

A Hybrid Approach for Heart Disease Prediction

Kumaresan Angappan¹, N. Meenakshi², Joel Evans E³, Harshitha Bharanika⁴, Suganya Jothi⁵

{kummaresan@gmail.com¹, abu221985@gmail.com², itsjoeydale@gmail.com³,
harshithabaranka@gmail.com⁴, suganya192002@gmail.com⁵}

Department of Information Technology, Hindustan Institute of Technology and Science,
Chennai, India^{1, 2, 3, 4}

Department of ECE, Mailam Engineering College, Mailam, India⁵

Abstract- An estimated count of 17900000 people, i.e., 31% of deaths is related to cardiac diseases every year. This number is projected to rise to 22 million by the end of this decade, making cardiac diseases among the most common global sources of death. The only effective method is ECG tests among various other limited methods for the detection of heart diseases. With regard to cardiac diseases, early diagnosis has the potential to produce better treatment outcomes. Hence, this paper discusses a Machine Learning methodology to detect heart diseases using the Data Set for Heart Disease by the repository of UCI Machine learning. This system is developed based on classification algorithms such as Support Vector Machines, K-Nearest Neighbour, Naïve Bayes, Decision Trees and Random forest classifiers. We define a hybrid stacking method and genetic algorithm which increases the accuracy achieved by the basic individual data mining techniques of classification.

Keywords— Data cleaning, Feature Scores, Machine Learning, Stacked Generalization, Genetic Algorithm

1 Introduction

Human hearts beat 1.5 lakh times each day, pumps around 2 thousand gallons of blood everyday and is the most vital organ of a human body. The heart's inability for supplying enough perfusion to target tissues and organs at physiological filling pressures to meet the metabolic demands and causes cardiac diseases which are chronic and a progressive condition. Despite the progress in treatment and medical industry, approximately 50% is the survival rate of the population affected with cardiac diseases for a period of 5 years.

In the typical case, most patients do not display symptoms and signs such as breathlessness, congestion etc., and are later admitted to receive intravenous medication treatments to help supply enough perfusion. 18.2 million adults around age 20 and older are detected with cardiac diseases and about 1 in every 4 deaths are related to cardiac diseases. According to the WHO, cardiac diseases account for 26% of the deaths in India and costs the economy a total of \$2 trillion. Typically, two heart sounds are heard, S1 and S2 in healthy subjects. The closure of mitral and ventricular valve at the early systole is termed as S1; The closure of the aortic and pulmonary valves is S2 in the beginning of the diastole.

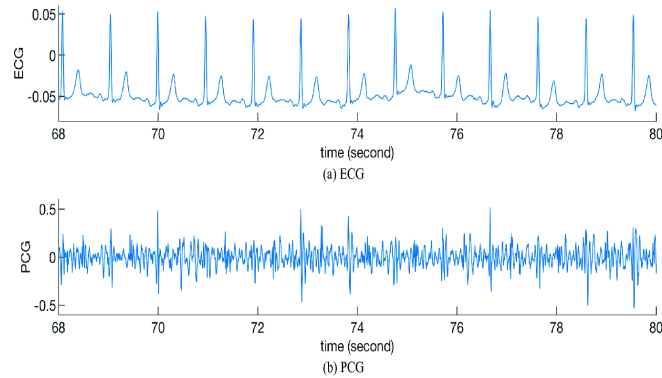


Fig. (1) ECG and PCG Recordings

In the phase of a cycle of a heart, the interval between S1 and S2 is systole, and diastole is interval between S2 and S1 also known as contraction phase. Fig. (1) shows the ECG and PCG recordings over time. ECG works with electrical activities of the heart and PCG is produced by the mechanical activities. In certain heart conditions that are never known to be normal, additional sounds S3 and S4 are heard. S3 appears to come 0.1-0.2 seconds after S2 as shown in Fig. (2).

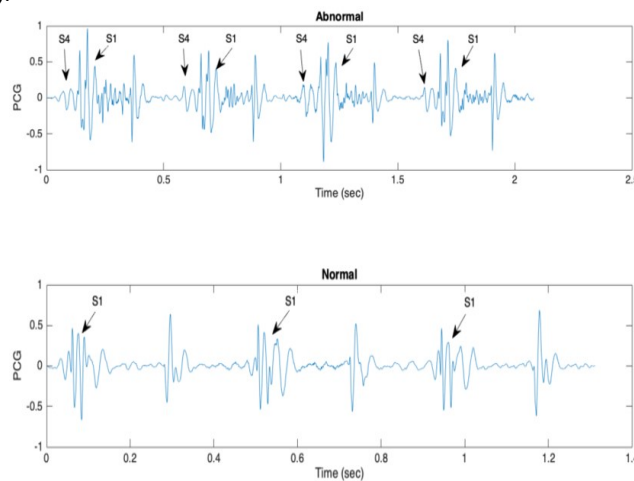


Fig. (2) Normal and abnormal sound waves of heart

The recordings of PCG that sound unhealthy to certain experts might sound healthy to others. Hence, only heart sounds cannot be used to diagnose and a holistic view of the patient that includes laboratory tests, medical history, blood pressure, age etc. is used.

Hand-in-hand, with the evolution of Machine Learning with digital era, an eruption of data from every single region of the world in all forms is brought together.

AI-ML learns to draw conclusions from data that is both unstructured and unlabelled. It has achieved breakthrough performance in image processing, speech and audio processing, pattern recognition, sensor data processing and natural language processing. While men are twice as likely to have heart diseases when compared to women, the risk persists equally with both genders. The signs of a woman developing the risk of cardiac diseases or a heart attack are much less noticeable when compared to men and hence, early detection can help reduce the

number of hospitalizations that happen due to worsening of the condition and has the potential to decrease the logistic and financial burden that's buried on the patient, hence improving the quality of life of patients. A successful approach of Machine Learning techniques such as stacking, boosting etc. with classification algorithm can surmount each single approach and can better help predict Cardiac Diseases.

2 Related Works

Cardiac diseases affect people in a way that can't be cured easily and hence diagnosing them at the right time is tough. K.G Dinesh et al. [1] worked on different techniques of machine learning for forecasting the uncertainty levels of cardiac diseases with the attributes from heart disease detection dataset by the UCI repository. Using the R environment, machine learning techniques such as Naive Bayes, Logistic Regression, SVM, Random Forest and Gradient boosting were used. An overall accuracy rate of 0.86 was given by the logistic regression model, however, ensemble techniques with more parameter settings can be used for better performance of these algorithms.

A hybrid method was given by H. A. Esfahani et al.[2], to outperform the results of individual machine learning methods. On the UCI dataset, pattern discovery algorithms like Neural Networks, Decision Trees, Support Vector machines, Rough set and Naive Bayes were applied and later, accuracy of their prediction was compared. F-measure is an important metric as it combines precision and sensitivity into a single value, hence among various measures, considering the F-measure metric, Rough set, Naive Bayes and Neural network achieved the highest performance. These classifiers were combined by the fusion strategy for better performance. The results later indicated that the fusion of outputs can improve the performance of the classifiers and hence, a trainable combining method such as a Bayesian combiner is unforeseen.

