

SUPER LEARNER MODEL IN PREDICTION OF HEART ATTACK BASED ON CARDIAC BIOMARKERS

Anuradha. P

Research Scholar, Dept. of Computer Science,
Avinashilingam Institute for Home Science and Higher Education for Women, deemed to be University,
Coimbatore, T.N, India
anujith72@gmail.com

Dr. Vasantha Kalyani David

Professor and HOD, Dept. of Computer Science,
Avinashilingam Institute for Home Science and Higher Education for Women, deemed to be University,
Coimbatore, T.N, India
vasanthadavid@gmail.com

Abstract

Unstable angina and/or a heart attack is caused when restricted flow of blood to the heart occurs due to the narrowed or blocked coronary arteries. On observing Electro cardiogram (ECG), ST segment Elevation Myocardial Infarction (STEMI) can be diagnosed but ECG might not show variation for Non-ST Segment Elevation Myocardial Infarction (NSTEMI). So, cardiac biomarkers could be tested in patients presenting chest pain to confirm whether heart attack or Acute Myocardial Infarction (AMI) is onset or not. Myoglobin, Troponin-I and CK-MB are sensitive biomarkers for diagnosing heart attack/ AMI within specific time frames. In this work, a novel real dataset from a hospital comprising cardiac biomarkers' values of patients was taken and Machine Learning (ML) classifiers namely Support Vector Machine, Logistic Regression (LR), XGBoost (XGB), CatBoost, Random Forest (RF), Decision Tree Classifier, Gaussian Naïve Bayes (GNB), Majority Vote Ensemble Classifier comprising of LR, XGB, GNB, RF were applied on the dataset. Then a Super Learner was designed by taking a novel combination of these classifiers. The comparison of these classifiers resulted in Super Learner outperforming the other ML classifiers. Subsequently, a graphical user interface prediction tool using the Super Learner model was designed which would guide those who have chest pain due to AMI, to undergo emergency medical care and thereby save lives.

Keywords: Heart attack; Biomarkers; Machine Learning; Super Learner.

1. Introduction

Coronary Artery Disease (CAD) is a foremost cause of death universally [Hanson *et al.* (2013)]. CAD occurs due to the accumulation of plaque (cholesterol and other deposits) in the coronary arteries which prevents oxygen rich blood from flowing to the heart. CAD includes Acute Myocardial Infarction (AMI), angina pectoris, silent cardiac ischemia and sudden cardiac death [Hanson *et al.*, (2013)].

A common indicator of AMI/ Heart Attack is Angina/Chest pain. Angina can be of stable and unstable types. Stable angina may arise due to stress or strenuous activities and would subside with rest or medication. When the blood clots in the arteries obstruct the blood supply to the heart, Unstable angina occurs. The symptoms occur while doing very little or resting. The pain often radiates to the left shoulder, neck, or arm. It aggravates over a period of a few minutes [Mythili and Malathi, (2015)].

Non-cardiac chest pain (NCCP) may occur due to gastroesophageal reflux disease, depression, anxiety, stress or stomach problems [Schey *et al.*, (2007)]. As NCCP symptoms are similar to ischemic heart disease, a thorough check-up by a cardiologist would be required [Schey *et al.*, (2007)].

ST segment Elevation Myocardial Infarction (STEMI) is caused by clogging in the coronary arteries. Changes on the Electrocardiogram (ECG) will be noticeable for this type of heart attack. STEMI causes damage to large area of the heart muscle [my.clevelandclinic.org, (2021)]. The levels of cardiac biomarkers in the blood will raise indicating an AMI.

In Non-ST Segment Elevation Myocardial Infarction (NSTEMI), changes may not be noticeable on Electrocardiogram (ECG) but there may be partial or temporary blockage and relatively small damage to the heart muscle compared to STEMI [my.clevelandclinic.org, (2021)]. Damage of the heart muscle raises the levels of the chemical markers in the blood with the help of which NSTEMI can be identified [my.clevelandclinic.org, (2021)].

Inappropriate diagnosis of patients with chest pain often leads to unfortunate admission of patients without AMI [Mythili and Malathi, (2015)]. This can be avoided by testing the Cardiac biomarkers' levels to confirm or rule out an AMI. Biomarkers have helped clinicians in rapid diagnosis of heart attack and saving the life of patients thereby reducing the mortality rate [Mythili and Malathi, (2015)].

1.1 Cardiac Biomarkers

1.1.1 Troponin-I

Cardiac troponin-I (cTnI) is a preferred biomarker for early diagnosis of AMI [Anderson *et al.*, (2007)]. cTnI is abnormal or may not rise for the first 4 to 8 hours after the onset of chest pain. It peaks at 12 to 16 hours so the test has to be repeated after 8hrs of the onset of chest pain. Also, cTnI remains high for 5 to 9 days [Larue *et al.*, (1993)] [Dolci and Panteghini, (2006)] [Jaffe *et al.*, (1996)]. Troponin-I (cTnI), when used together with other clinical information, provides a valuable diagnostic test for AMI, thereby assisting in informed clinical decision making [9].

1.1.2 Creatine Kinase- Myocardial Band

Creatine Kinase Myocardial Band (CK-MB), a cytosolic carrier protein for high-energy phosphates, is less sensitive and less specific for AMI than cardiac troponins. Healthy persons have low levels of CK-MB in the blood and if damage occurs to skeletal or cardiac muscle, CK-MB levels would elevate [Larue *et al.*, (1993)] [Dolci and Panteghini, (2006)] [Tsung, (1981)] [Surya *et al.*, (1999)].

