A Machine Learning Model for Predicting Heart Disease using Ensemble Methods

Jasjit Singh Samagh, Dilbag Singh
Department of Computer Science and Engineering
Chaudhary Devi Lal University, Sirsa, India

***Abstract*—There is the continuous increase in death rate related to cardiac disease across the world. Prediction of the heart disease in advance may help the experts to suggest the pre-emptive measures to minimize the death risk. The early diagnosis of heart disease symptoms is made possible by machine learning technologies. The existing machine learning models are inefficient in terms of simulation error, accuracy and timing for heart disease prediction. Hence, an efficient approach is needed for efficient prediction of heart disease. In the current research paper, a model based on Machine learning techniques has been proposed for early and accurate prediction of heart disease. The proposed model is based on techniques for feature optimization, feature selection, and ensemble learning. Using WEKA 3.8.3 tool, the feature selection and feature optimisation technique has been applied for irrelevant features elimination and then the pragmatic features are tested using ensemble techniques. Further, the comparison of the proposed model is made with the existing model without feature selection and feature optimisation technique in terms of heart disease prediction effectiveness. It is found that the results of proposed model gives the better performance in terms of simulation error, response time and accuracy in heart disease prediction.**

Keywords*—Heart disease; diagnosis; ensemble; optimization; prediction

I. INTRODUCTION

Globally, coronary disease is the main cause of death. 15% to 20% of fatalities were thought to be caused by ischemic heart disease and strokes. Investigations including ECGs, chest radiographs, and echocardiograms are typically done at the patient's bedside for the diagnosis of these disorders, although more involved procedures like cardiac catheterization, nuclear scanning, CT scans, and MRIs may also be done. The data that are gathered as a result of these examinations take a long time to analyse, and it takes a long time to deliver medications, which could be harmful to the patient. Doctors, pathologists, and other medical professionals may find that machine learning can shorten test times while improving test accuracy. Data mining and machine learning facilitate quick extraction and helps in fast extraction of results from a large size data in comparison with manual analysis [1].

Data mining plays vital role in health care systems using machine learning. Data mining is regarded as a crucial work that must be carried out precisely and competently since it aims to address real health issues in the diagnosis and treatment of disease. Heart disease prediction model, an electronic approach for detecting the heart illnesses based on earlier data and information, can be used to determine the sickness and its impact on patients. In order to aid medical professionals in making early forecasts of heart disease, this research aims to replace the time-consuming method with a quick one. Early diagnosis, prompt treatment, and decreased likelihood of casualties [2].

The objective of the present study is to develop and improve ML model which may help medical workers to extend accurate and quick medical help to the needy. For the said purpose different machine learning ensemble algorithms are compared using optimization algorithms. Due to their adaptability, optimization algorithms are frequently employed in numerous research domains. These algorithms are created by simulating or illuminating specific natural processes. GA and PSO are employed for this purpose. GA resolves the given process by replicating the natural process of evaluation. While PSO, a computer method, optimises the problem by iteratively attempting to enhance an optimal solution with regard to a specific number of features.The key features are produced by applying these optimising strategies one after another [10]. The drawbacks of using a single optimization algorithm to solve complicated problems include limited accuracy and generalizability.

In the research, GA and PSO are integrated, which means that exploitation and exploration capabilities are merged for binary and multi-class heart disease in order to further explore the application of optimization in bioinformatics. The wrapper-based feature selection approach is utilised in the study to eliminate redundant and unnecessary features, the GA optimization results are taken into account as the PSO's initial values, and finally the classification model for heart disease is built. Ensemble algorithms including Bagging, Boosting, Randomization, and Hybrid (Integration of Bagging, Boosting, and Randomization) are utilised to conduct the research. For the prediction and classification of the heart disease dataset, it has been found that randomization and hybrid models are better algorithms.

The remaining part of the paper is organized as follows: The problem statement for the research is covered in Section II. The relevant literature is discussed in Section III. The proposed methodology employed for the current study endeavor is described in detail in Section IV. The data mining tool used to conduct the research is discussed in Section V. The dataset for heart disease is presented in Section VI. The experimental findings of the proposed model for the prognosis of heart disease are presented in Section VII. The conclusion and Future Scope is covered in the final part.

II. Problem Statement

It has been observed in earlier studies that the research is only used to predict and categorise cardiac disease using machine learning approaches. However, the research does not focus on improving these techniques using optimisation techniques, simply on the unique consequences of various machine learning algorithms.

To carry out the work the wrapper-based feature selection method is applied as an initial step, which is also called the pre-treatment step. The mining-relevant attribute selection attributes are chosen from among all the original attributes once all continuous attributes have been discretized. The pre-processing stage of feature selection in machine learning is extremely successful at reducing dimensionality, removing irrelevant data, boosting learning accuracy, and enhancing comprehension of outcomes. The best attributes of the data set are chosen in the following stage by using PSO and GA. The classification will be more accurate thanks to these ideal features. The last phase, which diagnoses heart disease using ensemble approaches, assesses the performance of the suggested procedures by measuring classification accuracy.

