UNIVERSITY OF NAIROBI
DEPARTMENT OF COMPUTING AND INFORMATICS

A STACKED PREDICTIVE MODEL FOR CARDIOVASCULAR DISEASE DIAGNOSIS

NAME: OCHARO MARTIN OBARE
REG NO: P52/38785/2020

SUPERVISOR: STEPHEN NGANGA MBURU

A Project Report Submitted in Partial Fulfillment of the Requirements for the Award of Master of Science in Computational Intelligence of the University of Nairobi

May 2022

DECLARATION

This research project is my original work and has not been submitted in any university for examination, a degree, or publication.

Signature Date

Ocharo Martin Obare 08/06/2022

P52/38785/2020

This project has been submitted for examination towards fulfillment for the award of Master of Science in Computational Intelligence with my approval as the University of Nairobi supervisor:

Signature Date

Dr. Stephen Mburu for 22/08/2022

Department of Computing and Informatics

University of Nairobi

DEDICATION

This project is dedicated to my mother, Callen Kwamboka Ocharo, for her steadfast support and financial assistance throughout my studies.

ACKNOWLEDGEMENTS

First, I thank the Almighty God for great health and strength throughout the study period. The completion of this study could not have been possible without the expertise of Dr. Stephen Mburu, my project supervisor. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. I would profoundly like to thank Prof. Agnes Wausi, Dr. Christopher Moturi, and Dr. Samuel Ruhiu for sitting on my panel and taking the time to read my project.

A debt of gratitude is owed to Dr. Edward Ombui and Dr. Wanjiku Nganga for earlier pointing in the right direction in machine learning and providing the guides in the Python Programming Language.

I express my gratitude to Dr.Aaron Oirere of the School of Computing and Information Technology at Murang'a University of Technology and Eng. James Ondari of the Department of Civil Engineering at Kenyatta University for their valuable comments and criticism on the project. I would like to thank Nada Alay of the Tax and Customs Authority at Zakat for the HTML template input.

Finally, I would like to thank my family and classmates, Sharon Obanda and Ebbie Awino, who provided stimulating discussions.

God bless you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS AND ACRONYMS

CVD	Cardiovascular Disease
DT	Decision Tree
LDA	Linear Discriminant Analysis
LR	Logistic Regression
ML	Machine Learning
NB	Naive Bayes
RF	Random Forest
SVC	Support Vector Classifier
XGB	Extreme Gradient

ABSTRACT

The pervasiveness of Cardiovascular disease has rapidly led to it becoming a global threat in the past few years. The pathology causes 18.6 million deaths annually, and its projections are predicted to hike to more than 23 million deaths annually by 2030. The research aimed at developing a CVD predictive model that was more accurate and robust than conventional models. Stacking is one of the efficient methods in machine learning classification tasks that has been widely utilized to fight CVD. The stacking technique offers better solutions by providing a good trade-off between variance and bias. Stacking gives more accurate and robust results. The study compared seven conventional with stacked algorithms and evaluated the algorithms' performance with four evaluation metrics; accuracy, precision, recall, and f1 measure. The better-stacked algorithm was cross-validated with 10 K-folds. The proposed model was achieved: Data description, retrieval, pre-processing, partitioning, normalization, feature selection, stacking, and model evaluation. The stacked algorithms outperformed the conventional algorithms in classification accuracy with 73.62%, recall with 71.24%, and F1 measure with 72.86%. However, in precision, Decision Tree was better performing with 77.41%. Cross-validating, the stacked model with K-Fold, improved the accuracy from 72.76% to 74.71%. The proposed model can be utilized in the primordial prevention strategy of the World Heart Federation. Additionally, medical practitioners can use it as a CVD diagnostic tool. In future works, more data can be retrieved, investigated in multi-level stacking and deep learning, to research the pros and cons of the proposed model.

1.0 INTRODUCTION

1.1 Background Information

The World Health Organization defines CVD as a group of blood vessels and heart disorders involving diseases like pulmonary embolism, peripheral arterial, congenital heart, cerebrovascular, deep vein thrombosis, rheumatic heart, and coronary heart diseases. Cardiovascular disease attacks and strokes are usually extreme events caused by a preventive blockage that inhibits blood flow into the brain or the heart. The renowned justification for the blockage is fatty deposits build up on the intrinsic anatomical blood vessel walls that flow blood to either the human heart or the brain. Heart strokes are caused by either blood vessel bleeding in the cerebrum or blood clots (WHO,2021).

The pervasiveness of cardiovascular disease has rapidly led to it becoming a global health problem in the past few years. The World Heart Federation estimates that the disease causes 18.6 million deaths per year. Cardiovascular diseases constitute 31% of the global death percentile and thus making it the world's leading human disease killer. The federation approximates that 520 million human beings live with the deadly pathology (Aryal et al., 2020). Cardiovascular disease projections are predicted to hike to more than 23 million deaths per year by 2030(Liu et al., 2022).

Computational intelligence and machine learning techniques have been widely utilized to fight the CVD globally threatening disease (Liu et al., 2022). Stacking ensemble learning is one of the efficient methods in machine learning classification and regression tasks (Pavlyshenko, 2018). The stacking learning technique offers better solutions by providing a good trade-off between variance and bias. Hence, providing a more effective overall performance of a model. Combining various algorithms' advantages is crucial in providing more accurate and robust results (Rajagopal, 2020).

The stacked ensemble learning approach is a significant concern in inpatient management, prediction, medical diagnosis, and relatable healthcare administration issues (Hu et al., 2020).

The disease's intense social and economic impacts render the stacking learning approach one of the major priorities in healthcare research.

There is an uptrend due to CVD risk factors like physical activities, smoking, and alcohol consumption (Hu et al., 2020). The factors were used as the independent variables to predict the presence or absence of CVD pathology. Thus, attempts were made to the stacked machine learning approach in the proposed research. The method was investigated in CVD diagnosis by analyzing the disease risk factors. (Johnson et al., 2018; Cao et al., 2018). The study suggested a more accurate and robust stacked learning model that ought to provide credible results in pathology prediction.

1.2 Statement of the Problem

Currently, there has been explosive datasets growth. The growth has consequently made the development an integral part of scientific research in the biomedical discipline. The biomedical field includes datasets in genomics, medical images, and heart diseases. Thus, to handle biomedical data effectively in the healthcare field, data processing and computational tools must interpret, transform, and analyze the data. (Asgari & Mofrad, 2015; Sharma, 2016). The dataset enormously aids in research advancements of CVD. Whereby stacked algorithms are applied to heart disease. These biomedical advancements will ultimately progress prognosis, prediction, diagnosis, and early treatment of CVD, a human-threatening pathology. (Alizadehsani et al., 2019)

The integral approach to extracting knowledge in heart disease datasets is by using machine learning approaches in CVD research (Cao et al., 2018; Alizadehsani et al., 2019). There is an accuracy and robustness challenge in the review of related work by researchers, in healthcare applications, for disease prognosis, prediction, diagnosis, and early treatment (Sharma, 2016; Abdelaziz et al., 2018; Chowdhury et al.,2021; Ghosh et al., 2021). To overcome the CVD global threat, challenges faced by conventional and conventional models are of crucial concern to scholars (Wang, 2019). Integrating multiple predictors increases the accuracy and robustness of the model(Wang, 2019). Another challenge is that the less accurate single algorithms potentially led to the misdiagnosis of the patients (Narain, 2018). For this reason, the research suggests a predictive model of machine learning algorithms that address the challenges (Cao et al., 2018; Pandey et al., 2019).

A conventional model doesn’t capture all the dataset properties(Wang, 2019). Fortunately, the stacking ensemble solves the conventional problems by combining the different single algorithms, as indicated in figure 4.5. Additionally, combining different level 0 algorithms reduces the generalization error. The level 1 classifiers rectify the prediction errors of the base classifiers and acquire optimal results for the disease prediction. The stacking method improves the generalization capabilities of the algorithms by preventing the overfitting of a model and thus giving a higher prediction precision(Rajagopal, 2020). Using a meta classification approach and whenever single algorithms are combined, the performance of individual algorithms is enhanced. The stacking of different algorithms improves prediction accuracy. Combining conventional algorithms leads to optimized levels of superior predictions(Rajagopal, 2020). There is the potential development of more accurate predictive models by combining conventional algorithms, and this yields better results. (Hu et al., 2020). The research proposes a stacking ensemble predictive model using 7 supervised learning techniques on CVD datasets(Rajagopal, 2020).

The purpose of the research is to compare the performance of conventional supervised learning and stacked algorithms in CVD diagnosis. The research goal is to perform predictive analysis using seven stacked supervised learning classifiers on CVD diagnosis. The diagnosis approach will be based upon machine learning techniques in stacking and an HTML predictive graphical user interface.

1.3 Research Aim

The main research aim of the proposed study is to develop a predictive model that is more accurate and robust than conventional models. The research will find classifiers with the highest accuracy in predicting heart disease.

1.4 Objectives of the Research

The general objective of the proposed study is to obtain an accurate and robust CVD prediction model using the stacking technique.

The specific objectives are

1. To stack seven conventional machine learning algorithms with seven different meta-classifiers
2. To evaluate the performance of the seven conventional machine learning algorithms and seven stacked machine learning algorithms with four evaluation metrics; classification accuracy, precision, recall, and f1 measure.
3. To develop a stacked model prototype from the best performing machine learning algorithms for CVD diagnosis
4. To perform ten K-fold cross-validation to fine-tune the stacked prototype model
5. To compare the CVD prediction capability of the stacked prototype model with a previous study

1.5 Research Question

According to section 1.2, the research examines one comparison question.

Question: Which are the best performing ML algorithms for CVD prediction modeling, between conventional and stacked ensemble algorithms?

1.6 Expected contribution

To develop a robust machine learning model to predict and support medical practitioners as a diagnostic tool.

2.0 LITERATURE REVIEW

This chapter covered various recent works reviewed on the proposed model

Ahmad et al. (2015) reviewed and analyzed various data mining techniques. The authors described the significance of the methods in changing the big voluminous data, using five stages of the knowledge discovery process, into meaningful information. These stages include data pre-processing, selection, mining, interpretation, transformation, and evaluation. Additionally, the study showed the significant role of machine learning in data mining of several pathologies, such as breast cancer, Alzheimer's disease, diabetes, heart disease, etc. Furthermore, they discussed the two main categories of supervised learning; regression and classification. The supervised learning algorithms ranged from Decision Tree, Support Vector Machine, Neural Network, etc. Additionally, the challenges researchers face while conducting data mining in the healthcare industry were mentioned. One of the challenges is the heterogeneity of dataset formats by different healthcare organizations. The authors suggested the combination of various machine learning algorithms in future works to bring about more efficiency, improve care, decrease healthcare costs and overcome challenges.

Nilashi et al. (2017) suggested a knowledge-based system to predict various pathologies. The system utilizes clustering, an unsupervised learning technique, data cleansing by removing noises, and prediction techniques. Additionally, the knowledge-based system used fuzzy-based rules created by Classification and Regression Trees (CART), a method in supervised learning. The proposed approach was tested on several real-world public medical datasets with Mesothelioma, StatLog, and Cleveland datasets depicting an increased pathology prediction accuracy. The improved accuracy resulted from the concatenation of unsupervised learning-clustering, CART, data cleansing, and the use of fuzzy rules.

Kavakiotis et al. (2017) showed the assortment efficacy of 85% supervised learning and 15% unsupervised learning techniques in biological and clinical datasets. The reviewed literature resources were obtained from two renowned organizations: Digital Bibliography & Library Project (DBLP) and PubMed on computer science. It was evident that the most successful ML algorithms techniques were the support vector machines in both biomedical databases. The primary aim of the research was to review applications of data mining, ML, and tools in the diabetes research sphere for diabetic complications, prediction, diagnosis, and healthcare management, whereby prediction and diagnosis turned out to be the most popular. The authors investigated the pros and cons of various ML algorithms and data mining techniques. The majority of the literature resources insisted on performing data pre-processing to prevent discrepancies; these will remove noises, outliers, and data redundancies and reduce the execution time. In supervised ML, particularly classification, the authors recommended the cross-validation method in the classification process. For no or missing information, unsupervised ML algorithms, clustering would come in handy. Eventually, the authors proposed a combination of association, regression, classification, and clustering to improve performance.

Yassin et al. (2018) conducted a systematic review of 154 studies out of 320 research studies retrieved from PubMed, Science Direct, IEEE Explore Digital Library, and Springer Link scientific databases. The overall driving force of the reviews was to aid researchers in creating and innovating Computer-Aided Design (CAD) systems to help in the early treatment and diagnosis/detection of breast cancer pathology in the healthcare sphere. From the research study, the authors were able to unravel that the increased utilization of computational intelligence methods in classification, automated feature extraction, and detection schemes in helping medic experts in their field is a result of its efficacy. Furthermore, the research paper recommended uniformity of public image datasets with disparate image modalities. These image databases should support several image modalities for a similar case in both categorizations and join relevant information from many views. It will be of more importance if the cases contain Deoxyribonucleic acid sequences. The sequences will help CAD systems provide credible results since the systems will depend on different perspectives of sequences and modalities. Additionally, the authors recommended swarm intelligence research as it was seldom used in the researched CAD system publications. The joining of several image modalities is a necessity in creating ML-CAD systems.

Chauhan (2018) accurately compared and predicted heart diseases using five different blends of standardized ML algorithms and calculations. Classification techniques of supervised ML were used. Additionally, the following classification algorithms; Decision Trees, K-Nearest Neighbor, Support Vector Machine, Random Forest, and Logistic Regression in the predictive and overall risk analysis of the world's number one killer disease. The dataset was acquired from the Kaggle website, having 14 input variables. Additionally, it had a combination of 4238 data records of Framingham and Massachusetts. Entity relationships were applied with labels having either presence or absence of coronary heart disease. The prediction aim was to determine whether a the patient is at a 10-year risk of future heart pathology. Relevant features were selected using the process of backward elimination by considering their probability value. Data analysis was performed on Jupyter Lab using Python programming language, a flexible and powerful data mining application software. Logistic regression outperformed the other four supervised ML algorithms. The research recommended future work in predicting heart pathologies by the utilization of more advanced techniques and data, through classification ML algorithms for higher accuracy rates and with few times complexity.

(Adede et al., 2019) predicted drought and vegetation conditions in four Kenyan counties in the North Eastern part. The dataset was retrieved from the satellite and it was of 17 years. There were 16 selected variables, 244 trained Artificial Neural Network (ANN), and 111 support vector regression (SVR) models. The stacking ensemble of ANN and SVR outperformed the conventional models by an R^2 of 0.94. The conventional models' ANN attained an R^2 of 0.83, SVR attained 0.78. Suggestions were put forward for stacking ensembling of models for drought prediction since it was evident that the stacking was a good investment.

(Li, 2020) predicted heart disease using seven stacked models; Neural Network, decision tree, GBDT, Random forest, AdaBoost, Naive Bayes, and XGBoost. The dataset contained fourteen attributes as explanatory variables and a label having the presence or absence of the heart disease. SHAPley Additive exPlanations were used to explain the output of each single machine learning model and then the models were stacked together to achieve an improved prediction ability of the stacked model. The evaluation metrics used included the algorithmic stability, runtime, Accuracy, prediction ability, False Negative Ratio, confusion matrix, and Area Under the Curve (AUC). The technique of cross-validation was applied to the models for parameters tuning. The conventional models achieved an AUC of between 0.83 and 0.92, whereas the stacked models had an exact 0.92. Stacking was put forward as a recommendation in future works, has a superior prediction capability.

(Rajagopal, 2020) proposed an ensemble stacking approach for detecting network intrusion on two heterogeneous datasets. One was a packet-based emulated dataset -UNSW NB-15, and the other was a real-time, flow-based, UGR '16 dataset. The purpose of using the two different types of datasets was to build more secure applications. The persistent network cyber attacks compel scholars to devise more robust models to overcome the challenges of inefficiency and robustness of the conventional machine learning algorithms. The approach used a combination of four algorithms namely; random forest, support vector machine, logistic regression, and K nearest neighbor. The hybrid ensemble meta classification approach improved the performance by providing more reliable results. Seven performance evaluation metrics were conducted; false alarm rate, false-positive rate, false-negative rate, accuracy, F1 measure, precision, and recall. The stacking ensemble approach obtained superior performance on a real-time dataset with an accuracy of 97 % and 94 % accuracy in the emulated dataset.

Mehmood et al. (2021) proposed a framework called CardioHelp for predicting cardiovascular diseases at an earlier stage and comparing its result with state-of-the-art methods. CardioHelp uses a combination of the UCI repository dataset and deep learning CNN to aid medical practitioners. The proposed techniques used the least absolute shrinkage for regularization and independent variable selection to predict patients' heart health. Additionally, a correlation was used in the independent variables of the 14 attributes of the UCI dataset repository, and thus, data was classified. Additionally, a confusion matrix was obtained and displayed for visual inspection and multi-class classification. The proposed model was validated by metrics like Precision, Recall, and F-Score. Finally, results showed that the investigated technique outperformed the current methods, and suggestions were put forward to use more advanced machine learning techniques for cardiovascular diseases, cancer, and other diseases.

Liu et al., 2022 aimed at solving the challenges faced by conventional algorithms through validating the stacking model. The stacking technique was used to improve accuracy of the proposed model. It used publicly available datasets of UCI machine learning algorithms. The datasets were further partitioned into 80 % and 20 % then later validation was done on another heart disease dataset. For feature selection, the model used Lloyd Shapley's game theory to identify the relevant features. The stacking used ten conventional machine learning algorithms; K-Nearest Neighbor, Support Vector Machine, LR, RF, Extra Trees, Gradient Boosting Decision Trees, Extreme Gradient Boosting, LightGBM, CatBoost, and Multilayer Perceptron. The research used four performance evaluation metrics; accuracy, precision, recall, and F1 Score. Further works were suggested to validate the stacking model's pros and cons.

2.1 Research Gap

From the related works, articles, journals, and other publication reviews, the proposed approach can be applied to various diseases - supervised learning classification problems that involve databases having the same nature utilized. Additionally, there is an uptrend in risk factors that influence heart diseases like smoking, physical activities, alcohol intake, etc. There are plenty of research studies on ensemble learning to diagnose heart pathologies. The combination of various algorithms is to exploit ensembling usefulness and its possible potential in the CVD diagnosis.