1.1.3 Myoglobin

Myoglobin rises within 1 hour of the onset of infarction. As it also rises up in other conditions such as skeletal muscle injury and in case of renal impairment, its specificity for AMI is low [Hanson *et al.* (2013)] [Anderson *et al.*, (2007)] [Surya *et al.*, (1999)] [Wong *et al.*, (2004)] [Kruger *et al.*, (2002)] [Kontos *et al.*, (2007)] [Hamm and Katus, (1995)]. However, myoglobin can be used for diagnosis and risk stratification in the early hours after symptom onset [Dolci and Panteghini, (2006)] [Hamm and Katus, (1995)] [Apple, (1992)] [Ruzich, (1992)] [Aldous, (2013)].

1.1.4 Brain Natriuretic Peptide (BNP)

BNP testing is used to diagnose or rule out heart failure, including diastolic heart failure [Levin *et al.*, (1998)]. BNP levels are greater than 100 pg/ml when there is cardiac damage [Cowie *et al.*, (2003)] [Valli *et al.* (2001)]. Heart failure is highly unlikely for values of Brain Natriuretic Peptide (BNP) less than 100 pg/ml [20]. BNP provides prognostic information following an MI [Levin *et al.*, (1998)], Cowie *et al.*, (2003)] [de Lemos *et al.*, (2001)] [Arakawa *et al.*, (1996)]. Increased levels of BNP concentrations may also be due to cardiac ischemia [Cowie *et al.*, (2003)].

1.1.5 D-Dimer

D-Dimer and fibrinogen levels were found to be higher in patients with acute ischemic events (myocardial infarction and unstable angina) than non-ischemic patients [Bayes-Genis, A *et al.*, (2000)]. D-Dimer level > 500 microg/L had a self-determining diagnostic value for myocardial infarction [Bayes-Genis, A *et al.*, (2000)].

With reference to specific time frames, the biomarkers CK-MB, Troponin-I, and Myoglobin yield satisfactory diagnostic sensitivity [Chiu *et al.*, (2000)]. D-Dimer level > 500 microg/L had a self-determining diagnostic value for myocardial infarction [Bayes-Genis, A *et al.*, (2000)]. On combining these biomarkers, clinicians get valuable information to provide proper treatment to AMI patients [Chiu *et al.*, (2000)].

For a patient diagnosed with AMI, medication is given to reduce pain, prevent blood clots and to help blood flow through the coronary arteries. Oxygen may be given to make sure that the heart, lungs, organs and tissues stay healthy. The damage to the heart can be considerably reduced or prevented, if the treatment for heart attack is given as soon as Myocardial Infarction (MI) symptoms are first experienced.

In this work,

- A novel real-world dataset from Specialist hospital in Bengaluru, India, comprising of cardiac biomarkers was used.
- Machine Learning (ML) models, both base models and ensemble models were used to estimate the probability of Myocardial Infarction (MI) for an individual patient.
- A Super Learner consisting of novel combination of classifiers was also used in MI prediction.
- The performance of all the classifiers was evaluated on the real-world dataset comprising the biomarkers, Cleveland heart dataset from UCI ML repository and Cardiovascular Disease dataset from Mendeley data.

- A novel GUI prediction tool with the Super Learner as predictor was designed. This tool on providing the biomarker values, helps patients to know whether there is an onset of MI and take emergency treatment and thereby prevent deaths. It could be used as a support tool in making clinical decisions in patients with suspected MI.

The remaining part of this paper is structured as follows: Section 2 introduces the Machine Learning algorithms used in this work, Section 3 briefs the related work, Section 4 portrays the methodology adopted, Section 5 describes the dataset used, Section 6 discusses the results, Section 7 contains limitations and section 8 concludes the paper.

2. Machine Learning Algorithms

2.1. Support Vector Machine (SVM)

SVMs developed by Vladimir Vapnik, is a machine learning technique, used for both linear and non-linear classification. In the SVM algorithm, each data item is plotted as a point in n-dimensional space, where n is a number of features [Uddin *et al.*, (2019)]. By transforming low dimensional input space to a higher dimensional space, SVM kernel converts non-separable problem to separable problem. Then, the hyper-plane that separates the two classes is identified and classification is done.

2.2. Logistic Regression

Logistic Regression is used in classification of binary or multiple classes by employing the cost function called as sigmoid function,

$$f(x) = 1 / (1 + e^{-x}). \quad (1)$$

It is a predictive analysis algorithm based on the concept of probability.

2.3. Naive Bayes classifier

Naive Bayes classifier, based on Bayes' theorem, is a probabilistic classifier. It assumes that each predictor variable/feature contributes independently and equally to the class/outcome/target variable [Uddin *et al.*, (2019)].

2.4. Random Forest

Random forest (RF) is a tree-based ensemble classifier, which consists of a group of decision trees. Various sub-samples of the dataset are given as input to the decision trees. When a new sample has to be classified, each DT gives a classification outcome by considering different parts of that input vector and the final output of RF is made based on majority voting or average of all the trees in the forest [Uddin *et al.*, (2019)][Breiman L. (2001)].

