Cardiovascular Disease Recognition using Machine Learning with Genetic and Particle Swarm Optimization

Rachhpal Singh
P.G. Deptt. Of Computer Sc. And Applications
Khalsa College,
Amritsar

ABSTRACT

In today's world Cardio Vascular Disease (CVD) is very common and is the prominent reason for worldwide deaths. To find these diseases related problems, there is a need to study and analyse the related previous data. Cardio Vascular arena has very large data having several techniques for processing and finding accurate results. Various attributes and methods are used for prediction. A system should be to predict various possibilities of this disease-related and the output measured in terms of a high percentage of accuracy. Machine Learning (ML) has a classification approach by studying the patient's ailment and its state. Further optimization is the better approach for handling such non-linear complex problems having very good adaptability and flexibility. Classification operation will be performed based on various classification mechanisms like Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Naïve Bayes (NB), Artificial Neural Network (ANN) and Random Forest (RF) optimized by Genetic Algorithm (GA) combined with Particle Swarm Optimization (PSO) approaches. Hybrid proposed applied to Cardio Vascular dataset for classification giving output that determines the efficiency, accuracy and robustness. Here various ML algorithms examine and make comparisons by using various performance measures by considering the major parameter accuracy. Proposed optimized techniques GA and PSO have a maximum of 99.65% classification accuracy. Outcomes give the proposed system performance is better than 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 Cardio Vascular Disease is a popular perilous infection that cause heart problems with fast speed. Human body's vibrant organ is heart and if that organ not work, then all the operations of body not work properly and then it will impact on brain, kidney etc. organs. WHO's statistical data showed that 1/3rd worldwide population died by heart failure. In 2016 it was mentioned that more than 18 million peoples effected by the CVD [1]. Also more than 84% peoples died due to heart failure or heart stroke. It was noticed that around 3/4th deaths occurred by CVD in some dejected based nations. In a study done in 2015, many people having age around 70 effected by noninfectious and cause heart failure from which 82% belongs to discouraging based nations. From this data, it was observed that heart related brought via CVD [2]. Deaths by CVD tends to a factor of discernible hazards having some examples as physical dormancy, tobacco use, wrong eating habits, destructive liquor utilization and some bad habits related with health . Peoples having very high level cardiovascular problems or CVD affected individuals or any other hazards factors having some examples like diabetes, hypertension, high sickness and hyperlipidemia require discussion at early stage. Note that CVD puffing with greasy stores has a development inside atherosclories and blood clusters [3]. CVD harms organs like eyes, kidneys, heart and mind. High range deaths were occurred in UK/USA due to CVD [4]. Coronary strokes and episodes are intense occasions that create a blockage which effect bloodstream of mind or the heart. The mostly recognized aim behind CVD development is the greasy stores and affects the veins. CVD strokes and failures reason is due to some blend hazard factors as heftiness, tobacco use and unfortunate eating regimen. Circulatory system of heart used the blood vessels for blood pumping. Such blood flow remove metabolic wastage in heart and provide nutrients and oxygen to the various parts of heart. In case of insufficient flow of blood create problems in the different organs of the body that can further create heart failure. Various risk factors for CVD are as:

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

Coronary illness Symptoms of Heart attack:

- Shortness of breath
- Discomfort or pain in chest
- Fatigue
- Irregular heart beat
- Unsteadiness
- Heart burn

Various CVD classifications are:

- Cardiac arrest
- Coronary artery disease
- Congestive heart failure
- Stroke

For diagnose CVD many tests in medical history are suggested. In healthcare sector data mining is a best tool as an intelligent system for diagnosis and prediction regarding diseases. Number of methods like Stress Test (ST), Magnetic Resonance Imaging (MRI) and Electrocardiogram (ECG) are used for heart prediction except than doctor's checkup. Major purpose is prediction regarding the CVD problem by using the term accuracy, efficiency and percentage. Classification of data using data mining to the available dataset with ML technique is best way for finding the prediction process [5]. Further Naïve Bayes or Decision Tree (DT) models are used in ML for classification for finding and enhancing the accuracy level with the use of programming language python [6].

