icmr.nic.in/final/causes_death/NCD.pdf.