Much attention should be paid to the disease classification, diagnosis, early treatment, and prediction datasets by using supervised learning and stacking approaches. Thus, there is a compelling need to validate the stacked method by using evaluation metrics on the publicly available datasets . The need is so as to depict the effectiveness of the proposed model for CVD data analysis. Additionally, future investigations can be made on how the applicability of the proposed model can be extended to other medical domain datasets.

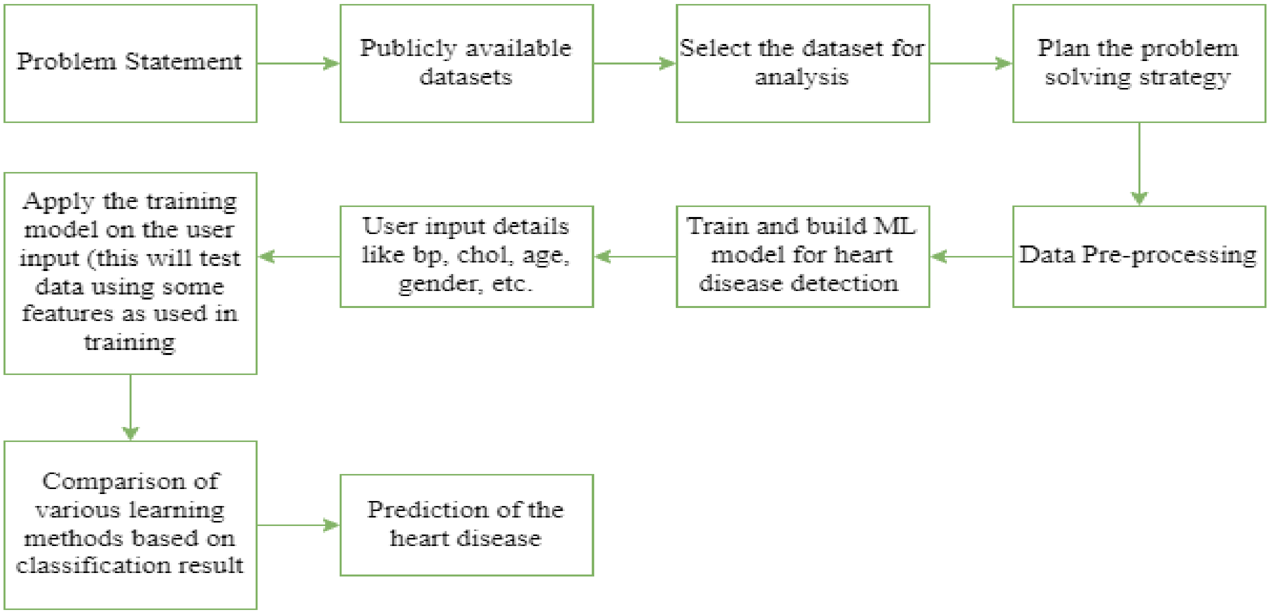


Figure 2.1: Implementation flow of the conceptual design of the proposed model (Ware et al., 2020)

From section 2.1, the proposed model will be designed as depicted in figure 2.1. The implementation flow of the conceptual design has been used in multiple literature resources and will be validated through train/ split test and K-fold cross-validation (Chauhan, 2018; Shah et al., 2020; Ghosh et al., 2021).

3.0 METHODOLOGY

The research investigated the stacking technique in a CVD dataset. The methodology was achieved in the following ways: Data description, retrieval, pre-processing, partitioning, normalization, feature selection, stacking, and model evaluation in chronological order. Data processing and stacking of supervised learning techniques were applied, particularly in the CVD classification. The prediction step was achieved by selecting the more accurate meta-classifier in the preceding stacking model selection. The proposed model can aid medical practitioners as a robust and more accurate clinical data analytical tool in their healthcare sphere.

There was a random partition of the dataset into three parts, having 70% training, 20% testing, and 10 % cross-validation. To validate the stacked algorithms, a K-fold cross-validation approach was employed. The optimal classifier was considered and chosen from the best performing classifier in performance evaluation metrics. The corresponding trained features were then utilized to categorize any new data from the patients' reports in the final prediction.

3.1 Implementation flow of the Proposed Model

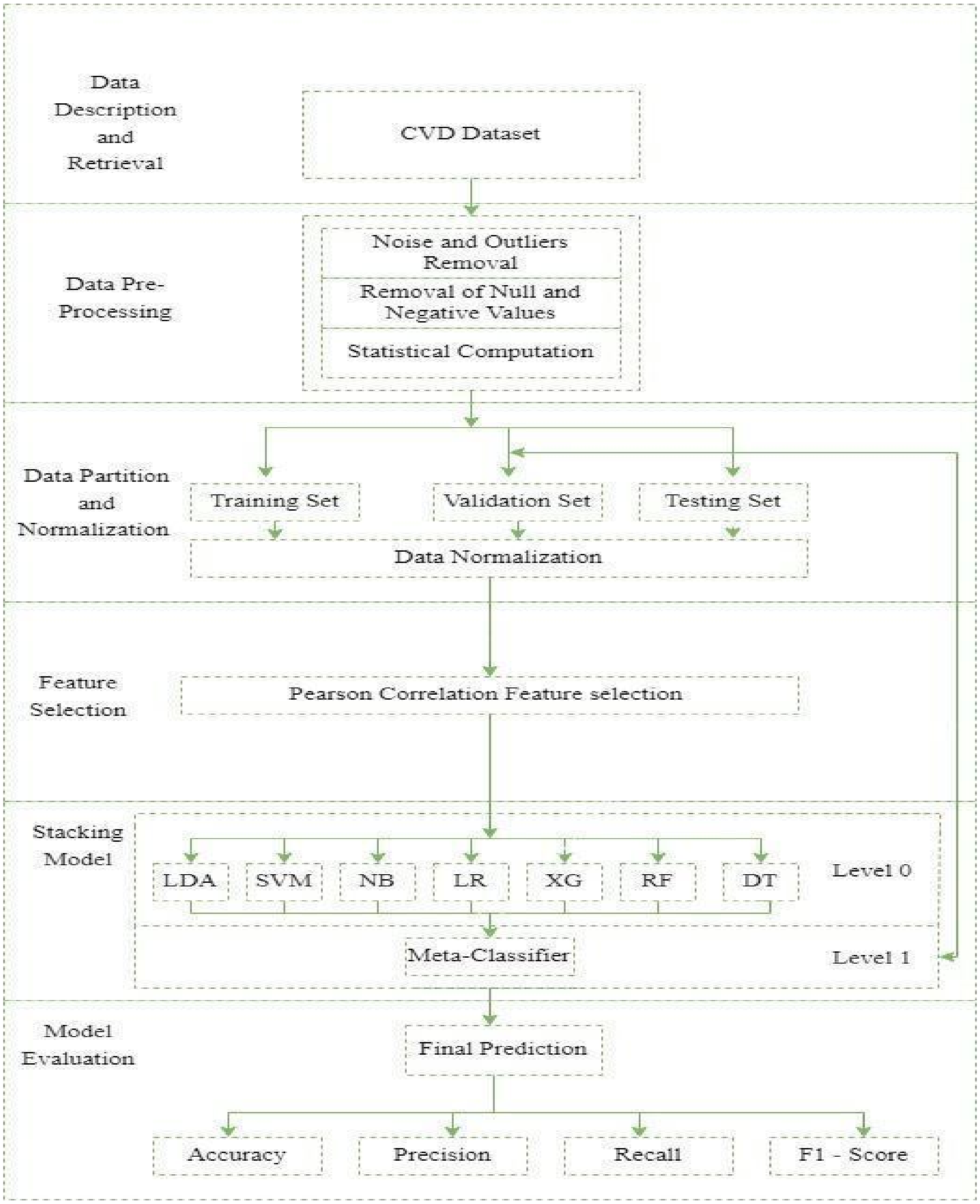


Figure 3.1: Implementation flow of the proposed model.

The proposed model combines conceptual frameworks used by other researchers (Kavakiotis et al., 2017; Narain, 2018; Ware et al., 2020; Nadeem et al., 2021). The suggested model has further been described in the remaining chapter 3:Data description, retrieval, pre-processing, partitioning, normalization, feature selection, stacking, and model evaluation.

3.1.1 Data Description and Retrieval

The selection was used to identify and choose the relevant dataset (Ahmad et al., 2015). The research dataset was identified and obtained from a free online database repository known as the Kaggle website. Kaggle is a Google subsidiary company, and its data has been widely used in both machine learning and data mining. Features from the data include the independent and dependent variables used to train the algorithms (Maharjan, 2020). The dataset had 270 medical records of patients with 13 attributes and a class label, whereby the target output indicates whether a patient has CVD or not (Chithambaram & Gowsalya, 2020). The algorithms were trained, tested, and validated by all the attributes. The CVD dataset is further described in Table 3.1. Additionally, the aforementioned CVD dataset can be downloaded from the University of California, Irvine, Machine Learning Repository (Khan et al., 2018; Muibideen, 2019).

Table 3.1: Attributes of the Kaggle CVD dataset (Dua & Graff, 2019)

Attribute	**Description**	**Domain values**
Age	The length of a patient's lifetime. Numerical data type.	29 - 77
Sex	A patient’s gender. Binary data type	female-0, male - 1
Cp	Nominal data type. Chest pain type with 4 being more likely to get CVD.	1 = typical angina, 2 = atypical angina, 3 = nonanginal pain, 4 = asymptomatic
Trestbps	Resting blood pressure. Numerical data type with concerns when the range is between 120 - 140 mmHg	94 - 200
Chol	Serum cholesterol in mg/dl with mg/dl greater than 200 being a concern. Numerical data type	126 - 564
Fbs	Fasting blood sugar. Binary data type with 120	0 = false

	mg/dL being prone to the pathology	1 = true
Restecg	Resting electrocardiographic result. Nominal data type with 2 being a concern. Patients with left ventricular hypertrophy(mild to severe symptoms) are likely to have the disease.	0 = normal 1 = having ST-T wave abnormality 2 = left ventricular hypertrophy
Thalach	Maximum heart rate achieved. Numerical data type. A patient having more than 120 is likely to have the disease.	71 - 200
Exang	Exercise-induced angina. Binary data type.	0 - No angina-induced exercise 1 - Having the CVD disease
Oldpeak	Depression reduced by exercise relative to stress. Continuous numerical data type.	0 - 6.2
Slope	Slope of peak exercise. Nominal data type with 2 more likely to get CVD.	1 = upsloping 2 = flat 3 = downsloping
Ca	Number of major vessels. Nominal data type with 0 being more prone to the disease.	0 - 3
Thal	Defect type. Nominal data type with 7 more prone to the pathology than the rest	3 = normal, 6 = fixed defect, 7 = reversible defect
Presence	CVD disease. Binary data type.	1 = No CVD 2 = CVD presence

mmHg is mercury measured in millimeters, and **mg/dL** is milligrams per decilitre. The dataset was then passed to the pre-processing step.

3.1.2 Data Preprocessing

The stage converts raw data into a cleansed dataset in the CVD diagnosis. The preprocessing involved; noise removal, checking and elimination of outliers by z-score, eliminating negative or zero data values, and data sampling and filtering through statistical computing methods like; standard deviation, average, mode, minimum, and maximum (Sharma, 2021; Chauhan, 2018). This stage will make the CVD data to be consistent.

3.1.3 Data Partitioning and Normalization

There are many transformation methods (Azevedo, 2019). The consistent data was dimensionally reduced into the generalized format required (Ahmad et al. 2015). Normalization of data transforms the dataset into a data-scaled format. The reduced data was ideal for CVD diagnosis.

The normalized dataset was divided into training, testing, and validation sets (Schmidt et al., 2019). To increase the performance of the algorithms, the training data was partitioned to 70 % and 30 % validation and testing data, which is the best performing data ratio partition in machine learning (Nguyen et al., 2021). The CVD dataset was trained, tested, and validated using the K-Fold cross-validation to produce better results in the CVD prediction.

Some of the commonly utilized data normalization techniques are Decimal scaling, Min-Max scaling and Z-Score normalization (Patro, 2015). The Min-Max normalization technique helped feature scaling by normalizing all the unstructured and structured data values in the CVD medical diagnosis (Maghdid, 2019). The normalization scaled the data values into a range of 0 and 1, as depicted by equation 3.1.

$$N_d = (N_i - N_{min}) / (N_{max} - N_{min}) \qquad \textbf{Equation 3.1}$$

Whereby i = 1, 2, 3, ……., n.

$$0 \leq Nd \leq 1$$

N_d is normalized data values, N_i is actual data values, N_{min} is minimum values, and N_{max} is the maximum values.

The normalization reduced the numeric calculation difficulties by standardizing small and big data values (Maghdid, 2019).

There are various machine learning validation techniques. Cross-validation, also known as, the re-sampling procedure is one of the techniques and was applied in the research (Maharjan, 2020). The model used 10 cross-validation. The 10 cross-validation has been proven to be the best metric for model evaluation selection (Li, 2020).

The validation technique separated the CVD training datasets into K equal partitions. One of the partitions was chosen as a test set. The remaining K-1 partitions were used once as training sets for each round of iterations. The iterations were from 1 to K and in the model used for this research, K = 10.

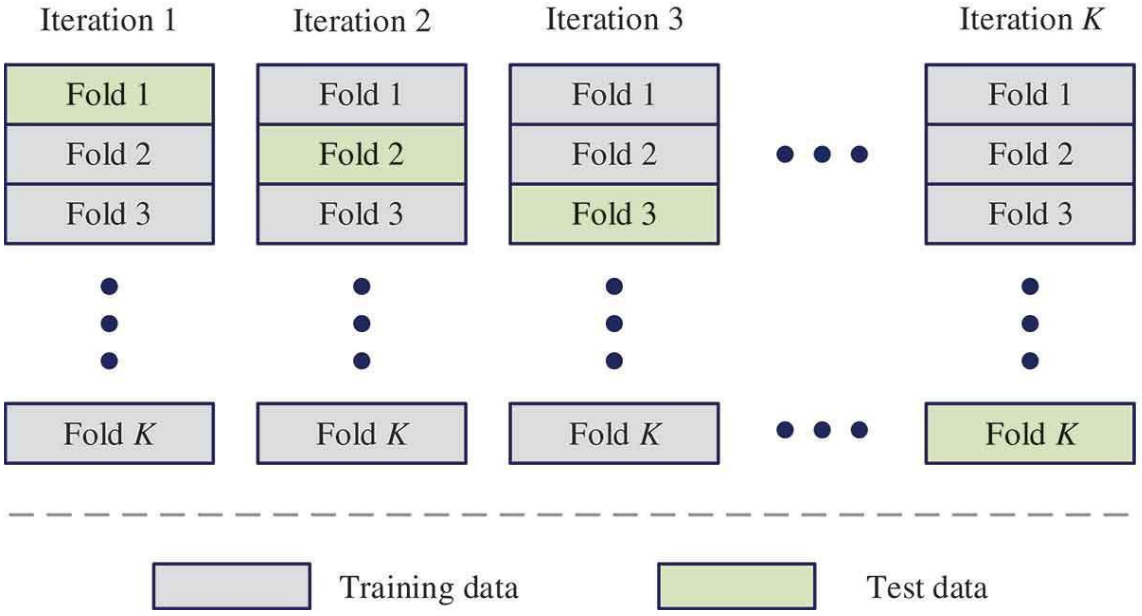


Figure 3.2: K-fold cross-validation (Ren et al., 2019)

The advantage of cross-validation is that all the data is used for both training and testing, as indicated in figure 3.2. It learns the parameters and patterns from different data points in the CVD dataset and turn, tunes it to fit the models (Maharjan, 2020).

3.1.4 Feature Selection

Selecting attributes is of great importance in data feature redundancy. The feature selection methods can be broadly classified into wrapper, embedded, and filter techniques. (Wang et al., 2020). Pearson correlation is a filter technique that is fast in execution and relies on the correlation between the attributes and the class label (Jain & Singh, 2018). The Pearson correlation formula is shown in equation 3.2 (Cui et al., 2020).

$$R(i) = \frac{\sum_{k=1}^{m}(x_{k,i} - \bar{x}_i)(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{m}(x_{k,i} - \bar{x}_i)^2 \sum_{k=1}^{m}(y_k - \bar{y})^2}}$$

Equation 3.2

Whereby, *R(i)* is the correlation coefficient, which ranges between -1 to 1

M, the number of instances

X_k is the feature values k >0

X_i are the features. i = 0,1,2…..13.

Y_k is the feature value,

$\bar{X}$ is identifying X variables

$\bar{Y}$ is the mean value of Y variables.

When R(i) indicated a value of -1, it showed a negative correlation in the variables. +1 correlation indicated a positive correlation. Additionally, the modulus of R(i) >0.3 depicted a weak correlation, 0.3 > modulus of R(i) > 0.5 low correlation, significant correlation 0.5 > modulus of R(i) > 0.8, and a high correlation when 0.8> modulus of R(i) >1 (Cui et al., 2020).

3.1.5 Stacking

The seven supervised learning algorithms are made of different mathematical formulas, functions, and techniques. The three create the prediction accuracy and interpretability difference. Every algorithm has its merits and demerits in the CVD diagnosis (Devi et al.,2016). A stacking ensemble technique was applied to combine the merits of the models to attain more accuracy and a good trade-off between variance and bias error. Different ML algorithms were joined multiple times to form a more robust model (Sharma, 2021).

The stacking technique combines heterogeneous algorithms of level 0 and learns their prediction through a meta-classifier of level 1 (Maini et al., 2021). The meta-classifier prevents overfitting (Liu et al., 2022). The meta-learner is a hybridization post-mortem technique identifying how a predictive model reached the classification decision. The proposed model illustrated a two-level stacking technique process, as depicted in figure 3.3.