GA's are heuristic based learning approaches derived from principle of natural oriented evolution for selective breed. In this a structure having population of datasets that well belongs to the candidate solution for better prediction and gave accurate results using a fitness function after computing some fitness values from the given population set. So average population fitness rapidly improved by using this mechanism. These were used for identification quickly in some high performance areas having complex searches. Actually GA work with local search approaches for creation of hybrid high search performance and it helps for survey and prediction in ML to solve any complex datasets.

Kennedy et al. (1995) introduced PSO after studying the swarm intelligence based on optimized population stochastic search originated from CS (computer simulation) deals with individuals (whether they are living organisms or particles) like either fish school or bird flock. Goal of this process is to find food i.e. to search the target based on natural behavior globally. Objective is to find global optima from the set of many populations having some non-linear or multidimensional datasets. After the use of PSO with GA exponential growth occurs in finding the exact solution and prediction process becomes fast in ML. Another objective is for convergence of global optima in the system from the randomly distributed search space. It is also known as ''survival of the fittest'' in Evolutionary Algorithms (EAs). Here success rate probability rises and optimized predicted accurate solution will be achieved.

Here in my proposed methodology optimization and classification techniques are used in a systematic way for achieving desired solutions by considering different optimizations algorithms with different ML approaches. Here a hybrid technique involved for exploiting Fast Correlation-Based Feature Selection (FCBF) technique for redundant feature filtrations that will improve quality of heart ailments and classify it. For this some classification techniques like SVM, NB, ANN, RF and KNN with GA and PSO optimization techniques used.

2. LITERATURE SURVEY

Heart disease prediction done in this paper for classification of male patient by discussing some parameters like risk factors, common types etc. using WEKA data mining tools (Decision Tree , Artificial Neural Networks and Naive Bayes) that had much impact in the field of Bioinformatics [7]. Costas Sideris et al. presented health Monitoring remote based system that had outcomes success prediction based on Baseline Intervention and First Month data [8]. Here illness was reduced and cost saved by using effective RHS systems by upgrading the Wanda- CVD and RHM framework based on cell phone that provide all the instructions by remote that was helpful for social members [9]. DhafarHamed et al. discussed ID3 technique for diseases similarities predictions in mobile phones and televisions having concealed and programmed techniques for design recognition of coronary illness that reduced the death rate by counting the affected people [10]. Mai Shouman et al. used the data mining K-

means clustering for disease Prediction naming MAFIA approach. It is also called Maximal Frequent Item set algorithm) for disease prediction classification showing maximum accuracy [11]. Borkar developed coronary heart problem related an intelligence System using algorithm K-Star explain the infection in heart and framed a work system for neural calculation based on learning vector quantization. It acknowledged 13 clinical facts that gave prediction and provide information for finding coronary illness [12]. Omar [13] reconnoitered crucial signs having discernible patterns of contextualized data with information from major clinical databases from the mobile phones. It will decrease the framework execution using some gadget having the mixture information collected to SVM which will be helpful for CVD.

A prediction was done by using ML by considering one year data of CVD having serious DCM [14]. Clinical information from 32 highlights contributed and assigned to ML algorithm for taking various CVD databases for information gain. This technique is better for heart problems related expectation using hybrid ML methodology. This proposed hybrid model forecast coronary illness using K means technique with arbitrary RF classifier in ML model that create a confusion matrix for demonstrating robustness [15]. Dinesh Kumar studied some strategies for anticipating CVDs that helps for finding progressions and makes settlement in high-chance patients and controls the danger in the patients. It make exclusion of missing information, removal of noise in data, attribute grouping for prediction and default value modification at different levels during preprocessing using various techniques[16]. Considering medication process in CVD patients work was done. It had two ML procedures from di Fisiologiab Clinica foundation and taking National organization American dataset of Stomach and Diabetes related with Kidney Diseases [17].

Anticipating and characterizing atherosclerosis illnesses using ML techniques with KNN, ANN as classifiers for foreseeing non-attendance and nearness of infection related with atherosclerosis [18]. Berina et al. elaborated ML approach for diabetics and CVD patients arrangement based on Artificial Neural Networks and Bayesian Networks as classifiers [19]. Using ANN the heart infections was anticipated with pattern matching and ML approach for fixing heart problems [20].

