International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

A Novel Intelligent Frame Work for Predicting Heart Diseases Using Machine Learning

Neha Madame[1], Prof. Mashhood Siddiqui[2]

[1]Research Scholar (M.Tech) Department of CSE, Bhabha Engineering Research Institute, Bhabha, University Bhopal, Madhya Pradesh, India

[2]Department of CSE, Bhabha Engineering Research Institute, Bhabha, University Bhopal, Madhya Pradesh, India

Article Info

Publication Issue :
Volume 8, Issue 5
September-October-2022

Page Number : 72-85

Article History
Accepted: 01 Sep 2022
Published: 22 Sep 2022

ABSTRACT

Heart disease is one of the most critical human diseases in the world and affects human life very badly. In heart disease, the heart is unable to push the required amount of blood to other parts of the body. Accurate and on time diagnosis of heart disease is important for heart failure prevention and treatment. the diagnosis of heart disease through traditional medical history has been considered as not reliable in many aspects. To classify the healthy people and people with heart disease, noninvasive-based methods such as machine learning are reliable and efficient. In the proposed study, we developed a machine-learning-based diagnosis system for heart disease prediction by using heart disease dataset. We used popular machine learning algorithms, three feature selection algorithms, the cross-validation method, and seven classifiers performance evaluation metrics such as classification accuracy, specificity, sensitivity and execution time. In this paper, we exploited the Modified Fast Correlation-Based Feature Selection (FCBF) method to filter redundant features in order to improve the quality of heart disease classification. Then, we perform a classification based on different classification algorithms such as K-Nearest Neighbour, Support Vector Machine, Naïve Bayes, Random Forest and a Multilayer Perception Artificial Neural Network optimized by Modified Particle Swarm Optimization (MPSO) combined with Ant Colony Optimization (ACO) approaches. The proposed mixed approach is applied to heart disease dataset; the results demonstrate the efficacy and robustness of the proposed hybrid method in processing various types of data for heart disease classification. Therefore, this study examines the different machine learning algorithms and compares the results using different performance measures, i.e. accuracy, precision, recall, f1-score, etc. A maximum classification accuracy of 99.65% using the optimized model proposed by FCBF, PSO and ACO. The results show that the performance of the proposed system is superior to that of the classification technique presented above.

Keywords : Heart Diseases, PSO (Particle Swarm optimization), ACO (Ant colony optimization)

I. INTRODUCTION

Although there are many potentially fatal diseases, heart disease has been the primary focus of medical investigation. While it's no easy task, a correct diagnosis of heart disease can lead to an automatic prognosis of the patient's heart status, which in turn improves the likelihood of a successful treatment. Physical manifestations of cardiovascular disease are used to make a diagnosis. Factors that increase the likelihood that you'll develop heart disease include never exercising, being overweight or obese, having high blood pressure, and not quitting smoking. A person's heart is a vital organ. That's where all the blood in the body meets. Using a system of veins and capillaries, the heart acts as a pump, distributing oxygen and nutrients to the body at all times. When the heart stops beating and blood stops being pumped around the body, life support systems shut down and the person dies quickly. As a result, more and more medical professionals are turning to IT tools to help them make more informed decisions. It aids medical professionals in disease management, medicine, and the identification of patterns and linkages in data gathered from diagnoses. Many patients who could benefit from preventative treatment for cardiovascular risk are missed by current methods, while others undergo unneeded intervention, because of how risk is currently calculated. To further enhance precision, machine learning can take advantage of intricate connections between risk factors. To see if machine learning can enhance cardiovascular risk prediction, we tested this hypothesis. Experts at Nottingham University's National Institute for Health Research School agree that cardiovascular disease is the main cause of sickness and death worldwide [1]. According to the results of our research, AI has the potential to greatly aid in the fight against the disease by helping more individuals be correctly diagnosed as being at high risk and facilitating early intervention by clinicians to avert major occurrences like cardiac arrest and stroke. The findings demonstrate the importance of AI and ML methods in tailoring patients' risk management plans.

II. Machine learning prediction technique

Supervised learning : Unlike unsupervised learning, with supervised learning the target is known in advance. The overarching goal is to create a model that can correctly anticipate the outcomes of future, as-yet-unknown situations. To train a computer, we provide it with data in the form of questions and their corresponding answers, and then we watch as it discovers patterns in the data. Dependent on the target of the prediction, supervised learning methods can be further broken down into regression and classification. It is a Regression problem if the anticipated value is continuous. It is said to as a Classification problem if the classes to which the variable belongs are themselves independent.

Support Vector Machine: To divide the training data into distinct groups, it builds a hyper plane, depicted in Figure 2, which is a plane in an infinite-dimensional space. Minimizing the error function is a crucial step in building a good decision plane for categorization. These methods can be further broken down into the more specific categories of linear SVMs, polynomial SVMs, sigmoid SVMs, and radial SVMs based on the shape of the error function.

Hence, if you have a set of data points that can be separated linearly, the goal of SVM is to find the hyper plane that maximises the separation. To be more precise, data points closest to the boundary are the ones that are referred to as "support vectors," because they are the most difficult to categorise. The margin around the hyper plane is the statistic that suggests the hyper plane is ideal. Therefore, we are now dealing with an optimization problem. According to what we know now, the best hyper plane can be found by using a maximum margin classifier that has been learned and derived from the training data. Achieving this involves reinterpreting the maximal margin classifier as the inner product (sum of multiplication of pair values) of two given data points. Then, we can use the following definition [3] to characterise the general kernel function: $f(x)=B0+\Sigma(ai\times(x,xi))\ i$ In this case, x is the new input vector, while B0 and ai are the coefficients that need to be estimated using training data.