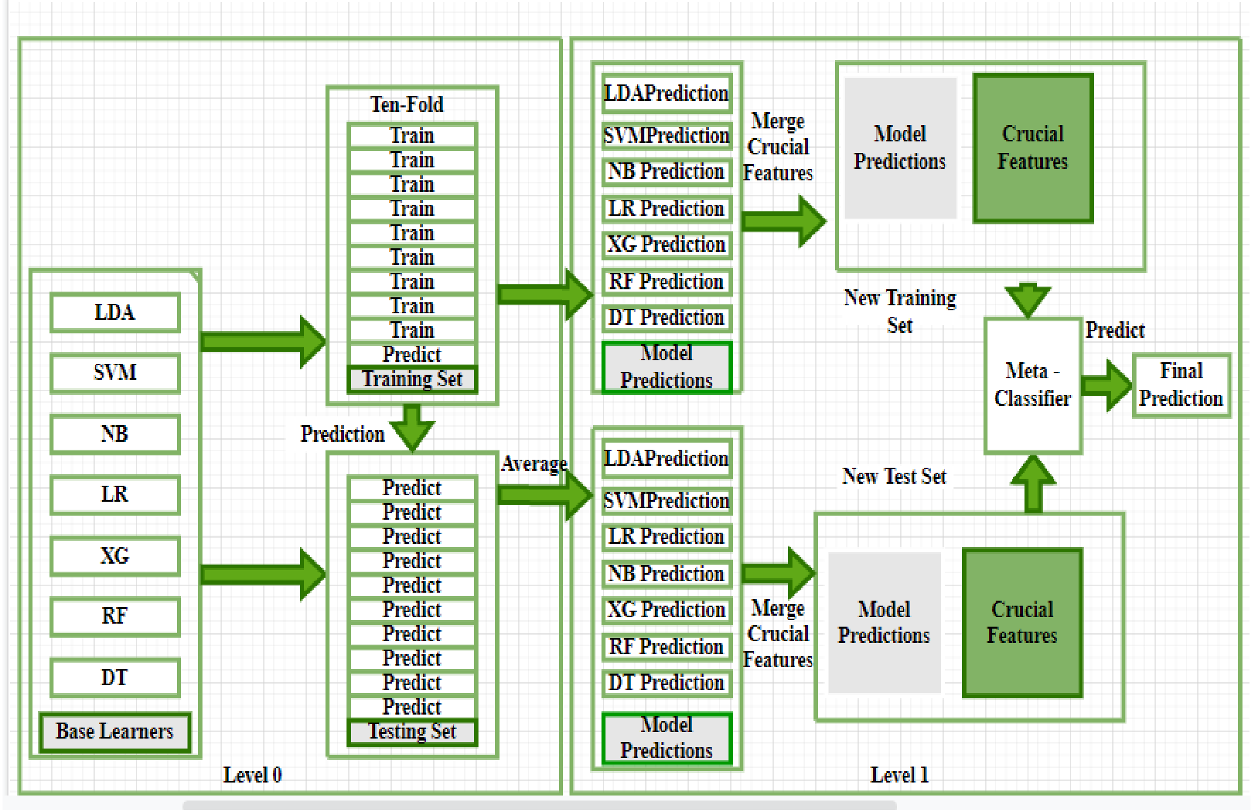


Figure 3.3: The proposed stacked model

The seven classifiers in the weak/base learners generated outputs that were later fed into the meta-classifier (Li, 2020). The CVD dataset in level 0 was cross-validated using ten k-fold hyperparameter tuning in both the training and testing dataset. New features were generated for the training and testing datasets by averaging the prediction results produced by the cross-validated classifiers in the earlier training. Seven different prediction features of the conventional classifiers were generated after that. The mean ensures the splitting ratio is maintained. The Cross-validation in level 0 was to ensure data consistency, avoid variance and bias, and prevent overfitting. In level 1 of the stacking, seven attributes that merged the crucial parameters were incorporated to generate new training and test datasets. Merging of crucial features fitted the stacked algorithms into a one dimensional matrix. The new crucial features included CVD features used by weak learners having high feature importance for predicting the presence or absence of the pathology. Seven different meta-classifiers were selected and used to train the final classifier for the CVD final prediction. The stacking learning model is further described in pseudocode 1.

Pseudocode 1 Stacking Learning Model

```
DEFINE: A training dataset U and a testing set dataset V .
The final prediction html_prediction.
$U_{kt}$ is the training dataset in cross-validation.
$U_{kv}$ is the validation dataset in cross-validation.
the weak learners $\xi l$, and the training dataset $U^l$ of $\xi l$, the meta classifier $UF_{meta}$. whereby
for $l$= { LDA, SVM, NB, LR, XGB, RF, DT }.
$U^l_{cf}$ is the crucial features of $U^l$ .
for $l$= {LDA, SVM, NB, LR, XGB, RF, DT} do
    for $k$=1,2,…,10 do
        $\xi l \leftarrow U^l_{kt}$ (use $U^l_{kt}$ to train $\xi l$ .)
        $train_k \leftarrow \xi l \leftarrow U^l_{kv}$ (use $\xi l$ to predict $U^l_{kv}$ to get $train_k$ .)
        $test_k \leftarrow \xi l \leftarrow V^l$(use $\xi l$ to predict $V^l$ to get $test_k$)
    end for
    $train_l = (train_1 + train_2 + \dots + train_k)$
    $test_l = (test_1 + test_2 + \dots + test_k)/10$ (calculate the mean of testing set.)
end for
$train_{new} = [train_{LDA}, train_{SVM}, \dots, train_{DT}]$
$test_{new} = [test_{LDA}, test_{SVM}, \dots, test_{DT}]$
for $l$= {LDA, SVM, NB, LR, XGB, RF, DT} do
    $train_{new} = train_{new} + U^l_{cf}$
    $test_{new} = test_{new} + V^l_{cf}$
```

```
    combines the crucial parameters of each weak learners.
end for
$UF_{meta}$ ← $train_{new}$(use $train_{new}$ to train $UF_{meta}$)
html_prediction ← $UF_{meta}$ ← $test_{new}$(use $UF_{meta}$ to predict $test_{new}$ to get final prediction)
return html_prediction
```

There is a difference in the classifiers' theories, functions, and concepts; hence to address the diversity and heterogeneity of the classifiers, a meta-classifier combined the seven classifiers. The different classifiers were built in level 1 of the stacking learning model that adopted the advantages of weak learners and produced better predictions.

The stacked classification algorithms were used to predict the presence or absence of the CVD pathology (Ahmad et al., 2015). The proposed model used supervised learning algorithms like the LDA, SVM, NB, LR, XG, RF, and DT described as

1. LDA

The LDA is a supervised classification algorithm that is used in the binary classification of medical diagnosis datasets (Almeida et al., 2017).

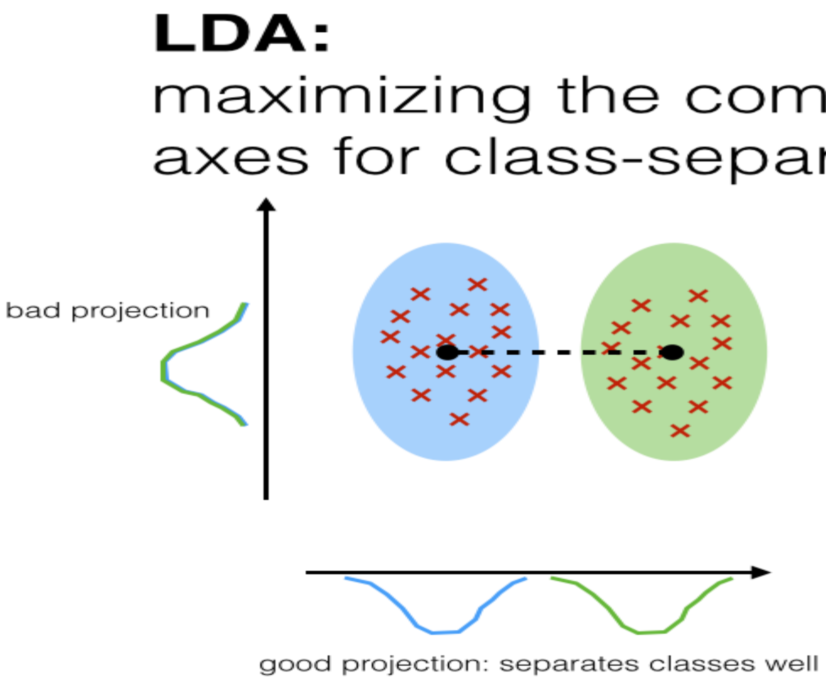


Figure 3.4: Linear Discriminant Analysis (Vincent, 2020).

The algorithm is a probabilistic classification technique that computes in-class variance and within-class separation. The data patterns that have a high probability score are grouped together. It determines the classes' mean and uses eigenvectors and values to find an optimal solution in disease classification (Osho & Hong, 2021). The algorithm aimed to maximize a good projection by ensuring the class labels were separated well, as shown in figure 3.4.

The algorithm is governed by equation 3.3 (Ricciardi et al., 2020)

$$Y=a_1x_{i1}+a_2x_{i2}+a_3x_{i3}+\cdots+a_qx_{iq} \qquad \textbf{Equation 3.3}$$

whereby, Y is the binary class label of having either CVD or not

$a^T=[\{a_1,a_2,\dots,a_q\}]$ is the coefficient vector to be determined

$x_i=[x_{i1},x_{i2},\dots,x_{iq}]$ = number of patients

$x_j=[x_{j1},x_{j2},\dots,x_{jq}]$ = number of features

$1\leq x_{i\leq}270$

$1\leq x_{j\leq}13$

The algorithm is suitable for high dimensional datasets and it detects a projection vector to be used in class mean and variance.

2. SVM

The algorithm depicted great results for predicting various diseases and other healthcare problems (Vincent, 2020). It can be utilized as a predictor, analyzer, and classifier (Chauhan, 2018; Ware et al., 2020). SVM predicted the occurrence of CVD disease by plotting the 13 features in a multi-dimensional hyperplane. The hyperplane provides a margin between the target attributes and classes, with a maximum margin proving an optimal hyperplane (Devi et al.,2016).

SVM is used for both binary, multi-class classification, and prediction. The predictions are dependent on support vectors which are the subset of the training data (Maharjan, 2020).

Suppose n is the number of attributes and m is the support vectors. In that case, the algorithm lies between a range of

$$O(n * m^2) \text{ and } O(n * m^3) \qquad \textbf{Equation 3.4}$$

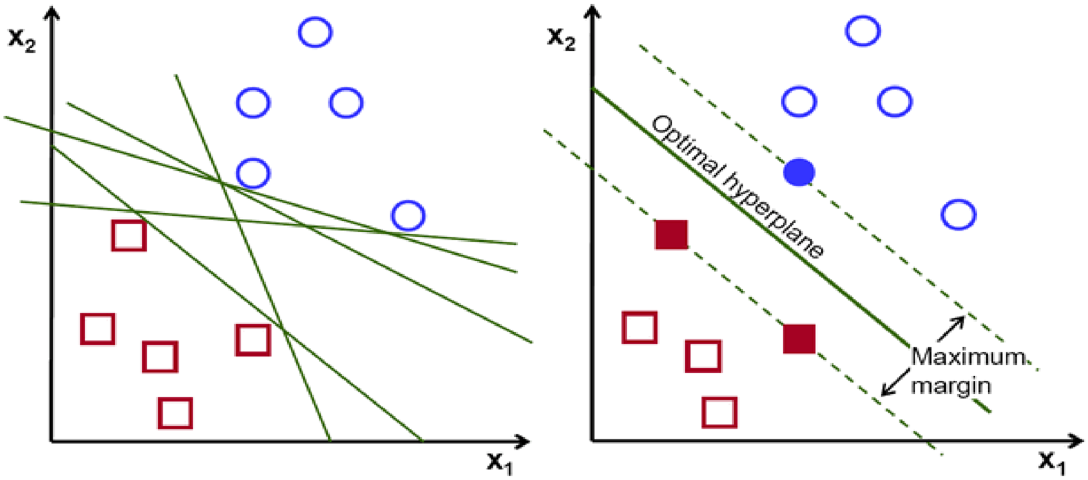


Figure 3.5: Support Vector Machine

Source:https://towardsdatascience.com/support-vector-machine-introduction-to machine-learning-algorithms-934a444fca47

The main objective of the classifier was to pick an optimal hyperplane (Ware et al., 2020). It computes more accurate results since it uses various kernel methods, for example, polynomial, linear, and sigmoid, which optimizes the algorithm (Devi et al.,2016).

3. NB

The algorithm uses the Bayes Theorem, factor, and conditional probabilities of attributes (Ware et al., 2020). The theorem applied in this classifier is represented by Equation 5 and assumes that features are independent of each other (Vincent, 2020). The 13 attributes in our dataset were computed as independent of one another. A new data sample to be classified was grouped on either of the classes based on the highest conditional probability value it attains per class.

$$P(A|B) = P(B|A) * P(A)/P(B)$$ **Equation 3.5**

The equation is a uni-denominal used for both forecasting and prediction (Yatsko,2017).

$P(A|B)$ is the posterior probability. It is the probability that A is true given that B is true.

$P(B|A)$ is the likelihood of B being true, given that A is true.

$P(A)$ is the prior knowledge and probability of A being true.

$P(B)$ is the marginalization and the probability of B being true.

This supervised and data mining classifier performs great classification (Vincent, 2020). Data processing aided in finding patterns and knowledge of past historic patients' records (Devi et al.,2016). The algorithm then predicts statistically the probability of CVD diagnosis based on the observation of the patient's medical history.

4. LR

LR is a classification and predictive analysis method. The algorithm is ideal for CVD diagnosis since it is suitable for binary classification (Ware et al., 2020). The algorithm uses a sigmoid/logistic function represented by equation 3.6.

$$LR = Log(P_i / (1-P_i)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + ... + \beta_k X_k$$ **Equation 3.6**

Whereby, P_i is the probability of occurrence of having CVD Disease

$(1-P_i)$ is the probability of not having the pathology

$\beta_k X_k$ is the independent variable, $1 \leq \beta_k X_k \leq 13$

The LR classifier can be transformed into; (Swamy, 2021).

$$p = \frac{e^{Z}}{1+e^{Z}} = \frac{1}{1+e^{-Z}}$$

(Sigmoid Function or Logistic Function)

where 'Z' varies from -∞ to +∞ 'p' varies from 0 to1

The threshold to classify a patient's diagnosis will be based upon a probability value (Uddin, 2019). This powerful algorithm has high accuracy since it uses different techniques to increase the accuracy (Chauhan, 2018).

5. XGB

XGB is the current evolution that developed from DT to RF to boosting to gradient boosting, a gradient descent algorithm. It minimizes errors/residuals through the algorithm and performs parallel computing (Jiang, 2020). The algorithm uses second-order Taylor series expansion methods (Chang, 2019). XG has various hyperparameters that govern its implementation. These include the number (α), depth of decision trees (β), the number of samples and features (κ), the learning rate/scaling factor, and auto-pruning (γ). Additionally, it has objective functions; regularization (λ) and training loss (Zoghi, 2020). The proposed model used $\alpha = 100$, $\beta = 5$, $\eta = 0.3$, $\kappa = 13$, and $\lambda = 0.3$.

The model learned and was validated by minimizing the errors from the base estimators. The model then modified the split error using the similarity score depicted in equation 3.7. The similarity score determined the splitting of the next split candidates.

The similarity score of errors = (Sum of Errors)^2 /(Number of Errors + λ) **Equation 3.7**

The algorithm is part of an open-source library, scalable and portable. It prevents overfitting through γ, shrinkage technique and feature subsampling that reduces bias (Ahlin & Ranby, 2019).

6. RF

The random forest is a classification ML algorithm based on constructing many decision trees and choosing the most voted tree (Maharjan, 2020). RF is not biased and works well where the dataset has both numerical and categorical attributes (Sharma, 2021). It uses two hyperparameters, the randomly selected attributes, denoted as *m,* and the number of trees from 0 to *n,* as shown in figure 3.6. If any new data is fed into the algorithm, the RF chooses the best solution to predict the CVD disease. Despite depending on the number of trees, the algorithm is faster when compared with other ML algorithms (Maharjan, 2020).

The algorithm's time complexity is calculated as;

$$O(m*log)$$ **Equation 3.8**

The decision for major voting is based upon equation 3.8;

$$H(x)=arg_Y max \sum^{k}_{i} I(h_i(x,L_i)=Y)$$ **Equation 3.9**

Whereby, x is the input variable, H(x) is the RF combined model, h_i is a single tree, *I(.)* is an indicator function, and *Y* is the CVD output to be diagnosed (Ren et al., 2019).

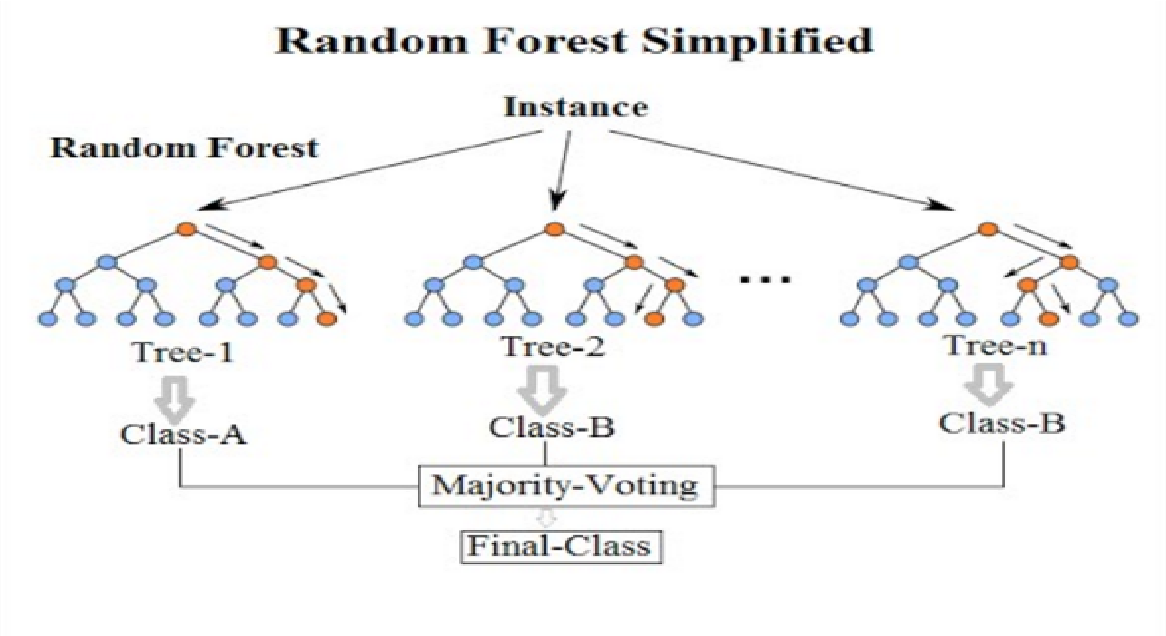


Figure 3.6: Random Forest (Maharjan, 2020)

The accuracy of RF is directly proportional to the number of trees, the more the number of trees, the better the algorithm (Ware et al., 2020). The increased number of trees reduced the CVD model overfitting problem.