Forecasting of CVD hazard depending on investigation of retinal vessel using ML by applying state-of-the-art strategies and oversampling showed some optimized outputs with some models like Qrisk and entrenched Framingham [21]. Martin projected constant CVD breakdown identification for diagnosing heart problems using ML classifiers foreseeing segmentation, filtering, ML and feature extraction [22].

A prediction on dynamic mortality concocted using ML approach for handling CVD related problems to improve the basic leadership on educational mode and gathering essential information by using some lab test like Red Platelets (RBC), Hemoglobin (HGB), aspartate transaminase (AST), Alanine Transaminase (ALT), platelet(PLT), glucose and creatinine levels as indicators[23]. Balasubramanian intended SVM conformal indicators to find danger in confusions in coronary eluting stent medication methods that remove the danger in patients having post-DES complexities [24]. ML improved the coronary supply in route ailment using NB classifiers [25]. CVD with ML for percutaneous coronary problems resolved by using SVM, Neural Network, light boosting machine and extreme gradient boosting for prediction [26]. Manpreet et al. proposed a structured model for CVD malady expectation using fuzzy cognitive map (FCN) and Structural equation modeling (SEM) [27]. ML approach for CVD forecasting was done using auto prognosis tools and algorithms that help in ML model pipelines on the basis of calibration algorithm, feature processing and data imputation for better results [28]. Karman et al. addressed a new cosmology with ML and ontology for CVD visualization to handle complex clinical datasets and remove chest related problems [29].

To determine Coronary Artery Disease with ML technique using N2 Genetic optimizer agent for getting best identical outputs related with heart diseases [30]. A work was done by studying in one-year CVD data with ML approach using NB classifiers for fabrication [31]. Bhuvaneswari et al. discussed CVD infection using GA and NN for framework preparation [32]. A classification technique on continuous arrhythmia heartbeats was done using the Rotated Linear-Kernel SVM and Parallel Delta Modulations [33]. An investigation on Photonic crystal by enhancing fluorescence imaging immunoassay for CVD biomarker screen with ML investigation was done by considering partial least squares regression (PLSR) algorithms, SVM, advanced ML and Principal component analysis (PCA) for better characterization [34]. Coronary artery problem on the basis of ML discussed by examining test sizes, datasets, areas of information accumulation, some highlights, applied ML and execution measurements for finding deficiencies in diagnosis [35].

ML classifiers discussed in investigation and anticipating hepatitis using RF classifiers for better examination of heart related problems[36]. A comparisons was done using ML techniques on non-small cell lung cancer that ensemble some multi-model mechanisms for chronic kidney disease [37], diabetes mellitus [38], Optimized random forest for diabetes mellitus [39], hybrid machine learning classifier [40] for finding infectious and chronic. Many experiments were done on medical datasets with some

feature classification techniques and multiple classifiers for heart problems dataset to get better accuracy [41]. Two hybrid ML algorithm as SVM and GA with wrapper technique using WEKA and LIBSVM data mining tool for output analysis discussed. It applied as experiment on diabetes disease, Iris, hepatitis, breast cancer disease and heart disease five datasets from Irvine UC ML storage system that gave around 85% accuracy [42].

A Coronary artery disease system for follow-up and analysis using UCI dataset having 76 features and 303 cases by applying three algorithms (BN, SVM and FT) on two tests for detection purposes defined. Also detection done by using tool WEKA and achieved 88.3% accuracy [43]. Heart problem using automatic learning on diabetic patients with SVM and NB approaches using WEKA tool identified. For this purpose 500 patients' datasets taken from Research Institute, Chennai. SVM and Naive Bayes output 94.60% and 74% accuracy respectively [44].

Heart related problem using WEKA data mining with ML algorithm discussed. Also bagging, J48 and Naive Bayes applied on this sets using UCI ML data set having only 76 attributes from which further 11 taken for prediction. Naïve, J48 and bagging provide 82.31%, 84.35% and 85.03% accuracy from which Bagging considered as best with good classification rate [45]. Heart problem identified with Naive Bayes based on independence principle by taking 500 patients data from research institutes, Chennai using WEKA that gave 86.42% accuracy [46]. Heart disease related hybrid classification method based on the ReliefF-Rough-Set (RFRS) approach and giving 92.59% accuracy for classification diagnosed [47].

