ISSN: 2708-7123 | Volume-02, Issue Number-04 | December-2021
LC INTERNATIONAL JOURNAL OF STEM
Web: www.lcjstem.com | Email: editor@lcjstem.com

Ensemble Model for Heart Disease Prediction

Bamanga, Mahmud Ahmad[1], Ahmadu, Asabe Sandra[2], Musa, Yusuf Malgwi[3]
[1]Computer Science Department, Federal University of Lafia, Nasarawa State (mabamanga@gmail.com)
[2]Computer Science Department, Modibbo Adama University, Yola (ahmaduasabe@mautech.edu.ng)
[3]Computer Science Department, Modibbo Adama University, Yola (yumalgwi@mautech.edu.ng)

DOI: 10.5281/zenodo.6412438

ABSTRACT— For the identification and prediction of different diseases, machine learning techniques are commonly used in clinical decision support systems. Since heart disease is the leading cause of death for both men and women around the world. Heart is one of the essential parts of human body, therefore, it is one of the most critical concerns in the medical domain, and several researchers have developed intelligent medical decision support systems to enhance the ability to diagnose and predict heart disease in human. However, there are few studies that look at the capabilities of ensemble methods in developing a heart disease detection and prediction model. In this study, the researcher looks at how to use ensemble model, which proposes a more stable performance than the use of base learning algorithm and these leads to better results than other heart disease prediction models. The University of California, Irvine (UCI) Machine Learning Repository archive was used to extract patient heart disease data records. To achieve the aim of this study, the researcher Bagging meta-algorithm. The ensemble model is a superior solution in terms of high predictive accuracy and diagnostics output reliability, according to the results of the experiments. An ensemble heart disease prediction model is also presented in this work as a valuable, cost-effective, and timely predictive option with a user-friendly graphical user interface that is scalable and expandable. From the finding, the researcher suggests that Bagging is the best ensemble classifier to be adopted as the extended algorithm that has the high prediction probability score in the implementation of heart disease prediction.

Keywords— Ensemble Model, Heart, Heart Disease, Classifiers and Dataset.

I. INTRODUCTION

In modern healthcare information system, detecting and predicting solution model for heart diseases is extremely challenging. Any clinical attempt made through modern computing to provide economical solutions and options is worth taking to medical experts. Therefore, the need for computerized models capable of performing various tasks such as data analysis, predictive and detective analysis using effective data mining tools is vital. The application of such tools can assist medical practitioners to make prompt and right medical decision. A lot of computation applications are designed purposely to assist physicians in making effective decision about the various diseases in patients. Despite distinct soft computing innovative effort in place, predicting and detecting heart diseases models stills remained an issue of grave medical concern. Data mining techniques provide lot of algorithm to detect and predict disease but the challenges associated with each algorithm is still the matter that needs proper consideration (Sarwar, Sharma & Gupta, 2015). Moreover, artificial intelligence plays a key role in every aspect of modern classified intelligence. In medical perspective, artificial intelligence assist medical doctors and other health practitioners in performing expert diagnosis as well as detection or prediction of diseases in a patient more efficiently and accurately. Besides, artificial intelligence algorithms have great potential for exploring the hidden patterns in the datasets of the various disease related subjects by adjusting the data mining model for utilizing such patterns for clinical diagnosis (Nikookar & Naderi, 2018).

This research work is proposes to developed an ensemble model for predicting heart disease using object oriented approach that would outlined the whole models to be adapted in the processes. Also, the research intended to use some dataset from Adamawa State Specialist Hospital which was located in Yola North Local Government Area of Adamawa State. The sample of patient database needed to carry out the detection and prediction of heart using common machine learning algorithms. This research will demonstrates the capability of computerizing three popular classifiers which are Support Vector Machine (SVM), Random Tree and Naïve Bayes of the collective dataset captured in the application. The model would also use Bagging as extended ensemble models or fuser's classifiers to finally perform decisive detection and prediction of heart diseases.

A. Research Motivation

The motivation for this study is the fact that human life is dependent on efficient working of heart because heart is an essential part of our body, and heart disease has become very common disease around us as number one killer disease worldwide. Heart disease remains the leading cause of death in most countries worldwide. Early detection and treatment of Heart disease can set the stage for a lifetime of better heart health. In medical perspective, artificial intelligence assist medical doctors and other health practitioners in performing expert diagnosis as well as detection or prediction of diseases in a patient more efficiently and accurately. In addition, the motivation of this study is to improve the research works of scholars, who have contributed to the field of study.

B. Statement of the Problem

The World Health Organization has estimated that 12 million deaths occurs worldwide, every year due to the Heart diseases (Soni, Ansari, Sharma & Soni, 2011). The continuous rising of this tragic loss of lives made it necessary for medical practitioners to find adaptive solutions to this epidemic. A conventional prediction method of heart disease is daunting especially in many developing countries where there are limited Doctors and voluminous data (Chadha & Shubhankar, 2016). The machine learning procedures are incorporated using data mining effective tools for identification and detection of diseases in patients.

The need for a better approach to handling detection and prediction of heart diseases using different techniques to generate particular detection output is what our research aims to offer.

