# Detection of Cardiovascular Disease using Machine Learning, Genetic Algorithms and Particle Swarm Optimization

Manyala Naga Sailaja (Asst. Professor)
Department of Computer Science & Engineering
CMR College of Engineering & Technology
Hyderabad, Telangana

***Abstract*— In today's world, cardiovascular disease (CVD) is prevalent and is the number one cause of death worldwide. Finding issues related to these diseases require relevant historical data to be examined and analyzed. Cardio Vascular Arena has a lot of features. Big data with multiple techniques for processing and retrieving accurate results. Various attributes and methods are used for this forecast. We need a system that predicts the different probabilities of this disease and measures output against high values, percent precision. Machine learning (ML) has a classification approach by looking at patient diseases and conditions. Another optimization is a better approach to handle such non-linear complex problems with very good flexibility and adaptability. Flexibility. Classification operations are performed based on various classification mechanisms such as support vector machines. (SVM), K Nearest Neighbors (KNN), Naive Bayes (NB), Artificial Neural Networks (ANN), and Random Forest (RF) through a genetic algorithm (GA) combined with a particle swarm optimization (PSO) approach. Proposed hybrid applied to cardiovascular classification dataset that provides results that determine efficiency, accuracy, and robustness. The algorithm considers the accuracy of key parameters and examines and compares them using various performance indicators. The maximum classification accuracy of the proposed optimization methods GA and PSO is 99.65%. The result is the suggested. The performance of the system outperforms other traditional classification techniques.**

***Keywords*— Genetic Algorithm, Optimization techniques, Machine Learning, Particle Swarm Optimization, Heart Disease, Cardio Vascular Disease, Support Vector Machine, Random Forest, Naïve Bayes, Classification techniques, Feature Selections.***

## 1. INTRODUCTION

In today's world, cardiovascular disease is a widespread and dangerous infectious disease that rapidly causes heart problems. The pulsating organ of the body is the heart, and if this organ does not work, all the functions of the body will not work properly and will be affected. Organs such as the brain and kidneys. According to WHO statistics, one-third of the world's population dies of heart failure. 2016, the time has come over 18 million people say they are affected by her CVD. More than 84% died of heart failure or heart attack. About three-quarters of deaths from cardiovascular disease were found to have occurred in some down countries. In a 2015 survey. Many people around the age of 70 have non-infectious heart failure, and from the age of 82 he suffers from heart failure until the age of 20 base country. From these data, it was observed that cardiac disease was caused by her CVD. Deaths from cardiovascular disease tend to be due to recognizable hazards, including physical rest, tobacco use, poor eating habits, and destructive alcohol consumption, to name a few bad habits related to health. People with very advanced cardiovascular problems or affected by CVD, or All other risk factors, including some examples of diabetes, hypertension, hypertension, hyperlipidemia, need discussion.in the early stages. Note that CVD puffs with adipose stores have evolved within atherosclerosis and blood clusters. It damages organs such as the eyes, kidneys, heart, and mind. Severe deaths from CVD have occurred in the UK/US. coronary stroke. An episode is a severe event that causes a blockage that affects the heart or blood circulation in the heart most recognized. The goal behind CVD development is fat accumulation and venous effects. Causes of CVD strokes and failures are due to a mixture of several risk factors such as obesity, tobacco use, and unhealthy diet. The heart's circulatory system uses blood vessels to send out blood. Such blood flow removes the heart's metabolic waste products and delivers nutrients and oxygen to different parts of the heart. Inadequate blood flow can cause problems in various organs of the body and lead to further heart failure. Various risk factors for CVD are:

| | | | |
|---|---|---|---|
| High blood pressure | High Cholesterol | Diabetics | Overweight |
| Lot of alcohol consumption | Smoking | CVD in family history | Hypertension |
| Ethnicity | Diet | Gender | Physical Inactivity |
| High LDL and low HDL | Diet | More stress | HB more than 16.5 for long time |

Heart Attack Coronary Artery Disease Symptoms:

- Shortness of breath
- Breast discomfort or pain
- Fatigue
- Irregular heartbeat
- Discontinuity
- Heartburn

The various CVD classifications are:

- Cardiac arrest
- Coronary artery disease
- Heart failure
- Stroke

A number of history tests have been proposed to diagnose CVD. Data mining as an intelligent tool is the tool of choice in healthcare. Disease diagnosis and prediction system. Number of methods such as stress test (ST), magnetic resonance imaging (MRI) and electrocardiogram (ECG) are used to predict the heart, except for physical examinations. Main purpose is prediction. On the problem of CVD using the terms accuracy, efficiency and percentage. Data classification by data mining available datasets using ML techniques are the best way to find prediction processes. Simpler Bayesian or Decision Trees (DT) models are used in ML for classification, using programming to find and improve levels of accuracy. Speech Python, GA is a heuristic-based learning approach derived from the principle of nature-oriented evolution for selective breeding. this A structure containing the population of records that gave a good match and accurate result to the candidate solution for better prediction Use the fitness function after calculating some fitness values from a given population set. Fast average population fitness It can be improved by using this mechanism. These were used for quick identification in some complex high power areas demand. In fact, GA works in conjunction with local search approaches to produce hybrid high search performance to aid investigations. Prediction in ML to solve complex datasets. Kennedy et al. (1995) introduced PSO after studying swarm intelligence based on optimized probabilistic population searches. It comes from CS (computer simulation) and deals with solids such as fish (whether living organisms or particles) flock of birds. The goal of this process is to find food. That is, looking for goals globally based on natural behavior.