For solving both, constrained and unconstrained optimization problems to generate high quality solutions, Gino Sophia et al. [3] describes the approach of a proposed intelligent regressor system using Genetic operations and ANN. As Genetic algorithms look for optimal combination of solutions for brute force search problems, objective functions such as Crossover, Mutation, Fitness scaling, Selection and migration were used for increasing optimization performance of system developed and a Regression accuracy of about 98% was achieved by the model.

Cardiac diseases are critical health issues which occur with symptoms that are common such as shortness of breath, weakness of physical body and feet generally being swollen [4]. The need for efficient techniques for the detection of cardiac diseases has increased since existing diagnosis methods are not effective for early identification due to execution time and its accuracy. Hence, a system was developed based on classification algorithms such as KNN, Naive Bayes, Logistic Regression, Decision Trees and Artificial neural network [5]. A mutual information feature selection algorithm (FCMIM) was proposed. Hence, the results later showed the accuracy of SVM with FCMIM as 92.37%, which outperforms the other individual classifiers however, other feature selection algorithms with optimization methods can be used to further increase the performance of diagnosis.

A method combining classic Machine Learning and Deep Learning for detecting Chronic Heart Failure (CHF) from heart sounds was given by J. P. Li et al. [6] which uses the recordings from six publicly available datasets of 947 subjects. The FCMIM feature selection

algorithm with mutual information and it was used for selection of features. Later, the model was trained by the Random Forest classifier. A Fully Connected Neural Network is used in the Deep Learning model, where the high-level abstractions from raw data are learnt by the FCNN model. While an accuracy of 93.2% was achieved, only 44 samples of 947 subjects were used for chronic heart failure detection, and hence the risk of overfitting in the final experiment exists making it a limitation for the approach. Hence, a solid base for further development is needed.

Zeinab Arabasadi et al. [7] proposed a system that increases performance of the neural network system by about 10% with the implementation of genetic algorithm and an accuracy of 92% was achieved on the Z-Alizadeh Sani dataset. While the system showed a great performance, other algorithms could have been tested and compared for the performance of the system with new dataset and its unique features.

3 The Method

The heart disease dataset from UCI Machine Learning repository consists of 4 databases that extract information from VA Long Beach, Hungary, Cleveland and Switzerland. It contains a total of 920 records with a total 76 attributes, but a subset of 14 attributes is used for experiments. These attributes are categorical, real and integer values.

The records classify into two columns i.e., training and testing dataset. The below table shows the list of 14 attributes that are significant for the experiment.

#	FEATURE	DESCRIPTION
1.		<i>age</i> Age (years)
2.		<i>sex</i> Gender [0: Female; 1: Male]
3.		<i>cp</i> Type of chest pain
		Value 1: Typical angina
		Value 2: Atypical angina
		Value 3: Non- anginal pain
		Value 4: Asymptomatic
4.		<i>trestbps</i> Resting BP (in mm Hg during admission at hospital)
5.		<i>chol</i> Serum cholesterol in mg/dl
6.		<i>fbs</i> Fasting blood sugar > 120mg/dl [0= False; 1= True]
7.		<i>restecg</i> Resting electrocardiographic results
		Value 0: Normal
		Value 1: Having ST-T wave abnormality
		Value 2: Showing definite or probable left ventricular hypertrophy

8.	<i>thalach</i> Maximum heart rate achieved
9.	<i>exang</i> Exercise induced angina [0= False; 1= True]
10.	<i>oldpeak</i> ST Depression induced by exercise relative to rest
11.	<i>slope</i> The slope of the peak exercise ST segment
	Value 1: Upsloping
	Value 2: Flat
	Value 3: Downsloping
12.	<i>ca</i> Number of major vessels (0-3) colored by fluoroscopy
13.	<i>thal</i> [3: Normal; 6: Fixed defect; 7: Reversible defect]
14.	<i>num</i> Predicted attribute i.e., diagnosis of heart disease

Table. (1) List of attributes

The locations most often used to listen to the heart sounds, are named with respect to the positions where the valves are best heard:

1. Pulmonic area - along the left sternal border, the 2nd intercostal space.
2. Aortic area - centered at the second right intercostal space.
3. Mitral area - at cardiac ape.
4. Tricuspid area - along left sternal edge, the 4th intercoastal space.

A. Data analysis

Data analysis performs major tasks such as data cleaning and integration, identification of missing values and removing the redundant data. Pandas recognizes all the features as numerical either as integers or floats. But some of the attributes are categorical in the dataset as their numbers represent mappings to a real-world phenomenon of the human heart. Hence, Exploratory Data analysis is performed where first, the list of all the features which are numerical is saved in a column in order to not mix in the analysis with categorical variables. Using Pandas Profiling, distribution of each variable is visualized and a detailed profile report is generated with inference type, quantile statistics, descriptive statistics etc. for exploratory data analysis.

The interactions between each variable is defined with their correlations for a measure of linear correlation between two variables using the Pearson's coefficient of correlation; Spearman's coefficient for detecting the nonlinear monotonic correlations between the variables and Cramer's V coefficient for association measure of nominal random variables.

Statistical tests were used for picking the features that seem to contain best relationship with performance variable.. The SelectKBest class from the scikit-learn library is used to select features of different statistical tests. For non-negative features to select best features from the dataset, Chi-squared test has been used in the proposed methodology and then the ranks of features according to their importance is given, shown in Fig. (3).

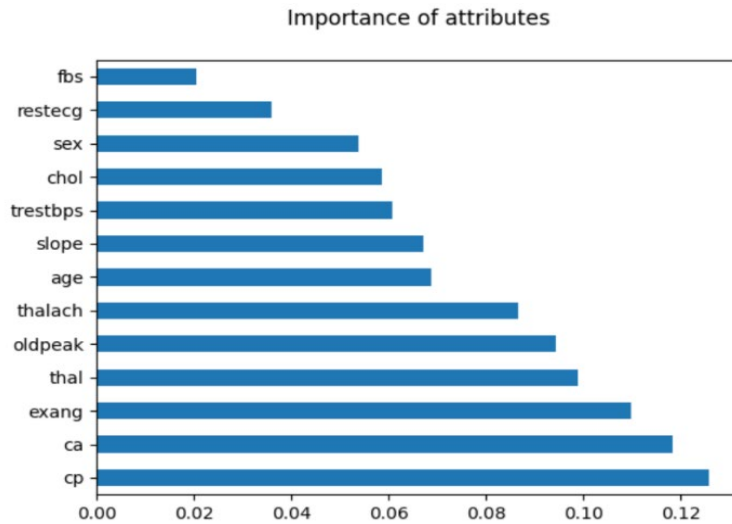


Fig. (3) Feature Importance

Using the Model Characteristics property, feature significance is gained. For extraction of the high scored features, the Extra Tree Classifier is used. It is a built-in class with Tree Based classifiers. A score for every function is given by the Feature value and it determines the significance of our performance variable.