C. Aim and Objectives

This research work is aimed at developing ensemble model for detecting heart disease using python. The specific objectives are:

i. To define the dataset

ii. To design a model for heart disease prediction using python and MongoDB

iii. To validate the proposed model using K-Fold Cross-Validation.

iv. To develop a graphical user interface that captures the patient feature.

v. To present the diagnosis analysis according to probability score point of the patient data.

II. LITERATURE REVIEW

According to (Chen, Huang, Hong, Cheng, & Lin, 2011), which exhibited a coronary illness forecast framework that can help specialists in anticipating coronary illness status dependent on the clinical information of patients? Thirteen significant clinical highlights, for example, age, sex, chest torment type were chosen. The indecisive neural organize calculation was utilized for characterizing coronary illness dependent on these clinical features. Data was collected from machine learning repository of University of California, Irvine (UCI) Machine Learning repository. The artificial neural network model contained three layers i.e. the input layer, the hidden layer and the output layer having 13 neurons, 6 neurons and 2 neurons respectively. Learning Vector Quantization (LVQ) was used in this study. LVQ is a special case of an artificial neural network that applies a prototype-based supervised classification algorithm. C programming language was used as a tool to implement heart disease classification and prediction trained via artificial neural network. The system was developed in C and C# environment. The accuracy of the proposed method for prediction is near to 80%.

However, (Manpreet, Martins, Joanis & Mago, 2016) proposed a heart disease prediction system based on

Structural Equation Modeling (SEM) and Fuzzy Cognitive Map (FCM). They utilized Canadian Community Health Survey (CCHS) 2012 dataset. Here, twenty critical qualities were utilized. SEM is utilized to create the weight grid for the FCM model which at that point predicts a plausibility of cardiovascular infections. A SEM model is characterized with relationship between CCC 121 (a variable which characterizes whether the respondent has coronary illness) alongside 20 traits. To build FCM a weight framework speaking to the quality of the causal connection between ideas must be developed first. The SEM characterized in the past area is presently utilized as the FCM in spite of the fact that they have accomplished the necessary fixings (for example weight grid, ideas and causality).80% of the informational index was utilized for preparing the SEM model and the remaining 20% for testing the FCM model. The precision got by utilizing this model was 74%.

According to (Hannan, Mane, Manza & Ramteke, 2010), many applications have been used to detect and predict heart diseases. However, the use of Radial Basis Function (RBF) to predict the medical prescription for heart disease, in the process of carrying out the experiments, there were about 300 patient's data collected from the Sahara Hospital, Aurangabad. RBFNN (Radial Basis Function–Neural Network) can be described as a three-layer feed forward structure. The three layers are the input layer, hidden layer and output layer. The hidden layer consists of a number of RBF units (nh) and bias (bk). Each neuron on the hidden layer uses a radial basis function as a nonlinear transfer function to operate on the input data. The most often used RBF is usually a Gaussian function. Designing a RBFNN involves selecting centres, number of hidden layer units, width and weights. The different methods for choosing the focuses are arbitrary subset determination, k-implies bunching and others. The technique was applied in MATLAB. Gotten results show that outspread premise capacity can be effectively utilized (with an exactness of 90 to 97%) for recommending the medications for heart disease illness.

The research study made by (Mujawar & Devale, 2015), discovered that the anticipated heart diseases can be detected and predicted using the anticipated coronary illness utilizing changed k-implies and Guileless Bayes. Finding of coronary illness is a mind boggling task and requires extraordinary abilities. The dataset is gotten from Cleveland Heart Disease Database. The property "Infection" with a worth '1' shows the nearness of coronary illness and a worth '0' demonstrates the nonattendance of coronary illness. Altered k-implies takes a shot at both straight out also, combinational information which we experience here. Utilizing two introductory centroids we get two most remote groups. It at last gives a reasonable number of bunches. Gullible Bayes' makes a model with prescient abilities. This indicator characterizes the class to which a specific tuple ought to have a place with. This indicator has 93 % precision in anticipating a coronary illness and 89% exactness in situations where it recognized that a patient doesn't have a coronary illness.

(Prajakta, Vrushali & Kajal, 2016), have made a comprehensive research worked on an intelligent heart attack prediction system using big data. According to him, heart attack needs to be diagnosed timely and effectively because of its high prevalence. The objective of this research article is to find a prototype intelligent heart attack prediction system that uses big data and data mining modeling techniques. This system can extract hidden knowledge (patterns and relationships) associated with heart disease from a given historical heart disease database. This approach uses Hadoop which is an open-source software framework written in Java for distributed processing and storage of huge datasets. Apache Mahout produced by Apache Software Foundation provides free implementation of distributed or scalable machine learning algorithms. Record set with 13 attributes (age, sex, serum cholesterol, fasting blood sugar etc.) was obtained from the Cleveland Heart Database which is available on the web. The patterns were extracted using three techniques i.e. neural network, Naïve Bayes and Decision tree. The future scope of this system aims at giving more sophisticated prediction models, risk calculation tools and feature extraction tools for other clinical risks.

According to (Sudhakar & Manimekalai, 2014), the heart illness expectation can be utilized using data mining and information system. The information created by the human services industry is colossal and "data rich". All things considered, it can't be deciphered physically. Information mining can be viably used to anticipate ailments from these datasets. In this paper, various information mining methods are dissected on coronary illness database. Order methods, for example, Decision tree, Naïve Bayes and neural system are applied here .Associative arrangement is another and productive procedure which incorporates affiliation rule mining and characterization to a model for expectation and accomplishes greatest exactness. Taking everything into account, this paper examines and looks at how changed arrangement calculations take a shot at a coronary illness database.