III. Literature Review

Dilbag Singh [11] et al. compared and reviewed already existed models and extracted some major key attributes which are highly useful in creating and building an effective model. The effective model will therefore aid in the early diagnosis of cardiac problems, which will aid medical professionals in the patient monitoring process.

Jasjit Singh Samagh [10] et al. suggested a machine learning model for the heart disease prediction. The model is an integration of wrapper-based feature selection and GA and PSO based feature optimisation technique which is tested on ensemble machine learning techniques. Ensemble techniques itself being the combination of two or more classification techniques resulted in predicting an efficient and more effective model.

Youness Khourdifi [12] et al. associated the algorithms with different performance metrics with the help of machine learning. Different methods were used for prediction. Artificial Neural Networks, Random Forest, and K-Nearest Neighbor provide the greatest outcomes in this investigation. Then the research combines the algorithms and attempted to test the model effectiveness to see if it would be more effective or not. The results were later applied to a data set of heart disease, and his suggested models produced greater accuracy.

K.Vembandasamy [13] et al. analysed some parameters and suggested a data mining-based technique for predicting cardiac disease. Naive Bayes is utilised in the study since it has a strong independence principle for the diagnosis. 500 patients' worth of data from Chennai's premier diabetes research institute were utilised in the study. The tool used WEKA and achieved accuracy of 86.419%.

Vikas Chaurasia [14] et al. suggested the use of data mining tools for heart disease detection. Due to its open source nature and inclusion of machine learning techniques for mining, the author believes WEKA to be the greatest tool. The J48, bagging, and Naive Bayes techniques were applied in the current investigation. The author employed only 11 of the 76 features in the UCI data set, which was used to predict heart disease. Naive is accurately documented with 82.31%. J48 is recorded with an accuracy of 84.35%, and bagging records accuracy at an accuracy of 85.03%.

On the basis of literature review, it can be concluded that to build an effective model for the heart disease prediction it should be an integration of feature selection, feature optimisation and ensemble method.

IV. Proposed Methodology

The research is carried out in stages for heart disease prediction. The process begins with the collection of data for the dataset. The creation of dataset for the heart disease prediction is the crucial part as ML techniques performance depends upon it. The raw data is collected from electronic source present in form of patient's medical history, observations and laboratory tests. Redundancy is eliminated during pre-processing, which follows data collection, in order to get the information ready for heart disease prediction. The dataset is entered into the data mining tool after it is prepared to assess performance.

The Fig. 1 explains the proposed method for the heart disease prediction. In Proposed method, heart disease training dataset will go through feature selection and feature optimisation technique which will eliminate the low ranked attributes and will result in providing the predominant features. The performance of these predominant features for heart disease prediction is assessed using ensemble approaches [9]. The results predicted by this method are exceptionally best.

A. Feature Selection

An attribute of a data collection that is used in machine learning is called a feature. Some machine learning experts hold the belief that features should only be applied to attributes that have relevance to a given machine learning problem, although this belief should not be taken at face value. A branch of feature engineering that has attracted a lot of research attention is the selection of the subset of useful features for machine learning. The most important pre-processing step in any machine learning project is probably feature selection. It aims to choose a subset of system information or attributes that contributes most significantly to a machine learning activity [3].

Fig. 2 explains the process of the feature selection that is the internal structure of subset evaluation and resulting in getting the best features.