With seaborn library, correlation matrix with Heatmap is generated for indicating the features and their relation with one another and as well as to target variable. Fig. (4) shows the correlation values between the variables, which are positive or negative in nature. It has been noticed that the attribute highly related to target variable is 'cp' i.e., chest pain and hence contributes the most for prediction of cardiac diseases.

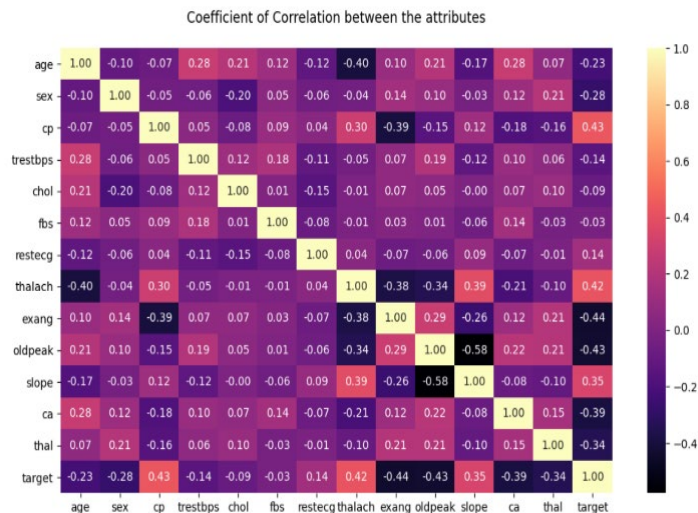


Fig. (4) Correlation Coefficient of attributes

In general, cardiac related diseases show symptoms such as pain in the center of the chest, sudden pressure discomfort or feeling of fullness in women and they're most likely less noticeable whereas in men, the typical symptoms such as chest pain, shortness of breath, sweating, discomfort, stress and other noticeable signs are seen. Hence, a detailed analysis was performed on this Heart Disease dataset and it shows that men are more susceptible to have a heart disease when compared to women. Fig. (5) shows the role played by Gender with respect to the Target variable where we can see that the ratio of men being affected by a heart disease when compared to women is 4:1.

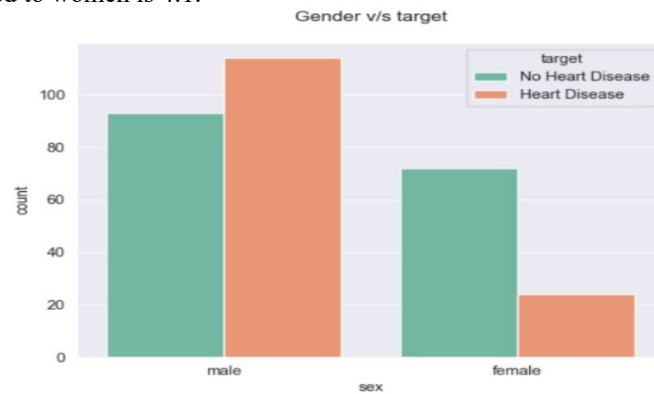


Fig. (5) Gender v/s Target



Fig. (6) Chest Pain Type v/s Target

Further, the dataset from the UCI Repository for cardiac diseases shows that 4 types of chest pain are generally produced- asymptomatic, atypical angina, non-angina and typical angina. Fig. (6) indicates the effect of these common chest pain types with respect to the target variable for both men and women. Most of the patients who are likely to have Cardiac

Diseases seem to have an asymptomatic pain which involves blood flow blockage to the heart, possibly damaging the heart muscle.

Age, high cholesterol, diabetes, high BP, obesity etc. are a variety of risk factors for asymptomatic pain. This puts a person at risk, as they have the possibility of developing a cardiac disease with serious complications. Other cardiac problems such as congestive heart failures, abnormal heart rhythms are associated with thalassemia- an inherited blood disorder which makes body have low levels of haemoglobin when compared to the normal count. Fig. (7) shows the comparison between the gender and thalassemia affected patients.

Male patients are more frequently affected than females, while it has little or less relationship with gender and mostly an inherited disease, it is most common in men.

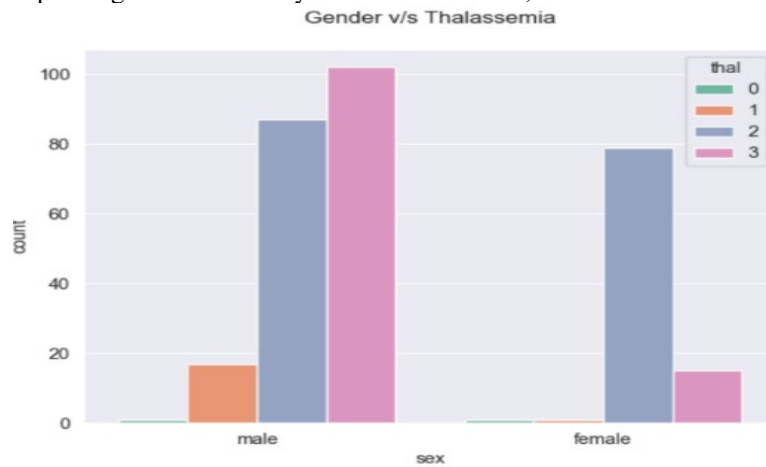


Fig. (7) Gender v/s Thalassemia

Heart Diseases affect both men and women irrespective of the age group but is very common in the age group above 60. For adults, total cholesterol levels are measured by milligram per deciliter (mg/dL). Cholesterol levels less than 200 (mg/dL) are desirable and considered. Borderline levels are considered around 200-239 (mg/dL) and high levels are considered to be above 240 (mg/dL). The LDL (Low Density Lipoprotein) must contain less than 100 mg/dL of cholesterol. Fig. (8) and Fig. (9) show the age and cholesterol of patients with heart disease.

For healthy individuals 100 mg/dL are appropriate but are found to be relevant on individuals with cardiac problems and other heart diseases. The borderline moderate levels range around 130 and 159 mg/dL while moderate measures range around 160 and 189 mg/dL. Any reading above 190 mg/dL is considered high and risky. High density lipoprotein or HDL levels can only be measured with a maximum of 60mg/dL. Borderline levels to be maintained are near to 41mg/dL and 59mg/dL as the risk factor for cardiovascular diseases is a reading around 40mg/dL and less.