The proposed prototype of heart disease prediction using data mining techniques was done by (Kamal & Kanwal, 2014), namely Naïve Bayes. Naive Bayes is a measurable classifier which allocates no reliance between the characteristics. The back likelihood should be augmented for deciding the class. Here, Naïve Bayes classifier additionally performs well. In likelihood and constant master framework, Naïve Bayes gives off an impression of being the best model for malady expectation pursued by neural system and Decision trees.

(Dhanashree & Manimekalai, 2013), developed a classifier technique for the heart sickness prediction and likewise they've confirmed how Naïve Bayes can be used for the classification purpose .They categorized clinical knowledge to five distinct classes namely no, low, normal, excessive, very excessive. If any unknown sample is discovered, the method will classify it into respective class label. The dataset used here is the Cleveland medical institution ground work coronary heart disease set which contains 303 observations and 14 parameters. The system works in two phases i.e. coaching phase and testing phase. In the coaching segment, the classification is supervised .The checking out segment involves the prediction of the unknown knowledge or the lacking values. The Naïve Bayes algorithm is used which is based on the Bayesian theorem. The outcome proves that the accuracy has been obtained by altering the number of occasions within the given dataset (Dhanashree & Manimekalai, 2014).

Recently, (Das, Ibrahim & Abdulkadir, 2009), proposed an ensemble of neural networks model for the diagnosis of heart disease and achieved 89.01% accuracy, 80.95% sensitivity, and 95.91% specificity. (Samuel, Grace, Arun, Peng, & Guanglin, 2017), developed a novel hybrid system based on ANN and fuzzy analytic hierarchy process i.e., Fuzzy-AHP technique for the diagnosis of heart disease. The ANN and Fuzzy-AHP based system achieved prediction accuracy of 91.10%. Most recently, (Arabasadi et al., 2017) checked the feasibility of adaptive weighted fuzzy system ensemble method for heart disease detection problem and achieved classification accuracy of 92.31% .

III. METHODOLOGY

This section presents the research design, research site, participants, instrumentation, research ethics protocol, data collection and statistical technique used in the study.

IV. DATA ANALYSIS & RESULTS

A. Source of Data and Data Collection

The sources of data for this study will be medical records of patients diagnosed with heart disease problem from Specialist Hospital, Yola Adamawa State. The researcher will also use data from UCI Machine Learning repository to generalize it all.

B. System Design

System design encompasses the processes of designing and specifying the architecture of the system, system modules, user interface and data needed to satisfy the system requirement (Alan, Barbara & David, 2015). To design a system, the functional parts need to be evaluated at each phase or development. Therefore, we specify our data and users at every phase of development. We break down the system into application architecture, data flow diagram (DFD), the application interface, database model and the code. The development tools used in this research are presented in table 1.

Table 1: Development Tools

Programming Languages	IDEs	Test Tools	Database
Python, JavaScript, HTML, CSS	Pycharm, Jupyter Lab	POSTMAN, Google Chrome, Microsoft Edge. Opera	MongoDB

C. Raw Data

Raw data typically refers to tables of data where each row contains observation and each column represents a variable that describes some property of each observation. Data in this format is sometimes referred to as tidy data, flat data, primary data, atomic data, and unit record data. Sometimes raw data mention to data that has not yet been processed. In this study our raw data contains features that are waiting to be extracted for our model.

D.Transformed Data

This is the ready data that the researcher achieved after from raw data. It is the cleaned and transformed data waiting for feature extraction.

E. Data Pre-processing

Data Pre-processing is that step in which the researcher transformed data in other to bring it to such a state that the machine can easily parse it. In other words, the features of the data can now be easily interpreted by the algorithm. At this phase, the data that will be taken from the dataset will be cleaned by generating structured format and treating missing values.

F. Feature Extraction

Feature extraction is a process at which the researcher identifies important features or attributes of the data. This is useful in reducing the number of attributes that describe the dataset.

G. Training and Test Dataset

The researcher, at this phase will try to create a model to predict the test data. So, we use the training data to fit the model and testing data to test it. The models generated are to

predict the result unknown which is named as the test set. As we pointed out; the dataset is divided into train and test set in order to check accuracies, precisions by training and testing it.

H. Validation

At this phase, the researcher will evaluate the trained data with a testing data set. The testing data set is a separate portion of the same data set from which the training set is derived. The main purpose of using the testing dataset is to test the generalization ability of a trained model.

I. Ensemble Model

The researcher will aggregate the prediction of each base classifiers and results in once final prediction for the unseen data. The motivation for using ensemble models is to reduce the generalization error of the prediction.

J. Sample Data

The dataset to be use in this study will include dataset from Kaggle repository, precisely a UCI Machine Learning Repository. The UCI dataset will contain a total of 14 attributes with 303 instances.