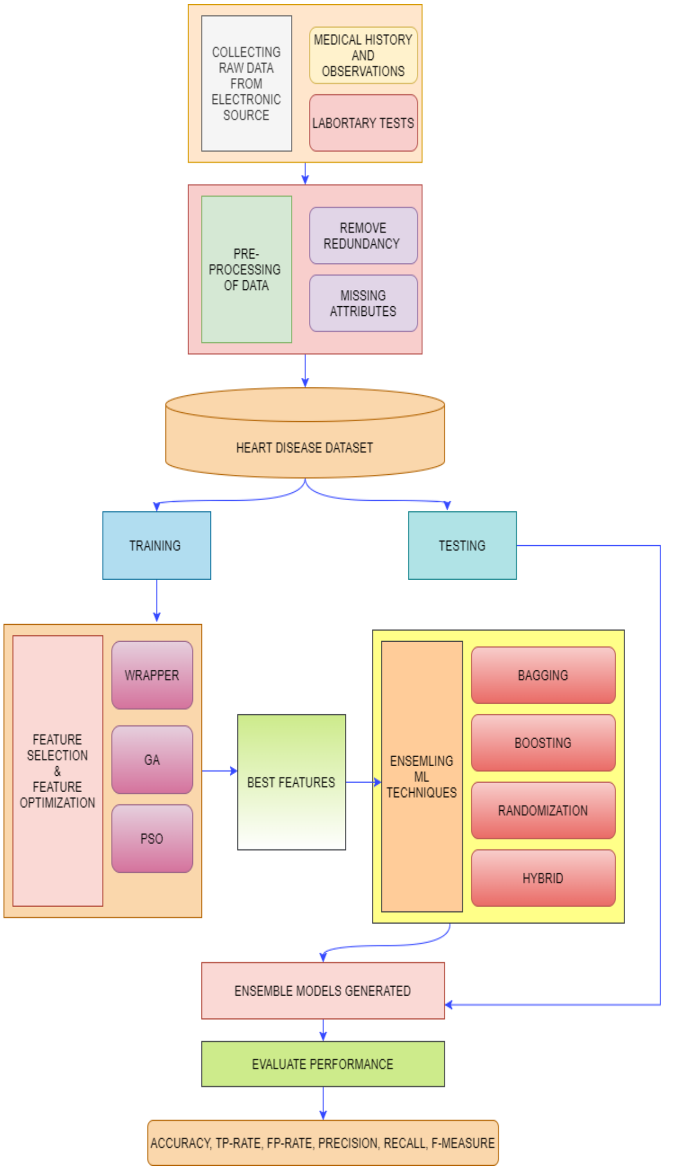

Fig. 1. Proposed Method (with Feature Selection and Feature Optimization).

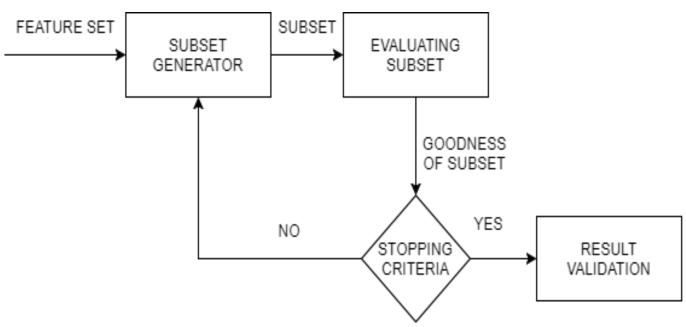

Fig. 2. Process of Feature Selection.

B. *Feature Optimization*

The goal of optimization is to achieve the best outcomes possible in any situation. We use optimization to either maximise the expected benefit or reduce the amount of effort needed. Decision variables can be stated as a function of the work necessary or the intended benefit. Discovering the circumstances that maximise or reduce a function is, thus, optimization. The highest value of the negative of the function is at point x, which corresponds to the smallest value of the function f(x). As a result, since finding the minimum of a function's negative allows us to find its maximum, optimization can be thought of as a minimization problem. Any optimization problem cannot be solved with a single technique. Consequently, we employ several techniques [4].

- Particle Swarm Optimization (PSO)

Every member of the population is referred to as a particle in particle swarm optimization. Particles in typical PSO update their velocity and location at each iteration based on their own experience, which is their personal best (best), and with the best experience of all the nearby particles, which is the global best (best), after the population has been initialised [7, 16].

A workflow diagram makes it simple to comprehend particle swarm optimization. Fig. 3 depicts this. It demonstrates in detail how PSO operates [8].

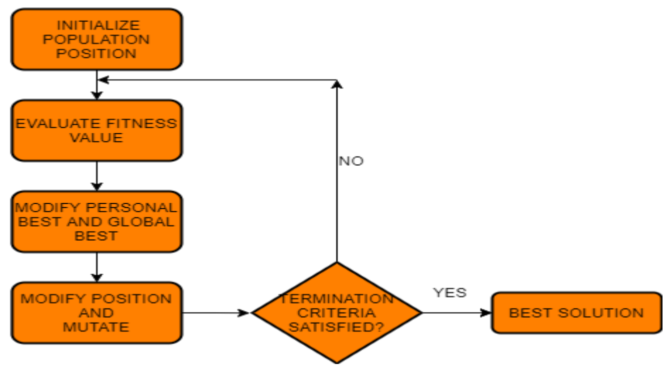

Fig. 3. Workflow of Particle Swarm Optimization (PSO) for Feature Optimization.

- Genetic Algorithm (GA)

The genetic algorithm, an evolutionary computational technique that has gained popularity recently, was created by Holland in the early part of 1975 and later improved by Goldberg [9]. It is a method of searching that resolves a specific issue by simulating the course of evolution. An algorithm that uses the idea of "survival of the fittest" and is based on Darwin's theory is known as a genetic algorithm [5]. The use of fresh and better optimal solutions by a genetic algorithm is non-presumptive like continuity. The genetic algorithm as a method has enormous potential, and as a result, it has been applied in a variety of industries, including gaming and financial research. The Genetic Algorithm has grown to be highly sought-after in various industries since it can manage a variety of characteristics. Instead of taking a lifetime to solve the problem, it has an ideal solution or one that is very close to the ideal [9].