7. DT

It is a graphical representation model that is explicitly to be comprehensible and interpretable by a user (Jiang, 2020). DT displays the outcome of inputs in a tree-like structure. It is usable in all types of supervised learning problems (Ahlin & Ranby, 2019). DT utilizes simple decision rules inferred from the data attributes to predict the presence or absence of disease (Chauhan, 2018; Tian, 2019).

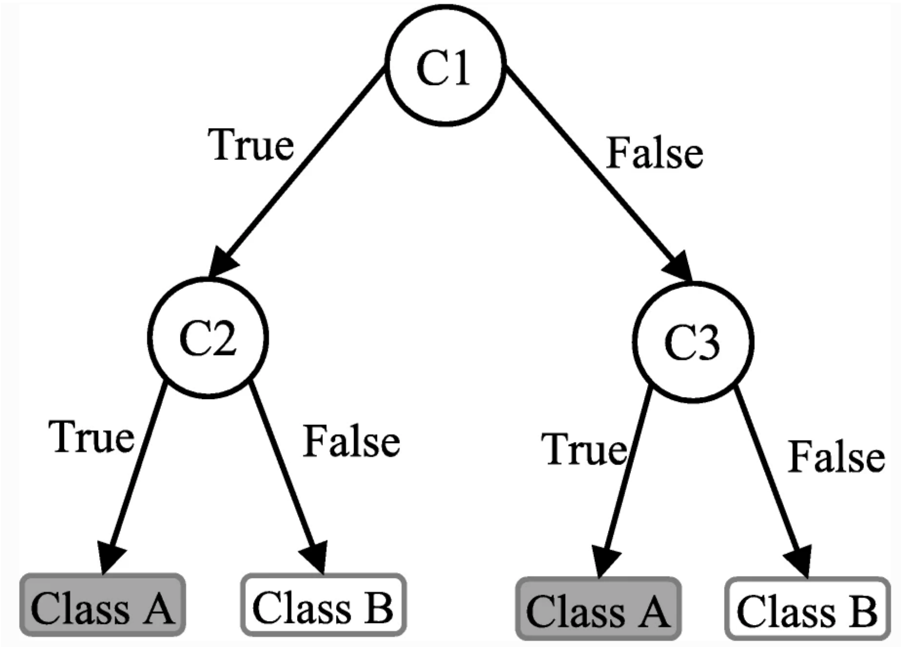


Figure 3.7:Decision Tree structure. (Uddin, 2019)

C1 indicates the Root node. C2, C3 to Cn depicted the test on the 13 independent variables. Class A and B are the leaf/terminal nodes that indicate the presence or absence of CVD disease. The true and false branches enable the algorithm to make decisions in an algorithm of DT. Splitting of the tree stops when data can't be split any further. There are three steps in the implementation of DT. Patient data is to be classified as input from the parent node and traversed through these three steps; data partitioning, classification, and category selection (Devi et al.,2016). The methodology in DT involves partitioning using the Gini index in equation 4.0, gain ratio, and information gain in equation 4.0. The information retrieved from the methodology is then pruned through post or pre-pruning to get a well-trimmed decision tree.

$$\text{Gini Index} = 1 - \Sigma_{i=1,n} (Pi)^2$$ **Equation 4.0**

$$IG(D_p, a) = I(D_p) - (N_l/N)I(D_l) - (N_r/N)I(D_r)$$ **Equation 4.1**

whereby IG is the information gain, a is the attribute to be split on, D_p is the dataset of the parent node, D_l is the child left node dataset, D_r is the child right node dataset, N_l, N_r is the number of samples in the left and right child dataset respectively.

3.1.6 Model Evaluation

The robustness and accuracy of our proposed model can be evaluated by performance indicators (Ghosh et al., 2021). Four aspects were used to compare the performance of the classifiers. Classification Accuracy, Precision, Recall, and F1 score comparison evaluation metrics, were used to test the credibility of the achieved results (Rajagopal, 2020; Premsmith & Ketmaneechairat, 2021). Classification accuracy measures the overall performance of the model classifier. It is the most commonly utilized performance metric in binary classification problems (Li, 2020). It measures the relationship closeness of the predicted to the actual value. Recall depicts the ratio of correctly identified true positives (Shultana et al., 2020). Precision measures the closeness of the measurements (Gupta, 2021). F1 score is the harmonic average of precision and recall (Yu, 2022).

The formula of the four aspects of the evaluation comparison metrics is as shown:

i. $\text{Accuracy} = (W + X) / (X + Y + Z + W)$ **Equation 4.2**
ii. $\text{Precision} = (W) / (W + Y)$ **Equation 4.3**
iii. $\text{Recall} = (W) / (Z + W)$ **Equation 4.4**
iv. $\text{F1 Score} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$ **Equation 4.5**

Whereby X is True Negative, W is True Positive, Y is False Positive, and Z is False Negative. These are further described in table 3.2.

Table 3.2: Attributes of the performance metrics

Acronyms	**Explanation**

True Negative (X)	The total quantity of human beings that are not healthy of heart pathology but are diagnosed as healthy
True Positive (W)	The total quantity of human beings who are diagnosed with heart pathology accurately
False Positive (Y)	The total quantity of human beings who are healthy but are diagnosed with the heart disease
False Negative (Z)	The total quantity of human beings who have the heart pathology but are diagnosed as healthy

The level 0 classifiers were compared with the level 1 classifiers using comparison evaluation metrics i) to iv) as described in table 3.2.

3.1.7 Final Prediction

Web programming application provided a user interface for CVD disease diagnosis (Premsmith & Ketmaneechairat, 2021). The web interface was built using a HyperText Markup Language (HTML). The best-performing model in the comparison section, 3.1.6, was integrated with the HTML file through an application programming interface. To predict the presence or absence of CVD pathology, a liveware will input data values of the 13 data attributes as described in Table 3.1.

3.2 Data Analysis

The research used comparative and quantitative data analysis to find patterns and draw meaningful conclusions (Li, 2020). The proposed model was implemented using the Python programming language. Python is free, open-source, and easy to utilize (Python Software Foundation, 2022). Jupyter Notebook will be the platform where the python program was coded (Chauhan, 2018). The software tools to be utilized are the python functions and libraries (Sharma, 2021). The python libraries and functions include;

i) Python libraries

a) **Numpy -** The library will facilitate advanced mathematics and statistical operations on data. It is fast because of the fewer code lines and less debugging time (NumPy Developers, 2022). Since the libraries were implemented, it was useful in CVD data analysis and acted as a base stack for other ML libraries (Muibideen, 2019).

b) **Pandas -** The library has various reader and writer functions that were input/output tools for comma-separated value files and other dataset format descriptions (the pandas' development team, 2021). It was utilized to read, view, and write on the CVD comma-separated value data. Pandas is BSD-licensed for data structures and analysis (Muibideen, 2019).

c) **Matplotlib -** The technique created a data visualization platform that was utilized in plotting data points and for visual inspection(2002 - 2012 John Hunter, Darren Dale, Eric Firing, Michael Droettboom, and the Matplotlib development team; 2012 - 2021 The Matplotlib development team, 2002–2021). It created a basis for other library implementations like Seaborn (Waskom, 2012–2021).

d) **Seaborn -** It is stacked on top of Matplotlib for drawings and statistical graphics presentation(Waskom, 2012–2021). Additionally, it created high-level interfaces for data mining. The library was utilized with Matplotlib to draw meaningful conclusions in the dataset (Muibideen, 2019).

e) **Scikit-learn -** The library played a crucial role in the proposed model. It is a popularly used python library(Muibideen, 2019). The technique was utilized in the classification of CVD. It is an efficient and simple tool commonly used in data prediction analysis(Bengfort, 2019). Scikit-learn called all the seven ML algorithms. Additionally, it performed evaluation metrics and provided validation tools and other algorithms like the Stacking Classifier (Li, 2020). Sklearn reduced the unnecessary CVD dimensionality and allowed for the heart disease data preprocessing that eventually aided in identifying whether a patient has the pathology or not.

f) **Stacking -** A python package for combining any number of machine learning models. It is convenient for prediction and provides a lightweight application Interface that utilizes stacked generalization techniques (Vecstack, 2019).

g) **Flask -** A simple web framework application for web server gateway interface between the proposed model and the Hypertext Markup Language (Flask, 2022).

ii) Python functions included calling the seven ML classifiers, specifying the evaluation metrics, hyperparameter tuning, and implementing other layers of the proposed model for CVD disease diagnosis.

3.3 Prototype Design

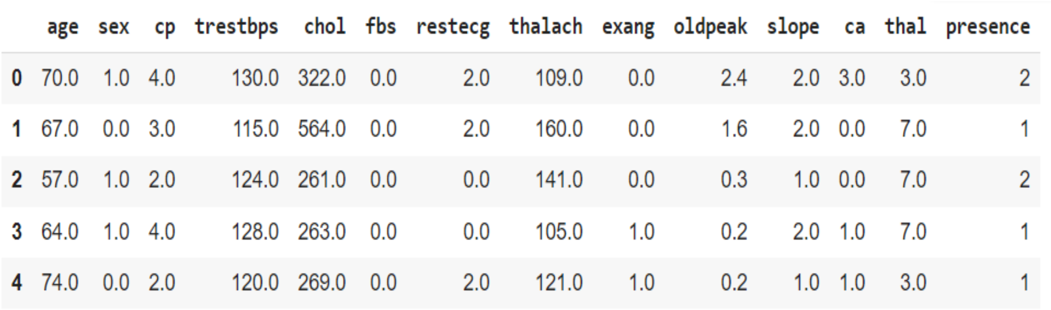
To demonstrate the significance of the proposed model, experimental results and discussion were analyzed in this chapter. The proposed model was implemented in the Jupyter Notebook python programming platform due to challenges faced by the limitation of computation power by the personal computer. The python code had to be implemented in the Google Colab platform having a graphic processing unit and TensorFlow processing unit.

The Python version of the Colab environment was 3.7.13, having libraries described in chapter 3.2 like the Numpy version 1.21.6 for numerical calculations, the pandas 1.3.5, Matplotlib 3.2.2, Seaborn 0.11.2, and Scikit-Learn 1.0.2 for the prediction of the CVD. Additionally, Python provided other libraries for data preprocessing, partitioning, normalization, and feature selection using Pearson's correlation equation. The machine learning stacking library imported was vecstack version 0.4.0. The stacking library combined the seven algorithms with a meta-classifier. The Scikit-Learn aided in providing the metrics for the model evaluation like the recall, accuracy, precision, and f1 measure. There was a 10 K-fold cross-validation for hyperparameter tuning of the proposed model prototype. Finally, the Pickle library stored the proposed model prototype in the final prediction while Flask 1.1.4 executed and displayed the presence or absence of CVD in an HTML layout.

4.0 RESULTS AND DISCUSSION

4.1 Data Description and Retrieval

The CVD dataset was downloaded from the University of California, Irvine, Machine Learning Repository. Figure 4a and 4b indicates the 13 attributes that were used to develop a robust stacked model for CVD prediction.



	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	presence
0	70.0	1.0	4.0	130.0	322.0	0.0	2.0	109.0	0.0	2.4	2.0	3.0	3.0	2
1	67.0	0.0	3.0	115.0	564.0	0.0	2.0	160.0	0.0	1.6	2.0	0.0	7.0	1
2	57.0	1.0	2.0	124.0	261.0	0.0	0.0	141.0	0.0	0.3	1.0	0.0	7.0	2
3	64.0	1.0	4.0	128.0	263.0	0.0	0.0	105.0	1.0	0.2	2.0	1.0	7.0	1
4	74.0	0.0	2.0	120.0	269.0	0.0	2.0	121.0	1.0	0.2	1.0	1.0	3.0	1

Figure 4a: The CVD Head Data Sample

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	presence
265	52.0	1.0	3.0	172.0	199.0	1.0	0.0	162.0	0.0	0.5	1.0	0.0	7.0	1
266	44.0	1.0	2.0	120.0	263.0	0.0	0.0	173.0	0.0	0.0	1.0	0.0	7.0	1
267	56.0	0.0	2.0	140.0	294.0	0.0	2.0	153.0	0.0	1.3	2.0	0.0	3.0	1
268	57.0	1.0	4.0	140.0	192.0	0.0	0.0	148.0	0.0	0.4	2.0	0.0	6.0	1
269	67.0	1.0	4.0	160.0	286.0	0.0	2.0	108.0	1.0	1.5	2.0	3.0	3.0	2

Figure 4b: The CVD Tail Data Sample

The instances were 270, numbered from 0 to 269. The presence column indicates the target attribute, whereas the 13 attribute shows the independent variables. Figure 4a indicates the 1st five rows of the dataset, while Figure 4b indicates the CVD dataset's last five rows. The dataset was used to train, test, and validate the suggested CVD model

4.2 Data Preprocessing

There was the checking for null in figure 4.2a and negative values in figure 4.2b in the columns of the CVD dataset**.**

```
age         0
sex         0
cp          0
trestbps    0
chol        0
fbs         0
restecg     0
thalach     0
exang       0
oldpeak     0
slope       0
ca          0
thal        0
presence    0
dtype: int64
```

```
age         0
sex         0
cp          0
trestbps    0
chol        0
fbs         0
restecg     0
thalach     0
exang       0
oldpeak     0
slope       0
ca          0
thal        0
presence    0
dtype: int64
```

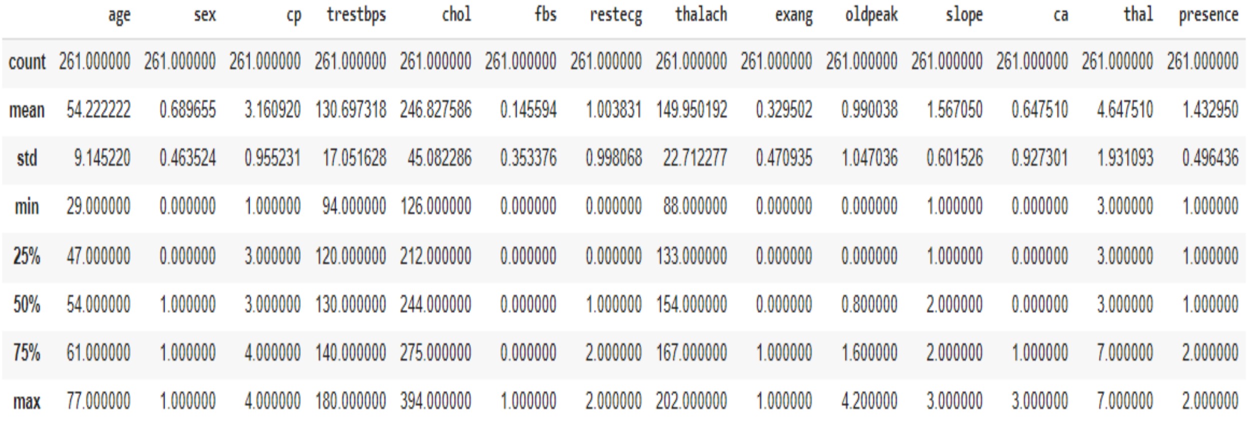
Figure 4.2a: Null Values

Figure 4.2b: Negative Values

The dataset had neither negative nor null values. The proposed model was trained on a clean dataset, ensuring it is more robust and accurate than the models in chapter 2 of the research study.

The z-score of the columns was calculated, and it was difficult to determine which data points were outliers in the two by 2 number arrays. Thus, a threshold of 3 was defined, and 9 data points were detected and eliminated. The removal of outliers helped in getting better results.

There were basic statistics computations applied to the model. They included the count, mean, standard deviation, minimum, and maximum data entries per column.



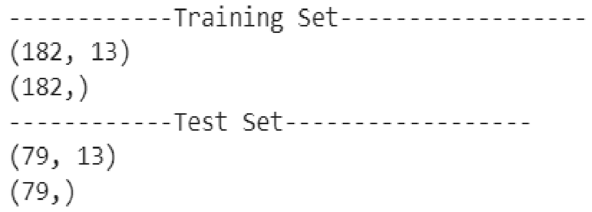
	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	presence
count	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000
mean	54.222222	0.689655	3.160920	130.697318	246.827586	0.145594	1.003831	149.950192	0.329502	0.990038	1.567050	0.647510	4.647510	1.432950
std	9.145220	0.463524	0.955231	17.051628	45.082286	0.353376	0.998068	22.712277	0.470935	1.047036	0.601526	0.927301	1.931093	0.496436
min	29.000000	0.000000	1.000000	94.000000	126.000000	0.000000	0.000000	88.000000	0.000000	0.000000	1.000000	0.000000	3.000000	1.000000
25%	47.000000	0.000000	3.000000	120.000000	212.000000	0.000000	0.000000	133.000000	0.000000	0.000000	1.000000	0.000000	3.000000	1.000000
50%	54.000000	1.000000	3.000000	130.000000	244.000000	0.000000	1.000000	154.000000	0.000000	0.800000	2.000000	0.000000	3.000000	1.000000
75%	61.000000	1.000000	4.000000	140.000000	275.000000	0.000000	2.000000	167.000000	1.000000	1.600000	2.000000	1.000000	7.000000	2.000000
max	77.000000	1.000000	4.000000	180.000000	394.000000	1.000000	2.000000	202.000000	1.000000	4.200000	3.000000	3.000000	7.000000	2.000000

Figure 4.3: Statistical computation

Figure 4.3 illustrates eight statistical computation rows for every column in the CVD dataset. The count is the total number of entries in each column after the dataset was cleansed in figures 4.2a and 4.2b. The count depicts no missing values, and no outliers in the dataset used to train the proposed stacked model. The mean depicted the average of the data entries in each column. Mean values calculated the standard deviation, and it measured how numerically the data values were spread out in the CVD dataset. The maximum and minimum values assisted in designing the HTML prediction layout in that the input values had to be within the set range. The percentiles, 25%, 50 %, and 75 %, were explained analogously, and they helped in the exploratory data analysis of the CVD dataset.