ISSN: 2708-7123 | Volume-02, Issue Number-04 | December-2021
LC INTERNATIONAL JOURNAL OF STEM
Web: www.lcjstem.com | Email: editor@lcjstem.com

Proposed Model

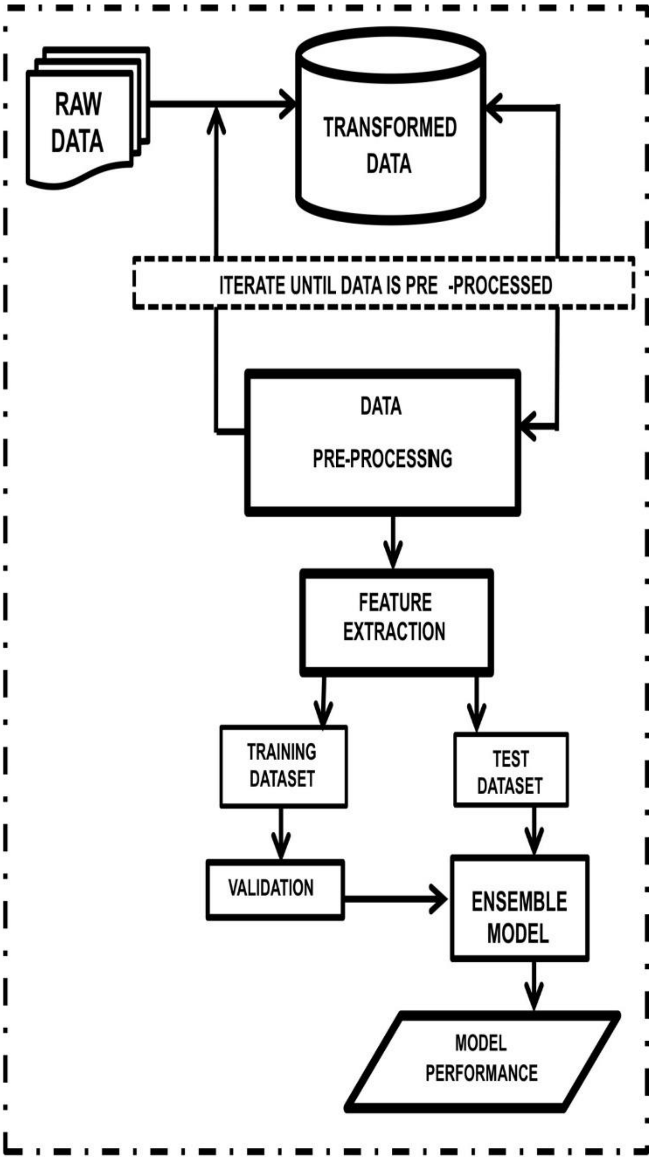

Figure 1: Proposed Model for Heart Disease prediction

Method of Data Sampling

After determination of the sample data, the researcher will use stratified sampling for this study.

A. Data Dictionary

The attributes description is produced here from the heart-disease dataset;

Table 2: Data Dictionary

Attributes	Description	Data Type	Character Length	Required?	Accept Null Value?
First Name	Patient first name	String			
Surname	Patient surname	String			
Other Name	Patient other name	String			
Age	Age in years	Integer			
Gender	Gender (1 = male; 0 female)	Integer			
Cp	Chest pain type	Integer			
Trestbps	Resting blood pressure (in mm Hg on admission to the hospital)	Integer			
Chol	Serum cholesterol in mg/dl	Integer			
Fbs	Fasting blood sugar > 120 mg/dl (1 = true; 0 = false)	Integer			
Restecg	Resting electrocardiographic results	Integer			
Thalach	Maximum heart rate achieved in beats per minute (bpm)	Integer			
Exang	Exercise induced angina (1 = yes; 0 = no)	Integer			
Oldpeack	ST depression induced by exercise relative to rest	Integer			
Slope	The slope of the peak exercise ST segment	Integer			
Ca	Number of major vessels (0-3) colored by fluoroscopy	Integer			
Thal	Thalassemia inherited blood disorder 3 = normal; 6 = fixed defect; 7 = reversible defect	Integer			

B. Model Performance

The researcher will evaluate the performance of the model as one of the core stages in the data science process. It indicates how successful the prediction of a dataset has been by a trained model.

C. Unified Modeling Language

According to (Habeeb, 2018), the Unified Modeling Language (UML) is a family of graphical notations, backed by a single meta-model that help in describing the designing software systems, particularly software systems built using the object oriented style. For the researcher to achieved his aim and have better understanding system; the researcher will use Use-Case and Sequence Diagram. The Use-Case and Sequence Diagram clearly show the interaction between the users and the system.

D. Use Case

This is an effective technique that helps for communicating system behavior in the user's terms by specifying all externally visible system behavior as depicted in figure 2.