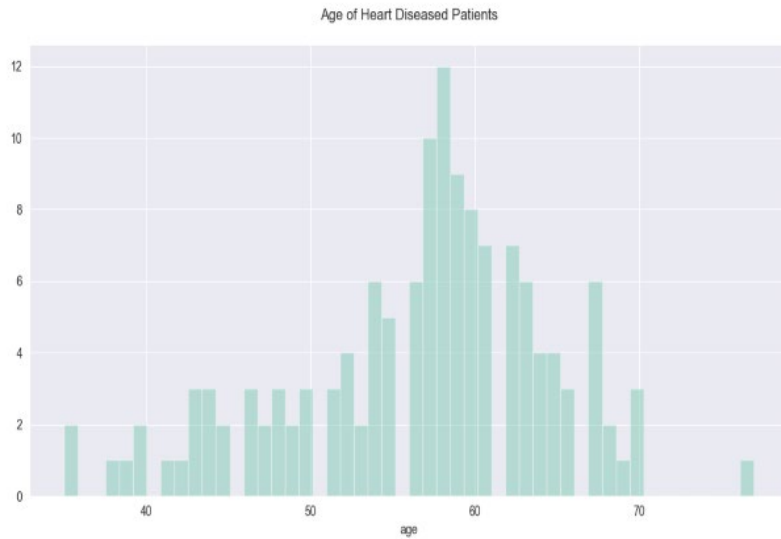


Fig. (8) Age of Patients who are affected with heart disease

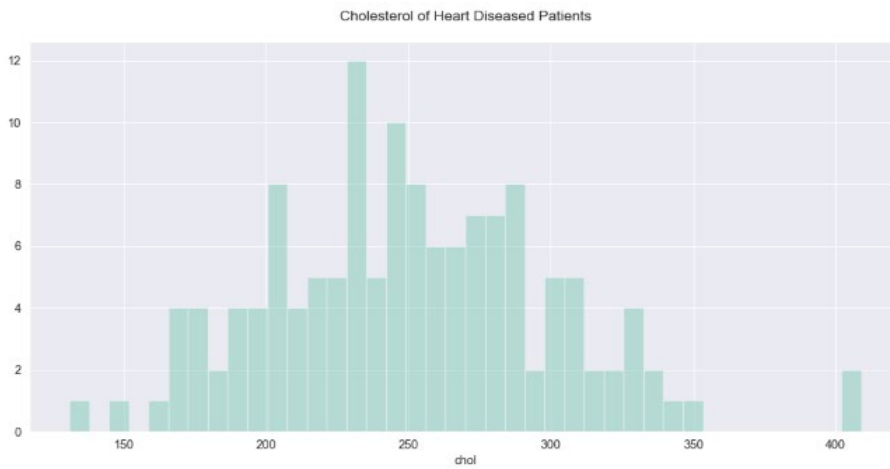


Fig. (9) Cholesterol in patients affected with heart disease

B. Experimental Analysis

Genetic Algorithm With SVM

For both classification and regression tasks, A supervised algorithm known as Support Vector Machine can be used. It is most commonly used in classification problems. As a point in n-dimensional space, each data item is plotted (n is the number of features of the dataset after EDA). The value for the co-ordinate is given by the value of each feature and then

classification is performed by finding the hyperplane that differentiates the two train and test classes.

For n features in M dimensions, the equation for the hyperplane is given as...

$$\begin{aligned}
 y &= w_0 + w_1x_1 + w_2x_2 + w_3x_3 \dots \\
 &= w_0 + \sum_{i=1}^m w_i x_i \\
 &= w_0 + w^T X \\
 &= b + w^T X
 \end{aligned}$$

here,

W_i = input vectors ($W_0, W_1, W_2, W_3, \dots, W_m$)

b = biased component (W_0)

X = variables.

Algorithm 1: Training an SVM

Input: X : The train set. δ : Threshold value

Output: x_R : $x_R \in X$

$|x_R| \lll |X|$.

Begin

Train one decision tree T ;

$x_R \leftarrow \text{NULL}$

For each leaf L_i of T , follow

1. for each opposite class neighbour L_j do
2. if entropy of L_j is low then
3. Select closest examples.
4. Use L_i and L_j to build X^+ ;
5. Compute
6. Add $x_i \in L_j$ to x_R
7. end for
8. else
9. Add all the elements in L_j to x_R
10. $x_R \leftarrow x_R \cup L_j$;
11. end if: end for return x_R

End

After the columns are gathered and the data is split with test size 20% and train size 80%, it is fit into the SVM classifier by importing SVC i.e., Support Vector Classifier from sklearn library.

By relying on biological operators such as mutation, crossover and selection, to generate high quality solutions to optimization and search problems, Genetic algorithm has been used with SVM classifier. It is a search heuristic algorithm. The fittest individuals are selected and they are made to produce off springs which inherit the characteristics of parent individuals and are later, added to the next continuous generation. The same process iterates and at the end, fittest individuals of a generation are found.

Five phases are considered:

1. Initial Population: A given set of individuals form a population where each individual is distinguished by a set of parameters that are called genes.

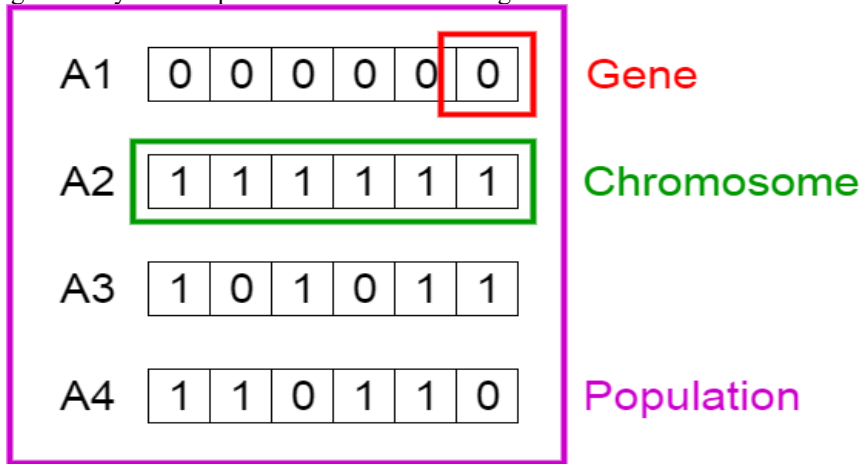


Fig. (10) Genetic Algorithm

To form a chromosome, these genes join as a string which are ideally solutions to each problem that needs to be solved. Genes usually are represented in terms of an alphabet by a string and it is said that genes of the population are encoded into chromosome.

2. Fitness Function: It determines the ability of an individual by evaluating and providing a fitness score for a particular individual who compete with other individuals and then get selected for reproduction based on their fitness scores.

3. Selection: Based on fitness scores of each individual, the selection phase takes place. Individuals with higher fitness scores have higher probability of getting selected for reproduction. This criterion helps the pair of individuals to pass on the genes for next generation.

4. Crossover: A Crossover is a stage used in generic algorithm during which a particular crossover point is picked at random from the genes of the parents. Then the genes are swapped among the parents until the reach of crossover point in order for the offspring to be produced.