2.5. XGBoost (Extreme Gradient Boosting)

In boosting, a strong classifier is built from several weak classifiers. Extreme Gradient Boosting, in short XGBoost, developed by Tianqi Chen and Guestrin, is an implementation of gradient boosted decision trees which gives improved performance [Tianqi Chen and Carlos Guestrin, (2016)] [Jason Brownlee, (2016)]. In XGBoost, Objective function is the sum of Training Loss and Regularization, where Training Loss is the differentiable convex loss function which computes the difference between the target y_i and the prediction \hat{y}_i . The regularization term is added in order to smooth the final learnt weights and thereby prevent over-fitting [Tianqi Chen and Carlos Guestrin, (2016)].

2.6. CatBoost

Yandex researchers and engineers developed CatBoost, a gradient boosting algorithm, which has both CPU and GPU implementations use binary decision trees as base predictors. As it does efficient vector representation of categorical data, CatBoost shows high performance when applied on categorical data [Liudmila Prokhorenkova *et al.*, (2018)] [Alexander Marz, (2020)]. With the default values of the hyper parameters, i.e., when hyperparameters are not tuned, Catboost classifier performs better than XGBoost.

2.7. Majority Vote Ensemble

Majority vote ensemble model combines the predictions of several other models considered in the ensemble. The predictions of each model in the ensemble for the respective class label is counted and the final prediction of the ensemble is the class label with the maximum votes.

3. Related Work

Lan Shou *et al.*, determined that resting rate pressure product (rRPP), a measure of cardiac workload, can be predicted using combinations of biomarkers in the blood. XGBoost model outperformed other ML models to predict rRPP [Lan Shou *et al.*, (2021)].

Asan Agibetov et al., used the XGBoost model to diagnose Cardiac amyloidosis among patients having symptoms of Heart Failure based on laboratory parameters [Agibetov *et al.*, (2020)].

In order to detect undiagnosed Heart failure with reduced ejection fraction (HFrEF), Mathis et al., used random forest, XGBoost and L1 regularized logistic regression models based on perioperative data consisting of 628 preoperative and 1,195 intraoperative features. XGBoost was found to yield AUROC value better than other models. However, as the model has low positive predictive results because of the disease being less prevalent, the authors are of the opinion that the model prediction needs to be followed by confirmatory testing with echocardiography or cardiac biomarkers [Mathis et al., (2020)].

Weng et al., applied gradient boosting machines, neural networks, random forest and logistic regression, on routine medical data of patients from UK family practices. They found that neural networks performed better in identifying patients who could benefit from preventive treatment [Weng *et al.*, (2017)].

Meeshanthini V dogan et al., used machine learning techniques on datasets from Intermountain Healthcare (IM) and Framingham Heart Study (FHS) in order to propose an integrated genetic-epigenetic model for predicting 3-year incident CHD. They showed that their proposed model performed better in identifying patients at high risk for a heart attack, compared to Framingham Risk Score (FRS) and atherosclerotic cardiovascular disease Pooled Cohort Equation (PCE) [Meeshanthini *et al.*, (2021)].

C. Beulah Christalin Latha et al., had experimented on Multilayer Perceptron, SVM, PART, Bayes Net, C4.5 and Naive Bayes as individual classifiers and also used a combination of these classifiers in ensemble techniques namely stacking, boosting, majority voting and bagging. The dataset used was Cleveland heart dataset taken from the UCI machine learning repository. On comparing the accuracy yielded by all the techniques, it was observed that the Majority voting classifier yielded an improved accuracy of 7.26% [Latha *et al.*, (2019)].

Martin.P. Than et al., proposed myocardial-ischemic-injury-index, a machine learning model, which used gradient boosting on dataset which included paired high-sensitivity cardiac troponin-I concentrations of patients along with traditional factors. The proposed model was designed to diagnose type 1 myocardial infarction [Than *et al.*, (2019)].

Alaa AM et al., used Auto Prognosis, an automated ML tool for CVD risk prediction on the UK Biobank population. On comparing the tool with traditional techniques like Framingham score and Cox PH model, the automated tool resulted in yielding improved accuracy [Alaa AM et al., (2019)].

Van der Laan et al., had proposed a Super Learner algorithm in their work in [Van der Laan et al., (2007)], where selected classifiers were modelled on v-fold cross validation and the predicted value for each validation set was stored as meta-x values and the actual target values as meta-y. Then the classifiers were modelled on the whole dataset. The target values predicted by the selected classifiers when applied on the test data was saved as meta-x values. The meta learner or Super Learner then predicted the y-values for these meta-x values and the performance was evaluated [Van der Laan et al., (2007)].

4. Methodology

The methodology adopted in this work is depicted in the flowchart given in Figure 1.