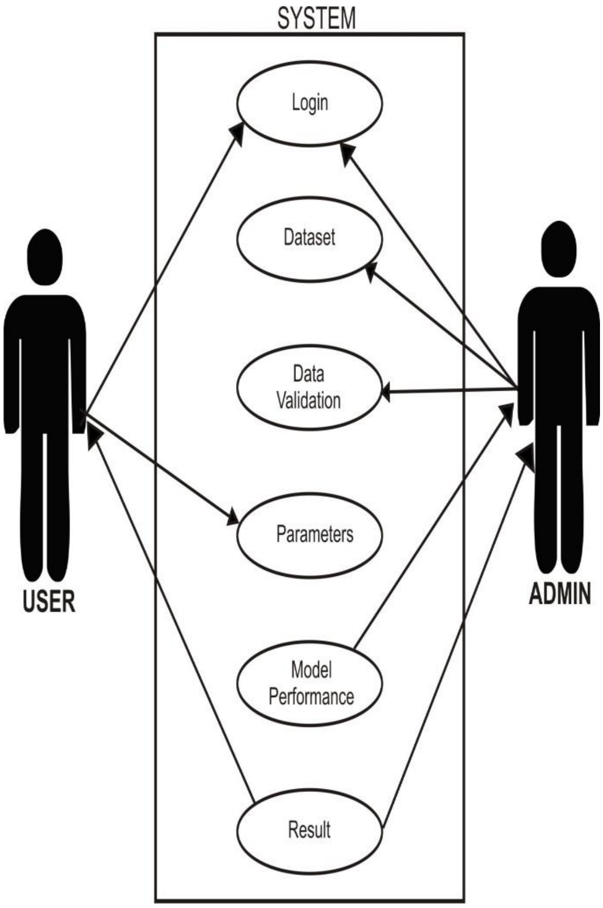

Figure 2: Use Case - Ensemble Model for Heart Disease

D.1 Description of the System Function (Use Case)

Login: This authenticates the access for the user into the system

Dataset: The Admin collect and store object record that made the dataset

Data Validation: At this case, the admin generalized the ability of the trained model

Parameters: The user inputs the behavioral entities from the learning model to predict the presence of heart disease

Model Performance: The system evaluates the classifiers and returned its performance to the admin such as the accuracy of the model.

Result: The system returned the prediction outcome based on the parameters issued.

D.2 Sequence Diagram

The diagram represented in figure 3 describes the integration between the users of the system;

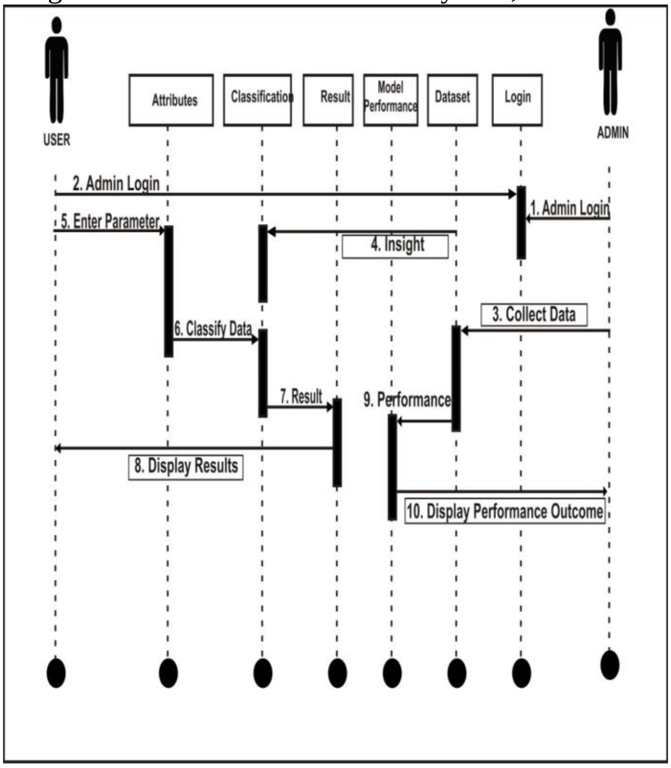

Figure 3: Sequence Diagram - Ensemble Model for Heart Disease

D.2 Description of the System Function (Sequence)

Login: This authenticates the access for the user into the system

Dataset: The Admin collect and store object record that made the dataset

Data Validation: At this case, the admin generalized the ability of the trained model

Parameters: The user inputs the behavioral entities from the learning model to predict the presence of heart disease

Model Performance: The system evaluates the classifiers and returned its performance to the admin such as the accuracy of the model.

Result: The system returned the prediction outcome based on the parameters issued.

E. Implementation of the Diagnostic System

During the implementation of this work, the system interfaces, along with their source codes, is a result of the deployment phases at the realization.

E.1 The Sign-Up Interface

This interface enables system users to register their credentials by creating account that will be use within the system at the authentication phase.