Decision Tree: Decision trees are a type of supervised learning algorithm that may be applied to the prediction of both categorical and numeric outcomes [4]. They use a tree diagram to show each data instance and its associated class name. The tree can be used to infer a set of rules for assigning the output value to the unknown data record. The internal node is put through a test for a certain attribute. In this illustration of the test's outcome, a tree branch with a class label at its leaf node stands in for the test's passing or failing status. This method involves categorising every data point or sample into two or more groups with similar characteristics. The dividing line is set according to the parameter or factor that proves to be the most useful in drawing distinctions.

Naïve Bayes: Rather than a single classifier it actually is a combination of multiple classifiers all working on the basic Naïve Bayes principle of independent features [5]. Hence each feature is assumed to be independent and autonomous contributing individually to the training data point's probability of belonging to a particular class. As per the Bayes theorem [5], $P(c|x)=P(x|c)P(c)P(x)$ $P(c|X)=P(x1|c)P(x2|c)\times\ldots\times P(xn|c)P(c)$ Here $P(c|x)$is the posterior probability of class given predictor $P(c)$is the prior probability of class $P(x|c)$is the likelihood probability of predictor given class $P(X)$is the prior probability of predictor

Artificial Neural Network: Modules of a dynamic system connected to a learning rule or learning algorithm, these are used to model or simulate the distribution, functions, or mappings among variables. Neuron simulators and the synapse simulating interconnections between these modules across layers make up what is commonly referred to as an ANN here [6]. An ANN is characterised by the function carried out by each individual neuron and the learning method for the dynamic weights provided to the interconnections between neurons. ANN stands out because of its capacity to mimic human thought processes while also learning, growing, and evolving on the fly. In addition, it can process a huge number of parameters and a sizable quantity of noisy data while maintaining a high degree of precision.

Unsupervised learning: In unsupervised learning, our aim is to find unknown trends. The data has no associated labels, but we want to organize the data into groups or clusters[28]. Unsupervised learning techniques are further classified as Cluster Analysis and Dimensionality Reduction. In Cluster Analysis, data is grouped according to similarities or distances between them. In Dimension Reduction, duplicated or unnecessary variables are removed to produce a smaller subset of the original data[29].

III. Literature Review

This paper [1] reviews the most up-to-date articles covering a wide range of diseases and conditions, including but not limited to: heart attack, asthma, diabetes, dengue, allergy, AIDS, vertigo, acne, fungal infection, etc., culled from studies published in Excerpta Medical Database, Google Scholar, WOS, SSRN, PubMed, and Scopus up until January 2022. Thus, we evaluate the outcomes according to a set of quality metrics, and we look ahead to the many problems and obstacles that must be solved before ML approaches can be effectively applied to illness diagnosis. Due to the stresses of modern living, heart disease has risen in prevalence and is now a leading cause of death. We are on the cutting edge of learning about and investigating a more delicate disease. There is a mountain of information and studies on healthcare. As a result, by employing and analysing novel and commendable methods, we may make or forecast the defect of a person who is at risk for developing heart-related disorders and thus aid in the early detection, prevention, and treatment of such conditions.

The author of paper [2] presented a solution to these problems using Machine Learning (ML) and Data Mining (DM) techniques, both of which have been shown to be useful in the field of medicine. The proposed study aims to overcome the limits of prior research by examining risk factors that contribute to negative outcomes like heart disease, as well as novel methods for detecting, predicting, and preventing heart disease.

The cardiovascular patient risk model is outlined in paper [3], which makes use of the author's work on Personal health records (PHRs) available in the data lake in addition to Electronic health records (EHRs). Features gathered from clinical health records form the basis of the proposed risk model. The XML-formatted PHR is part of the preprocessing that will eventually be combined with the patient master database. Several distinct categorization algorithms are employed in this paper [3]. K-Nearest Neighbor, Naive Bayes Classifier, and Random Forest Classifier are used as machine learning algorithms on the data. Machine learning algorithms are used to predict and categorise people with heart disease.

In paper [4], the author employs four Machine Learning algorithms—a Decision Tree, a Logistic Regression, a Random Forest Classifier, and a k-N Neighbor—to make predictions and categorise patients with heart disease. Changing the model parameters may improve the precision with which a heart attack can be predicted given a person's characteristics. This is, thus, a highly helpful strategy. The proposed model's effectiveness became quite satisfactory because it predicted evidence of having a heart illness in a person using Logistic Regression, Random Forest, and Decision Tree, all of which displayed high accuracy in comparison to k-N Neighbor.

The primary objective of the study reported in paper [5] is to utilise ML classification algorithms for cardiac illness prognosis. Cloud-based IoMT diagnostics have been proposed for cardiovascular illness. Rapid analysis of patient data using ML classification methods is made possible by the fog layer. Internet of Things (IoT) technologies are expanding rapidly across many industries, including healthcare services, because they provide significant promise in terms of technology, economy, and society. Improvements in the quality and satisfaction of healthcare services are possible with the help of IoT capabilities like as remote patient monitoring, real-time diagnosis of medical concerns, and more. Wearable devices and the many health monitoring apps they enable are driving growth in the Internet of Medical Things (IoMT). Because of its ability to detect diseases at an early stage, the IoMT plays a crucial role in reducing mortality rates. Among the most difficult tasks involved in analysing clinical datasets is the prediction of cardiovascular disease.