GA execution can easily be understood with the help of a workflow diagram. Fig. 4 shows the flowchart of GA, and it explains the steps it takes in execution [6].

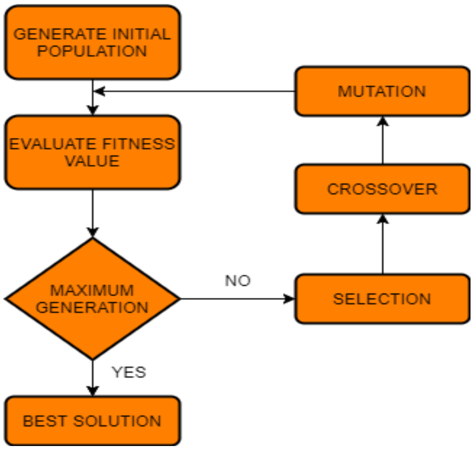

Fig. 4. Workflow of Genetic Algorithm (GA) for Feature Optimization.

C. Ensemble Techniques

An ensemble is a strategy that combines various models with varied strengths. Ensemble methods combine weaker learners to create stronger ones. Ensemble learning models are used to improve the predictive performance of statistical learning techniques by constructing a linear mixture of the simpler base learner. Ensemble learning models use decision trees as the base learner; in the case of random forest, many boosting and bagging implementations have been proposed [17].

One of the earliest and the most popular ensemble models is bootstrap aggregating or bagging. Bagging uses bootstrapping to generate multiple training datasets, and utilizing the same learning process, a collection of models are created using these training datasets [15].

Boosting is another crucial ensemble-based approach, similar to bagging. With boosting, weaker learning models are trained on resampled data, and the results are blended based on the performance of many models and a weighted voting mechanism. A specific type of boosting algorithm is adaptive boosting, often known as AdaBoost. Another well-liked ensemble learning strategy for creating prediction models is randomization [15, 17].

V. Data Mining Tool

When doing numerous experiments on datasets for machine learning, the data mining tool is crucial. WEKA 3.8.3 is the data mining tool utilised to carry out the current study. Waikato Environment for Knowledge Analysis is abbreviated as WEKA. It is created in the New Zealand's University of Waikato. Its platform is Java-based [21].

WEKA is a collection of AI algorithms used for data mining jobs. There are two ways to apply an algorithm when using WEKA, either directly on a dataset or by calling it from Java code. Numerous built-in features in WEKA allow for the prediction of the model's correctness [19].

VI. Heart Disease Dataset

Data could be gathered for the research's objective from a variety of sources. The UCI machine learning repository and Kaggle are the two most frequently used sources. In 1987 at Irvine, David Aha and a few other students founded the UCI repository [22], whereas the Kaggle is Google Subsidiary founded by Anthony Goldbloom in April, 2010 at United States.

Both UCI machine learning repository [18] and Kaggle contains number of datasets related to healthcare sector. Machine learning enthusiasts, from novices to experts, frequently use the dataset available from these sources to comprehend data empirically.

Further, to carry out the present research the dataset from Staglog (heart) dataset from UCI repository is used which is provided by university hospital, Basel, Switzerland by Matthias Ptisterer, M.D. The data is then pre-processed to check for missing attributes and to eliminate duplication. Machine learning is used to forecast heart disease using a later dataset. The heart disease dataset utilised in the study has features that can number in the tens of thousands, but it also contains 14 attributes that are listed below [20].

TABLE I. HEART DISEASE DATASET

Attributes	**Types**	**Explanation**
Age	Numeric	Age in years (29 to 77)
Sex	Numeric	Sex (1 = Male, 0 =Female)
Chest Pain Type (cp)	Numeric	Chest pain type (1: typical angina, 2: atypical angina, 3: non-angina pain, 4: asymptomatic)
Rest Blood Pressure (trestbps)	Numeric	Resting blood pressure (in mm Hg on admission to the hospital) [94, 200]
Serum Cholestoral (chol)	Numeric	Serum cholesterol in mg/dl [126, 564]
Fasting Blood Sugar (fbs)	Numeric	Fasting blood sugar > 120 mg/dl (1 = True, 0= False)
Resting Electrocardiographic (restecg)	Numeric	Resting Electrocardiographic(ECG) results values(0: normal, 1: having ST-T wave abnormality, 2:showing probable left ventricular hypertrophy)
Maximum Heart Rate (thalach)	Numeric	Maximum heart rate achieved [71, 202]
Exercise Induced (exang)	Numeric	Exercise induced angina (1: Yes, 0: No)
Oldpeak	Numeric	ST depression induced by exercise relative to rest [0.0,62.0]
Slope	Numeric	The slope of the peak exercise ST segment (1: upsloping, 2: flat, 3: downsloping)
Major Vessels (ca)	Numeric	Number of major vessels values(0-3) colored by flourosopy
Thal	Numeric	Defect types: value 3: normal, 6: fixed defect, 7: irreversible defect
Class	Nominal	Prediction of heart disease (1: Absence, 2: Presence)

Table I is the Heart disease dataset used in predcting the heart disease. It consisits of the 14 attributes with its explanantion used for predicting the model.