My goal is to find global optima from a large set of populations using several nonlinear or multidimensional data sets. After using PSO GA experiences exponential growth in finding the exact solution, while ML speeds up the prediction process. another goal is on the convergence of the global optima in systems from randomly distributed search spaces. Also known as "survival". Optimal in Evolutionary Algorithm (EA)". Here the probability of success is increased, optimizing the predicted correct solution reached. The methodology proposed here systematically uses optimization and classification techniques to achieve the desired goal. A solution by considering different optimization algorithms with different ML approaches. This is a hybrid technology. Leverage fast correlation-based feature selection (FCBF) techniques for redundant feature filtering to improve quality Classify heart problems and classify them. In addition, several classification techniques such as SVM, NB, ANN, RF, and ANN with GA and PSO Optimization techniques used.

## 2. LITERATURE REVIEW

The heart disease predictions made in this paper were based on risk factors, Common types using many WEKA data mining tools (Decision Trees, Artificial Neural Networks, Naive Bayes) etc. Implications in the field of bioinformatics. Kostas Sideris et al. Distinctive remote health monitoring system Prediction of outcome and success based on baseline intervention and first month data. Sickness reduced and cost saved here use an effective RHS system by updating the Wanda CVD and RHM frameworks based on mobile phones and helpful remote instructions for social members. Duffer Hamed et al. Discussed ID3 technology for disease. Mobile Phone and TV Similarity Prediction Using Hidden Programmed Design Recognition Techniques. Coronary heart disease that reduced mortality by counting affected people.

Mai Shouman et al. used data mining k-means clustering for disease prediction, named the MAFIA approach. This is also known as the maximum frequency itemset algorithm i.e. classification of disease prediction with maximum accuracy.

Borkar developed coronary artery disease related to an intelligence system that uses the K-Star algorithm to explain heart infections and create a working system for neural computation about learning vector quantization. Identified 13 clinical facts that gave predictions and provided information for searching Coronary artery disease.

Omar investigated critical signs using recognizable patterns in informative, contextualized data from leading clinical databases from mobile phones. Using gadgets reduces framework execution. The mixed information was gathered in his SVM to help CVD. Considering 1-year data from CVD with severe DCM, predictions were made using ML. Clinical information from32 highlights contributed and were assigned to ML algorithms to leverage different CVD databases for information retrieval. This technique Suitable for heart problems related to expectations using hybrid ML methodology. This proposed hybrid model predicts coronary arteries disease using K-Means with Arbitrary RF Classifiers in an ML Model to Create a Confusion Matrix for Demonstration Robustness.

Dinesh Kumar explores some cardiovascular disease prediction strategies that help find advances and brands Probable patient resolution and patient risk management. Elimination and deletion of missing information Noise in data, grouping attributes for prediction, and changing default values at various levels during preprocessing Various techniques. Work is being done on the medication process for CVD patients. There were two ML steps from Extraction of American datasets from the Fisiologiab Clinica Foundation and the National Organization for Stomach- and Kidney-Related Diabetes illness predict and characterize atherosclerosis using ML techniques with ANNs, using ANNs as classifiers to predict the presence and proximity of atherosclerosis-associated infections.

Bellina et al. Developing an ML approach for diabetics, Placement of CVD patients based on artificial neural networks and Bayesian networks as classifiers. ANN in mild infections were predicted with pattern matching and ML approaches to correct heart problems applying modern strategies to predict CVD risk depending on examination of retinal vessels using ML, Oversampling has shown optimized output with several models such as Qrisk and the established Framingham. Projected by Martin segmentation, filtering, ML and feature extraction. Dynamic Mortality Predictions Created Using ML Approaches to Address Cardiovascular Disease-Related Issues to Improve Fundamentals Guided by educational mode, gather important information using laboratory tests such as red blood cells (RBC), hemoglobin (HGB), aspartate transaminase (AST), alanine transaminase (ALT), platelets (PLT), glucose and creatinine levels, etc.

Balasubramanian aimed to find an SVM-compliant index to detect the risk of confounding coronary eluting stents. A medication regimen that eliminates the risk of patients with complications after DES. ML improves coronary care. Root disease using the NB classifier. SVM, his CVD using ML for percutaneous coronary problems solved using neural. Networks for prediction, light boosting machines, extreme gradient boosting.

Manpreet et al. proposed a structured model. His CVD disease prediction using fuzzy cognitive map (FCN) and structural equation modeling (SEM) ML approach. CVD prediction was performed using automated prediction tools and algorithms useful for calibration-based ML model pipelines, algorithms, feature processing, and data imputation for better results.

Kalman et al. Working on a new cosmology with ML. A CVD visualization ontology for processing complex clinical datasets and eliminating breast-related problems. Determine coronary artery disease with ML technology using the N2 Genetic Optimizer for best identical results. It is associated with heart disease. The work was performed by examining his 1-year CVD data with the ML approach using the NB classifier manufacturing.