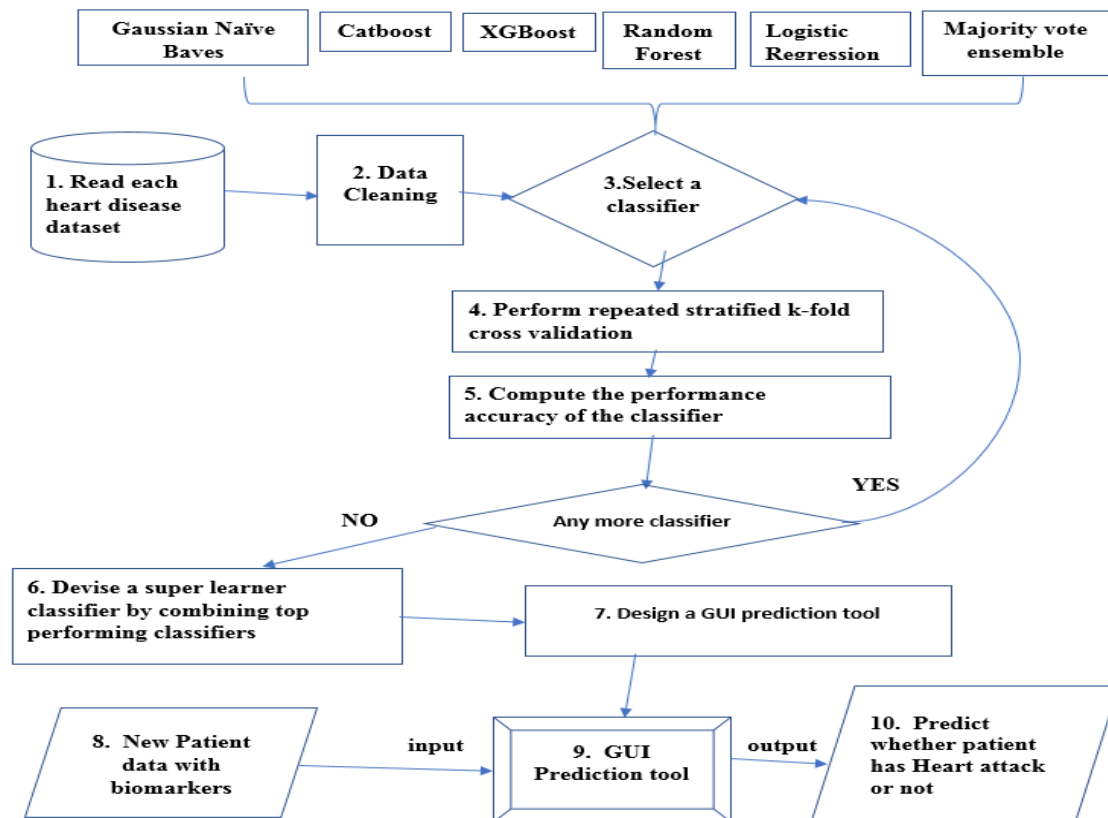


Fig. 1. Methodology

5. Dataset

On reviewing related work, it is clear that biomarkers play a vital role in predicting heart diseases. This work is aimed at developing ML model to predict heart attack/MI using a dataset containing cardiac biomarker values which would be useful to predict or rule out MI in patients presenting chest pain, especially in cases of NSTEMI. The first dataset used in this work is the data of patients who presented chest pain and had undergone lab tests, at Specialist hospital, Bangalore, in order to confirm whether they had an MI or not. The personal details of the patients were not disclosed. The data collected consists of Age, Gender, CKMB, Myoglobin, Troponin-I, BNP, D-Dimer, ACS_types, Disease, shown in Table 1, consists of 192 instances.

The normal and abnormal range for each of these biomarkers are as follows: Normal range of CKMB is 0.0 - 4.3 ng/mL and abnormal CKMB is a value >4.3 ng/mL. Normal range of Troponin-I is 0.0 - 0.4 ng/mL and abnormal Troponin-I is a value >0.4 ng/mL. Normal range of BNP is 0.0 - 100 ng/mL and abnormal BNP is a value >100 ng/mL. Normal range of D-Dimer is 0.0 - 400 ng/mL, abnormal D-Dimer is a value >400 ng/mL. Normal range of myoglobin 20-80 ng/mL and abnormal myoglobin is a value >80 ng/mL.

Two other datasets which were also used for evaluating the performance of the Super Learner are: (i) Cardiovascular disease dataset taken from Mendeley data [Doppala *et al.*, (2021)]. Table 2 exhibits the features of this dataset which has 1000 instances. (ii) Cleveland heart disease dataset at UCI Machine Learning Repository, created by Robert Detrano, M.D., Ph.D., V.A. Medical Center, Long Beach and Cleveland Clinic. This dataset has 303 instances. Table 3 displays the features of Cleveland heart dataset.

Features	Features	Features
1. Age	1. Age	1. Age
2. Gender: 0-female, 1-Male	2. Gender	2. Sex
3. CKMB- Creatine Kinase Myocardial Band	3. Chest pain	3. Fbs-fasting blood sugar
4. Myoglobin	4. Resting BP	4. Cp-chest pain type
5. Troponin-i	5. Serum cholestrol	5. Chol-serum cholesterol
6. BNP- Brain Natriuretic Peptide	6. Fasting blood sugar	6.
7. D-Dimer	7. Resting relectro	7. Trestbps- resting blood pressure
8. ACS_types- heart disease types	8. Max heart rate	8. Restecg -ecg at rest
9. Target: Disease: 0- no AMI, 1- AMI, 2- heart problems	9. Exercise angia	9. Thalach-maximum heart rate
	10. Old peak	10. Exang-exercise induced angina
	11. Slope	11. Ca-number of major vessels colored
	12. No of major vessels	12. Slope- slope of the peak exercise st segment
	13. Target	13. Thal -defect type
		14. Oldpeak-st depression induced by exercise
		15. Target: num

Table 1. Cardiac biomarkers-Heart Dataset

Table 2. Cardiovascular Disease Dataset (Mendeley Data)

Table 3. Cleveland Heart Dataset (UCI)

6. Results and Discussion

On exploring the cardiac biomarkers dataset using the WEKA tool, we observe in Figure 2 that the dataset consists of 81 females and 111 males. From figure 3, we can visualize that the age of the patients lie between 21 and 90, and more data is concentrated around the median, 55 years of age. Also, both male and female patients are found in all age groups. More people have undergone lab tests for MI between the age groups of 45 to 75.