A. Feature Selection Feature Optimisation and Ensembling

The databases for cardiac disease can contain up to tens of thousands of features. However, only roughly 14 qualities are necessary to forecast cardiac problems because a large number of irrelevant and redundant attributes frequently produce unreliable results, incur expensive costs, and take a lot of time. The performance in predicting heart disease would be better if the qualities were less. On the other hand, Feature Optimization techniques [25] used to identify the optimal solution by minimising or maximising the objective function without going against resource limitations.

By adjusting parameters, optimization methods can be used to enhance the performance of classifiers in the prediction of heart disease. GA and PSO are employed for this purpose [8]. While PSO, a computer method, optimises the problem by iteratively attempting to enhance a candidate solution with regard to a specific number of features, GA resolves the given process by replicating the natural process of evaluation. These optimization strategies are repeatedly used to produce the key features [17].

After applying the wrapper, Genetic Algorithm and Particle swarm optimisation to the heart disease dataset following features got extracted:

- Chest Pain Type (cp).
- Resting Electrocardiographic (restecg)
- Maximum Heart Rate (thalach).
- Exercise Induced (exang).
- Oldpeak.
- Major Vessels(ca).
- Thal.
- Class.

The number of features reduced from 14 attributes to the 8 predominant features. These features are then tested on ensemble techniques that is Bagging, Boosting, Random Forest and Hybrid (a combination of Bagging, Boosting and Randomization) methods of machine learning to evaluate the results regarding prediction of heart disease.

VII. EXPERIMENTAL RESULTS

To evaluate the performance, comparison of various ensemble machine learning techniques is made on various criteria.

The Results are calculated on the basis of following norms:

- Results without using Feature selection and Feature Optimization Techniques.
- Results after using Feature Selection and Feature Optimization Techniques (Proposed method).

A. Simulation Error of Ensemble's

Simulation error is also taken into consideration in the study's execution to enhance the ensemble learning model's performance. The prediction model's effectiveness is described by simulation errors. The five aforementioned evaluation criteria are used in the current study to assess the simulation error, as indicated in Table II and III, respectively.

Table II describes the simulation error for the heart disease prediction model using ensemble technique. Simulation error tell us how effective the model is in predicting the accuracy. In this, result evaluated is computed without using feature selection and feature optimisation technique.

Table III is the simulation error table for the heart disease prediction model using ensemble technique. In this table the results are computed using feature selection and feature optimisation technique. It is seen that the error rate of the simulation model are reduced after using feature selection and feature optimisation technique that is proposed method.

B. Confusion Matrix

The confusion matrix reveals how a predictive model operates internally. It provides insight information on classes that are accurately predicted, classes that are wrongly forecasted, and also about the different forms of faults. The confusion matrix produced for a two-class classification issue with negative and positive classes is the most basic type of confusion matrix. Each cell in the figure has a distinct and clear display, as illustrated. [23].

Fig. 5 explains the insight of the confusion matrix, that how the internal functional of the predictive model are classified in following classes.

TABLE II. SIMULATION ERROR WITHOUT USING FEATURE SELECTION AND FEATURE OPTIMISATION

Evaluation Criteria	Bagging	Boosting	Randomization	Hybrid
KS	0.58	0.595	0.6244	0.6694
MAE	0.2934	0.2374	0.2696	0.2668
RMSE	0.3774	0.3807	0.3587	0.362
RAE (%)	59.4167 %	48.0607 %	54.5794 %	54.0189 %
RRSE (%)	75.956 %	76.6224 %	72.1837 %	72.8601 %

TABLE III. SIMULATION ERROR WITH FEATURE SELECTION AND FEATURE OPTIMISATION

Evaluation Criteria	Bagging	Boosting	Randomization	Hybrid
KS	0.7657	0.6772	1	0.7963
MAE	0.2271	0.2121	0.0867	0.1753
RMSE	0.3001	0.3296	0.1346	0.2474
RAE (%)	45.9852 %	42.9527 %	17.5484 %	35.4954 %
RRSE (%)	60.3894 %	66.3318 %	27.0926 %	49.7841 %

Where KS- Kappa Statistic, MAE- Mean Absolute Error, RMSE- Root Mean Squared Error, RAE- Absolute Error and RRSE- Root Relative Squared Error.