Bhuvaneswari et al. Discussed his CVD infection using GA and NN for scaffold preparation classification. We performed a technique on continuous arrhythmic heartbeats using rotated linear kernel SVMs and parallel delta modulation. Photonic crystal investigation by improving fluorescence imaging immunoassay for CVD biomarker screening by ML. The study was conducted considering the Partial Least Squares Regression (PLSR) algorithm, SVM, Advanced ML, and Principal.

Component analysis (PCA) for better characterization. Coronary Artery Problem Considered Based on ML. Find out the size of the test, the dataset, the area where the information is accumulated, some highlights, the ML applied, and the running measurements and diagnostic flaws. ML classifiers are discussed in this study, and RF classifiers are used to predict hepatitis for better study of heart disease problem. Comparisons were made using ML methods for non-small cell lung cancer involving several multi-models, mechanisms of chronic kidney disease, diabetes, random forests optimized for diabetes, hybrid machine learning classifier for finding infectious and chronic. Many experiments have been performed on medical datasets using several feature classification techniques and multiple cardiac problem dataset classifiers to achieve better accuracy. Two hybrid MLs algorithms described as SVM and GA with wrapper method using WEKA and LIBSVM data mining tools for output analysis. That we applied his five datasets from Irvine UC ML as experiments on diabetes, iris, hepatitis, breast cancer, and heart disease. A storage system that gave about 85% accuracy. The coronary artery disease tracking and analysis system using the UCI dataset with 76 features and 303 cases per application defined three algorithms (BN, SVM, and FT) with two tests for detection purposes. Detection is also done in the WEKA tool won 88.3 Curacy. Auto-Learning Cardiac Problems in Diabetic Patients Using SVM and NB approaches. The WEKA tool has been identified. For this purpose, his 500-patient dataset was obtained from the Chennai Institute. SVM and Naive Bayes, Output accuracy of 94.60% and 74%. Cardiac related problems discussed using WEKA data mining with the ML algorithm. There is also slack, applying J48 and Naive Bayes.

This set uses the UCI ML data set with only 76 attributes, of which an additional 11 are used for prediction. Naive, J48, Slack. It offers accuracies of 82.31%, 84.35%, and 85.03%. Of these, bagging is considered optimal with good classification rates. Heart acquired 500 patient data from research institutions, identify issues with Naive Bayes based on the principle of independence, Chennai using her WEKA which gave an accuracy of 86.42. A hybrid classifier related to heart disease, based on the Relief Rough-Set (RFRS) approach, yielded an accuracy of 92.59 in the diagnosed classification. Prediction using Cardiac-Based Effective Hybrid Methods for Extracting and Determining Unknown Knowledge About Cardiac Problems, Artificial neural networks and k-means clustering that provide 97% accuracy. heart-related prediction system using a Powerful Quantized Learning Vector Approach to Find Disease Infections by Computing Neural Functions in a Framework was designed. Classifiers using ML are based on several clinical factors such as accuracy score, correctness, DT, RF, SVM, and logistics, Regression and neural networks in disease research. CVD investigated failure rates based on variable analysis of CNN, distribution distance matrix, transit time of each. Heart rate pulse and her SVM for disease classification and detection. Understanding the Different CVDs to address ML classifiers for atrial disease with dilated cardiomyopathy and septal defect are automatically suggested at the time of ordering by regulated SVM

## 3. METHODOLOGY, EXPERIMENTAL WORK AND RESULT DISCUSSION

UCI is a database collection with real data sets from the ML repository for predicting heart disease, consisting of 300 instance data. This dataset includes chest pain types, blood pressure levels, Electrocardiogram results. A GA/PSO-based proposal system was used to handle classification and feature selection. Illustration 1shows the main structure of the proposed hybrid approach with machine learning using optimization techniques. This technique includes the Feature Selection Based Fast Correlation (FCBF) approach. We use a hybrid approach of GA and PSA that adopts feature selection-based optimization techniques and also SVM, K-Nearest based classification methods. Neighbors, Artificial Neural Networks, Random Forests, Naive Bayes. Important and common features are selected by application. After further optimizing the training dataset optimized using a hybrid approach combining GA/PSO and selecting the best features, Classification was performed using the WEKA data mining tool implemented by Java. Swarm Intelligence Metaheuristic PSO solves complex problems in a very simple way by Russel Eberhart in 1995.

Another name from the development of this optimization approach is social psychologist Jacob Kennedy. There is a separate collaboration that each particle is closest to in each iterative process identify the best positions for communication and create trajectory paths that change with each iterative process. it is based on the principle of optimal particle motion in better paths. PSO's weakness is high, submerged in local optima and dimensional space with a low rate of convergence in an iterative process. In PSO, each particle has:

- Position
- Speed
- Particles in motion (each particle changes its position during motion)
- Best Neighbors
- Best position
- Previous position
- Objective function for comparing current and previous positions