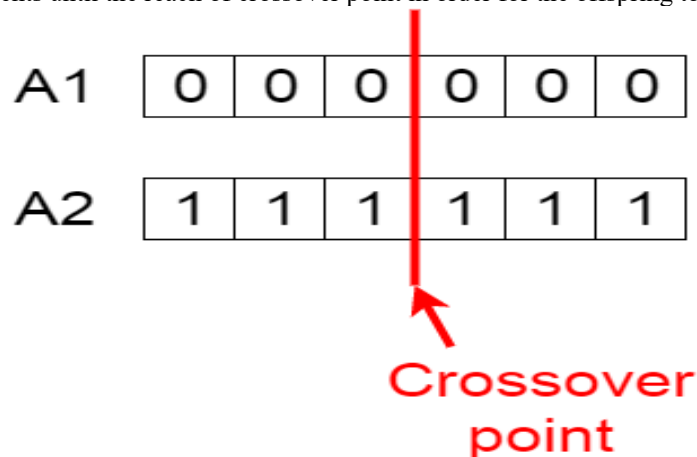


Fig. (11) Crossover phase

This is done in order for the offspring to consist of genes from both parents.

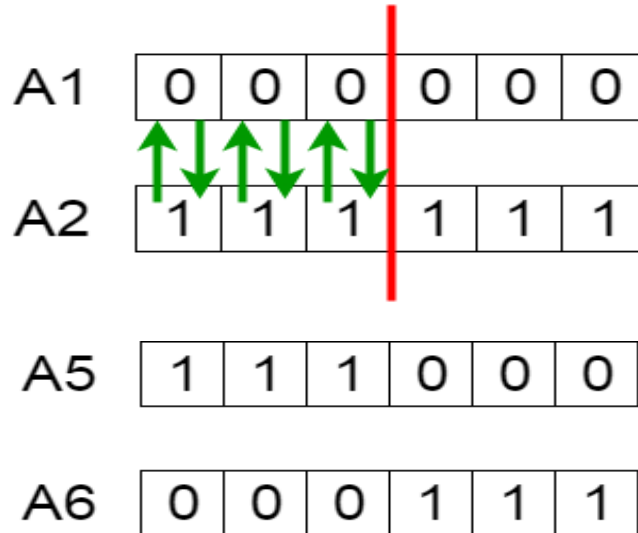


Fig. (12) Exchange of off springs for better performance results

5. Mutation: Sometimes for maintaining diversity within the population, a process known as mutation is performed by flipping a few of the bits in a bit string. Some of the genes in certain new offspring formed are subjected to mutation for a low discriminate probability.

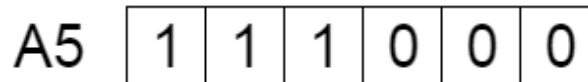


Fig. (13) Before Mutation

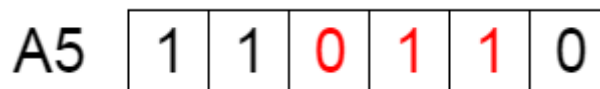


Fig. (14) After Mutation

This process also prevents premature convergence as well.

Algorithm 2: Genetic Algorithm

Start

1. Generating the initial population
 2. Computing fitness
- REPEAT

```

3. Select
4. Crossover
5. Mutate
6. Computation of Fitness
UNTIL POPULATION INTERSECTS

```

End

A set of solution is provided after the population has converged. Individuals with minimal fitness die when additional individuals which are better than previous generations are formed.

Stacked Generalisation

Stacked Generalization considers heterogeneous weak learners using a meta-model to give predictions depending on multiple predictions given as outputs by the weak models.

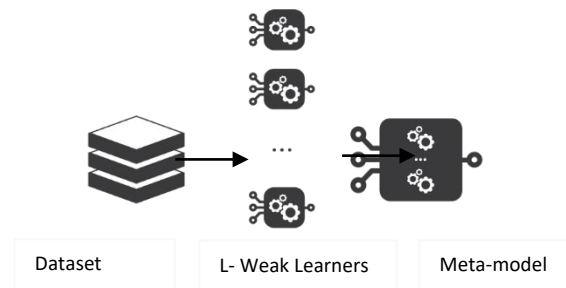


Fig. (15) Stacking concept

Multiple models are trained and combined to resolve a problem for better results. A paradigm of Machine learning, ensemble learning isworks on the main hypothesis which states that the weak models are combined to obtain more robust and accurate solutions and models. Depending on the volume of data, dimensionality of space, hypotheses distribution etc., the choice of model plays a major role in classification problems.

Base models or weak learners are models that don't perform so well as they have a high bias or too much variance to be robust. Hence, they can used as building blocks to combine and design complex models. These complex models are called strong models, which have less bias and variance and give better performance. The proposed stacking system works on the respective four classifiers:

1. K Neighbours classifier
2. Random Forest classifier
3. Gaussian NB
4. Decision Trees

Decision Trees acts as a meta model whereas the other classifiers act as the Base Learners.

Algorithm 3: Stacking

Input: Training Data $D = \{(x_i, y_i)\}_{i=1}^m$ ($x_i \in \mathbb{R}^n, y_i \in Y$)

Output: An ensemble learning classifier H

START

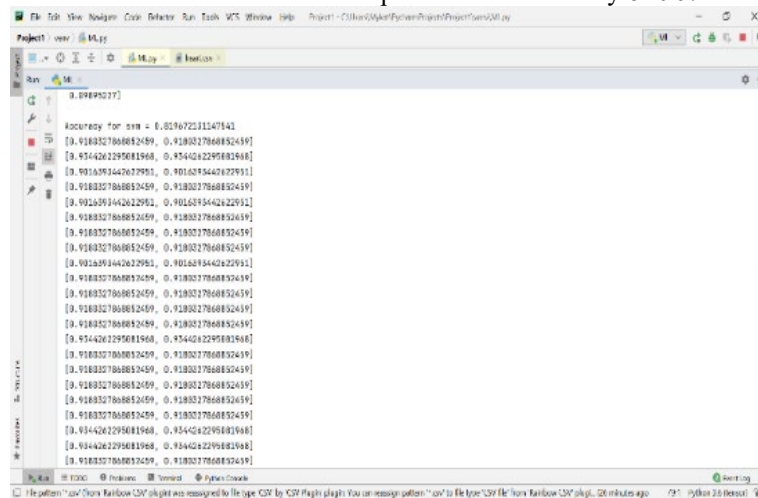
1. Learn classifiers of first level
 2. for $t = 1$ to T do
 3. based on D , learn a base classifier h_t
- End of for
4. Construct original data sets using D
 5. For $i=1$ to m do
 6. Construct new dataset that holds $\{x'_i, y_i\}$, where $x'_i = \{h_1(x_i), h_2(x_i), \dots, h_T(x_i)\}$
- End of for
7. Learn second level of classifier
 8. Based on the newly constructed data set, learn a new classifier h'
 9. Return $H(x) = h'(h_1(x), h_2(x), \dots, h_T(x))$

END

4 Results

For measuring the performance of the Classification model, ROC curve method was used. The true positive rate was plotted against the false positive rate, for the probabilities of predictions of classifiers.