Fig.2: Gender wise classification of data

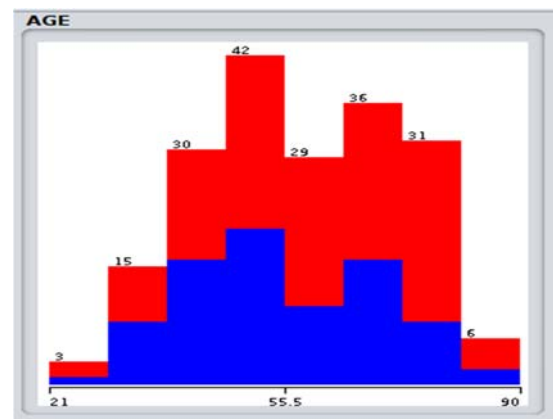


Fig.3: Age wise Distribution of patients

Figure 4 shows the count of patients with no MI(Myocardial Infarction) as 61, patients diagnosed with MI and Heart failure and in critical condition as 3, number of patients who had MI with heart muscle damage is 51, only cardiac muscle damage is 42, early MI was diagnosed in 18 patients, 7 patients had MI diagnosed early and also after 8 hrs, 1 patient was diagnosed of MI after 8 hrs, muscle/skeletal damage for 5 patients, higher level blood clots was diagnosed in 4 patients. Summing up the above data we get in figure 5, that the number of patients with no MI are 61 and belong to class 0, patients who had MI are 80 and classified as disease class 1, and those with cardiac muscle damage and clots are 51 in number and belong to class 2. The disease classes 0 and 2 have male and female patients of almost the same ratio but class 1 consisting of MI patients have more males compared to females. In the gender field, males were coded as 0 and females as 1.

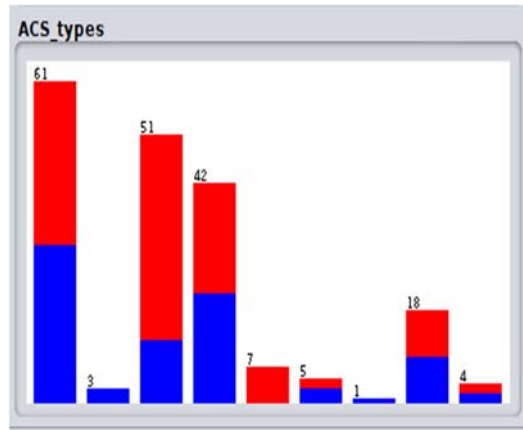


Fig.4: Heart disease types based on biomarkers' levels

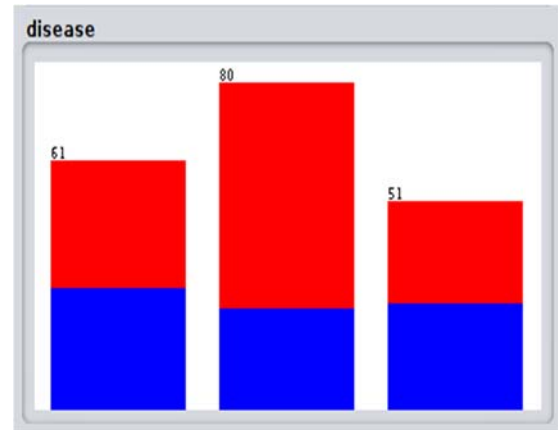


Fig.5: Heart disease types

The training and evaluation of the classifiers were done using Python on Jupyter notebook. The classifiers used were Support Vector Machine (SVM), Logistic Regression (LR), XGBoost (XGB), CatBoost (CB), Random Forest (RF), Decision Tree Classifier (DT), Gaussian Naïve Bayes (GNB), Majority Vote Ensemble Classifier comprising of LR, XGB, GNB, RF. The details of all these classifiers are found in section II. The Repeated Stratified K-Fold method was used to train and test the classifiers. The Stratified K-Fold Cross validator is repeated 3 times with different randomisation in each repetition. In this work, the Stratified K-Fold divides the entire data set into 10 splits. For each iteration it approximately maintains the same percentage of samples of each target class. In cross validation, for each iteration, 9 folds are considered as the training set and 1-fold is taken as the validation set. The number of iterations is 10, which is equal to the number of folds. The mean accuracy across all folds and all repeats is calculated for every classifier and a comparison of their performance is made. The training and testing process are repeated on Cleveland Heart dataset and Cardiovascular disease dataset. Table 4 shows the accuracy obtained by each classifier on performing repeated stratified K-fold cross validation on the three selected datasets.