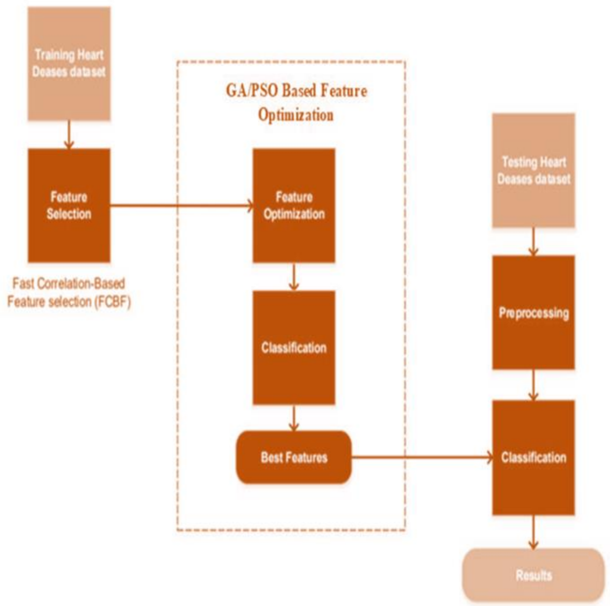


Figure 1: Process of Proposed Hybrid Approach

According to the hybrid technique depicted in the aforesaid Figure 1, the FCBF process was used for feature selection at the first level, and the output was then passed on to the GA and PSO applications at the second level, which produced a better result. ML was applied at the end of the third level through a categorization procedure that produced the best results possible. Table 1 below lists the many heart-related metrics.

| Parameters with notation | Description |
|---|---|
| Age –Ag | Age from 30 to 80 years |
| Sex-Sx | 1 for male, 0 for female |
| Chest pain – cp | 1 for angina, 2 for atypical angina, 3 for non- angina pain, 4 for asymptomatic |
| Restbloodpressure - rbp | Range from 90 to 220 in mm Hg |
| Serum_clestrol - chol | Range from 122 to 555 in mg |
| Fastingblood sugar – fbs | Fbs>120 starting danger zone |
| ECG – ecg | 0-normal, 1-abnormal |
| Maxheartrate - mhr | Range from 70 to 200 |
| Slope- sl | 1 for up, 2 for flat, 3 for down |
| Major vessal – mv | Range from 0 to 3 |
| Thal – th | 3 for normal, 6 for defect, 7 for irreversible defect |
| Class – cl | 1 for unhealthy, 2 for healthy |
| Oldpeak – op | Range from 0 to 60 |
| Exercise induced - Ei | 0 for No, 1 for Yes |

Table 1: Heart disease Parameters with description

PSO has a flowchart as shown in Figure 2. Here, each individual in the population is a particle. So after initializing the population, update calculation of each particle's position and velocity is done in each iteration process to generate pbest and globals. As an expression, the position as gbest of every particle. (1 & 2). After each iterative process completes its duration, the performance of all particles is evaluated against a fitness function or goal. A function called the cost function.

$v\ i\ [t + 1] = w.\ v\ i\ [t] + a1r1(p\ i,best[t] - p\ i\ [t] + a2r2(p\ q,best[t] - p\ i\ [t])$ .............. (1)

$p\ i\ [t + 1] = p\ i\ [t] + v\ i\ [t + 1]$ ........................................... (2) Where, $i = 1,2, \ldots, N$, Here,

- N is swarm population number.
- $v\ i\ [t]$ is velocity vector in $[t]h$ iteration.
- $p\ i\ [t]$ is current position of the ith particle.
- $p\ i,[t]$ is previous best position of ith particle
- $p$ q,$[t]$ is previous best position.
- w is for controlling global and local pressure
- a1and a2 are positive acceleration coefficients
- $r1$ and $r2$ are random number between 0 and 1.

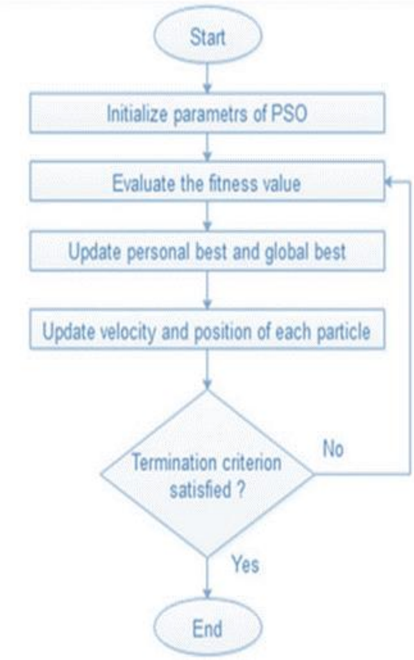


Figure 2: A Flow chart view of PSO

GA is a natural selection-based heuristic search concept (local search) that solves unconstrained and constrained optimization. Problems, derived from H. Charles Darwin's theory of biological evolution and natural evolution, functioning together Inspiration. This algorithm mimics the course of natural selection. The person with the highest score was selected first. This is further used to produce the next generation of offspring. GA finds approximate results. Optimization of search related problems and better results compared to other optimization algorithms. that is Effective and efficient machine learning approaches and optimization of problem solving. Weak against GA. Very low computational efficiency and premature convergence also occur. In today's daily life, it is widely used Engineering, natural sciences, economics. GA goes through various stages, including:

- Population initialization
- Fitness calculator
- Select
- Crossover
- Mutation

Every problem begins with population initialization, followed by a fitness function calculated as a fitness calculation or objective function. A further selection process can be used, which will lead to crossover and mutation, two operations that will optimize the solution. Figure 3 depicts a comprehensive GA flow chart:

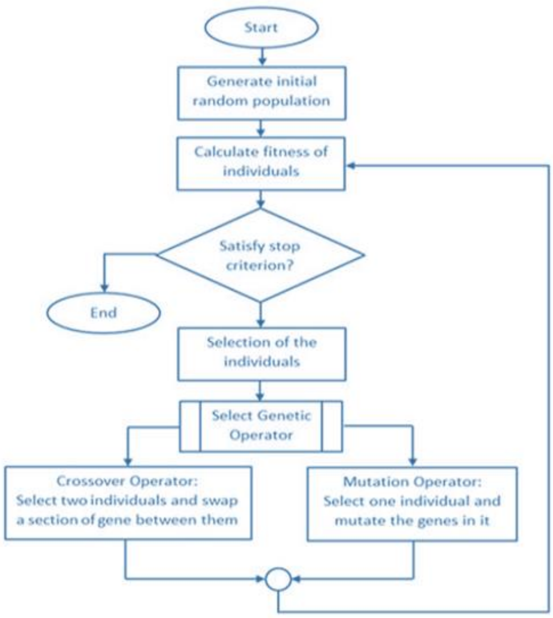


Figure 3: A flowchart representation of the Genetic Algorithm

I worked on a heart disease dataset obtained from a repository of UCI dataset from University of California (Irvine) and only 10 attributes from this datasets were used such as ag, sx, cp, rbp, fbs, cp, chol, ecg, mhr,sl, mv,th, cl, op, and ei with 304 instances as target described in Table I. First, the dataset is cleaned, and then processing begins with preprocessing methods such as data reduction, data transformation, data cleaning, and data integration using the pandas tool. Figure 4 depicts a complete architecture design for the machine learning process.

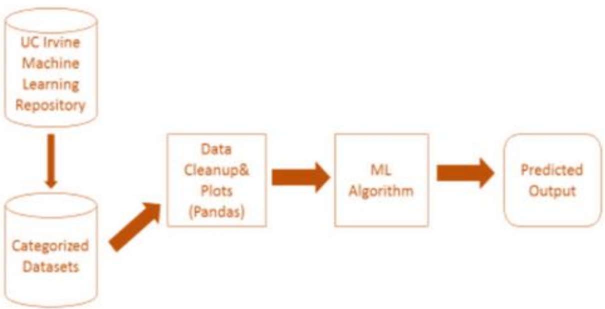


Figure 4: Architectural Design of Machine Learning Process

ML architecture passes through various layers such as data ingestion, data processing, data modeling, execution, and delivery. We also use a few key steps to transform the raw data into a training dataset to enable decision making system. CVD analysis and results are performed as experiments in this section. The experimental platform is actually an important part of the data setup and the outage of the infrastructure. UCI data related to CVD are used for experimental purposes on high-end machines. High speed and high primary storage network system with huge data storage devices plus a quad-core i5 processor. Some coprocessors with 8 GB RAM size or more, secondary storage media > 2 TB, and some tools. Like Ipytho, Pandas, Matplotlib SciPy, and Stats. Models use the environment Jupyter as a web application. Experiment by two people. The phase is the first cleaning of the dataset using the tool pandas, then his second phase is the classification of his CVD clean data using ML. A classifier to get better predictions and accuracy levels. Classifiers with different accuracies are shown in Table 2 and a comparison of accuracies is shown in Figure 5.

Table 2: A comparison report of various Classifiers with varying Accuracy levels

| Classifier | Accuracy (%) | Inaccuracy (%) |
|---|---|---|
| Random Forest | 85.71 | 14.29 |
| Decision Tree | 74.28 | 25.72 |
| Logistic Regression | 74.28 | 25.72 |
| Support Vector Machine | 77.14 | 22.86 |
| K-Nearest Neighbor | 68.57 | 31.43 |

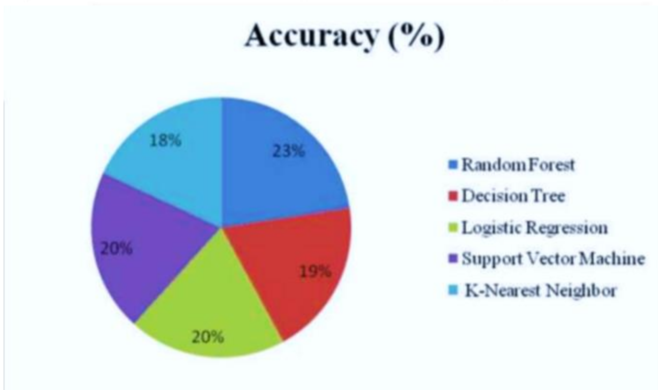


Figure. 5 Analysis of accuracy of classifiers

The goal of this mechanism is to classify CVD datasets using optimization techniques and machine learning methods. WEKA was used for the classification experiment, and some cross validations were performed on some selected features. Every experiment was repeated eight times for optimal accuracy comparisons in classification of selected featured data.

Effective evaluation of all classifiers in three phases using aspects such as accuracy computation, incorrectly classified instances, correctly classified instances, and model building as follows:

**Phase 1:** Non-optimized classifiers
**Phase 2:** FCBF-optimized classifiers

**Phase 3:** Classifier optimization using FCBF, GA, and PSO
Table 3 shows the results without optimization, Table 4 shows the results with FCBF optimization, and Table 5 shows the results with FCBF, GA, and PSO optimization. Also, simulation error was taken into account for improving classifier performance measurements, and its effectiveness and prediction were assessed using Kappa tools (it is an agreement in actual class and classifiers for correct measurement randomly, to find mean absolute error for prediction, error in root mean square, error in Root Relative Squared, and error in Root Relative Absolute). Figures 5, 6, and 7 depict the results.