SVM with Genetic Algorithm-After the columns are gathered and the data is split with test size 20% and train size 80%, it is fit into the SVM classifier by importing SVC i.e., Support Vector Classifier from sklearn library. Through evaluation, it is seen that the Support Vector Machine classifier on the heart disease predicts with accuracy 81.967%.



```
Accuracy for svm = 0.81967211147141
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
[0.9188327868852459, 0.9188327868852459]
```

Fig. (16) Accuracy of SVM classifier

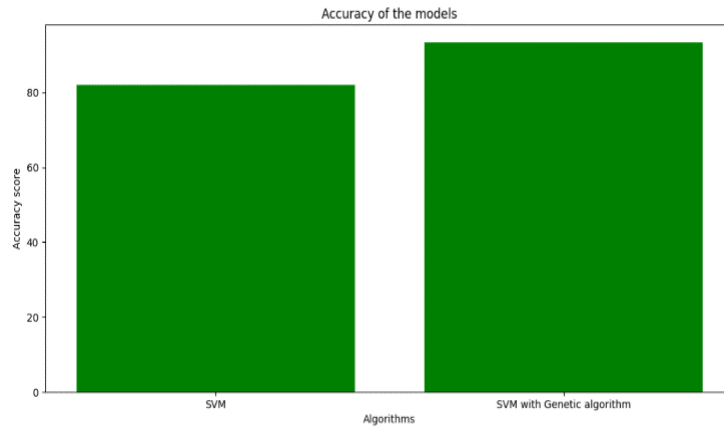


Fig. (17) Accuracy of SVM with Genetic algorithm

An accuracy of 93% has been achieved after implementing genetic algorithm on the heart disease dataset.

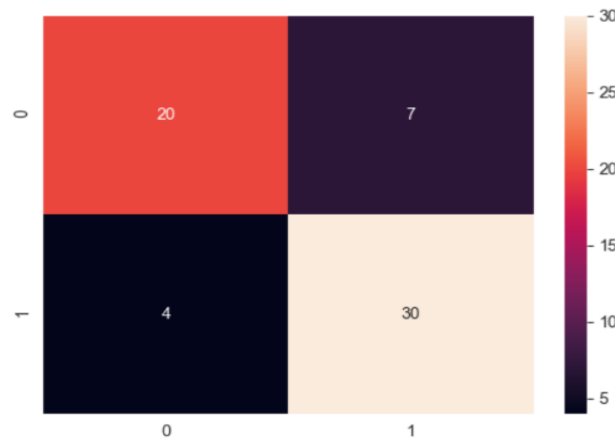


Fig. (18) Confusion Matrix for SVM with genetic algorithm

The Confusion Matrix shows True Positive rate to be 20 and True Negative as 30. The False Positive rate appeared to be 7 and False Negative with a total of 3. The ROC Curve with Genetic Algorithm for SVM is plotted and the Area under curve of the ROC is given to be 0.92 which says that the predictions are excellently made by the classifier.

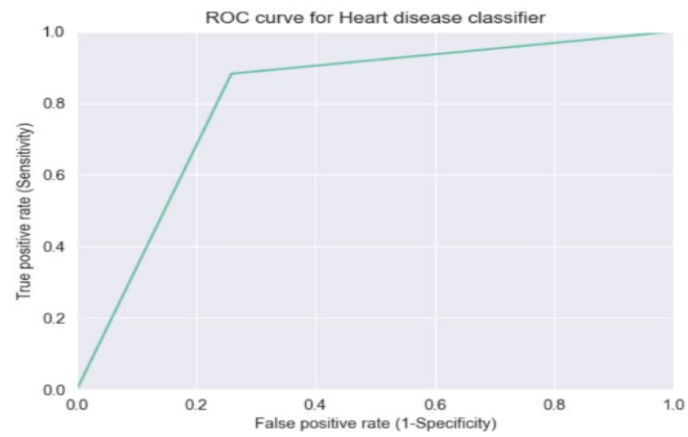


Fig. (19) ROC curve for SVM with genetic algorithm

Stacked Generalisation-After the base models were combined with a meta model as a learner, an accuracy of 86.88% was achieved. It surpasses the accuracy of each individual classifier.