Classifiers	Accuracy in the dataset		
	Cardiac biomarkers DS	Cleveland Heart DS	Cardiovascular disease DS
CatBoost (CB)	96.5%	82.5%	98.37%
XGBoost (XGB)	96.3%	80.9%	98.0%
Random Forest (RF)	95.7%	82.6%	97.8%
Decision Tree Classifier (DT)	95.7%	74.5%	96.3%
Majority Voting Classifier (VC)	92.7%	84.3%	96.1%
Support Vector Machine (SVM)	91.5%	82.9%	96.5%
Logistic Regression (LR)	86.1%	83.7%	95%
Gaussian Naive Bayes (GNB)	78.1%	83.5%	94.6%
* Super Learner (DT, CB)	97.9%	88.53%	98.8%
*Super Learner combining DT and CB classifiers and LR as meta learner			

Table 4. Performance Accuracy of the Classifiers

Then Super Learner technique proposed by Van der Laan et al., described in section 3, was tried with all combinations of the classifiers mentioned in Table 4 and Logistic regression as meta learner. The Super Learner with a combination of Decision Tree and CatBoost classifier with Logistic Regression as meta learner gave an accuracy of 97.9% on cardiac biomarkers real world dataset, 88.53% accuracy on Cleveland dataset and 98.8% accuracy on Cardiovascular disease dataset. Table 5 displays the precision, recall and fi-score and Accuracy of the Super Learner model on the three heart disease datasets.

Results show that the Super learner with DT and CB combination and LR as meta learner has consistently performed better than other classifiers in all the three datasets. The graphical form of the comparison of classifiers on the three heart disease datasets is shown in fig.6.

Datasets	class	precision	recall	f1-score	Accuracy
Cleveland Heart DS	No Disease	0.85	0.97	0.91	88.53%
	Has Disease	0.95	0.77	0.85	
Cardiac Biomarkers DS	No Heart Attack	1.00	1.00	1.00	97.9%
	Has Heart Attack	0.95	1.00	0.97	
	Has Disease	1.00	0.94	0.97	
Cardiovascular disease DS	No Disease	0.98	0.99	0.99	99%
	Has Disease	0.99	0.99	0.99	

Table 5. Classification Matrix of the Performance of the Super Learner on Heart disease datasets

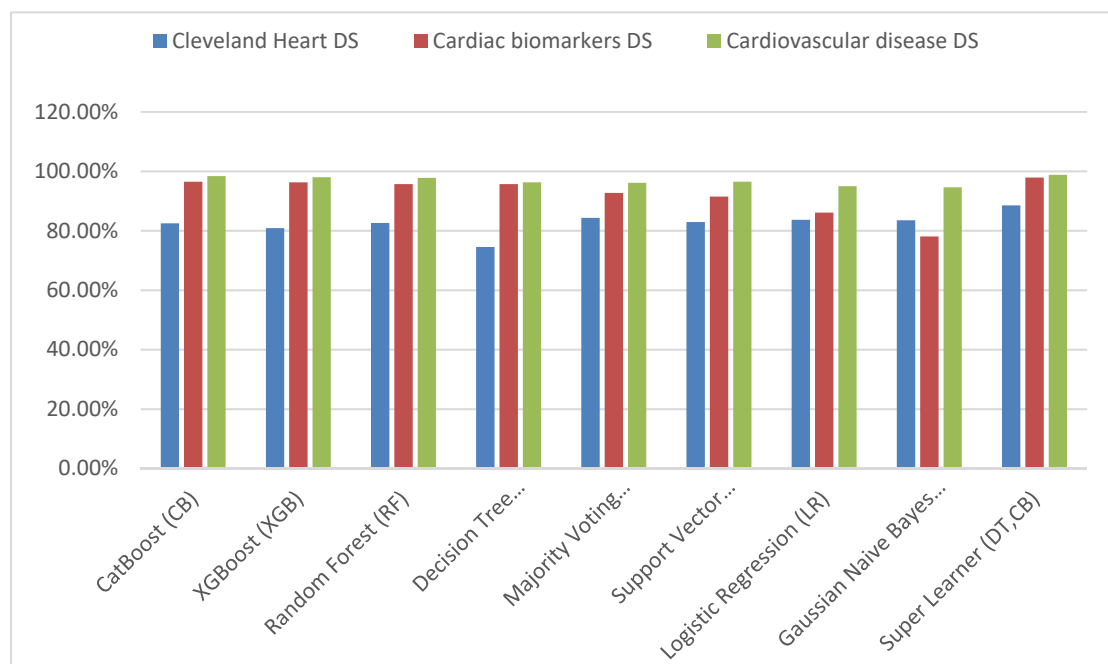


Fig.6: Comparison of Performance of the Classifiers on heart disease datasets

Using the Super Learner classifier, a GUI was created in python, using which, one can check whether he/she had chest pain due to heart attack or not, by entering the biomarkers' values. Fig.7 shows images of the prediction tool. As often people have lost lives by ignoring the chest pain, this tool can be used to check whether emergency medical care is required or not and thereby save lives. This tool can be useful especially in rural areas, where cardiac specialists are not available in the neighbourhood and the patient can rush for emergency treatment based on the prediction given in the tool. It is only a support tool which guides patients to approach the medical practitioners, who in turn would diagnose the illness and do the treatment required.

The figure displays four screenshots of the 'Heart disease prediction tool' GUI. Each screenshot shows a form with the following fields: Gender (radio buttons for MALE and FEMALE), Age, CKMB, Myoglobin, Troponin-i, BNP, and D-Dimer. A 'submit' button is located at the bottom of each form. The predicted result is shown at the bottom of each screenshot.

Gender	Age	CKMB	Myoglobin	Troponin-i	BNP	D-Dimer	Prediction
MALE	78	5	80	0.01	90	90	Predicted to have heart disease. pls consult doctor.
MALE	50	2	32	0.05	100	100	NO heart disease predicted
FEMALE	78	3	120	0.5	100	100	heart attack predicted.pls consult doctor immediately
FEMALE	69	2	67	0.01	300	200	Predicted to have heart disease. pls consult doctor.