In paper [6] main purpose of the author of to deals with the optimization of a fuzzy system. In fact, the Genetic Algorithm (GA) is applied mainly for tuning the membership and rules parameters. The Root Mean Square Error (RMSE) is considered as the cost function. Cardiac decision support system has become an effective tool for monitoring, classification and prediction of heart diseases. The common purpose of many researches is to improve its performances. In this paper, Genetic-Fuzzy hybrid approach ensuring accuracy is proposed. It is planned to classify ECG signals into five cardiac types, including, the Normal class (N), Paced class (P), Left Bundle Branch Block class (LBBB), Right Bundle Branch Block class (RBBB) and Premature Ventricular Contraction class (PVC). The main purpose deals with the optimization of a fuzzy system. In fact, the Genetic Algorithm (GA) is applied mainly for tuning the membership and rules parameters. The Root Mean Square Error (RMSE) is considered as the cost function.

In paper [7] author offer a unique ensemble architecture that uses a hard voting mechanism to improve performance. XGBoost (XGB), Logistic Regression, Random Forest, and the K-nearest neighbors (K-NN) algorithms are employed in the ensemble architecture. The proposed model scores were obtained with a 94% percent accuracy rate.

The goal of the author in paper [8] is to evaluate classification techniques to see which one is the most accurate in predicting HD using R software. Statistical analysis helps in mining and examining the important factors of HD and can aid in determining whether or not a patient has a cardiac condition. In the proposed architecture the potential of six classification techniques is used to predict heart failure. Namely, Logistic Regression (LR), Naïve Bayes (NB), Decision Tree (DT), K-Nearest Neighbor (KNN), Random Forest (RF), and Support Vector Machine (SVM). According to the results of the analysis, KNN outperforms the other classification techniques in HD diagnosis.

In paper [9] author tested Neural network based approaches for their ability to predict cardiac disease at this point. Based on six different categorization performance indices, evaluated these approaches. In addition, using the receiver operating characteristic curve was used to evaluate these techniques. The Neural network classification algorithm was found to have a maximum classification accuracy of 98.3%. It's no secret that heart disease is a significant source of public health concern. Heart patients are on the rise because of a lack of health awareness and unhealthy eating habits. Because of this, it is vital to establish a basis that can quickly identify the occurrence of cardiac disease in hundreds of samples.[9]

In Paper [10], author proposed an integrated model to use IoT and ML algorithms for a healthcare system. Tracking of patients' status can be done using some sensors such as lightweight, portable, and low-powered sensor nodes. These Sensors sense the patient's status and send the parametric data to the central controller, to take actions during the critical condition of the patients. The data sent to the controller always provided in secure and encryption form. At the same time, patient data is sent to doctors, so that they can provide the instructions to the caretakers of the patients with quick and proper solutions in real-time. For disease prediction, our model uses supervised machine learning algorithms, in order to get the efficient feature set and improve the better accuracy, and pre-processing techniques to eliminate features that are irrelevant, missing values and outliers from biomedical data which aids in better disease prediction. To further strengthen the proposed integrated model design is compared with various traditional classification algorithms to specify its improved accuracy and computational time for accurate prediction of the patient's disease and acts as a decision support system for health care assistants.

In paper [11] author works on SVM and DT algorithm where it has been grouped in 1 and 2, respectively with a g-power value of 80 percent and the heart images were collected from various web sources with recent study findings and threshold 0.05%, Confidence Interval 95% mean and standard deviation. The proposed system using SVM algorithm has achieved an improved accuracy of 72.54%, compared with DT algorithm with an accuracy of 70.32% with a significant value of two tailed tests is 0.027 ($p < 0.05$) with 95% confidence interval. Prediction of heart failure using the SVM algorithm appears to be significantly better than the DT with improved accuracy.

In paper [12] the goal of the proposed research is to make a 10-year prediction based on a person's lifestyle and medical data in order to assess his or her risk of developing coronary heart disease. To understand the gathered data, it has been visualized by the means of Data analysis. Machine learning (ML) techniques, such as Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Decision Tree (DT), are employed on the Framingham dataset to predict the chance of disease.

In paper [13] author developed a machine learning technique to predict whether or not a patient will be diagnosed with heart disease based on their medical history. Physical characteristics, clinical laboratory test results and Symptoms can be examined and used for predicting cardiac disease in electronic medical records. For predicting car di ac related disease algorithms such as Random Forest, Decision Tree, Logistic Regression were used Among these, logistic regression produces the most accurate results.

In this paper [14] , a comparative analysis of different machine learning classifiers based on dataset to predict the chance of heart disease with minimal attributes. The ML algorithms used for HD prediction are K- Nearest Neighbour, Gradient Boosting Classifier, Support Vector Machine, Naive Bayes, Logistic Regression and Random Forest algorithm. The paper also finds the correlation between different attributes and hence using them efficiently for prediction of heart attack. Prediction of heart disease is one of the most complex tasks in medical field and prior detection of heart disease become an area of research to save patient lives. During the pandemic period, the number of cardiac arrest cases at home has drastically increased due to inaccurate predictions and delay in seeking medical attention. The health care industry works on processing of huge data and the solution for this is machine learning.

In paper [15] author proposes an emotion-based music player, which suggests songs based on user's various emotions namely happy, sad, angry and neutral. The application captures the user's photo through web camera and processes the facial image to identify user's emotion using machine learning techniques. Based on the emotion detected, it selects some song to play. The proposed application is more accurate in determining human emotion than existing techniques People in the current world are suffering from lot of stress related diseases due to various reasons. High stress levels may lead to various health hazards like high blood pressure, heart attack etc. One of the stress relief activities is listening to music. If the music played does not suit the current emotion of the listener, it may aggravate stress of the user further. Emotion based music player is a music player based on machine learning techniques which suggests the songs of the playlist based on person's emotions.