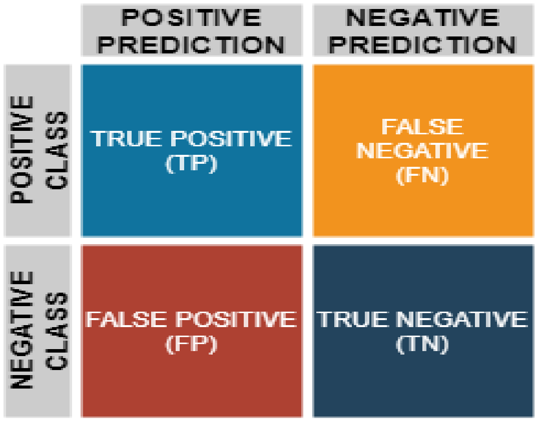

Fig. 5. Insight of Confusion Matrix.

- True Positive (TP): When both the truth and the test's prediction of a positive are true, the class is true positive. For instance, when a patient is ill and the test also detects this.
- True Negative (TN): When the test also predicts a negative outcome and the truth is negative, the class is true negative. For instance, when a test accurately detects that a person is healthy.
- False Negative (FN): When the truth is positive but the test predicts a negative outcome, the class is considered false negative. For instance, when a test falsely indicates that someone is healthy when they are actually ill. In statistics, this is also known as Type II mistake.
- False Positive (FP): The term "false positive" refers to a situation in which the test anticipates a positive result even though the reality is different. If a test falsely indicates that a person is ill even when they are not ill. In statistics, this is known as Type I mistake**.** [23, 24].

The Table IV indicates the Confusion Matrix of the Ensemble Machine learning models for prediction of the heart disease. The table shows detailed confusion matrix of machine learning ensemble model when created without using feature selection and feature optimisation technique that is the traditional method the proposed method that is using feature selection and feature optimisation techniques. This matrix depends upon four factors namely TP, TN, FP and FN [23, 24]

C. Accuracy Factors

The factors on which the accuracy of model is based are TP rate, FP rate, Precision, Recall and F measures [24].

Fig. 6 is about the calculations of accuracy factors. It shows the formulas used to calculate the accuracy factors Such as TP rate, FP rate, Precision, Recall and F-measure.

Following model creation, its accuracy is assessed by contrasting it against the following criteria.

- True positive rate: that a real positive will test positive is known as the true positive rate. It is calculated as TP/TP+FN and is often referred to as sensitivity.
- Precision: is the ratio of examples that genuinely belong to a class to all instances that are categorised in that class.
- Recall: a class's true total is equal to its actual fraction of instances classed as that class (equivalent to TP rate).
- F-Measure: is a composite measure for recall and precision that is calculated as 2 * recall / (precision + recall).

TABLE IV. CONFUSION MATRIX OF MODELS

		Absent	Present	Class
Without Feature Selection and Feature Optimisation	**Bagging**	122	28	**Absent**
		28	92	**Present**
	Boosting	123	27	**Absent**
		27	93	**Present**
	Randomization	126	24	**Absent**
		24	94	**Present**
	Hybrid	129	21	**Absent**
		23	97	**Present**
With Feature Selection and Feature Optimisation	**Bagging**	139	11	**Absent**
		20	100	**Present**
	Boosting	129	21	**Absent**
		22	98	**Present**
	Randomization	150	0	**Absent**
		0	120	**Present**
	Hybrid	140	10	**Absent**
		17	103	**Present**

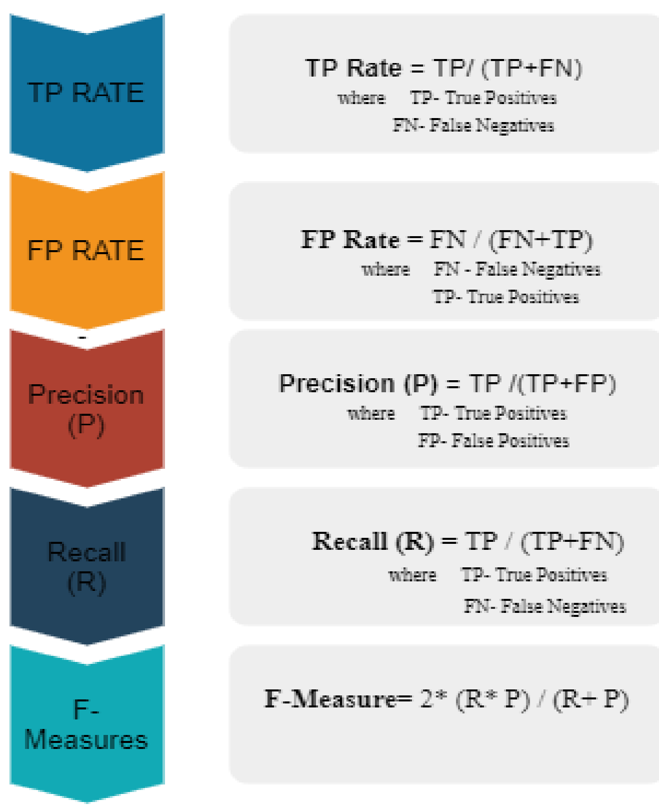

Fig. 6. Calculations of Accuracy Factors.