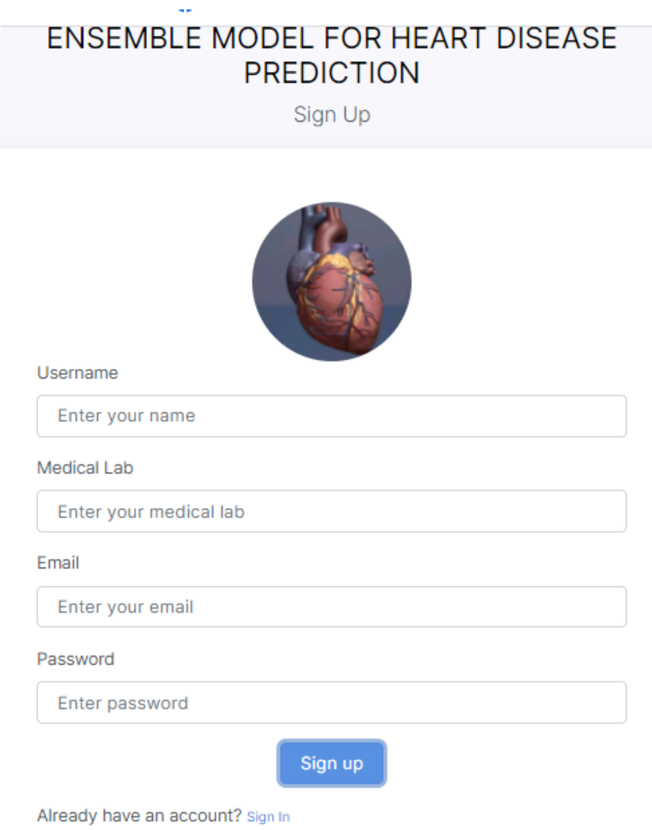

Figure 4: The Sign-Up Interface

E.2 The Authentication Interface

This interface enables system users to be identified by the identity of the user. It is a mechanism to combine an incoming request with a set of credentials. The credentials supplied are compared with those of the authorized user data in a database.

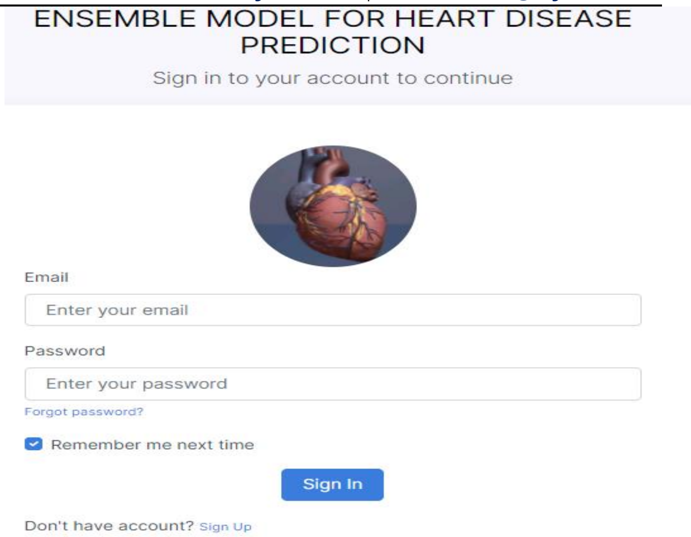

Figure 5: The Authentication Interface

E.3 Dashboard Interface

A dashboard in this work is used to visually tracks, analyzes and displays key performance indicators (KPI) and key data points to navigate within the system. The dashboard of this study has four (4) commands button (navigation interfaces) which shows in figure 6.

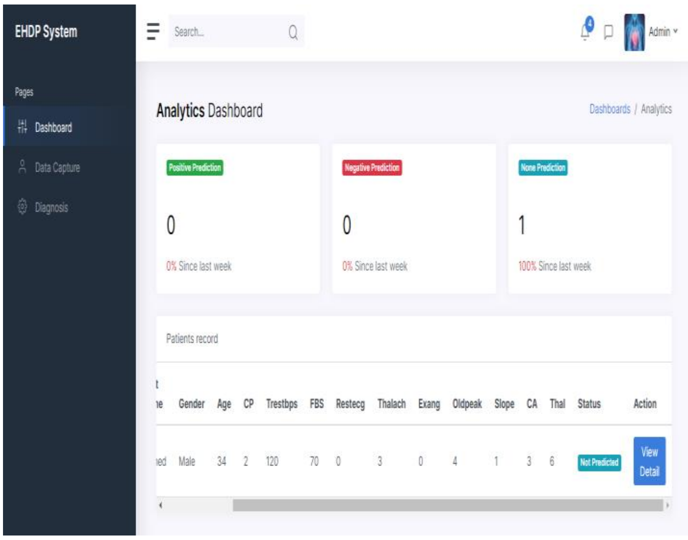

Figure 6: The Dashboard Interface

E.4 The Data Capture Interface

This interface allows user of the system to input disease attributes of patience for the prediction of heart disease presence.

ISSN: 2708-7123 | Volume-02, Issue Number-04 | December-2021
LC INTERNATIONAL JOURNAL OF STEM
Web: www.lcjstem.com | Email: editor@lcjstem.com

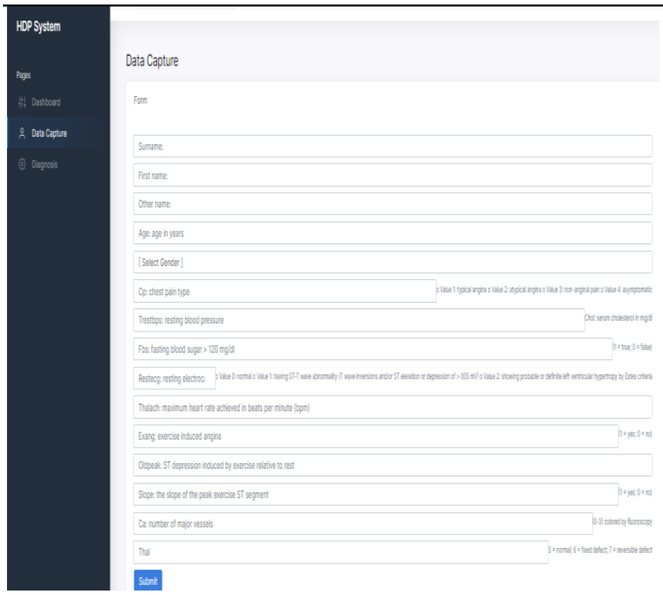

Figure 7: The Data Capture Interface

E.5 The Diagnoses Interface

This interfaces its uses for diagnosing the disease of patience by selecting on patience data as depicted in figure 7.