In paper [16] Two machine learning methods are employed by the author: DT and Naive Bayes. Heart disease detection and prediction can be improved by combining these two methods. Here are the components and steps: heart disease can be predicted using the Decision Tree algorithm and the Naive Bayes approach. Both the Decision Tree and the Naive Bayes algorithms employ machine learning to make predictions about heart disease. The proposed approach repeated 20 times to get the best results from heart disease images with a G power of 80 percent and a 0.05 percent threshold, the mean and standard deviation of which were in the 95 percent confidence interval (CI) 95 percent. This was necessary to get the best results.

In paper [18] author improve the efficiency of these blocks, several bio-inspired optimization algorithms are proposed by researchers. These include but are not limited to, Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Neural Networks (NN), etc. Each of these algorithms can be applied to optimize individual signal processing blocks, thereby improving overall system performance. Due to a large variety of available bio-inspired algorithms, it is ambiguous for system designers to select the best possible algorithmic combination for their medical disease classification design. In order to reduce this ambiguity, the underlying text evaluates performance of some of the most efficient bio-inspired algorithms, and statistically compares them on basis of their application. These applications vary w.r.t. identified disease, type of signal being processed, etc. This comparison will assist researchers and system designers to develop highly efficient medical disease classification systems for clinical use.

The authors of paper [19] present a method that uses machine learning to evaluate multi-modal bio-signals such as electrocardiogram (ECG) and photoplethysmography (PPG) to predict and understand stroke prognostic symptoms in the elderly in real time. We developed and deployed a real-time, on-the-go stroke illness prediction system using an ensemble structure that integrates convolutional neural networks (CNNs) with long short-term memories (LSTMs). The suggested system takes into account the elderly's desire for portability when designing the bio-signal sensors, and the signals were sampled at a rate of 1,000Hz per second from the three electrodes of the ECG and the index finger for PPG while the subject was walking. The experimental results showed that during senior people's walks, C4.5 decision tree had a prediction accuracy of 91.56%, whereas RandomForest had a prediction accuracy of 97.51%. Also, the prediction accuracy of the CNN-LSTM model trained with raw ECG and PPG data was 99.15%, which is quite good. Consequently, both the prediction accuracy and performance in real-time for older stroke patients were quite good.

There are three models used by the authors of study [20]: the Gaussian Naive Bayes Function, the Random Forest Classifier, and the Neural Network with One-Hot Encoding. Accordingly, the results of the authors' research will be useful in determining which model is most suited to identifying and classifying such flaws. The ECG is a useful diagnostic tool for a wide variety of heart conditions, including atrioventricular blockages. With the help of leads, which record electric impulses created by the heart, aberrant heart activity can be discovered. The superiority of machine learning models was demonstrated in an early investigation for single-lead electrocardiograms. To determine which machine learning algorithm or neural network[20] model can detect and categorise atrioventricular conduction abnormalities with the highest degree of accuracy, we conducted a comparative research utilising ECG-derived Data from the KURIAS-ECG database.

In paper [21] author describes the use of variety of methods in Machine learning classification techniques such as decision trees, random forests, Naive Bayes, SVM, and others can be employed in classification operations. Used the WEKA tool and applied multiple algorithms to the UCI dataset that generated a graph that helped the researcher understand it better. To increase the performance of the model, the researcher used the data science approach. The hybrid method was applied in the proposed model.

In paper [22] author proposed Dual disease prediction technique is user interactive based method. The proposed method observe inputs from the end user

with realistic data to predict heart and diabetes disease. In the presented work, we used Logistic regression model (LR) and Support vector machine (SVM) model for prediction of diseases. The proposed model works with 85 and 78 percent accuracy in prediction of heart and diabetes diseases respectively.
In paper [23] author work aims to determine heart diseases using medical parameters of cardiac patients to improve the accuracy rate to predict in advance. Materials and Methods: One of the Supervised Learning algorithms is Decision tree algorithm, the widely used one for regression as well as classification issues in the problems raised in machine learning. A subset of deep neural networks is CNN Convolutional neural network. Two algorithms used in this research work performed with five different datasets at each time to record five samples.

In paper [24] neural framework-based affliction examination is made to anticipate diverse Diabetic Cardiac disorders joins cardiovascular disease, for instance, angina and myocardial dead tissue (normally known as a heart assault), coronary supply course infirmities (CAD, for instance, stroke, and periphery course sickness incorporate atherosclerosis from the dataset of ECG accumulated from Physio Net Bank.

In paper [25] ML algorithms such as KNN, K-Means, DSC, RFC and NB were used to predict the diseases accurately. The DSC shows the maximum accuracy. This is best because it limits the ambiguity of complex decisions and assigns appropriate values to the results of different actions. The other best accuracy rate is given by KNN algorithm. The KNN classifier can compete with the most accurate models because it gives highly accurate predictions. Therefore, KNN algorithm can be used for applications that require high accuracy but that do not require a human-readable model. The third highest accuracy rate is shown by Naive Bayes algorithm. This is the fastest and easiest forecastable class of test dataset. It also performs better with multiclass forecasting when the assumption of independence is applied. It is superior to other models such as LR and requires less training data.