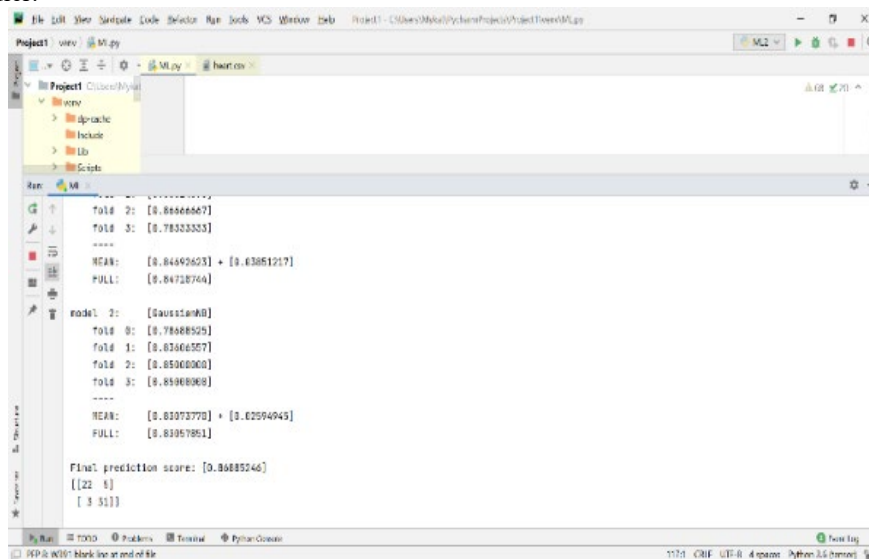


Fig. (20) Accuracy with stacked generalization

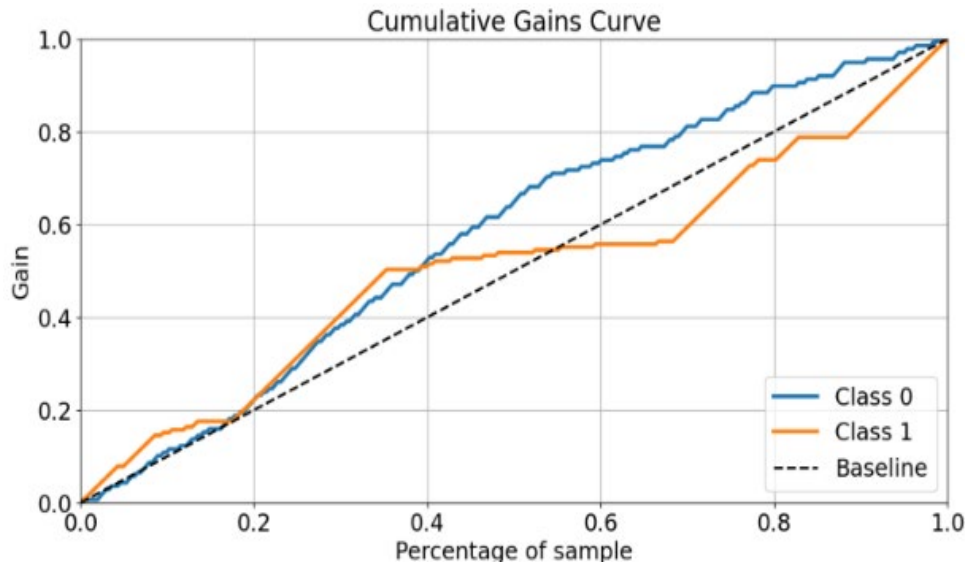


Fig. (21) Cumulative Gains Curve of the model

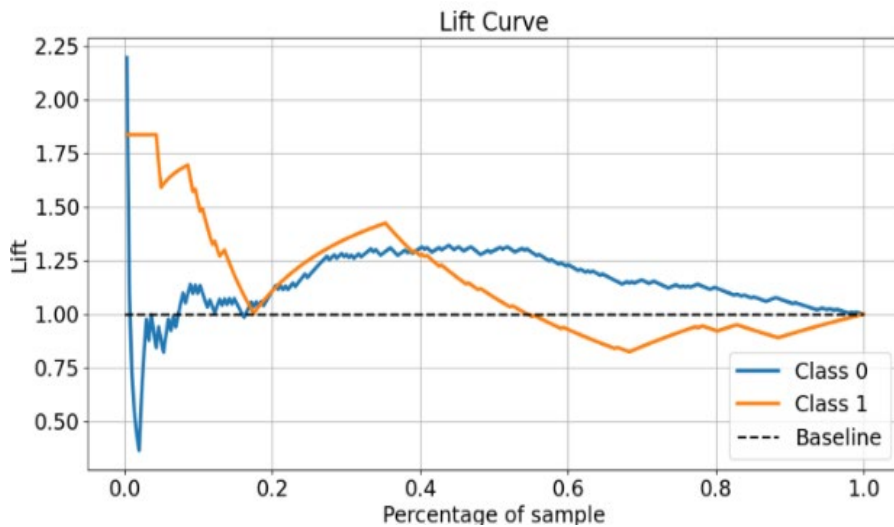


Fig. (22) Lift Curve for the model.

According to the UCI Heart Disease Dataset, 0 represents having affected by a heart disease and 1 represents not being affected by a heart disease. Hence, we can infer from Fig. (21) and Fig. (22) that the proposed system has gained a compelling accuracy in predicting the presence of heart disease.

With the generation of the classification report, it has been noticed that the precision rate for detecting presence of heart disease is 88.9% and for the absence of heart disease, it is

87.6%. Therefore, we can say that the stacking model performed well in predictions. The AOC under the ROC is calculated for stacked generalization method and a score of 0.86 is achieved which shows that the predictions are ranked with a good score.

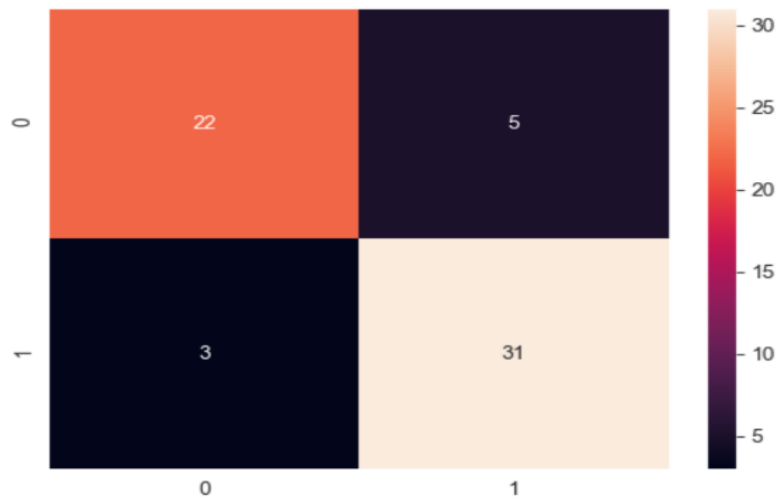


Fig. (23) Confusion Matrix for Stacking

Similarly, from Fig. (23) for the stacked method, Confusion Matrix displays True Positive rate to be 22 and true Negative with a rate of 31. The False Positive appeared to be 5 and False Negative as 3. The trade-off along the true positive rate and the false positive rate on the predictive model is summarized.