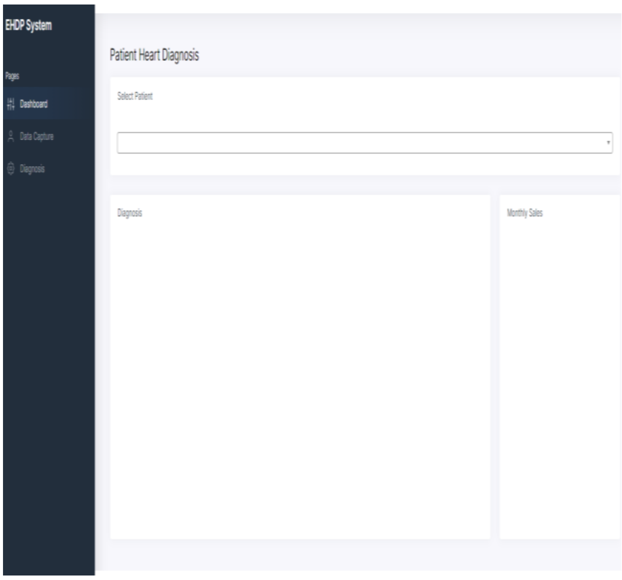

Figure 7: The Diagnoses Interface

E.6 The User Profile Interface

The user profile interface in this study enables the system user to update his/her profile information.

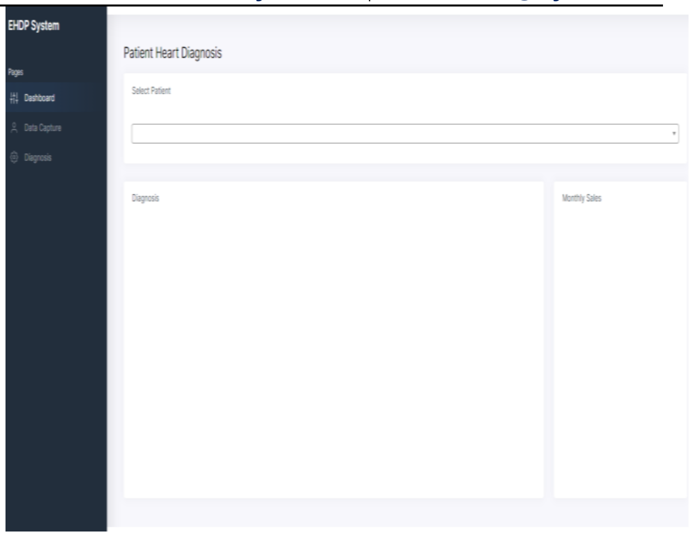

Figure 8: The User Profile Interface

E.6 Comparison Chart for Extended Classifiers

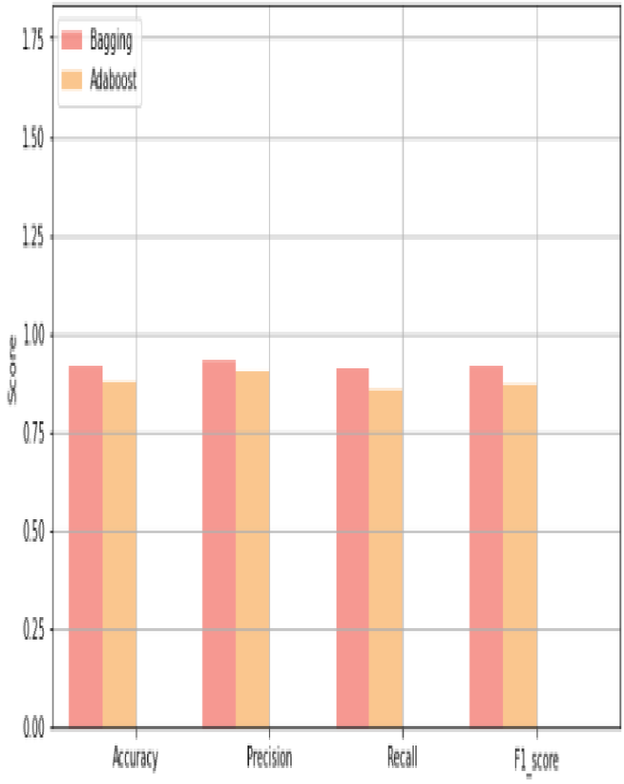

Figure 9: Comparison Chart for Extended Classifiers

Figure 9 above depicts the performance of the extended classifiers utilized and their predictive accuracy, from which the researcher chose bagging as the classier with the best predicting performance for the study's implementation.