The accuracy of SVM, MLP, RF, K-NN, and NB optimized by FCBF, GA, and PSO was compared to determine predictive model efficiency.

Table 3: Without Optimization Performance of Classifiers

| Evaluation criteria | K-NN | SVM | RF | NB | MLP |
|---|---|---|---|---|---|
| Time to build model (s) | 0.01 | 0.07 | 0.16 | 0.01 | 0.89 |
| Correctly classified instances | 202 | 226 | 220 | 226 | 222 |
| Incorrectly classified instance | 68 | 44 | 50 | 44 | 48 |

Table 4: With FCBF Optimization Performance of Classifiers

| Evaluation criteria | K-NN | SVM | RF | NB | MLP |
|---|---|---|---|---|---|
| Time to build model (s) | 0.01 | 0.09 | 0.55 | 0.01 | 0.58 |
| Correctly classified instances | 212 | 225 | 217 | 227 | 227 |
| Incorrectly classified instance | 58 | 45 | 53 | 43 | 43 |

Table 4: With FCBF, GA and PSO Optimization Performance of Classifiers

| Evaluation criteria | K-NN | SVM | RF | NB | MLP |
|---|---|---|---|---|---|
| Time to build model (s) | 0.01 | 0.05 | 0.03 | 0.01 | 0.4 |
| Correctly classified instances | 269 | 226 | 269 | 232 | 246 |
| Incorrectly classified instance | 1 | 44 | 1 | 38 | 24 |

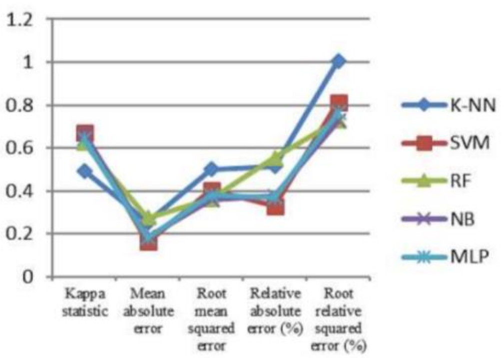


Figure 5: Error in Simulation without Optimization

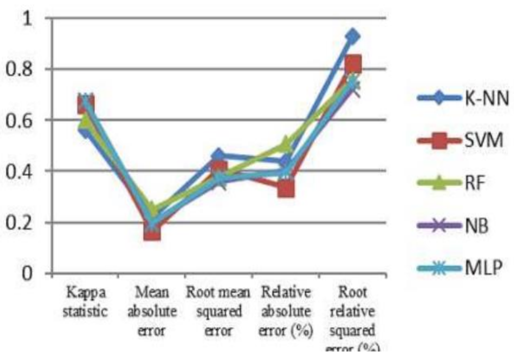


Figure 6: FCBF Optimization Simulation Error

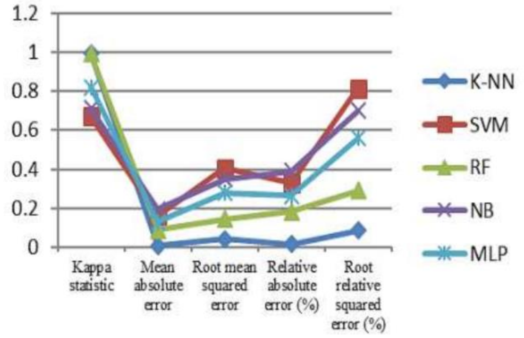


Figure 7: Error in Simulation with FCBF, GA, and PSO Optimization

Every algorithm performed better than the others. In terms of accuracy, RF outperformed SVM, and similarly elaborated comparisons performed better. Finally, a comparison of various techniques with hybrid proposed optimized algorithms using FCBF, GA, and PSO found K-NN to be the best with accuracy 99.65% compared to RF with accuracy 99.6%. As a result, proposed hybrid classifier models optimized using FCBF, GA, and PSO were compared to other methods for CVD problem classification, and a complete comparison was shown in Table 6 by comparing proposed methodology with previous research outcomes and some more classifier models.

Table 6: Performance of various research techniques

| Model | Techniques | Disease | Tool | Accuracy |
|---|---|---|---|---|
| Otoom et al. [11] | Bayes Net | Heart Disease | WEKA | 84.5% |
| | SVM | | | 84.5% |
| | Functional Trees | | | 84.5% |
| Vembandasamy et al. [14] | Naive Bayes | Heart Disease | WEKA | 86.419% |
| Chaurasia et al. [13] | J48 | Heart Disease | WEKA | 84.35% |
| | Bagging | Heart Disease | WEKA | 85.03% |
| | SVM | Heart Disease | WEKA | 94.60% |
| Parthiban et al. [12] | Naive Bayes | Heart Disease | WEKA | 74% |
| Tan et al. [10] | Hybrid Technique (GA + SVM) | Heart Disease | LIBSVM+WEKA | 84.07% |
| Proposed Hybrid Optimized Model by FCBF, GA and PSO | K-NN | Heart Disease | WEKA | **99.65 %** |
| | SVM | Heart Disease | WEKA | 83.55% |
| | RF | Heart Disease | WEKA | **99.6%** |
| | NB | Heart Disease | WEKA | 86.15% |
| | MLP | Heart Disease | WEKA | **91.65%** |

## 4. CONCLUSION AND FUTURE WORK

The purpose of this research work is to make comparisons on different models with different performance measures machine learning algorithms. Here, the given dataset is first preprocessed and then test predictions are applied. whatever the technique. It worked well in some cases and failed in others. ANN, NB MLP, RF and K-NN performed best on the assigned dataset. Experimental results showed an improved accuracy level of prediction using the proposed optimized hybrid method.