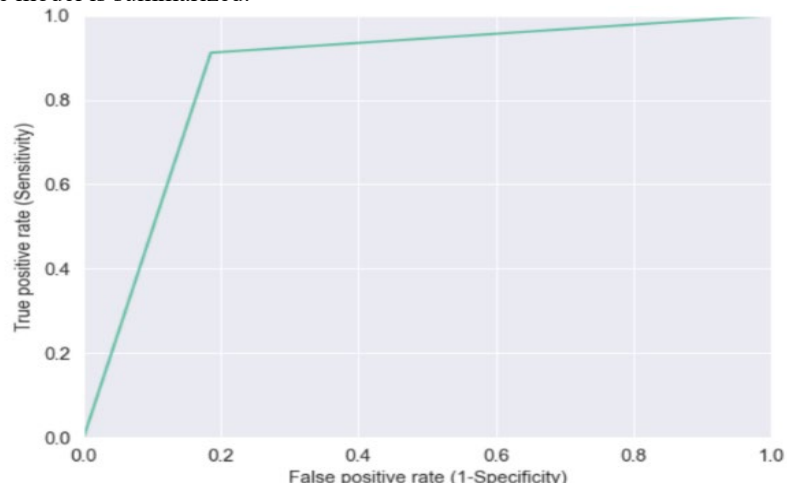


Fig. (24) ROC curve for Stacking

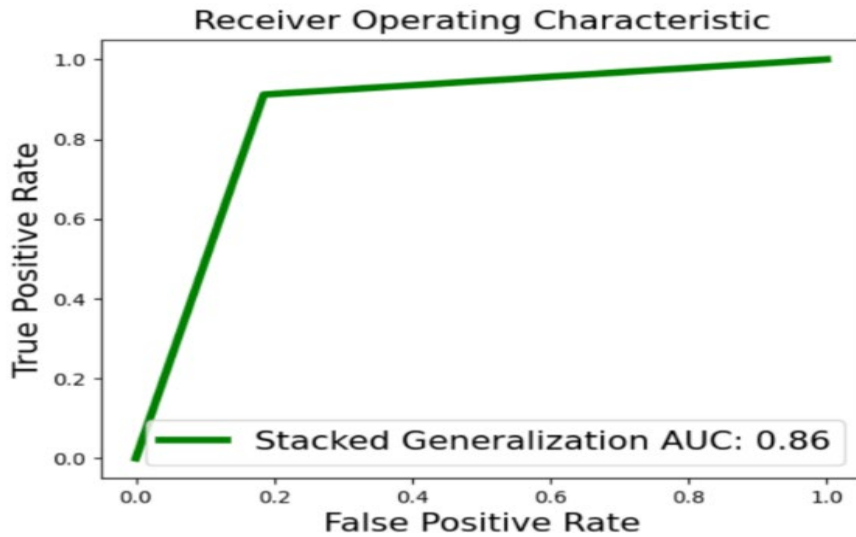


Fig. (25) AOC under ROC for stacked generalization

A web app has been deployed on the predictive model for the prognosis of heart disease. The web application takes the following features as input: Age, Gender, Chest pain type, Resting BP, Cholesterol etc. as they are considered to be the most important features in predicting the probability of developing a heart disease. After the input is gathered, the model runs and predicts the individual's risk of developing a heart disease in an effective manner.

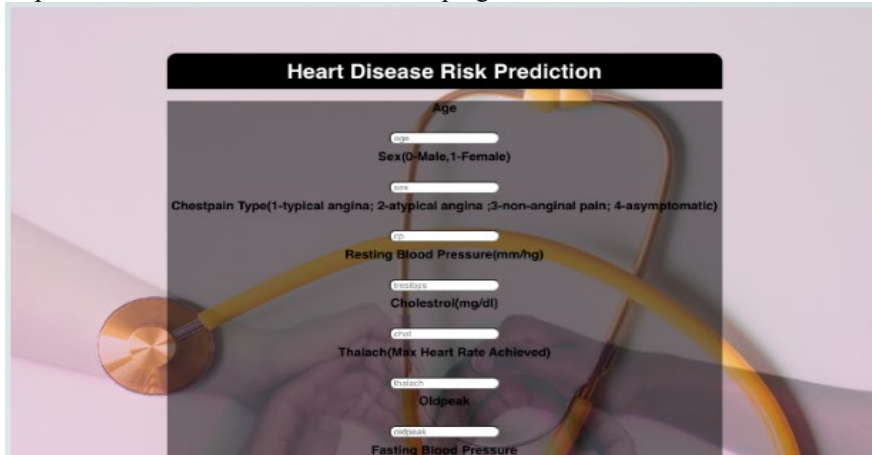


Fig. (26) Web Application



Fig. (27) Output from the web app

5 Conclusion And Future Work

With the aim to propose a hybrid method for prediction of cardiac diseases, the presented methodology achieved an accuracy of 86% with the concept of stacking and 93% accuracy for implementing the Genetic algorithm on the SVM classifier. This demonstrates the effectiveness of the proposed methodology for the identification of patients affected with heart disease or not. Although, there is a risk of overfitting since the dataset consists of 303 sample items, we trust that the results are encouraging and they represent a solid base for additional development of personalised models. Moreover, investigating other trainable combining methods is unforeseen.

Acknowledgements

The final outcome and success of this project required a lot of assistance and guidance and we are extremely privileged to have this along the completion of the project. All that we have done is only due to such assistance and supervision and we thank our guide Prof. Kumaresan Angappan for the support.

References

- [1] K. G. Dinesh, K. Arumugaraj, K. D. Santhosh and V. Mareeswari, "Prediction of Cardiovascular Disease Using Machine Learning Algorithms," 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, 2018, pp. 1-7.
- [2] H. A. Esfahani and M. Ghazanfari, "Cardiovascular disease detection using a new ensemble classifier," 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI), Tehran, 2017, pp. 1011-1014.

- [3] Gino Sophia, Ceronmani Sharmila, "Water management using Genetic algorithm-based machine learning" *Soft computing*, Vol. No. 24, Issue No.22, ISSN:1432-7643, pp. 17153-17165
- [4] Pavithra B., Rajalakshmi V. (2020) Heart Disease Detection Using Machine Learning Algorithms. In: Hemanth D., Kumar V., Malathi S., Castillo O., Patrut B. (eds) *Emerging Trends in Computing and Expert Technology. COMET 2019. Lecture Notes on Data Engineering and Communications Technologies*, vol 35. Springer
- [5] Kumaresan, A, Mohankumar, N., Sureshanand, M. and Suganya, J. (2016) "An effective recognition of partial speech using non audible murmur (NAM) for speech impaired children", *International Journal of Control Theory and Applications*, 9(28) 2016, pp. 165-174
- [6] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," in *IEEE Access*, vol. 8, pp. 107562-107582
- [7] ZeinabArabasadi, RoohallahAlizadehsani, MohamadRoshanzamir, HosseinMoosaei, Ali AsgharYarifard, Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm, *Computer Methods and Programs in Biomedicine*, Volume 141, 2017, Pages 19-26, ISSN 0169-2607
- [8] Fatma Zahra Abdeldjouad, MenaouerBrahami, and Nada Matta, A hybrid approach for heart disease diagnosis and prediction using machine learning techniques, *The Impact of Digital Technologies on Public Health in Developed and Developing Countries (Cham)* (Mohamed Jmaiel, MounirMokhtari, BessamAbdulrazak, HamdiAloulou, and Slim Kallel, eds.), Springer International Publishing, 2020, pp. 299–306
- [9] S. Babu et al., "Heart disease diagnosis using data mining technique," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2017, pp. 750-753
- [10] A. Singh and R. Kumar, "Heart Disease Prediction Using Machine Learning Algorithms," 2020 International Conference on Electrical and Electronics Engineering (ICE3), Gorakhpur, India, 2020, pp. 452-457
- [11] B. Gnaneswar and M. R. E. Jebarani, "A review on prediction and diagnosis of heart failure," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2017, pp. 1-3
- [12] Tougui, I., Jilbab, A. & El Mhamdi, J. Heart disease classification using data mining tools and machine learning techniques. *Health Technol.* 10, 1137–1144
- [13] YashJayeshChauhan, "Cardiovascular Disease Prediction using Classification Algorithms of Machine Learning", *International Journal of Science and Research (IJSR)*, https://www.ijsr.net/search_index_results_paperid.php?id=SR20501193934, Volume 9 Issue 5, May 2020, 194 - 200

- [14] S. S. Yadav, S. M. Jadhav, S. Nagrale and N. Patil, "Application of Machine Learning for the Detection of Heart Disease," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 2020, pp. 165-172
- [15] N. Louridi, M. Amar and B. E. Ouahidi, "Identification of Cardiovascular Diseases Using Machine Learning," 2019 7th Mediterranean Congress of Telecommunications (CMT), Fès, Morocco, 2019, pp. 1-6
- [16] David H. Wolpert, Stacked Generalization, Neural Networks, Volume 5, Issue 2, 1992, Pages 241-259