Fig. 7. Images of the GUI heart attack prediction tool

7. Limitations

The correlation of biomarkers with other parameters like family history, cholesterol, fasting blood sugar, blood pressure was not analysed in this work. In future work it can be done in order to predict heart diseases much earlier than the onset.



8. Conclusion

A common indicator of Acute Myocardial Infarction (AMI)/ Heart Attack is Angina/Chest pain. Angina can be of stable and unstable types. Stable angina may arise due to stress or strenuous activities and would subside with rest or medication. When the blood clots in the arteries obstruct the blood supply to the heart, unstable angina occurs. Troponin-I, myoglobin and CK-MB are sensitive biomarkers for diagnosing heart attack/ AMI within specific time frames. A patient presenting chest pain when checked for the above-mentioned biomarkers' levels will be given an emergency treatment if the tests confirm a heart attack. In this work, a novel real world dataset comprising cardiac biomarker values is used and Machine Learning models namely, Support Vector Machine (SVM), Logistic Regression (LR), XGBoost (XGB), CatBoost (CB), Random Forest (RF), Decision Tree Classifier (DT), Gaussian Naïve Bayes (GNB), Majority Vote Ensemble Classifier comprising of LR, XGB, GNB, RF were applied on the dataset by performing repeated stratified cross validation. Then a Super Learner was designed by taking combinations of these classifiers. The Super Learner with Decision Tree and CatBoost classifiers combination with Logistic Regression as Meta Learner gave 97.9% accuracy. Then the same set of classifiers and Super Learner model was applied on Cleveland Heart dataset and Cardiovascular disease dataset. The Super Learner with the novel combination outperformed other classifiers in all the three datasets. As many people by being reluctant to consult medical practitioners, had ignored the chest pain and have lost lives, a graphical user interface prediction tool was designed with the Super Learner classifier for predicting heart attack. The GUI prediction tool on entering the biomarkers' values would predict whether emergency medical attention is required or not. If patients take emergency treatment in case, they are predicted to have a heart attack, their lives can be saved.