Prediction of heart based effective hybrid method for extracting and determining unknown knowledge of heart problems using artificial neural network and K-means clustering that provide 97% accuracy [48]. Heart related prediction system using a powerful quantization learning vector approach that finds infection in the illness by computing the neural in the framework designed [49]. The classifiers using ML based on some clinical factors like precision values, legitimacy, DT, RF, SVM, logistic regression and neural network during disease examination [50].

CVD examined the breakdown rate on the basis of CNN, distribution distance matrix, variable analysis of transit time of every pulse of heart beat and SVM for classification and recognition of disease [51]. To understand the various CVDs using directed ML classifiers for dilated cardiomyopathy automatically and atrial ailment having Septal Defect proposed for arranging the regulated SVM [52].

3. METHODOLOGY, EXPERIMENTAL WORK AND RESULT DISCUSSION

UCI is database collection using ML repository real dataset for prediction of heart disease consisting of 300 instance data from this dataset considering 14 clinical heart problem related test parameters like chest pain type, blood pressure level and electrocardiographic result. GA/PSO based proposed system was taken processing classification and feature selection. Figure 1 showing major structure of hybrid proposed approach with machine learning using optimization techniques.

This technique has Feature-selection-based-Fast-Correlationion (FCBF) approach. It use hybrid approach having GA with PSA optimization techniques that is based on feature selection and also adopt classification procedure based on SVM, K-Nearest Neighbour, Artificial Neural Network, Random Forest and Naïve Bayes. Major and common features are selected by using training dataset that was further optimized by using hybrid combined GA/PSO approach and after selecting the best features, classification was done with data mining tool WEKA implemented by Java.

Swarm intelligence a meta-heuristic PSO solves complex problems in very simple way from 1995 by Russel Eberhart who is by profession an electrical engineer [56]. Another name from the development of this optimization approach is socio-psychologist James Kennedy. It has individual collaboration where every particle moves in every process of iteration showing the closest optimum position for communication and create trajectory path that modified in every process of iterations. It is based on principle of optimum move of particle in a better path [57]. Weakness of PSO is the falling it in local optima in a high dimensional space having low convergence rate in iterative process. In PSO every particles has:

- A position
- A speed
- Particle having movement (every particle changed his position during move)
- A best neighbor
- A best position
- Previous position
- An objective function for comparison purpose between current and previous position