TABLE V. DETAILED ACCURACY MEASURES 0F FACTORS WITH CLASS

		TP Rate	FP Rate	Precision	Recall	F-Measure	Class
Without Feature Selection and Feature Optimization	**Bagging**	0.813	0.233	0.813	0.813	0.813	**Absent**
		0.767	0.187	0.767	0.767	0.767	**Present**
	Boosting	0.820	0.225	0.820	0.820	0.820	**Absent**
		0.775	0.180	0.775	0.775	0.775	**Present**
	Randomization	0.840	0.217	0.829	0.840	0.834	**Absent**
		0.783	0.160	0.797	0.783	0.790	**Present**
	Hybrid	0.860	0.192	0.849	0.860	0.854	**Absent**
		0.808	0.140	0.822	0.808	0.815	**Present**
With Feature Selection and Feature Optimization	**Bagging**	0.927	0.167	0.874	0.927	0.900	**Absent**
		0.833	0.073	0.901	0.833	0.866	**Present**
	Boosting	0.860	0.183	0.854	0.860	0.857	**Absent**
		0.817	0.140	0.824	0.817	0.820	**Present**
	Randomization	1.000	0.000	1.000	1.000	1.000	**Absent**
		1.000	0.000	1.000	1.000	1.000	**Present**
	Hybrid	0.933	0.142	0.892	0.933	0.912	**Absent**
		0.858	0.067	0.912	0.858	0.884	**Present**

The Table V indicates the Acuuracy factors of the Ensemble Machine learning models for prediction of the heart disease. The table shows detailed accuracy measures of factors with class of machine learning ensemble model when created without using feature selection and feature optimisation technique, that is, the traditional method (the proposed method) to use feature selection and feature optimisation techniques.

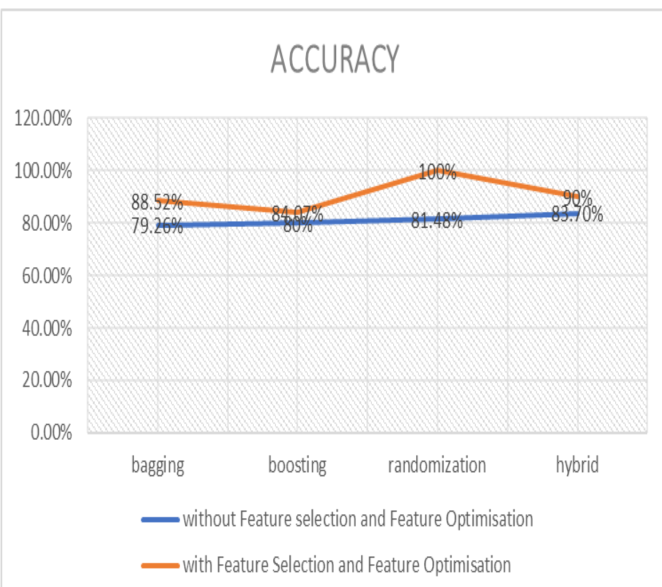

Fig. 7. Comparison of Performance of the Models with and without Feature Selection and Feature Optimisation.

Fig. 7 illustrates the comparison of performance of the model with and without feature selection and feature optimisation technique in heart disease prediction. It is depicted from the line chart that the model with feature selection and feature optimisation techniques have better results than the model without feature selection and feature optimisation technique. Moreover, an increase in accuracy in all the stated ensemble techniques, after the use of feature selection and feature optimization, is observed.

VIII. CONCLUSION AND FUTURE SCOPE

One of the leading causes of death in the modern world is heart disease. However, if it can be predicted beforehand, it can give clinicians crucial information for diagnosis and treatment. In order to prevent cardiac diseases, it is essential to keep track of any health issues. The effectiveness of machine learning in making these predictions has been demonstrated, and the dataset can be used to derive some interesting conclusions. Though there are many existing machine learning models, whose performance needs to be improved in terms of accuracy, consumes more time and are more prone to simulation errors. The prediction behaviour of the model depends upon these factors. A machine learning model built on ensemble learning and using feature selection and feature optimization techniques is suggested to reach this goal.

The experimental findings for the prediction of heart disease utilising ensemble machine learning approaches are presented in this work. To carry out the current study, exploratory, experimental, and applied research approaches were employed. Data about the patient is gathered from the UCI repository. Experiments are conducted using heart disease dataset which are applied on Weka3.8.3 a data mining tool to predict results. Experiments are conducted on two methods, one without using feature selection and feature optimisation method and another on proposed method, using feature selection and feature optimisation technique and later these methods are tested on Ensemble techniques to evaluate the results.