References

- [1] Hanson, MA, *et al.* (2013): Coronary artery disease. *Primary Care*, 40(1),1-16.
- [2] Sabesan Mythili; Narasimhan Malathi. (2015): Diagnostic markers of acute myocardial infarction (Review). *BIOMEDICAL REPORTS*, 3: 743-748.
- [3] Schey, R; Villarreal, A; Fass, R. (2007): Non Cardiac chest pain: current treatment. *Gastroenterology & hepatology*, 3(4), 255–262.
- [4] CAD: Acute Coronary Syndrome. Retrieved on September 20, 2021, from <https://my.clevelandclinic.org/health/diseases/16713-cad-acute-coronary-syndrome>
- [5] Anderson, J.L, *et al.* (2007): ACC/AHA 2007 guidelines for the management of patients with unstable angina/non-ST-elevation myocardial infarction. *J Am Coll Cardiol.*, 250, 1-157.
- [6] Larue, C, *et al.* (1993): Cardiac specific immunoenzymometric assay of troponin I in the early phase of acute myocardial infarction, *Clinical Chemistry*, 39(6), 972–979.
- [7] Dolci, A; Panteghini, M. (2006): The exciting story of cardiac biomarkers: From retrospective detection to gold diagnostic standard for acute myocardial infarction and more. *Clinica Chimica Acta*, 369, 179-187.
- [8] Jaffe, AS, *et al.* (1996): Comparative sensitivity of cardiac troponin I and lactate dehydrogenase isoenzymes for diagnosing acute myocardial infarction. *Clin Chem*, 42(11), 1770-6.
- [9] Collinson, P. O; Stubbs, P.J. (2003): Are troponins confusing?. *Heart (British Cardiac Society)*, 89(11), 1285–1287.
- [10] Tsung, SH. (1981): Several conditions causing elevation of serum CK-MB and CK-BB. *Am J Clin Pathol*, 75:711–5.
- [11] Surya, P Rao, *et al.* (1999): Cardiac troponin I and cardiac enzymes after electrophysiologic studies, ablations, and defibrillator implantations, *The American Journal of Cardiology*, 84(4), 470.
- [12] Wong, WM, *et al.* (2004): Population based study of noncardiac chest pain in southern Chinese: prevalence, psychosocial factors and health care utilization. *World J Gastroenterol.*, 10, 707–712.
- [13] Kruger, S; Graf, J, *et al.* (2002): Brain natriuretic peptide levels predict functional capacity in patients with chronic heart failure. *J Am Coll Cardiol*, 40, 718–22.
- [14] Kontos, MC, *et al.* (2007): Ability of myoglobin to predict mortality in patients admitted for exclusion of myocardial infarction. *Am J Emerg Med*, 25, 873–9.
- [15] Hamm, CW; Katus, HA. (1995): New biochemical markers for myocardial cell injury. *Curr Opin Cardiol*, 10, 355–60.
- [16] Apple, FS. (1992): Acute myocardial infarction and coronary reperfusion. Serum cardiac markers for the 1990s. *Am J Clin Pathol*, 97, 217–26.
- [17] Ruzich, RS. (1992): Cardiac enzymes. How to use serial determinations to confirm acute myocardial infarction. *Postgrad Med*, 85–9.
- [18] Aldous SJ. (2013): Cardiac biomarkers in acute myocardial infarction. *Int J Cardiol.*, 164(3), 282-94.
- [19] Levin, ER; Gardner, DG; Samson, WK. (1998): Natriuretic peptides. *N Engl J Med.*, 339: 321-328.
- [20] Cowie, M.R, *et al.* (2003). Clinical applications of B-type natriuretic peptide (BNP) testing, *European Heart Journal*. 24, 1710-1718.
- [21] Valli, N, *et al.* (2001): Assessment of brain natriuretic peptide in patients with suspected heart failure: comparison with radionuclide ventriculography data. *Clin Chim Acta*, 306, 19–26.
- [22] de Lemos, JA, *et al.* (2001): The prognostic value of B-type natriuretic peptide in patients with acute coronary syndromes. *N Engl J Med.*, 345: 1014-1021.
- [23] Arakawa, N; Nakamura, M; Aoki, H; Hiramori, K. (1996): Plasma brain natriuretic peptide concentrations predict survival after acute myocardial infarction. *J Am Coll Cardiol.*, 27: 1656-1661.
- [24] Bayes-Genis, A *et al.* (2000): D-Dimer is an early diagnostic marker of coronary ischemia in patients with chest pain. *Am Heart J*, 140(3), 379-84.
- [25] Chiu, A; Chan, W.K; Cheng, S.H; Leung, C.K; Choi, C.H. (1999): Troponin-I, myoglobin, and mass concentration of creatine kinase-MB in acute myocardial infarction. *Q J Med*. 92, 711–718.
- [26] Uddin, S, *et al.* (2019): Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inform Decis Mak*, 19, 281.
- [27] Breiman L. (2001): Random forests. *Mach Learn*, 45(1):5–32.
- [28] Tianqi Chen; Carlos Guestrin. (2016): “XGBoost: A Scalable Tree Boosting System,” KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794.
- [29] Jason Brownlee. (2016). XGBoost With Python: Gradient Boosted Trees with XGBoost and scikit-learn. Machine Learning Mastery.
- [30] Liudmila Prokhorenkova; Gleb Gusev; Aleksandr Vorobev; Anna Veronika Dorogush; Andrey Gulin. (2018): CatBoost: unbiased boosting with categorical features. *32nd Conference on Neural Information Processing Systems*. Montreal, Canada.
- [31] Alexander Marz. (2020): Catboost Lss an Extension of Catboost to Probabilistic Forecasting. *arXiv:2001.02121*.
- [32] Lan Shou, *et al.* (2021): Blood Biomarkers Predict Cardiac Workload Using Machine Learning. *BioMed Research International*, 6172815.
- [33] Agibetov, A, *et al.* (2020): Machine Learning Enables Prediction of Cardiac Amyloidosis by Routine Laboratory Parameters: A Proof-of-Concept Study, *Journal of Clinical Medicine*, MDPI.
- [34] Mathis, M.R, *et al.* (2020): Early Detection of Heart Failure With Reduced Ejection Fraction Using Perioperative Data Among Noncardiac Surgical Patients: A Machine-Learning Approach. *Anesthesia and analgesia*, 130(5), 1188–1200.
- [35] Weng, SF, *et al.* (2017): Can machine-learning improve cardiovascular risk prediction using routine clinical data. *PLoS ONE*, 12(4), e0174944.
- [36] Meeshanthini, V Dogan, *et al.* (2021): External validation of integrated genetic-epigenetic biomarkers for predicting incident coronary heart disease. *Epigenomics, Future Medicine*, 13(14).
- [37] Latha CBC; Jeeva SC. (2019): Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Informatics in Medicine Unlocked*, 16, 100203.
- [38] Than, M.P, *et al.* (2019). Machine Learning to Predict the Likelihood of Acute Myocardial Infarction. *Circulation*, 140(11), 899–909.
- [39] Alaa, AM, *et al.* (2019). Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. *PLoS ONE*, 14(5), e0213653.
- [40] Van der Laan, Mark J; Polley ,Eric C; Hubbard Alan E. (2007): Statistical Applications in Genetics and Molecular Biology. *The Berkeley Electronic Press*, 6(1).
- [41] Doppala, Bhanu Prakash; Bhattacharyya, Debnath. (2021). Cardiovascular_Disease_Dataset, Mendeley Data, V1, doi: 10.17632/dzz48mvjht.1.

Authors Profile

	<p>Ms. Anuradha P is a Research Scholar at Avinashilingam Institute for Home Science and Higher Education for Women, deemed to be University, Coimbatore, India. She has done MCA and is KSET and UGC NET qualified. She is Associate Professor and Head of the Department of Computer Science at Indian Academy Degree College (Autonomous), Bangalore, Karnataka, India. Her areas of interest include Object Oriented Programming, Analysis and design of algorithms, Database Management System, Operating System, Systems Programming, Advanced Java Programming and Data Science.</p>
	<p>Dr. Vasantha Kalyani David is Professor and Head in the Department of Computer Science at Avinashilingam Institute for Home Science and Higher Education for Women, deemed to be University, Coimbatore, India. Earlier a Mathematician with a Master of Philosophy in Mathematics and later did research in Computer Science. Dr. Vasantha Kalyani David has published many papers in areas of Soft Computing. Her research interests, include Neural Networks, Artificial Intelligence, Fuzzy Logic, Genetic Algorithms, Cellular Automata, Theoretical Computer Science, and Automata Theory. She has authored a book on “Pattern Recognition Using Neural and Functional Networks”.</p>