IV. PROPOSED WORK

This section includes our MPSO/ACO based feature selection and classification system. The main structure of the proposed system is shown in Fig. 1. Our system consists of feature selection-based Correlation-Based Feature selection (FCBF), feature selection based PSO (Particle Swarm Optimization) combined with ACO (Ant Colony Optimization), and classification components based on K-Nearest Neighbour, Support Vector Machine, Naïve Bayes, Random Forest and Artificial Neural Network. The training dataset is prepared according to a binary class classification problem. From the training dataset, features are selected, after that the best subset of features is optimized by our combined PSO/ACO algorithm and then, by using the selected best features, new features are classified with WEKA [17] data mining software implemented in Java. The components of our proposed system are explained in detail in the following subsections. Fig. 1 shows the proposed architecture.

Data Set and attributes

The data is collected from the UCI machine learning repository. The data set is named HeartDisease DataSet and can be found in the UCI machine learning repository. The UCI machine learning repository contains a vast and varied amount of datasets which include datasets from various domains. These data are widely used by machine learning community from novices to experts to understand data empirically. Various academic papers and researches have been conducted using this repository. This repository was created in 1987 by David Aha and fellow students at UCI Irvine. Heart disease dataset contains data from four institutions [18].

1. Cleveland Clinic Foundation.
2. Hungarian Institute of Cardiology, Budapest.
3. V.A. Medical Centre, Long Beach, CA.
4. University Hospital, Zurich, Switzerland.

For the purpose of this study, the data set provided by the Cleveland Clinic Foundation is used. This dataset was provided by Robert Detrano, M.D, Ph.D. Reason to choose this dataset is, it has less missing values and is also widely used by the research community [19].

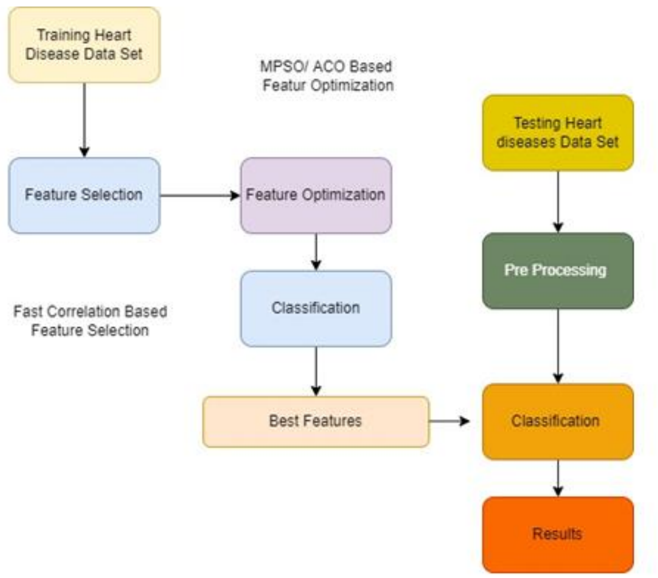

Figure 1: Propose Method Flow Diagram

A. Classification Task

From the perspective of automatic learning, heart disease detection can be seen as a classification or clustering problem. On the other hand, we formed a model on the vast set of presence and absence file data; we can reduce this problem to classification. For known families, this problem can be reduced to one classification only - having a limited set of classes, including the heart disease sample, it is easier to identify the right class, and the result would be more accurate than with clustering algorithms. In this section, the theoretical context is given on all the methods used in this research. For the purpose of comparative analysis, five Machine Learning algorithms are discussed. The different Machine Learning (ML) algorithms are K-Nearest Neighbour (KNN), Random Forest (RF), Support Vector Machine (SVM), Naïve Bayes and Artificial Neural Network (ANN). The reason to choose these algorithms is based on their popularity [20] [26].

B. Feature selection

In the heart disease datasets, the number of features can reach up to tens of thousands; the heart disease dataset has 14 attributes. Since a large number of irrelevant and redundant attributes are involved in these expression data, the heart disease classification task is made more complex. If complete data are used to perform heart disease classification.

Feature optimisation

Particle swarm optimization (PSO)

Swarm intelligence is a distributed solution to complex problems which intend to solve complicated problems by interactions between simple agents and their environment [28]. In 1995, Russel Eberhart, an electrical engineer and James Kennedy, socio-psychologist, were inspired by the living world to set up a metaheuristic: optimization by particle swarm. This method is based on the collaboration of individuals between them: each particle moves and at each iteration, the one closest to the optimum communicates its position to the others so that they can modify their trajectory. This idea is that a group of unintelligent individuals may have a complex global organization.

Due to its recent nature, a lot of research is being done on P.S.O.[27], but the most effective so far is the extension to the framework of combinatorial optimization.

V. Simulation and Result

In this paper, we applied machine learning algorithms on heart disease dataset to predict heart disease, based on the data of each attribute for each patient. Our goal was to compare different classification models and define the most efficient one. From all the tables above, different algorithms performed better depending upon the situation whether cross-validation, grid search, calibration and feature selection is used or not. Every algorithm has its intrinsic capacity to outperform other algorithm depending upon the situation. For example, Random Forest performs much better with a large number of datasets than when data is small while Support Vector Machine performs better with a smaller number of data sets. Performance of algorithms decreased after boosting in the data, which did not feature, selected while algorithms were performing better without boosting in feature selected data. This shows the necessity that the data should be feature selected before applying to boost. For the comparison of the dataset, performance metrics after feature selection, parameter tuning and calibration are used because this is a standard process of evaluating algorithms. The precision average value of the best performance without optimization it's for SVM and NB with 83.6% than RF with 81.4%. These shows SVM and NB are performing on average, after optimized by FCBF we find the best performance of precision it's for MLP with 84.2% than NB with 84% shown In Table 10. In the last stage, we compared the different algorithms with the proposed optimized model by FCBF, PSO and ACO, we find the best one is K-NN with 99.7 % than RF with 99.6 %.