It is seen that results are better with the proposed technique. Thus, the method can be utilised by medical professionals to forecast and detect cardiac illness early on, which helps to avoid problems.

In future, the same approach might be applied to various datasets of various diseases gathered from various sources.

REFERENCES

[1] John A.A. Hunter, Editor, "Davidson's Principles and Practice of Medicine", Churchchill Livingstone An Imprint of Elsevier Limited, (2002).

[2] Subha, R. (2016). Study On Cardiovascular Disease Classification Using Machine Learning Approaches. *International Journal Of Engineering And Computer Science*. doi: 10.18535/ijecs/v4i12.16.

[3] Alabdulwahab, S., & Moon, B. (2020). Feature Selection Methods Simultaneously Improve the Detection Accuracy and Model Building Time of Machine Learning Classifiers. *Symmetry*, *12*(9), 1424. doi: 10.3390/sym12091424.

[4] B. Dengiz, F. Altiparmak and A. E. Smith, Local search genetic algorithm for optimal design of reliable networks, IEEE Trans. Evol. Comput. 1 (1997).

[5] G. R. Harik, F. G. Lobo and D. E. Goldberg, IEEE Trans. Evol. Comput. 3 (1999), 287–297.

[6] C. Miles, S. J. Louis, N. Cole and J. McDonnell, Learning to play like a human: case injected genetic algorithms for strategic computer gaming, in: Proc. 2004 Congr. Evol. Comput. (IEEE Cat. No. 04TH8753), vol. 2, pp. 1441–1448, IEEE, Portland, OR, USA, 2004.

[7] Indu Yekkala, Sunanda Dixit and M.A. Jabbar, "Prediction of heart disease using Ensemble Learning and Particle Swarm Optimization", International Conference on Smart Technology for Smart Nation, (2017).

[8] Russell Eberhart and James Kennedy," A New Optimizer Using Particle Swarm Theory", IEEE (1995), pp.39-43.

[9] Twardowski, K. (1994). An associative architecture for genetic algorithm-based machine learning. Computer, 27(11), 27-38. doi: 10.1109/2.330041.

[10] Jasjit Singh Samagh, Dilbag Singh, "Machine Learning Based Hybrid Model for Heart Disease Prediction" Annals of the Romanian Society for Cell Biology, 2199, (2021).

[11] Dilbag Singh and Jasjit Singh Samagh, "A Comprehensive Review of Heart Disease Prediction using Machine Learning", Journal of Critical Reviews, vol. 7, no. 12, (2020), pp.281-285.

[12] Youness Khourdifi and Mohamed Bahaji, "Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization", International Journal of Intelligent Engineering and Systems, Vol. 12, no. 1, (2019).

[13] K.Vembandasamy, R.Sasipriyaand E.Deepa, "Heart Diseases Detection Using Naive Bayes Algorithm", International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 9, (2015).

[14] Vikas Chaurasia and Saurabh Pal, "Data Mining Approach to Detect Heart Diseases", International Journal of Advanced Computer Science and Information Technology (IJACSIT) Vol. 2, No. 4, Month Year, Page: 56-66, ISSN: 2296-1739, (2012).

[15] Soni, A. (2020). Diabetes Mellitus Prediction Using Ensemble Machine Learning Techniques. SSRN Electronic Journal. doi: 10.2139/ssrn.3642877.

[16] J. Kennedy, "Particle Swarm Optimization", in Encyclopedia of Machine Learning (Springer), C. Sammut and G. I. Webb, Eds. Boston, MA: Springer US, pp. 760–766, (2010).

[17] Saikat Dutt, Subramanian Chandramouli and Amit Kumar Das, "Machine Learning", Pearson India Education Services Pvt. Ltd, India, (2019).

[18] M. Lichman, "UCI Machine Learning Repositry", [Online]. https://archive.ics.uci.edu/, 2013.

[19] Fatma Zahra Abdeldjouad, Menaouer Brahami, and Nada Matta, "A Hybrid Approach for Heart Disease Diagnosis and Prediction Using Machine Learning Techniques", ICOST 2020, LNCS 12157, pp. 299–306, (2020).

[20] U. H. Dataset, "UCI Machine Learning Repository", [online]. https://archive.ics.uci.edu/ml/datasets/statlog+(heart).

[21] Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

[22] 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.

[23] Tyagi, N., 2022. What is Confusion Matrix? | Analytics Steps. [online] Analyticssteps.com. Available at: <https://www.analyticssteps.com/ blogs/what-confusion-matrix>.

[24] Brownlee, J. (2022). How to Calculate Precision, Recall, and F-Measure for Imbalanced Classification. Retrieved from https://machinelearnin gmastery.com/precision-recall-and-f-measure-for-imbalanced-classification/.

[25] H. Ceylan and M. G. H. Bell, Traffic signal timing optimisation based on genetic algorithm approach, including drivers' routing,Transport. Res. 38 (2004), 329–342.