4.3 Data Partitioning and Normalization

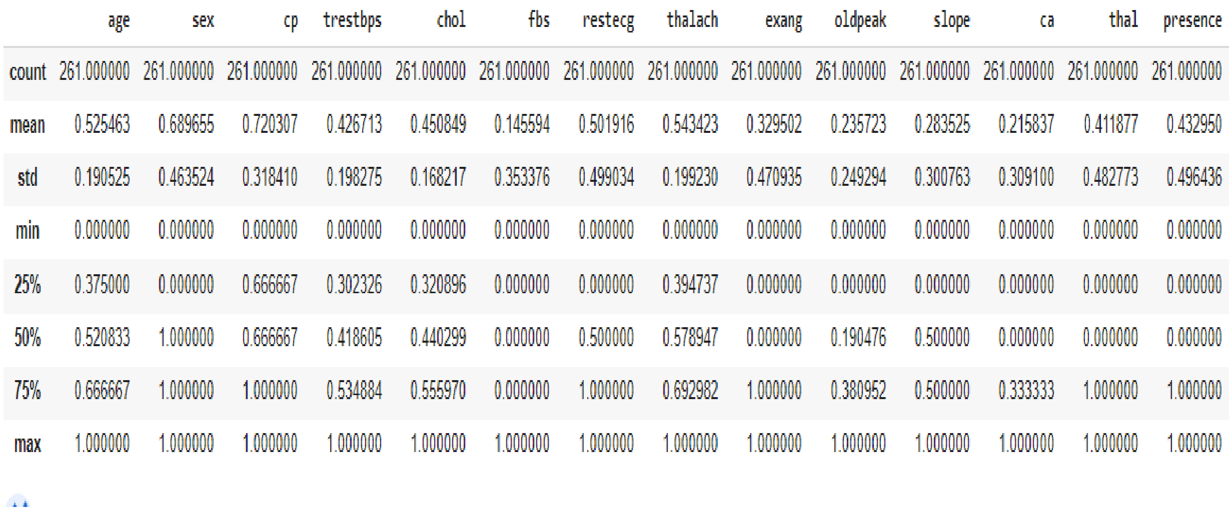
After the outliers were truncated and removed, the total number of instances was reduced to 261. They implied that 70% was used for training and 30 % for testing and validation. The calculation of the partition was 182 for training and 79 for testing, as indicated in figure 4.4.



```
------------Training Set-----------------
(182, 13)
(182,)
------------Test Set------------------
(79, 13)
(79,)
```

Figure 4.4: CVD Dataset partition

The train and test splitting prevented the model from overfitting or underfitting, reducing the generalization error and thus improving its prediction capability.



	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	presence
count	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000	261.000000
mean	0.525463	0.689655	0.720307	0.426713	0.450849	0.145594	0.501916	0.543423	0.329502	0.235723	0.283525	0.215837	0.411877	0.432950
std	0.190525	0.463524	0.318410	0.198275	0.168217	0.353376	0.499034	0.199230	0.470935	0.249294	0.300763	0.309100	0.482773	0.496436
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.375000	0.000000	0.666667	0.302326	0.320896	0.000000	0.000000	0.394737	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	0.520833	1.000000	0.666667	0.418605	0.440299	0.000000	0.500000	0.578947	0.000000	0.190476	0.500000	0.000000	0.000000	0.000000
75%	0.666667	1.000000	1.000000	0.534884	0.555970	0.000000	1.000000	0.692982	1.000000	0.380952	0.500000	0.333333	1.000000	1.000000
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Figure 4.5: Data Normalization

The dataset was normalized to a scale of 0 to 1 as shown in figure 4.5. This ensured the 13 features consistency and standardization for better training and testing of the proposed stacked model.

4.4 Feature Selection

The feature selection indicated the contribution of each feature in a scale of -1 to 1. With -1 having the highest negative correlation while 1 having the highest positive correlation.

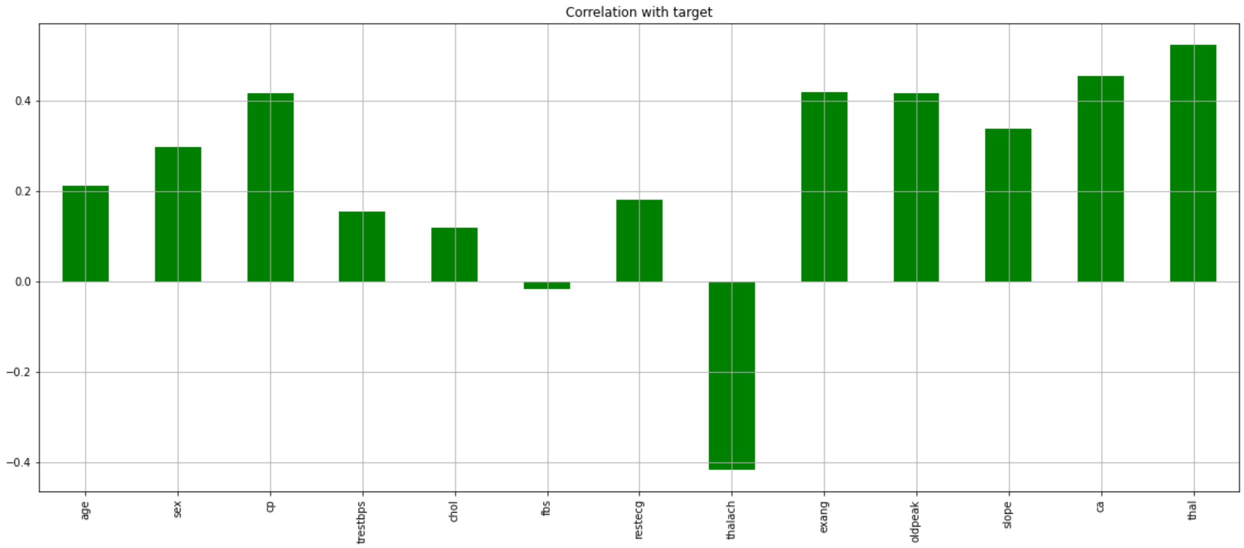


Figure 4.6: The Pearson Correlation

Fbs and thalach attributes are negatively correlated with the target variable, whereby thalach has the least correlation. Age, sex, trestbps,chol, and restecg fell into the category of weak correlation with the target attribute. Cp, slope, ca, exang and oldpeak were lowly correlated with the presence variable. The significance correlation was between the target attribute and thal. There was no high correlation in the 13 attributes to the prediction outcome variable.

The attained feature selection results can be used in future research as described in the conclusion section. However, the proposed model was implemented by all the features.

4.5 Machine Learning Algorithms Evaluation

There were 4 performance evaluation metrics as indicated in research specific objective ii). They included accuracy, precision, recall, and f1 measure. The metrics were used to determine the research question in section 1.5

1. **Accuracy**

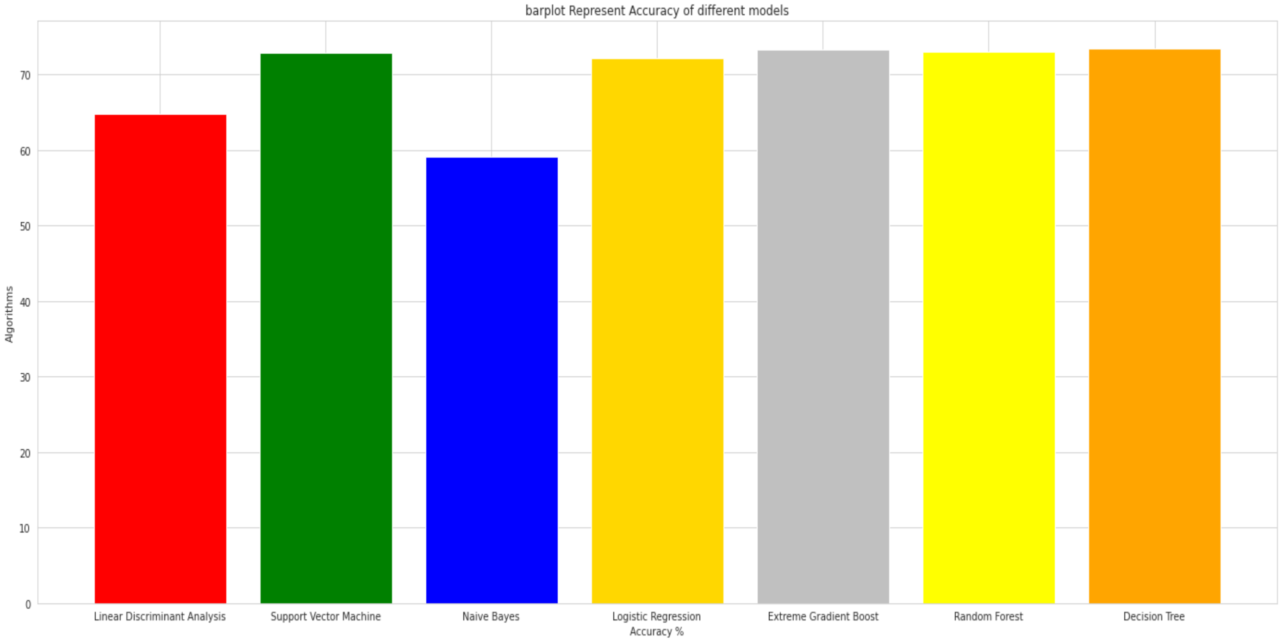


Figure 4.7a Accuracy of the 7 conventional algorithms

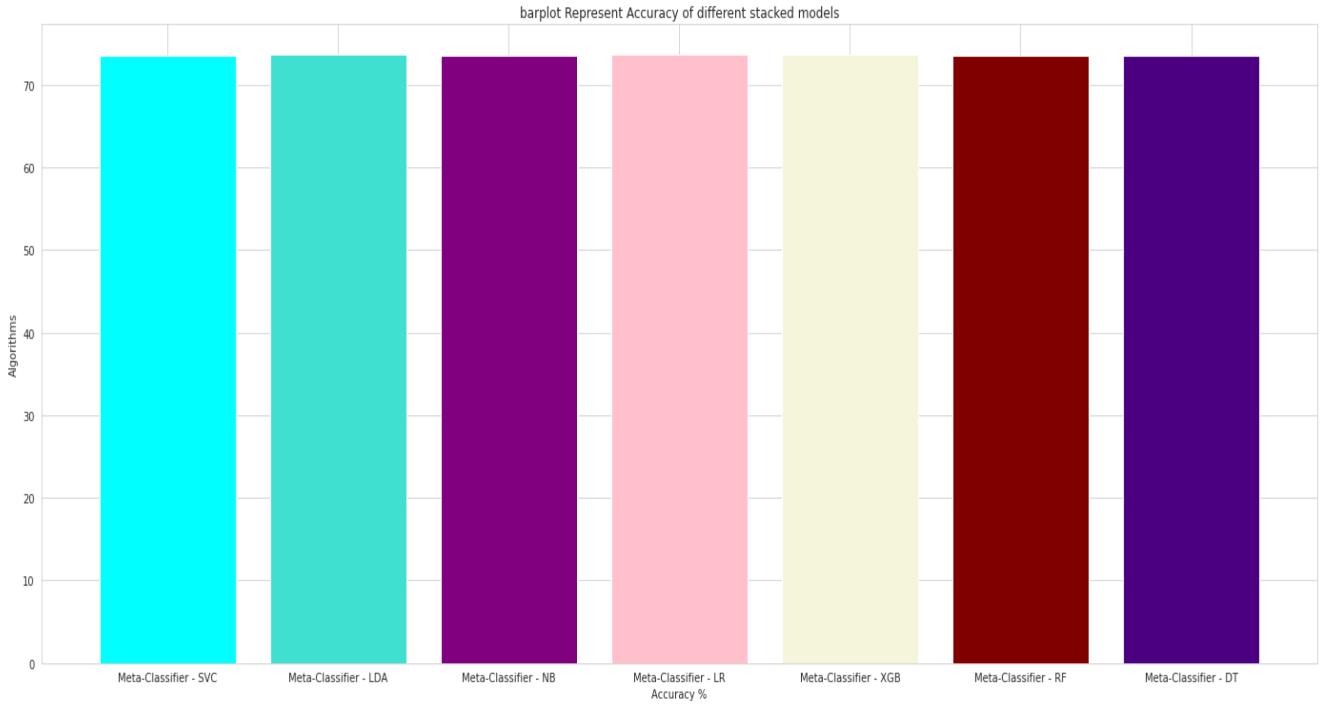


Figure 4.7b Accuracy of the 7 Stack algorithms

Figure 4.7a depicted the best performing conventional model to be DT with 73.39%, followed by XGB with 73.25%. RF had 73.02 %, SVC - 72.87%, LR - 72.0 %, LDA - 64.72 %, and NB - 59.02%.

From figure 4.7b Stacking with XGB had an accuracy of 73.62%, followed by stacking with either LDA or LR having 73.59%. Both SVC and DT stacking had an accuracy of 73.54%. Stacking with NB obtained an accuracy of 73.59%, and the least performing stacked model was with RF having 73.46%. Stacking with NB improved the accuracy of the conventional NB from 59.02 to 73.59%; this improved the model performance and applicability whereby the features were to be trained independently.

Table 4.1: % Accuracy of the 7 Stacked and Conventional ML Algorithms

No.	Conventional ML Algorithms	Conventional ML Accuracy	Stacked ML Accuracy	Increase/Decrease
0	LDA	64.719048	73.590476	+8.871428
1	SVC	72.866667	73.542857	+0.676190
2	NB	59.019048	73.504762	**+14.48571**
3	LR	72.090476	73.590476	+1.500000
4	XGB	73.247619	**73.619048**	+0.371429
5	RF	73.019048	73.457143	+0.438095
6	DT	**73.385714**	73.542857	+0.157143

In all the stacking with conventional models, there was an improvement in the prediction accuracy, making the stacked algorithm relevant and more efficient than conventional algorithms as indicated in Table 4.1.