A comparison of the proposed approaches was performed with a supervised algorithm with existing approximate classification. Accuracy and given dataset for performance evaluation. During analysis of the effectiveness of the proposed hybrid GA and PSO models. It has been proven for disease diagnosis in comparison with some existing techniques. Hybrid Proposal Optimization Approach has Accuracy of 99.65% for ANN and 99.6% for RF was achieved using FCBF, GA, and PSO. This research looks first at learning how to diagnose heart problems with automated learning methods and for scalable future research.

## REFERENCES

[1] Cardiovascular diseases (CVDs). (2021, June 11). Cardiovascular Diseases (CVDs). https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] Poli, R., Kennedy, J., & Blackwell, T. (2007, January 1). [PDF] Particle swarm optimization | Semantic Scholar. [PDF] Particle Swarm Optimization | Semantic Scholar. https://www.semanticscholar.org/paper/Particle-swarm-optimization-Poli-Kennedy/b7919bcfa38aa97514187501a23c983e8eb5482b

[3] Alshurafa, N., Sideris, C., Pourhomayoun, M., Kalantarian, H., Sarrafzadeh, M., & Eastwood, J. (2016, January 1). Remote Health Monitoring Outcome Success Prediction Using Baseline and First Month Intervention Data | Semantic Scholar. Remote Health Monitoring Outcome Success Prediction Using Baseline and First Month Intervention Data | Semantic Scholar. https://www.semanticscholar.org/paper/Remote-Health-Monitoring-Outcome-Success-Prediction-Alshurafa-Sideris/9f95d7c9055ea2f0475825255dd7586b7eb5ac1d

[4] Abd, D., Alwan, J. K., Ibrahim, M., & Naeem, M. B. (2021, January 1). The utilisation of machine learning approaches for medical data classification and personal care system mangementfor sickle cell disease | Semantic Scholar. The Utilisation of Machine Learning Approaches for Medical Data Classification and Personal Care System Mangementfor Sickle Cell Disease | Semantic Scholar. https://www.semanticscholar.org/paper/The-utilisation-of-machine-learning-approaches-for-Abd-Alwan/c6f8c761177a64b9b8ebe8da1a1793b1723a80df

[5] Shouman, M., Turner, T., & Stocker, R. (2020, January 1). [PDF] Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients | Semantic Scholar. [PDF] Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients | Semantic Scholar. https://www.semanticscholar.org/paper/Applying-k-Nearest-Neighbour-in-Diagnosing-Heart-Shouman-Turner/f7c7ffe100c096a4ae40601126966d01c02d6b26

[6] Borkar, S., & Annadate, M. N. (2020, January 1). Supervised Machine Learning Algorithm for Detection of Cardiac Disorders | Semantic Scholar. Supervised Machine Learning Algorithm for Detection of Cardiac Disorders | Semantic Scholar. https://www.semanticscholar.org/paper/Supervised-Machine-Learning-Algorithm-for-Detection-Borkar-Annadate/5b4a2d98771700b3566f341fb77d7a0a65a59bb5

[7] Boursalie, O., Samavi, R., & Doyle, T. (2020, January 1). [PDF] M4CVD: Mobile Machine Learning Model for Monitoring Cardiovascular Disease | Semantic Scholar. [PDF] M4CVD: Mobile Machine Learning Model for Monitoring Cardiovascular Disease | Semantic Scholar. https://www.semanticscholar.org/paper/M4CVD%3A-Mobile-Machine-Learning-Model-for-Monitoring-Boursalie-Samavi/e595413d15c971b786e8f5405a356b70083ac0d2

[8] Dinesh, K. G., Arumugaraj, K., Santhosh, K., & Mareeswari, V. (2021, January 1). Prediction of Cardiovascular Disease Using Machine Learning Algorithms | Semantic Scholar. Prediction of Cardiovascular Disease Using Machine Learning Algorithms | Semantic Scholar. https://www.semanticscholar.org/paper/Prediction-of-Cardiovascular-Disease-Using-Machine-Dinesh-Arumugaraj/a7b8d2eebe36878415972b51e47550a7ddbcd4f6

[9] Alić, B., Gurbeta, L., & Badnjević, A. (2020, January 1). Machine learning techniques for classification of diabetes and cardiovascular diseases | Semantic Scholar. Machine Learning Techniques for Classification of Diabetes and Cardiovascular Diseases | Semantic Scholar. https://www.semanticscholar.org/paper/Machine-learning-techniques-for-classification-of-Ali%C4%87-Gurbeta/509a8ee7cc9edbf24d9a2940ecbbf6362ccb3392