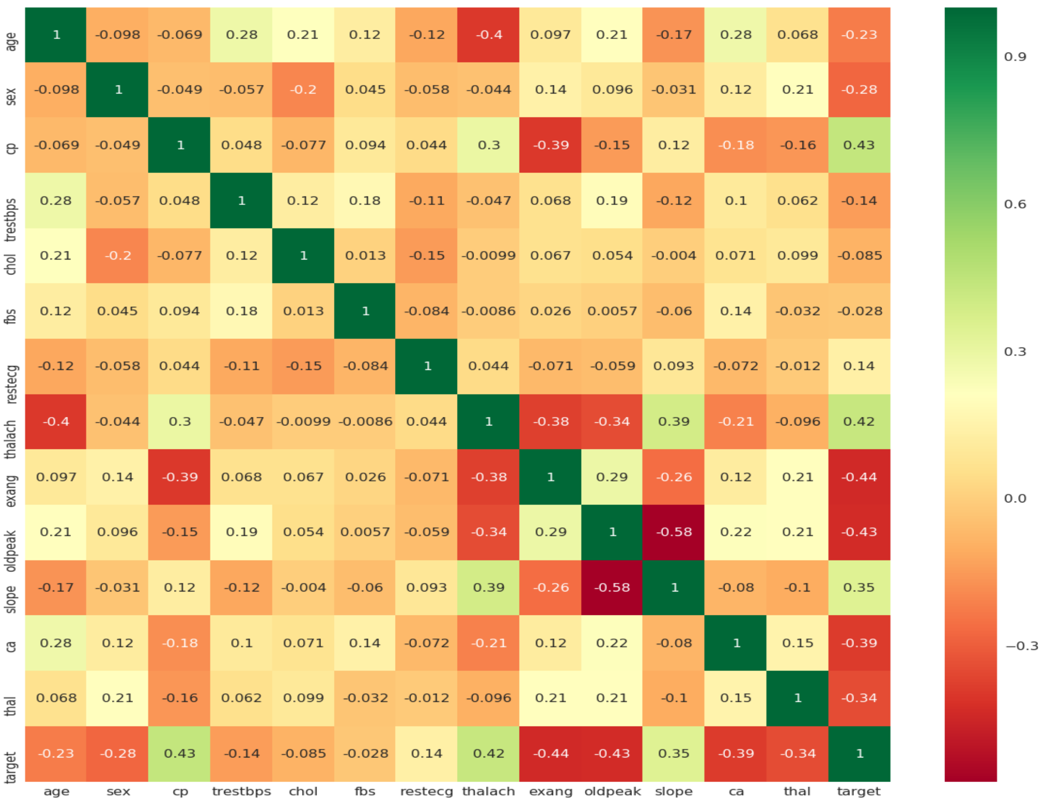

Figure 2 : Correlation matrix

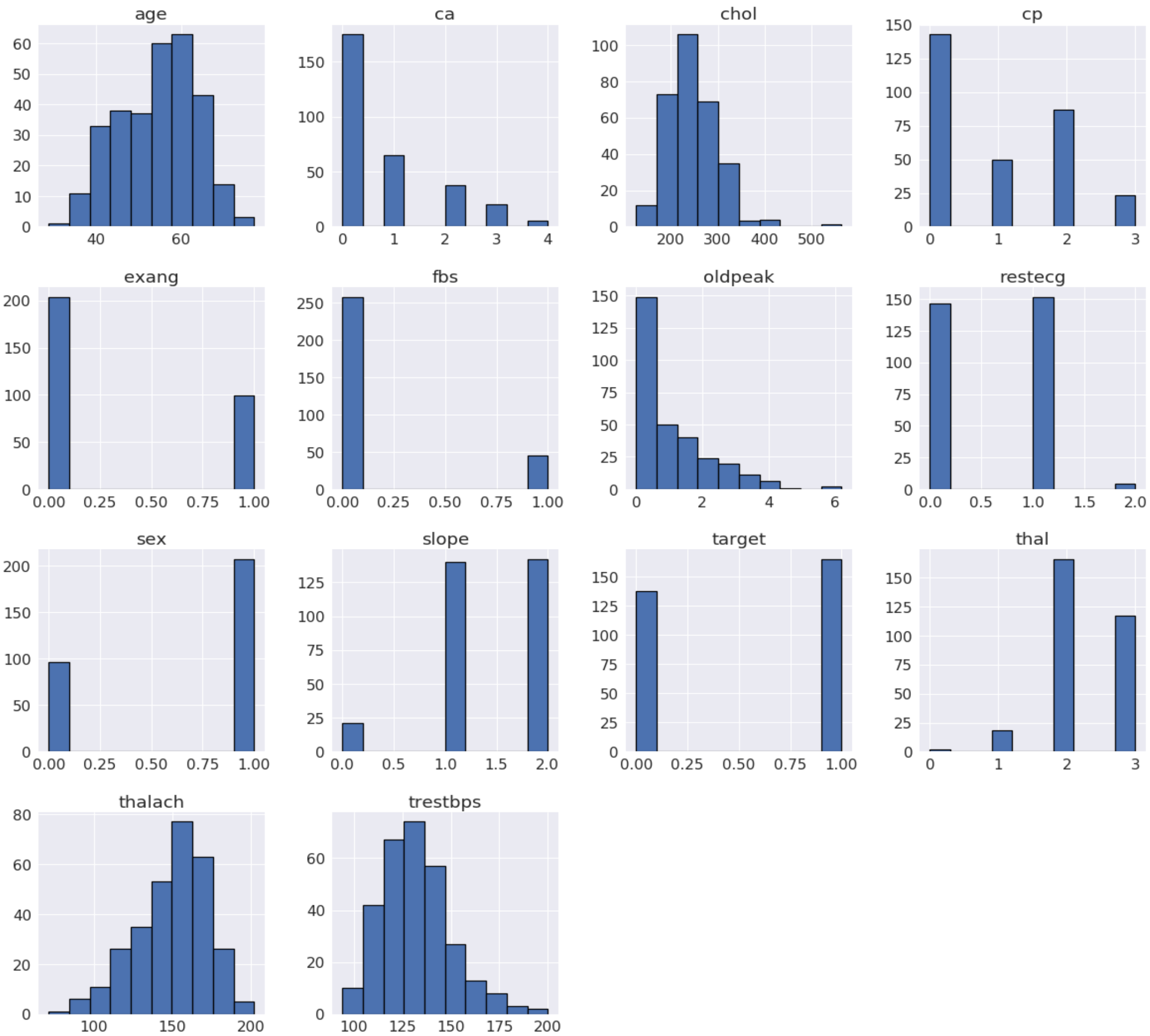

Figure 3: Data Visualization

Table 1: Performance Comparison of Different Methods

Model	Technique	Diseases	Tool	Accuracy
Obtemm et al [11]	Bayes Net	Heart Diseases	WEKA	84
Chaurasia et al [15]	Naïve Bayes	Heart Disease	Juptire Python	84
Parthibal et all[18]	Naïve bayes	Heart Disease	WEKA	74
Proposes Method	KNN	Heart Disease	Jupyter Note Book	99.65
	SVM			83.54
	RF			99.67
	NB			97.00
	NLP			91.65

VI. CONCLUSION

This study's goal was to use machine learning to evaluate and contrast various algorithms' performance under various metric sets. All raw data was put through a series of transformations before being put to use in a predictive test. To varying degrees, the algorithms were successful under different conditions. Preliminary results suggest that K-Nearest Neighbor K-NN, Random Forest RF, and Artificial Neural Network MLP perform well on the dataset utilised in this research. The optimization hybrid technique has been shown to improve the predicted accuracy of medical data sets in experiments. Classification accuracy metrics are used to evaluate the effectiveness of the suggested methods in comparison to supervised algorithms based on preexisting approximate sets. As a result, the analysis section provided conclusive evidence for the superiority of the hybrid MPSO and ACO method to disease diagnosis over the alternatives. Accuracy scores of 99.65 percent with KNN and 99.6 percent with RF are attained by the suggested optimised model with FCBF, MPSO, and ACO. This report has the potential to serve as a springboard for further study into the use of automatic learning in cardiac illness diagnosis. This research is limited in several ways: first, by the author's level of expertise; second, by the resources at hand (such as the speed of the computer); and third, by the time frame in which the research was conducted. This kind of research calls for cutting-edge tools and subject-matter experts.

VII. REFERENCES