2. **Precision**

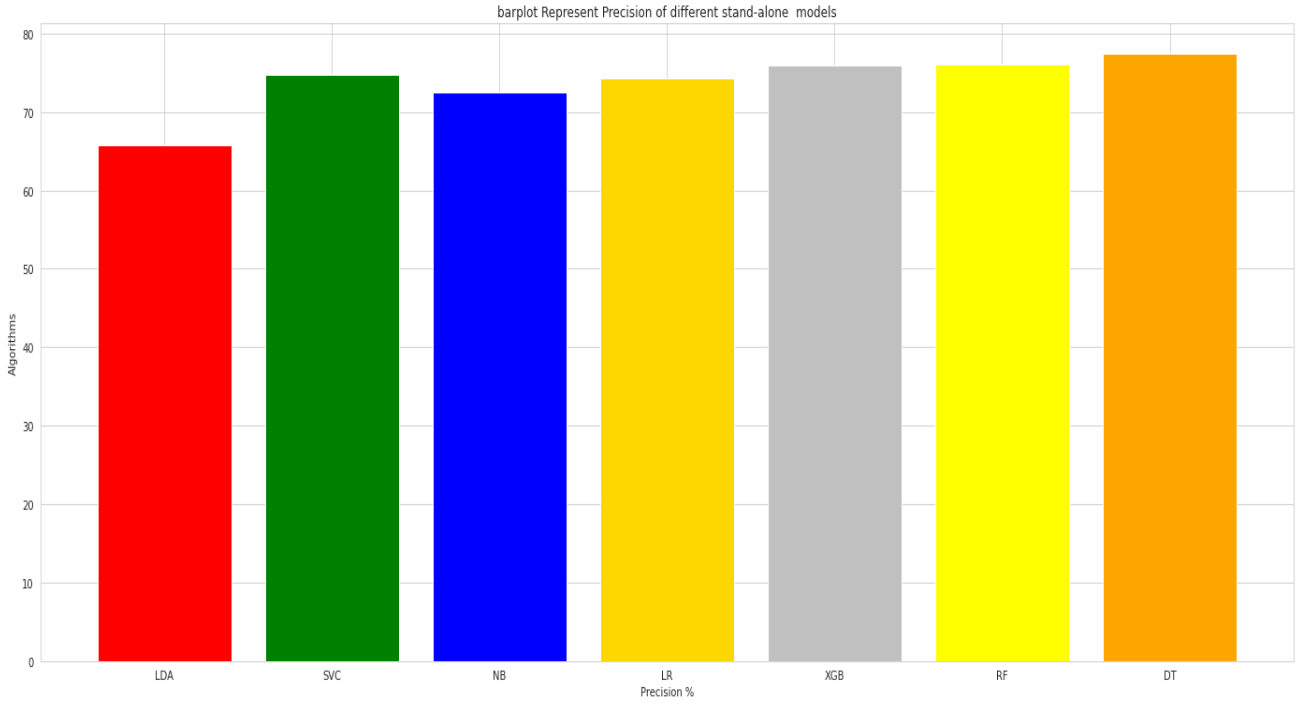


Figure 4.8a Precision of the 7 conventional algorithms

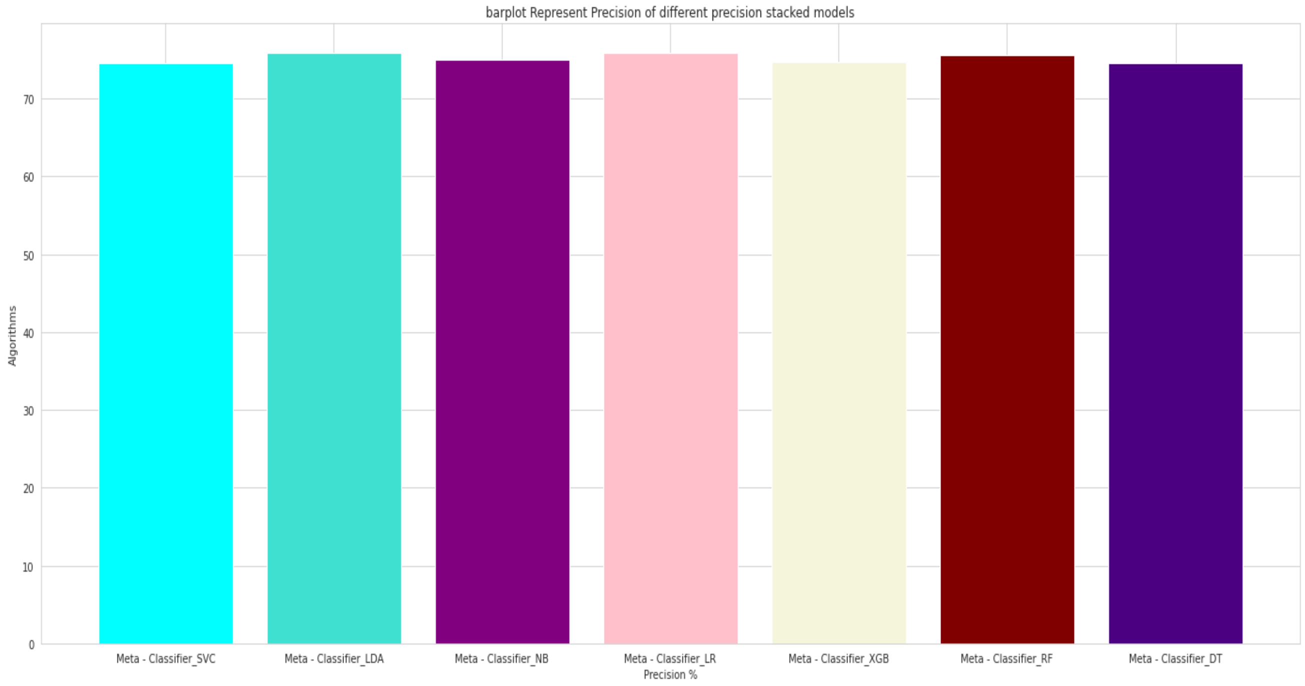


Figure 4.8b Precision of the 7 Stacked algorithms

Figure 4.8a and 4.8b indicated that the best performing precision algorithm is the conventional DT model, having 77.41 % Precision. Followed by conventional RF with 76.13 %, then XGB - 75.90%, stacking with either LDA or LR having 75.87% precision, stacking with RF-75.61 %, stacking with NB - 75.05 %, stacking with XGB - 74.72%, conventional SVC - 74.66%, stacking with SVC - 74.54 %, stacking with DT - 73.54 %, stand - alone LR - 74.29 %, conventional NB 72.47%, and conventional LDA - 65.80 precision.

Table 4.2: % Precision of the 7 Stacked and Conventional ML Algorithms

	Conventional ML Algorithms	**Conventional ML Precision**	**Stacked ML Precision**	**Increase/Decrease**
0	LDA	65.797101	**75.865698**	+10.068597
1	SVC	74.660868	74.537269	-0.123599
2	NB	72.473952	75.048593	+2.574641
3	LR	74.287223	**75.865698**	+1.578475
4	XGB	75.901166	74.716223	-1.184943
5	RF	76.125095	75.605156	-0.519939
6	DT	**77.413910**	74.527453	-2.886457

Stacking with LDA, NB, and LR proved an improvement, whereas stacking with SVC, XGB, RF, and DT decreased the performance. For precision, the proposed model recommends the use of conventional algorithms. However, stacking with LDA showed a significantly increased precision of 10.07 percent as shown in table 4.2.

3. **Recall**

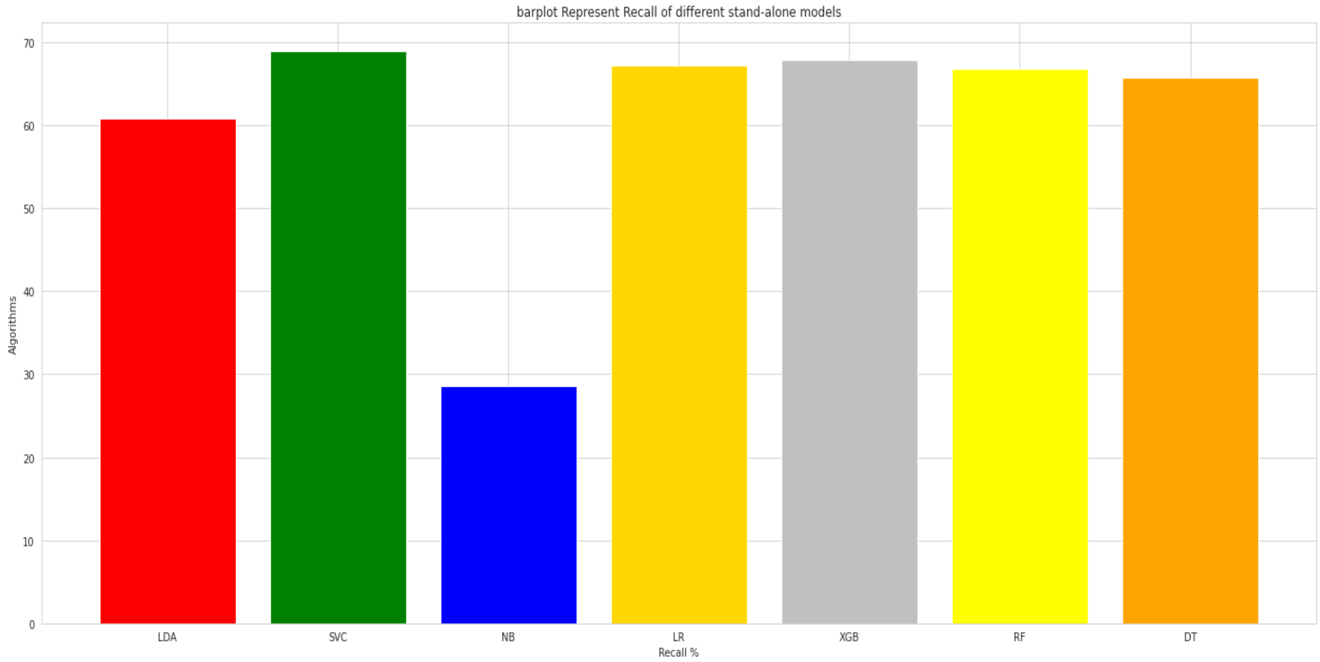


Figure 4.9a Recall of the 7 conventional algorithms

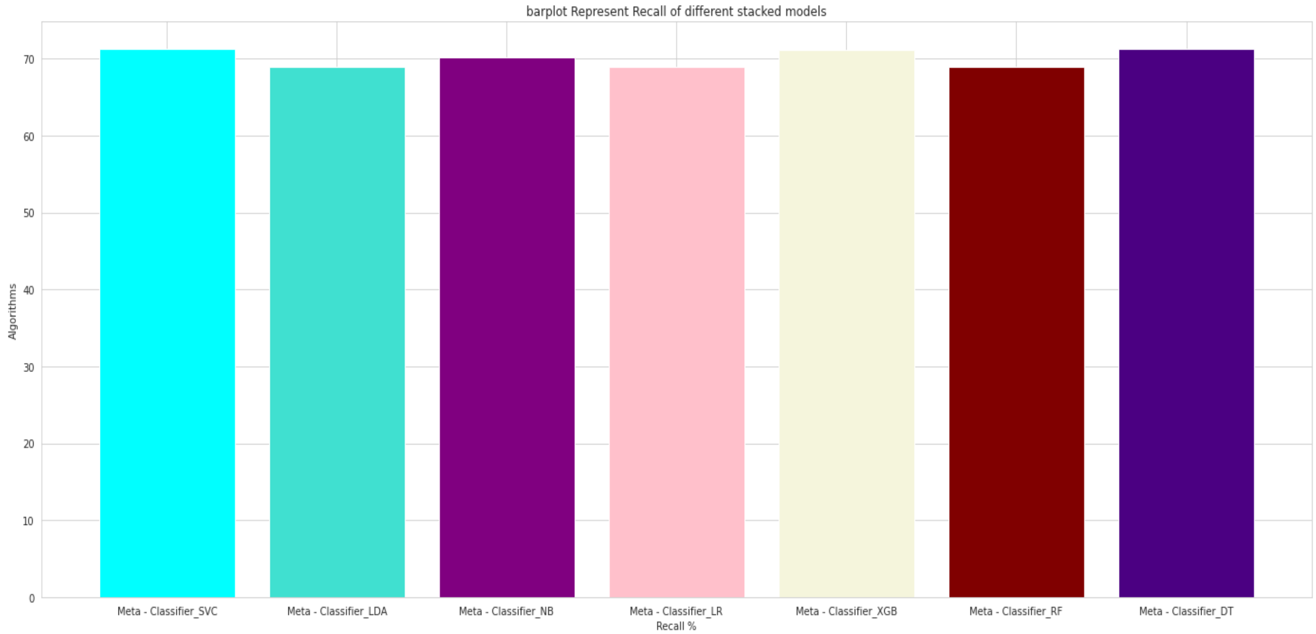


Figure 4.9b Recall of the 7 Stacked algorithms

Figure 4.9a and 4.9b indicated that the best performing recall was stacking with DT - 71.23%, followed by stacking SVC - 71.22% , stacking with XGB- 71.10%, stacking with NB - 70.13%, stacking with RF - 68.97%, conventional SVC - 68.92%, stacking with either LDA or LR - 68.90 %, conventional XGB - 67.83%, conventional RF - 66.78%, conventional LR - 67.25 %, conventional DT - 65.76 %, conventional LDA - 60.76 %, and conventional NB - 28.59 % in respective order of recall performance.

Table 4.3: % Recall of the 7 Stacked and Conventional ML Algorithms

	Conventional ML Algorithms	**Conventional ML Recall**	**Stacked ML Recall**	**Increase/Decrease**
0	LDA	60.759010	68.903547	+8.144537
1	SVC	68.922665	71.216901	+2.294236
2	NB	28.591913	70.127139	**+41.535226**
3	LR	67.249785	68.903547	+1.653762
4	XGB	67.832903	71.102189	+3.269286
5	RF	66.781378	68.970462	+2.189084
6	DT	65.758532	**71.236020**	+5.477488

All the stacking proved to have better recall results than conventional models. There was an improvement in every algorithm that was stacked. NB stacking had the most improvement with more than 41 % increase, with stacked LR slightly improving as shown in table 4.3. The stacked recall results indicated that stacking is effective and robust, specifically reducing false negatives.

4. F1 Score

The algorithm that had the highest F1 measure score was the stacked one with XGB - 72.86%, followed by stacking with either SVC or DT- 72.84 %, stacked with NB - 72.50 %, stacked with LR or LDA - 72.22 %, stacked with RF - 72.14 % as shown in figure 5a

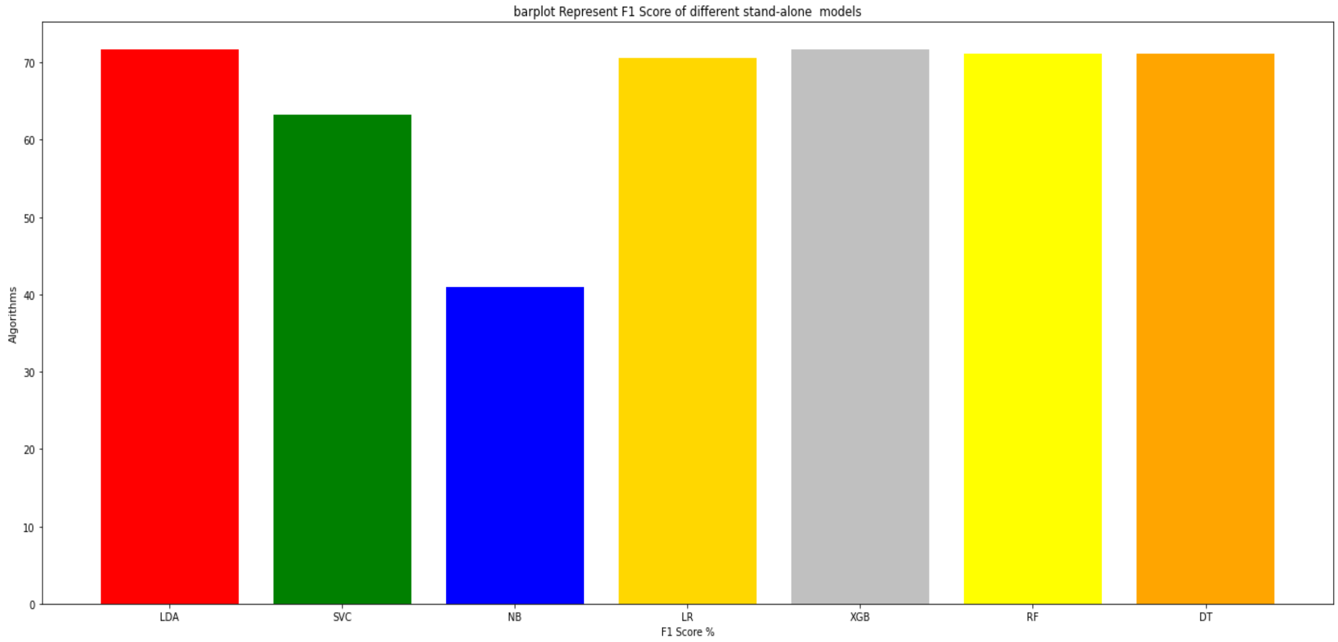


Figure 5a F1 measure of the 7 conventional algorithms

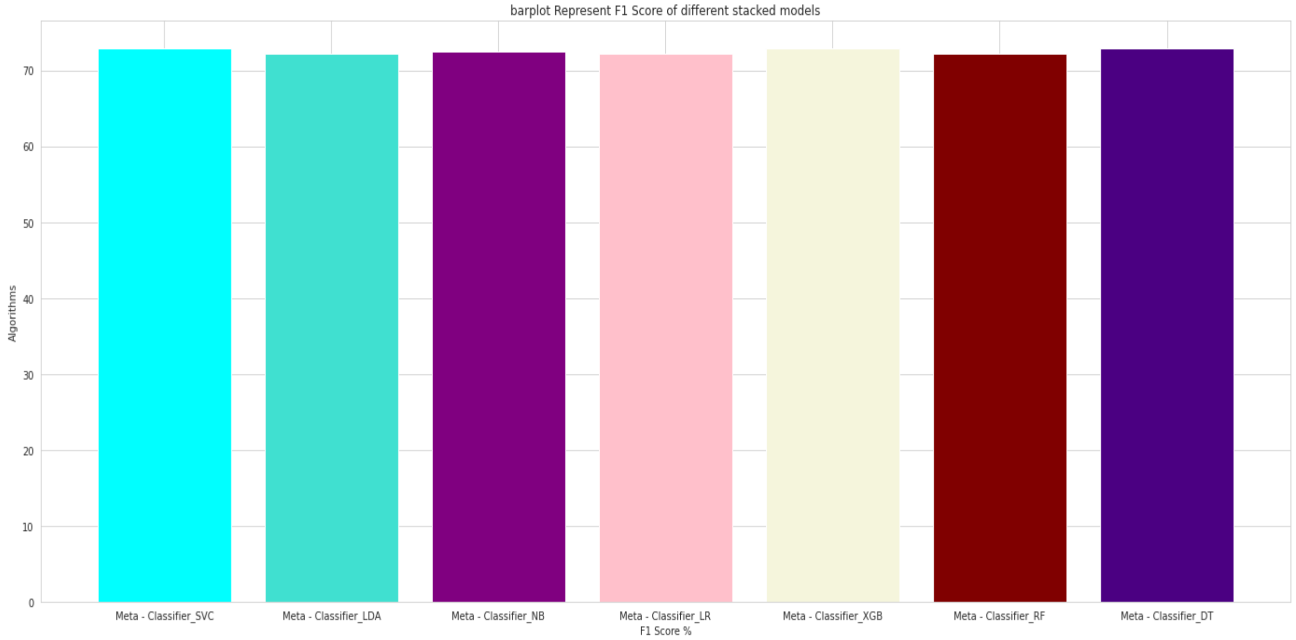


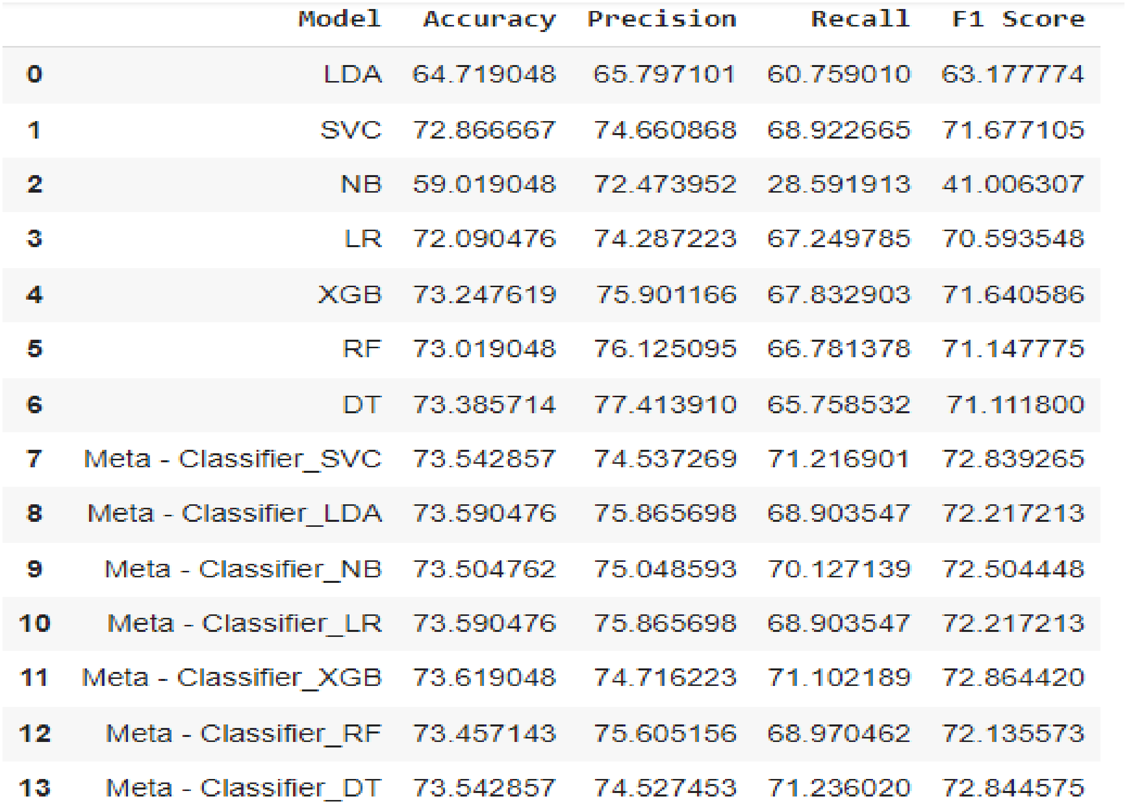
Figure 5b F1 measure of the 7 Stacked algorithms

For the conventional algorithms the f1 measure was conventional SVC - 71.68 %, conventional XGB - 71.64%, conventional RF - 71.15 %, conventional DT - 71.11 %, conventional LR - 70.59 %, conventional LDA - 63.18 %, and conventional NB - 41.01 % as shown in figure 5b.