II. DISCUSSION AND RESULTS

In this research, the researcher adopt Naïve Bayes, Decision Tree and Support Vector Machine as base learning algorithms in machine learning technology which contains a set of outputs and combines them into a strong and to enhanced prediction of the model by comparing the two extended classifiers or meta-algorithm.

The implementations of this ensemble model were applied on a dataset from UCI Machine Learning Repository for the training and testing of our model for prediction of heart disease. The data collected contains 303 instances of patients diagnosed of heart disease and each clinical instance of patience consists of 14 raw attributes.

A. Discussion

Ensemble technique is a well-proven methodology used in research for achieving extremely accurate data classification by integrating several machine learning algorithms in order to get more dependable and accurate prediction outcomes.

Based on the findings of this study, ensemble learning prediction models were built and created using Bootstrap Aggregating as one of the extended classifiers employed in the ensemble method implementation in order to diagnose and classify heart disease existence in patients as a diagnostic outcome and it is used in the ensemble model for heart disease prediction.

III. CONCLUSION

Prediction of heart disease has been the most promised approach that will help in the suitable reduction in the fertility rate and will assist in the development of the alerts on the large scale. The proposed model for heart disease prediction using ensemble method discussed in this research was successfully implemented and simulated using python and MangoDB. The model captures symptoms and risk factors to diagnose a patient with heart disease and present it in a Graphical User Interface (GUI) form where the user will keyed in the attribute and the system will diagnose the patience based on the trained dataset used in Bagging.

A. Finding of the Study

The study's findings suggest that Bagging is the best ensemble classifier to be adopted in the implementation of ensemble model for heart disease presence in patience. The study model developed; is dependable, efficient, accurate, and can be used in modern health care facilities for the prediction of heart disease.

ACKNOWLEDGMENT

Should be brief, mention people directly involved with the research project (i.e. advisors, sponsors, funding agencies, colleagues, technicians or statistical helper who have supported your work).

REFERENCES

[1] Alan, D., Barbara, H. W. and David, T. *Systems Analysis & Design: An Object-Oriented Approach with UML* (5th Edition). East Carolina University, 2015. ISBN 978-1-118-80467-4.

[2] Chadha, R. & Shubhankar, M. Prediction of heart disease using data mining techniques. *CSI Transactions on ICT, 2016*. 4. 10.1007/s40012-016-0121-0.

[3] Chen, A. H., Huang, S. Y. Hong, P. S., Cheng, C. H. & Lin, E. J. Heart disease prediction system. *Computing in Cardiology, Hangzhou*, 2011. Vol. 5(8), PP. 557-560.

[4] Das, R. Ibrahim, T. & Abdulkadir, S. Effective diagnosis of heart disease through neural networks ensembles. *Expert Systems with Applications, 2009.* Vol.36(4), PP. 7675-7680.

[5] Dhanashree, S. M., Mayur, P. B. & Shruti, D. D. Heart Disease Prediction System using Naive Bayes. *International Journal of Enhanced Research in Science Technology & Engineering, 2013.* Vol.2(3), PP. 2319-7463.

[6] Habeeb, A. *Introduction to Unified Modeling Language*, 2018. 10.13140/RG.2.2.28706.32963.

[7] Hannan, S., Alane, R., March, S., Salvatorem, T., Park, J., & Ram, S. Prediction of heart disease medical prescription using radial basis function. *IEEE International Conference on Computational Intelligence and Computing Research, Coimbatore*, 2010, PP. 1-6.

[8] Kamal, K. & Kanwal, G. Review of Heart Disease Prediction using Data Mining Classifications. *International Journal for Scientific Research & Development, 2014.* Vol.2(4), PP. 2321-0613.

[9] Manpreet, S., Martins, L., Joanis, P., & Mago, V. Building a Cardiovascular Disease predictive model using Structural Equation Model & Fuzzy Cognitive Map, 2016. 1377-1382. 10.1109/FUZZ-IEEE.2016.7737850.

[10] Mujawar , S. H. & Devale, P. R. Prediction of Heart Disease using Modified K-means and by using Naive Bayes. *International Journal of Innovative Research in Computer and Communication Engineering, 2015.* Vol.3(10), DOI: 10.15680/IJIRCCE.2015. 0310123.

[11] Nikookar, E. & Naderi, E. Hybrid Ensemble Framework for Heart Disease Detection and Prediction. *International Journal of Advanced Computer Science and Applications*, 2018.Vol.9(10), www.ijacsa.thesai.org

[12] Prajakta, G., Vrushali, G., Kajal, K. & Prajakta, D. Intelligent Heart Attack Prediction System Using Big Data. *International Journal of Recent Research in Mathematics Computer Science and Information Technology*, 2016. Vol.2(2), PP. 73-77.

[13] Sarwar, A., Sharma, V. & Gupta, R. Hybrid ensemble learning technique for screening of cervical cancer using Papanicolaou smear image analysis. *Personalized Medicine Universe*, 2015. Vol.4(2), PP. 54-62.

[14] Soni, J., Ansari, U., Sharma D. & Soni, S. Predictive data mining for medical diagnosis. An overview of heart disease prediction. *International Journal of Computer Applications*, 2011. Vol.17(8), PP. 43-48.

[15] Sudhakar, K. & Manimekalai, M. Study of Heart Disease Prediction using Data Mining. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2014. Vol.4(1), www.ijarcsse.com