[10] Gjoreski, M., Simjanoska, M., Gradišek, A., Peterlin, A., Gams, M., & Poglajen, G. (2022, January 1). Chronic Heart Failure Detection from Heart Sounds Using a Stack of Machine-Learning Classifiers | Semantic Scholar. Chronic Heart Failure Detection From Heart Sounds Using a Stack of Machine-Learning Classifiers | Semantic Scholar. https://www.semanticscholar.org/paper/Chronic-Heart-Failure-Detection-from-Heart-Sounds-a-Gjoreski-Simjanoska/d67cf159918d52757c522b28e06fac8d9856b5b1

[11] Gouripeddi, R., Balasubramanian, V., Panchanathan, S., Harris, J., Bhaskaran, A., & Siegel, R. (2009, January 1). Predicting risk of complications following a drug eluting stent procedure: A SVM approach for imbalanced data | Semantic Scholar. Predicting Risk of Complications Following a Drug Eluting Stent Procedure: A SVM Approach for Imbalanced Data | Semantic Scholar. https://www.semanticscholar.org/paper/Predicting-risk-of-complications-following-a-drug-A-Gouripeddi-Balasubramanian/c782523d4b26680912c7472b0742a8da25456169

[12] Singh, M., Martins, L., Joanis, P., & Mago, V. K. (2020, January 1). Building a Cardiovascular Disease predictive model using Structural Equation Model & Fuzzy Cognitive Map | Semantic Scholar. Building a Cardiovascular Disease Predictive Model Using Structural Equation Model & Fuzzy Cognitive Map | Semantic Scholar. https://www.semanticscholar.org/paper/Building-a-Cardiovascular-Disease-predictive-model-Singh-Martins/7c9c233198282f1bea931315fb5606ff196ed828

[13] Farooq, K., & Hussain, A. (2019, January 1). [PDF] A novel ontology and machine learning driven hybrid cardiovascular clinical prognosis as a complex adaptive clinical system | Semantic Scholar. [PDF] a Novel Ontology and Machine Learning Driven Hybrid Cardiovascular Clinical Prognosis as a Complex Adaptive Clinical System | Semantic Scholar. https://www.semanticscholar.org/paper/A-novel-ontology-and-machine-learning-driven-hybrid-Farooq-Hussain/bc7202627f98c3f6b804de131e302c8d2a314086

[14] Bhuvaneswari Amma, N. G. (2017, January 1). Cardiovascular disease prediction system using genetic algorithm and neural network | Semantic Scholar. Cardiovascular Disease Prediction System Using Genetic Algorithm and Neural Network | Semantic Scholar. Retrieved February 16, 2023, from https://www.semanticscholar.org/paper/Cardiovascular-disease-prediction-system-using-and-Amma/4be25c06b82f3f6d326913d4012544c6193dd777

[15] Kumar, N., & Vigneswari, D. (2020, January 1). Hepatitis-Infectious Disease Prediction using Classification Algorithms | Semantic Scholar. Hepatitis-Infectious Disease Prediction Using Classification Algorithms | Semantic Scholar. https://www.semanticscholar.org/paper/Hepatitis-Infectious-Disease-Prediction-using-Kumar-Vigneswari/bf960ea304ab51ce600d8e9a3495bbbb6442f7eb

[16] Groselj, C., Kukar, M., Fettich, J., & Kononenko, I. (1999, January 1). Machine learning improves the accuracy of coronary artery disease diagnostic methods | Semantic Scholar. Machine Learning Improves the Accuracy of Coronary Artery Disease Diagnostic Methods | Semantic Scholar. https://www.semanticscholar.org/paper/Machine-learning-improves-the-accuracy-of-coronary-Groselj-Kukar/e770e8f736905ed6eabecf31a87015dc597acec1

[17] Kumar, N., Vigneswari, D., Kavya, M., Ramya, K., & Druthi, T. L. (2020, January 1). Predicting Non-Small Cell Lung Cancer: A Machine Learning Paradigm | Semantic Scholar. Predicting Non-Small Cell Lung Cancer: A Machine Learning Paradigm | Semantic Scholar. https://www.semanticscholar.org/paper/Predicting-Non-Small-Cell-Lung-Cancer%3A-A-Machine-Kumar-Vigneswari/6d4091f5faf43dc1cf8b236f49837ed056b4ffb0

[18] Singh, R., & Singh, R. (2021, January 1). [PDF] Nature Inspired Job Scheduling For E-Health Services in Mobile Cloud Computing |

Semantic Scholar. [PDF] Nature Inspired Job Scheduling for E-Health Services in Mobile Cloud Computing | Semantic Scholar. https://www.semanticscholar.org/paper/Nature-Inspired-Job-Scheduling-For-E-Health-In-Singh-Singh/69e5b8c2bcf443e2d78b28213350433a4c2607a0

[19] Kumar, N. K., & Vigneswari, D. (2020, August 14). A Drug Recommendation System for Multi-disease in Health Care Using Machine Learning. A Drug Recommendation System for Multi-disease in Health Care Using Machine Learning | SpringerLink. https://doi.org/10.1007/978-981-15-5341-7_1

[20] Sharma, A., & Tivari, N. (2013, January 1). [PDF] A Survey of Association Rule Mining Using Genetic Algorithm | Semantic Scholar. [PDF] a Survey of Association Rule Mining Using Genetic Algorithm | Semantic Scholar. https://www.semanticscholar.org/paper/A-Survey-of-Association-Rule-Mining-Using-Genetic-Sharma-Tivari/8f240f2b69c05025b5c03d3ad89b6fddfbd1ef98