Table 4.4: % F1 Score of the 7 Stacked and Conventional ML Algorithms

	Conventional ML Algorithm	**Conventional ML F1 Score**	**Stacked ML F1 Score**	**Increase/Decrease**
0	LDA	63.177774	72.217213	+9.039439
1	SVC	**71.677105**	72.839265	+1.16216
2	NB	41.006307	72.504448	+31.498141
3	LR	70.593548	72.217213	+1.623665
4	XGB	71.640586	**72.864420**	+1.223834
5	RF	71.147775	72.135573	+0.987798
6	DT	71.111800	72.844575	+1.732775

All the stacked algorithms outperformed the conventional algorithms in the F1 score measure as shown in table 4.4. The best performing algorithm in the criterion was one stacked with XGB, the least performing stacked algorithm was one with RF, which was still higher than the best performing conventional algorithm. Every stacked algorithm, improved positively, with the highest improvement depicted by the stacked NB and the lowest positive improvement by stacked with RF. The criterion proved that the stacked algorithms provide a great trade-off between reducing false negatives and positives. Hence aided in making the proposed model to have a good trade-off between variance and bias.



	Model	Accuracy	Precision	Recall	F1 Score
0	LDA	64.719048	65.797101	60.759010	63.177774
1	SVC	72.866667	74.660868	68.922665	71.677105
2	NB	59.019048	72.473952	28.591913	41.006307
3	LR	72.090476	74.287223	67.249785	70.593548
4	XGB	73.247619	75.901166	67.832903	71.640586
5	RF	73.019048	76.125095	66.781378	71.147775
6	DT	73.385714	77.413910	65.758532	71.111800
7	Meta - Classifier_SVC	73.542857	74.537269	71.216901	72.839265
8	Meta - Classifier_LDA	73.590476	75.865698	68.903547	72.217213
9	Meta - Classifier_NB	73.504762	75.048593	70.127139	72.504448
10	Meta - Classifier_LR	73.590476	75.865698	68.903547	72.217213
11	Meta - Classifier_XGB	73.619048	74.716223	71.102189	72.864420
12	Meta - Classifier_RF	73.457143	75.605156	68.970462	72.135573
13	Meta - Classifier_DT	73.542857	74.527453	71.236020	72.844575

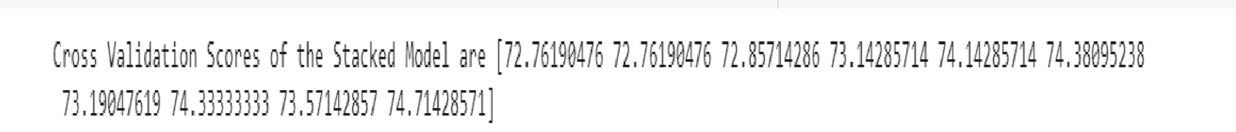
Figure 5.1: All the four performance metrics of the 7 conventional and Stacked algorithms.

Stacking improves the classification accuracy of the CVD prediction algorithms to a range of between 73.46% and 73.61%. The conventional models had a range of between 59.02% and 73.39 %. This showed an improvement with the best performing conventional algorithm having slightly lower accuracy than the lowly performing stacked algorithm. Stacking with XGB, LDA, and LR produced the highest accuracy, respectively, when compared to other conventional and stacked algorithms. LDA and LR stacking had similar evaluation metrics results. For precision, DT had the best performing metric with 77.41 %; however, other conventional algorithms performed poorly, with some algorithms like the LDA having a precision of 65.80 %. The conventional algorithm precision was improved to a range of between 74.53% and 75.87% by the stacking technique, indicating an improvement in the precision results of the research study.

In recall metrics, NB had 28.59%, and SVC had 68.93 %; this was the lower range of conventional algorithms, proving the fact that is stacking algorithms are superior predictors because stacking with either LDA or LR had the lowest performance of 68.90%, and stacking with DT obtained the highest score of 71.23%. With such high performance in the recall, the stacked algorithms gave more true positive results than the conventional algorithms. For the F1 measure, NB performed poorly with 41.01 %, and SVC had the best performance in the conventional category with 71.68%. This, however, was out-performed by the stacked algorithms. All the stacked algorithms had more than 72 % F1 measure, with the lowest stacked algorithm having 72.13%, stacking with RF, and the highest stacked algorithm, with XGB, having 72.86%.

Precision and Recall are inversely proportional to each other, and hence arithmetically hard to obtain an increase in both of them (Rajagopal, 2020). Precision is utilized when the algorithm's objective is to reduce false positives. In comparison, recall is used when the objective is to reduce false negatives. The f1 score was an ideal trade-off between precision and accuracy. A trade-off had to be made between reducing false positives or false negatives (Papageorgiou, 2018). When the model requires a reduction in false positives, then from the empirical results, then conventional DT is recommended. Whereas, when there is a need for reducing the false negative, stacking with DT suits best.

Stacking with XGB proved to be best in two out of the four performance evaluation metrics; classification accuracy and F1 measure. The stacked algorithms outperformed the conventional algorithms in classification accuracy, recall, and F1 measure. However, in precision metrics, DT was the best performing model. As such, choosing stacking with XGB was deemed appropriate.

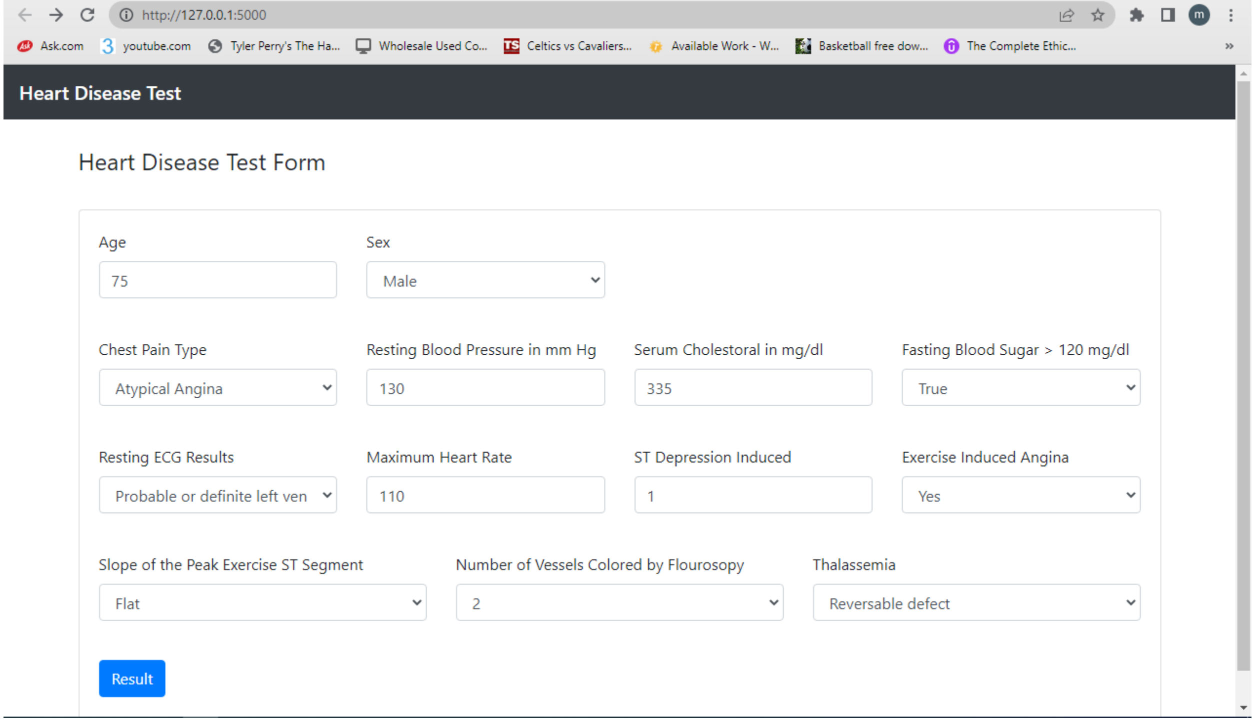


```
Cross Validation Scores of the Stacked Model are [72.76190476 72.76190476 72.85714286 73.14285714 74.14285714 74.38095238
 73.19047619 74.33333333 73.57142857 74.71428571]
```

Figure 5.2: 10 KFold Cross-Validation

The accuracy of the best chosen cross-validated algorithm improved from 72.76 to 74.71 after performing ten k fold iterations, as shown in figure 5.2. That made the model more credible, robust and accurate than the conventional algorithm.

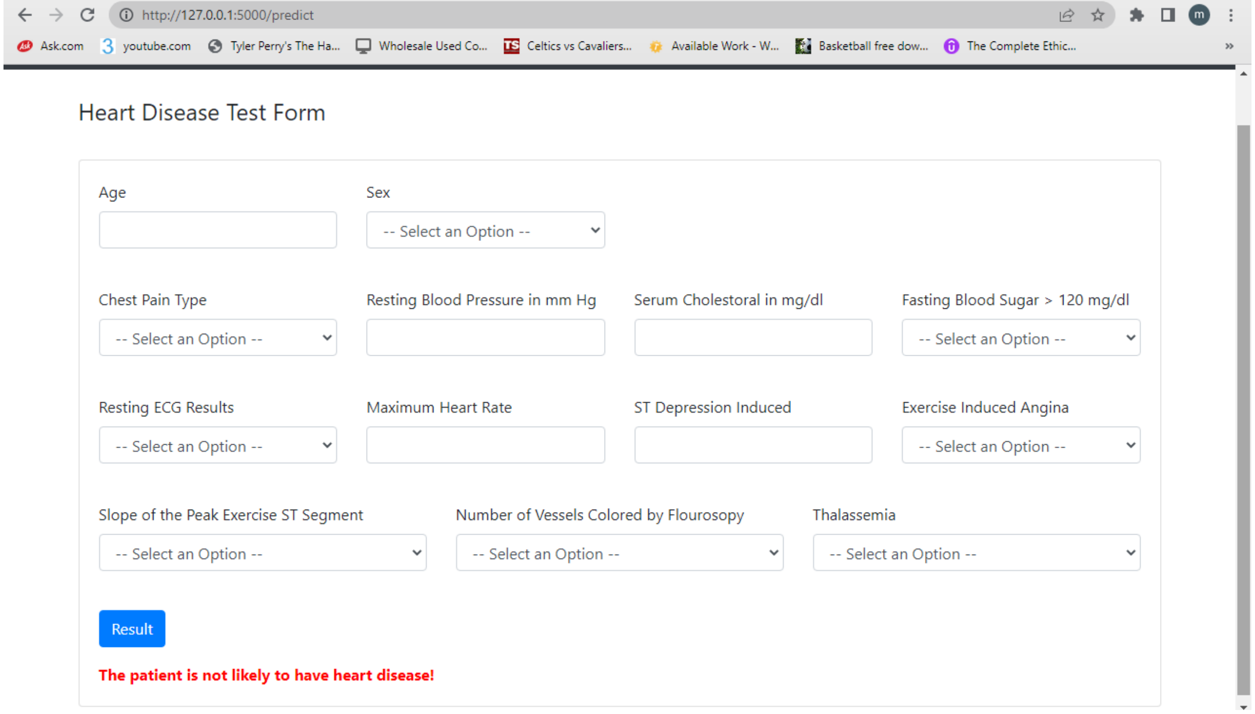
4.6 Final Prediction and Experimental results



Heart Disease Test Form

Age: 75
Sex: Male
Chest Pain Type: Atypical Angina
Resting Blood Pressure in mm Hg: 130
Serum Cholestoral in mg/dl: 335
Fasting Blood Sugar > 120 mg/dl: True
Resting ECG Results: Probable or definite left ven
Maximum Heart Rate: 110
ST Depression Induced: 1
Exercise Induced Angina: Yes
Slope of the Peak Exercise ST Segment: Flat
Number of Vessels Colored by Flourosopy: 2
Thalassemia: Reversable defect

Figure 5.3a Sample of Experimental results



Heart Disease Test Form

Age:
Sex: -- Select an Option --
Chest Pain Type: -- Select an Option --
Resting Blood Pressure in mm Hg:
Serum Cholestoral in mg/dl:
Fasting Blood Sugar > 120 mg/dl: -- Select an Option --
Resting ECG Results: -- Select an Option --
Maximum Heart Rate:
ST Depression Induced:
Exercise Induced Angina: -- Select an Option --
Slope of the Peak Exercise ST Segment: -- Select an Option --
Number of Vessels Colored by Flourosopy: -- Select an Option --
Thalassemia: -- Select an Option --

The patient is not likely to have heart disease!

Figure 5.3b Results of the sample experiment

The best performing algorithm, which was stacked with XGB, was saved on the local disk using the Python pickle library. The HTML template was created and stored in a templates folder on the homepage of the python code folder. An API framework was used to communicate the pickle algorithm and the HTML template. When the user submitted a sample request on the template, as shown in figure 5.3a, the API used the POST request to communicate with the stacked XGB model. Finally, the request was processed, and a result was returned, as indicated in Figure 5.3b indicating that the patient was not likely to have CVD pathology.

The graphical user interface provided the interface of a robust machine learning model that predicts and supports medical practitioners as a CVD diagnostic tool.

4.7 Comparative Analysis of Stacked Prototype Model

The research of (Liu et al., 2022) and this study both aim to solve the challenges faced by conventional algorithms through validating the stacking model, so the differences and similarities of the research by(Liu et al., 2022) are described in section 4.7

Table 4.5: Benchmark comparison of the proposed model with previous study

Study	**Technique**	**# Features**	**Conventional ML**	**Cross-Validation**	**Accuracy**	**Precision**	**Recall**	**F1 Score**
(Liu et al., 2022)	Stacking	10	10	Used 2 datasets for validation	84.62%	86%	86%	86%
This	Stacking	13	7	10	74.71%	74.72%	71.10%	72.86%

The proposed research study had some similarities with the proposed stacking ensemble of (Liu et al., 2022) as depicted in table 4.5. First of all, they both use the stacking technique to improve accuracy of the proposed model. They similarly used publicly available datasets of UCI machine learning algorithms. Although, (Liu et al., 2022) validated using another heart dataset. This is different from the 10 K-Fold cross validation.

The proposed model had a data partitioning of 7:3 while the model suggested by Liu et al had a partition of 8:2. For feature selection the proposed model used Pearson Correlation formula while Liu et al used Lloyd Shapley's game theory to identify the relevant features.

.

Another difference is that the classification models involved in the two studies are not identical since their stacking used ten conventional machine learning algorithms while the proposed study utilized seven of the ML algorithms. Besides, because more stacked algorithms were compared that out-performed the proposed stacked algorithms , this study put forward suggestions of using more stacked ML algorithms in future.

Both studies used similar performance evaluation metrics. Liu et al's stacking out-performed the proposed model in all the four metrics. This is as a result of differences in the ML algorithms that have different equations, formulas as was described in section 3.1.5.

5.0 SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1 Summary Of Findings

objective 1: To stack seven conventional machine learning algorithms with seven different meta-classifiers

LDA, SVM, NB, LR, XGB, RF,and DT were the seven ML algorithms that were investigated in the research that aided in building the proposed model. The seven ML algorithms were stacked with a meta-classifier, seven of them at a time. The meta-classifiers were the same ML algorithms stacking the seven algorithms in a chronological order, each one a time and they were not repeatedly stacked more than once. This meant that it was a two-level stacking technique.

Objective 2: To evaluate the performance of the seven conventional machine learning algorithms and seven stacked machine learning algorithms with four evaluation metrics; classification accuracy, precision, recall, and f1 measure.

The performance evaluation metrics depicted that the suggested two-level stacked algorithms outperformed the conventional algorithms in three out of the four metrics. This included classification accuracy, recall, and f1 measure. For precision, the conventional DT algorithm outperformed the stacked algorithms. This showed the inverse relationship between precision and recall.

objective 3: To develop a stacked model prototype from the best performing machine learning algorithms for CVD diagnosis

A total of 14 ML algorithms were evaluated and the best performing one chosen for stacking the CVD prediction model. The Stacking with XGB deemed appropriate as it outperformed all the 14 ML algorithms in the performance evaluation metrics

objective 4: To perform ten K-fold cross-validation to fine-tune the stacked prototype model

The accuracy of the best chosen cross-validated algorithm improved by approximately 2 %, after performing ten K fold iterations.. That made the model more credible, robust and accurate than the conventional algorithm.

objective 5: To validate the CVD prediction capability of the stacked prototype model with a previous study

Both the proposed model and the compared one used the same stacking technique. Other similarities included; similar performance evaluation metrics and the use of publicly available datasets. The differences between the two models, ranging from number of attributes used, the different kinds of ML algorithms, train and splitting ratio, cross validation techniques used, and feature selection techniques, brought about the differences in the results.