[1]. Deepanshu, K. Singh and S. Dhawan, "Diagnosing Multiple Chronic Diseases based on Machine Learning Techniques: Review, Challenges and Futuristic Approach," 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, 2022, pp. 1-6. doi: 10.1109/IC3IOT53935.2022.9767930

[2]. M. Sureshkumar, M. S. V. Sibi, R. P. Bharathi and P. Shanmugapriya, "Myocardial Prediction and Identification using Convolution Neural Networks," 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, 2022, pp. 1-4. doi: 10.1109/IC3IOT53935.2022.9767935

[3]. D. B. M, K. Ramesh, P. N. Renjith and B. Prabha, "A Risk Model Based Heart Disease Prediction Using Data Lake Architecture," 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Trichy, India, 2022, pp. 1-6. doi: 10.1109/ICEEICT53079.2022.9768452

[4]. M. M. Rahman, M. I. Sykat, G. R. Khan and A. Sattar, "Prediction of Heart Diseases Using Machine Learning Approaches," 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Trichy, India, 2022, pp. 1-4. doi: 10.1109/ICEEICT53079.2022.9768486

[5]. C. Chakraborty and A. Kishor, "Real-Time Cloud-Based Patient-Centric Monitoring Using Computational Health Systems," in IEEE Transactions on Computational Social Systems. doi: 10.1109/TCSS.2022.3170375

[6]. H. Lassoued and R. Ketata, "Genetic-Fuzzy Hybrid Approach for Arrhythmia Classification," 2022 5th International Conference on Advanced Systems and Emergent Technologies (IC_ASET), Hammamet, Tunisia, 2022, pp. 138-142. doi: 10.1109/IC_ASET53395.2022.9765913

[7]. R. Jani, M. Shariful Islam Shanto, M. Mohsin Kabir, M. Saifur Rahman and M. F. Mridha, "Heart Disease Prediction and Analysis Using Ensemble Architecture," 2022 International Conference on Decision Aid Sciences and

Applications (DASA), Chiangrai, Thailand, 2022, pp. 1386-1390. doi: 10.1109/DASA54658.2022.9765237

[8]. S. Alshakrani and S. Hilal, "A Comparative Study of Heart Disease Prediction Using Classification Techniques," 2022 International Conference on Decision Aid Sciences and Applications (DASA), Chiangrai, Thailand, 2022, pp. 11-16.