5.2 Conclusions

Research Question: Which are the best performing algorithms for CVD prediction modeling, between conventional and stacked ensemble algorithms?

The performance evaluation metrics depicted that the suggested two-level stacked model had ML algorithms that outperformed the conventional algorithms in three out of the four metrics. The underperformed metric resulted from the conventional need for a trade-off between recall and precision metrics. The good trade-off illustrated that the stacking of conventional algorithms combined their advantages to yield better prediction results.

5.3 Limitations

The main challenge of the research was the limitation of computational power and time by the personal computer, as indicated in section 3.3. The limitation was however solved by using the google colab platform which reduced the amount of time used to design the proposed model by the help of Graphical and Tensorflow Processing Unit.

5.4 Future Work

In future work, more data should be collected, retrieved, and researched in deep learning to investigate the pros and cons of the proposed model. There should be focus on building and designing neural multi-level stacking and deep learning algorithms for the model.

5.5 Recommendations

The proposed stacked model had slightly above average prediction capability, accuracy, and efficiency for identifying CVD pathology. The cross-validated proposed model can assist in the primordial prevention strategy of the World Heart Federation and medical practitioners as a diagnostic tool for CVD pathology.

References

2002 - 2012 John Hunter, Darren Dale, Eric Firing, Michael Droettboom, and the Matplotlib development team; 2012 - 2021 The Matplotlib development team. (2002–2021). Users guide — Matplotlib 3.5.1 documentation. Https://Matplotlib.Org/. https://matplotlib.org/stable/users/index.html

Abdelaziz, A., Elhoseny, M., Salama, A. S., & Riad, A. M. (2018). A machine learning model for improving healthcare services on cloud computing environment. *Measurement, 119,* 117-128

Adede, C., Oboko, R., Wagacha, P. W., & Atzberger, C. (2019, December 8). Model Ensembles of Artificial Neural Networks and Support Vector Regression for Improved Accuracy in the Prediction of Vegetation Conditions and Droughts in Four Northern Kenya Counties. MDPI. https://www.mdpi.com/2220-9964/8/12/562

Ahlin, M., & Ranby, F. (2019). Predicting Marketing Churn Using Machine Learning Models. Diva-Portal.Org. https://www.diva-portal.org/smash/record.jsf?pid=diva2:1335397

Almeida, R., Teodoro, A. C., Gonc¸alves, H., Freitas, A., Sa-Sousa, A., Jacome, C., & Fonseca, J. (2017, April 27). Forecasting Asthma Hospital Admissions from Remotely Sensed Environmental Data. Scitepress. https://www.scitepress.org/Papers/2017/63852/63852.pdf

Ahmad, P., Qamar, S., & Rizvi, S. Q. A. (2015). Techniques of data mining in healthcare: a review. *International Journal of Computer Applications, 120*(15).

Aryal, S., Alimadadi, A., Manandhar, I., Joe, B., & Cheng, X. (2020). The machine learning strategy for gut microbiome-based diagnostic screening of cardiovascular disease. *Hypertension, 76*(5), 1555-1562.

Asgari, E., & Mofrad, M. R. (2015). Continuous distributed representation of biological sequences for deep proteomics and genomics. *PloS one, 10*(11), e0141287.

Azevedo, A. (2019). Data mining and knowledge discovery in databases. In *Advanced Methodologies and Technologies in Network Architecture, Mobile Computing, and Data Analytics,* (pp. 502-514). IGI Global.

Alizadehsani, R., Roshanzamir, M., Abdar, M., Beykikhoshk, A., Khosravi, A., Panahiazar, M., Koohestani, A., Khozeimeh, F., Nahavandi, S., & Sarrafzadegan, N. (2019, October 23).

A database for using machine learning and data mining techniques for coronary artery disease diagnosis. *Scientific data*, *6*(1), 1-13.

Bengfort, B. (2019, March 24). Yellowbrick: Visualizing the Scikit-Learn Model Selection Process. Journal of Open Source Software. https://joss.theoj.org/papers/10.21105/joss.01075

Cao, C., Liu, F., Tan, H., Song, D., Shu, W., Li, W., Zhou, Y., Bo, X., & Xie, Z. (2018). Deep Learning and Its Applications in Biomedicine. *Genomics, proteomics & bioinformatics*, *16*(1), 17–32.

Cardiovascular diseases. Who.int. (2021). Retrieved 1 September 2021, from https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1.

Chang, W. (2019, November 7). A Machine-Learning-Based Prediction Method for Hypertension Outcomes Based on Medical Data. MDPI. https://www.mdpi.com/2075-4418/9/4/178

Chauhan, Y. J. (2018). Cardiovascular Disease Prediction using Classification Algorithms of Machine Learning. *International Journal of Science and Research (IJSR), ISSN,* 2319-7064.

Chithambaram, T., & Gowsalya, M. (2020). Heart Disease Detection Using Machine Learning.

Chowdhury, M. N. R., Ahmed, E., Siddik, M. A. D., & Zaman, A. U. (2021, April). Heart Disease Prognosis Using Machine Learning Classification Techniques. In *2021 6th International Conference for Convergence in Technology (I2CT)* (pp. 1-6). IEEE.

Cui, W., Sun, Z., Ma, H., & Wu, S. (2020). *The Correlation Analysis of Atmospheric Model Accuracy Based on the Pearson Correlation Criterion*. IOPSCIENCE. https://iopscience.iop.org/article/10.1088/1757-899X/780/3/032045/pdf

Devi, S. K., Krishnapriya, S., & Kalita, D. (2016). Prediction of heart disease using data mining techniques. *Indian Journal of Science and Technology, 9*(39), 1-5.

Dua, D. and Graff, C. (2019). *UCI Machine Learning Repository* [http://archive.ics.uci.edu/ml]. *Irvine, CA: University of California, School of Information and Computer Science*.

Flask. (2022, April 28). PyPI. https://pypi.org/project/Flask/

Ghosh, P., Azam, S., Jonkman, M., Karim, A., Shamrat, F. J. M., Ignatious, E., ... & De Boer, F. (2021). Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques. *IEEE Access, 9,* 19304-19326.

Gupta, A. (2021, December 14). Stacking Ensemble-Based Intelligent Machine Learning Model for Predicting Post-COVID-19 Complications. SpringerLink. https://link.springer.com/article/10.1007/s00354-021-00144-0?error=cookies_not_supported&code=7c87cdbd-5bb2-4632-a9f0-f9f7085a9cca

Home - World Heart Federation. World Heart Federation. (2021). Retrieved 1 September 2021, from https://world-heart-federation.org/.

Hu, Z., Qiu, H., Su, Z., Shen, M., & Chen, Z. (2020). *A Stacking Ensemble Model to Predict Daily Number of Hospital Admissions for Cardiovascular Diseases*. IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9149874

Hua, Y., Guo, J., & Zhao, H. (2015, January). Deep belief networks and deep learning. In *Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things* (pp. 1-4). IEEE.

Jain, D., & Singh, V. (2018, April 5). *Feature selection and classification systems for chronic disease prediction:* A review. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S1110866517300294

Jiang, S. (2020). Heart Disease Prediction Using Machine Learning Algorithms. Escholarship.Org. https://escholarship.org/content/qt6n98m5ck/qt6n98m5ck_noSplash_fc31320b5d7a89485160d91bfed7adaf.pdf?t=q78snk

Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., ... & Dudley, J. T. (2018). Artificial intelligence in cardiology. *Journal of the American College of Cardiology, 71*(23), 2668-2679.

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and structural biotechnology journal, 15*, 104-116.

Khan, M. M. R., Arif, R. B., Siddique, M. A. B., & Oishe, M. R. (2018, September). Study and observation of the variation of accuracies of KNN, SVM, LMNN, ENN algorithms on eleven different datasets from the UCI machine learning repository. In *2018 4th International Conference on Electrical Engineering and Information & Communication Technology (iCEEiCT)* (pp. 124-129). IEEE.

Kumar, S. (2018, December 19). Correlation Analysis to Identify the Effective Data in Machine Learning: Prediction of Depressive Disorder and Emotion States. MDPI. https://www.mdpi.com/1660-4601/15/12/2907

Li, J. (2020). Research on a Heart Disease Prediction Model Based on the Stacking Principle. Diva-Portal.Org. https://www.diva-portal.org/smash/record.jsf?pid=diva2:1454488

Liu, J., Dong, X., Zhao, H., & Tian, Y. (2022, April 13). *Predictive Classifier for Cardiovascular Disease Based on Stacking Model Fusion*. MDPI. https://www.mdpi.com/2227-9717/10/4/749/pdf

Maghdid, S. S. (2019). Intelligent System For Identification Heart Diseases. ResearchGate. https://www.researchgate.net/profile/Sozan-Maghdid/publication/343080708_Intelligent_System_For_Identification_Heart_Diseases/links/5f15c6594585151299aaef34/Intelligent-System-For-Identification-Heart-Diseases.pdf

Maharjan, A. (2020, May). Machine Learning Approach for Predicting Cancer Using Gene Expression - ProQuest. ProQuest. https://www.proquest.com/openview/ce8553c88a8337b4f3e75fe7b112f616/1?pq-origsite=gscholar&cbl=18750&diss=y

Maini, E., Venkateswarlu, B., Marwaha, D., & Maini, B. (2021, October 28). Upgrading the Performance of Machine Learning Based Chronic Disease Prediction Systems using Stacked Generalization Technique. University of Bahrain. https://journals.uob.edu.bh/handle/123456789/3953

Mehmood, A., Iqbal, M., Mehmood, Z., Irtaza, A., Nawaz, M., Nazir, T., & Masood, M. (2021). Prediction of Heart Disease Using Deep Convolutional Neural Networks. *Arabian Journal for Science and Engineering, 46*(4), 3409-3422.

Muibideen, M. A. (2019, August 8). Prediction of Heart Disease using Bayesian Network Model. AUST- African University of Science and Technology Research. https://repository.aust.edu.ng/xmlui/handle/123456789/4903

Nadeem, M. W., Goh, H. G., Ponnusamy, V., Andonovic, I., Khan, M. A., & Hussain, M. (2021, October 18). A fusion-based machine learning approach for the prediction of the onset of diabetes. MDPI. Retrieved November 24, 2021, from https://www.mdpi.com/2227-9032/9/10/1393.

Narain, R. (2018, July). Development of Cardiovascular Disease Prediction System. THAPAR INSTITUTE OF ENGINEERING & TECHNOLOGY. Retrieved November 24, 2021, from http://tudr.thapar.edu:8080/jspui/handle/10266/6093.

Nguyen, Q. H., Ly, H., Lanh, H. S., & Al-Ansari, N. (2021, February). Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. ResearchGate. https://www.researchgate.net/publication/349125804_Influence_of_Data_Splitting_on_Performance_of_Machine_Learning_Models_in_Prediction_of_Shear_Strength_of_Soil

Nilashi, M., bin Ibrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). An analytical method for diseases prediction using machine learning techniques. *Computers & Chemical Engineering, 106,* 212-223.

NumPy Developers. (2008–2022). What is NumPy? — NumPy v1.22 Manual. Numpy.Org. https://numpy.org/doc/stable/user/whatisnumpy.html

Osho, O., & Hong, S. (2021, April). An Overview: Stochastic Gradient Descent Classifier, Linear Discriminant Analysis, Deep Learning, and Naive Bayes Classifier Approaches to Network Intrusion Detection. ResearchGate. https://www.researchgate.net/publication/351103983_An_Overview_Stochastic_Gradient_Descent_Classifier_Linear_Discriminant_Analysis_Deep_Learning_and_Naive_Bayes_Classifier_Approaches_to_Network_Intrusion_Detection

Pandey, S. K., & Janghel, R. R. (2019). Recent deep learning techniques, challenges and its applications for medical healthcare system: a review. *Neural Processing Letters, 50*(2), 1907-1935.

Papageorgiou, V. (2018, June 1). The limitations of dermoscopy: falseâpositive and falseânegative tumours. Wiley Online Library. https://onlinelibrary.wiley.com/doi/abs/10.1111/jdv.14782

Patro, K. G. S. (2015, March 19). Normalization: A Preprocessing Stage. ArXiv.Org. https://arxiv.org/abs/1503.06462

Pavlyshenko, B. (2018). *Using Stacking Approaches for Machine Learning Models*. IEEE. https://ieeexplore.ieee.org/abstract/document/8478522

Premsmith, J., & Ketmaneechairat, H. (2021). A Predictive Model for Heart Disease Detection Using Data Mining Techniques. Semanticscholar. https://pdfs.semanticscholar.org/1392/0587a5794f92f39e0db2e71b845a85625617.pdf

Python Software Foundation. (2001-2022). The Python Tutorial — Python 3.10.2 documentation. Python.Org. https://docs.python.org/3/tutorial/

Rajagopal, S. (2020, January 24). A Stacking Ensemble for Network Intrusion Detection Using Heterogeneous Datasets. Hindawi. https://www.hindawi.com/journals/scn/2020/4586875/\

Ren, Q., Li, M., & Han, S. (2019, February 19). Tectonic discrimination of olivine in basalt using data mining techniques based on major elements: a comparative study from multiple perspectives. Taylor & Francis. https://www.tandfonline.com/doi/full/10.1080/20964471.2019.1572452

Ricciardi, C., Valente, A. S., Edmund, K., Cantoni, V., Green, R., Fiorillo, A., Picone, I., Santini, S., & Cesarelli, M. (2020, January 23). Linear discriminant analysis and principal component analysis to predict coronary artery disease. SAGE Journals. https://journals.sagepub.com/doi/full/10.1177/1460458219899210

Santos, S. G., Barros, R. S., & Júnior, P. M. G. (2015, November). Optimizing the parameters of drift detection methods using a genetic algorithm. In *2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 1077-1084). IEEE.

Schmidt, J., Marques, M. R., Botti, S., & Marques, M. A. (2019). Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials, 5*(1), 1-36.

Shah, D., Patel, S., & Bharti, S. K. (2020). Heart disease prediction using machine learning techniques. *SN Computer Science, 1*(6), 1-6.

Sharma, M., Singh, G., & Singh, R. (2017). Stark assessment of lifestyle based human disorders using data mining based learning techniques. *IRBM, 38*(6), 305-324.

Sharma, S. (2021, April 28). Heart diseases prediction using hybrid ensemble learning. Dublin Business School. https://esource.dbs.ie/handle/10788/4219.

Shultana, S., Moharram, M. S., & Neehal, N. (2020). Olympic sports events classification using convolutional neural networks. *In Proceedings of International Joint Conference on Computational Intelligence* (pp. 507-518). Springer, Singapore.

Swamy, H. (2021, December 11). How to Deploy a Logistic Regression Model in GCP - Analytics Vidhya. Medium. https://medium.com/analytics-vidhya/insiders-view-on-logistic-regression-and-how-do-we-deploy-regression-model-in-gcp-as-batch-c62a64563210

the pandas development team. (2008–2021). IO tools (text, CSV, HDF5, . . .) — pandas 1.3.5 documentation. Pydata.Org. https://pandas.pydata.org/docs/user_guide/io.html#csv-text-files

Tian, X. (2019, June 11). Using Machine Learning Algorithms to Predict Hepatitis B Surface Antigen Seroclearance. Hindawi.Com. https://www.hindawi.com/journals/cmmm/2019/6915850/

Uddin, S. (2019, December 21). Comparing different supervised machine learning algorithms for disease prediction - BMC Medical Informatics and Decision Making. SpringerLink. https://link.springer.com/article/10.1186/s12911-019-1004-8?error=cookies_not_supported&code=457c3465-59a2-4b34-b526-c46a7d598f77\

vecstack. (2019, August 12). PyPI. https://pypi.org/project/vecstack/#history

Vincent, A. (2020, December 11). Medical Decision Support Using Machine Learning. Http://Www.Theses.Fr/. https://www.theses.fr/2020LYSEM029\

Wang, J., Liu, C., Li, L., Li, W., Yao, L., Li, H., & Zhang, H. (2020, February 20). A Stacking-Based Model for Non-Invasive Detection of Coronary Heart Disease. IEEE Access. https://ieeexplore.ieee.org/abstract/document/9005221

Wang, M. (2019, December 26). A novel model for malaria prediction based on ensemble algorithms. PLOS ONE. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0226910#sec005\

Ware, S., Rakesh, S. k, & Choudhary, B. (2020, January). Heart Attack Prediction by using Machine Learning Techniques. Retrieved November 24, 2021, from https://www.ijrte.org/wp-content/uploads/papers/v8i5/D9439118419.pdf.

Waskom, M. (2012–2021). seaborn: statistical data visualization — seaborn 0.11.2 documentation. Seaborn.Pydata. https://seaborn.pydata.org/

Yassin, N. I., Omran, S., El Houby, E. M., & Allam, H. (2018). Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. *Computer methods and programs in biomedicine, 156,* 25-45.

Yatsko, V. A. (2017). Bayes theorem and a methodology for prediction of US presidential elections results. https://www.researchgate.net/profile/Viatcheslav-Yatsko-2/publication/321148606_Bayes_theorem_and_a_methodology_for_prediction_of_US_presidential_elections_results/links/5dce038892851c382f3e086f/Bayes-theorem-and-a-methodology-for-prediction-of-US-presidential-elections-results.pdf

Yu, Z. (2022). Prediction of Airway Management of Trauma Patients Based on Machine Le. SpringerLink. https://link.springer.com/chapter/10.1007/978-981-16-8656-6_12?error=cookies_not_supported&code=6bb9b631-17f2-4efc-905b-0f7a564e88f0

Zoghi, Z. (2020, August). Ensemble Classifier Design and Performance Evaluation for Intrusion Detection Using UNSW-NB15 Dataset - ProQuest. ProQuest. https://www.proquest.com/openview/7c41b65580bde34353f4a2d61523ab3f/1?pq-origsite=gscholar&cbl=18750&diss=y