[9]. N. Kumar Trivedi, S. Maheswari, H. Sharma, S. Jain and S. Agarwal, "Early Detection & Prediction of Heart Disease using Various Machine Learning Approaches," 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2022, pp. 793-797. doi: 10.23919/INDIACom54597.2022.9763188

[10]. T. Bikku, K. P. N. V. S. Sree, J. Jarugula and M. Sunkara, "A Novel Integrated IoT Framework with Classification Approach for Medical Data Analysis," 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2022, pp. 710-715. doi: 10.23919/INDIACom54597.2022.9763297

[11]. D. Dharmendra and S. M. S, "Prediction of Heart Failure using Support Vector Machine compared with Decision Tree Algorithm for better Accuracy," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2022, pp. 1535-1540. doi: 10.1109/ICSCDS53736.2022.9760989

[12]. D. Lakshmi Padmaja, B. Sai Sruthi, G. Surya Deepak and G. K. Sri Harsha, "Analysis to Predict Coronary Thrombosis Using Machine Learning Techniques," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2022, pp. 21-27. doi: 10.1109/ICSCDS53736.2022.9760765

[13]. H. K. Medasani, S. A. Kottamasu, S. Doppalapudi and S. Koneru, "Machine Learning Techniques for Cardiac Risk Analysis," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2022, pp. 91-95. doi: 10.1109/ICSCDS53736.2022.9760743

[14]. N. S. Kurian, K. S. Renji, S. Sajithra, Y. R, F. A. Jenefer and S. G, "Prediction of Risk in Cardiovascular Disease using Machine Learning Algorithms," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2022, pp. 162-167. doi: 10.1109/ICSCDS53736.2022.9760879

[15]. K. N. Prasanthi, G. S. S. V. S. Tejaswi, M. J. Sai, G. M. Reddy and L. Yasaswini, "Machine Learning Techniques based Audio Player to Soothe Human Emotions," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2022, pp. 174-178. doi: 10.1109/ICSCDS53736.2022.9760912

[16]. S. K. L. Sameer and P. Sriramya, "Improving the Accuracy for Prediction of Heart Disease by Novel Feature Selection Scheme using Decision tree comparing with Naive-Bayes Classifier Algorithms," 2022 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 2022, pp. 1-8. doi: 10.1109/ICBATS54253.2022.9758926

[17]. K. S. K. Reddy and K. V. Kanimozhi, "Novel Intelligent Model for Heart Disease Prediction using Dynamic KNN (DKNN) with improved accuracy over SVM," 2022 International Conference on Business Analytics for Technology and Security (ICBATS), Dubai, United Arab Emirates, 2022, pp. 1-5. doi: 10.1109/ICBATS54253.2022.9758996

[18]. S. Mujawar and J. Gupta, "A Statistical Perspective for Empirical Analysis of Bio-

Inspired Algorithms for Medical Disease Detection," 2022 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2022, pp. 1-7. doi: 10.1109/ESCI53509.2022.9758270

[19]. J. Yu, S. Park, S. -H. Kwon, K. -H. Cho and H. Lee, "AI-Based Stroke Disease Prediction System Using ECG and PPG Bio-Signals," in IEEE Access, vol. 10, pp. 43623-43638, 2022. doi: 10.1109/ACCESS.2022.3169284

[20]. K. Singh, V. Nair, M. Kumar, R. Shukla, G. S. Wander and A. K. Sahani, "Machine Learning Algorithms for Atrioventricular Conduction Defects Prediction using ECG: A Comparative Study," 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, 2022, pp. 1-5. doi: 10.1109/DELCON54057.2022.9753488

[21]. Puneet and Deepika, "Cardiovascular Disease Prediction Analysis using Classification Techniques," 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, 2022, pp. 1-6. doi: 10.1109/DELCON54057.2022.9753356

[22]. D. Sharathchandra and M. R. Ram, "ML Based Interactive Disease Prediction Model," 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, 2022, pp. 1-5. doi: 10.1109/DELCON54057.2022.9752947

[23]. Sateesh and R. Balamanigandan, "Heart Disease Prediction using Innovative Decision tree Technique for increasing the Accuracy compared with Convolutional Neural Networks," 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), Gautam Buddha Nagar, India, 2022, pp. 583-587. doi: 10.1109/ICIPTM54933.2022.9754196

[24]. D, K. M and M. N. K, "Discovering Cardiovascular infection of Diabetics through Machine Learning Using Physionet bank," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2022, pp. 1-7. doi: 10.1109/ACCAI53970.2022.9752488

[25]. B., S. K, N. Prabhu, A. R. V, S. T and B. S. V, "A Hybrid Approach for Predicting Diseases using Clustering and Classification Techniques," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2022, pp. 1-6. doi: 10.1109/ACCAI53970.2022.9752552

[26]. Khourdifi, Youness & Bahaj, Mohamed. (2019). 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. 12. 10.22266/ijies2019.0228.24.

[27]. Ankit, Pushpneel Verma, and Gaurav Dhiman. "A PSO enable multi-hop clustering algorithm for VANET." International Journal of Swarm Intelligence Research (IJSIR) 13.2 (2022): 1-14.

[28]. Temurnikar, Ankit, Pushpneel Verma, and Jaytrilok Choudhary. "Development of multi-hop clustering approach for vehicular ad-hoc network." International Journal on Emerging Technologies 11.4 (2020): 173-177.

[29]. Temurnikar, Ankit, Pushpneel Verma, and Jaytrilok Choudhary. "Securing vehicular adhoc network against malicious vehicles using advanced clustering technique." 2nd International Conference on Data, Engineering and Applications (IDEA). IEEE, 2020.

Cite this article as :

Neha Madame, Prof. Mashhood Siddiqui, "A Novel Intelligent Frame Work for Predicting Heart Diseases Using Machine Learning ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 5, pp. 72-85, September-October 2022.

Journal URL : https://ijsrcseit.com/CSEIT228519