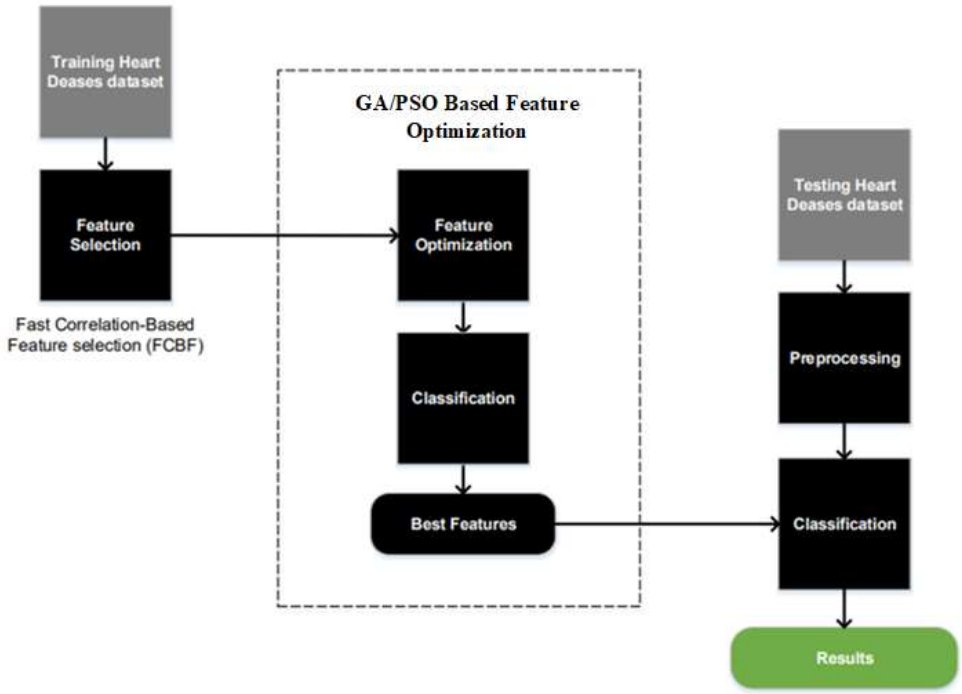

Figure 1: Process of Proposed Hybrid Approach

In above figure 1, hybrid approach explained us that at first level FCBF process applied for feature selection and then the output passed in second level where GA and PSO applied in a sequence giving better output. At end in the third level ML applied through classification process that gave optimized results. Below Table 1 show the different parameters related with heart problems.

Table 1: Heart disease Parameters with description

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

PSO has flow chart as shown in Figure 2. Here every individual in population is a particle. So after initialize population, updating of position and velocity of every particle takes place in every process of iteration based that give pbest and also give global location as gbest of all particles as showed Eqs. (1 & 2).

When every process of iteration finished its period, all particles performance evaluated using a fitness function or objective function called cost functions.

$v\ i\ [t+1] = w.\ v\ i\ [t] + c1r1(p\ i,best[t] - p\ i\ [t] + c2r2(p\ g,best[t] - p\ i\ [t])$..... (1)

$p\ i\ [t + 1] = p\ i\ [t] + v\ i\ [t + 1]$……. (2)

Where, $i = 1,2, \dots , N$,

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
- $p\ g,[t]$ is previous best position.
- W is for controlling global and local pressure
- c1and $c2$ are positive acceleration coefficients
- $r1$ and $r2$ are random number between 0 and 1.

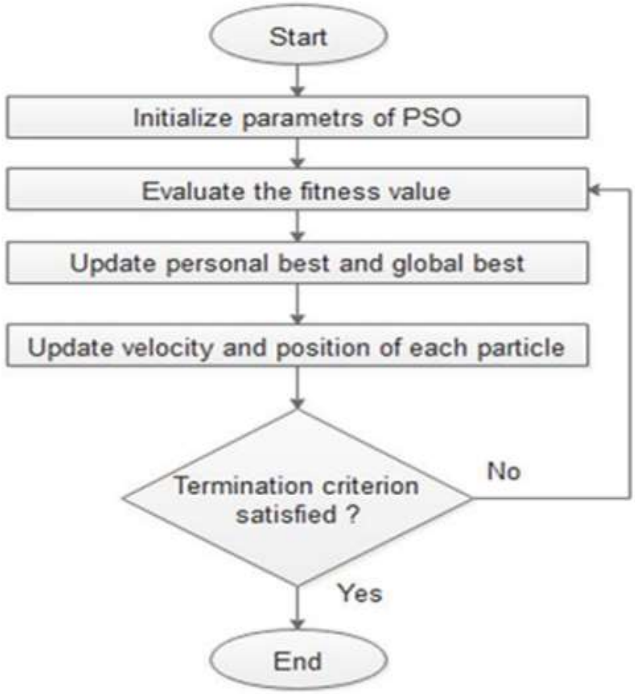

Figure 2: A Flow chart view of PSO

GA is a natural selection based heuristic search (local search) concept that solves unconstrained and constrained optimization problems i.e. it is derived from biological evolution process and natural evolution theory by Charles Darwin work as an inspiration [55]. Natural selection process reflection occurs in this algorithm. Individuals having fittest values were selected first for reproduction that further used for production of offspring for next generation [56]. GA finds approximate results for optimization the problem related with search and give better results as compared to other optimization algorithms. It is an effective and efficient, approach for machine learning and optimization the problem's solution [57]. Weakness of GA is having very low computation efficiency and a premature convergence also occurs [58][59]. In today's daily life it is widely used in engineering, scientific and business fields. GA passes through various phases as:

- **Population initialization**
- **Fitness calculation**
- **Selection**
- **Crossover**
- **Mutation**.

Every problem start from the population initialization, then a fitness function computed as fitness calculation or objective function. Further selection process can be applied which will move towards crossover and mutation two operations that make the solution more optimized. A complete GA flow chart is as shown in Figure 3 below:

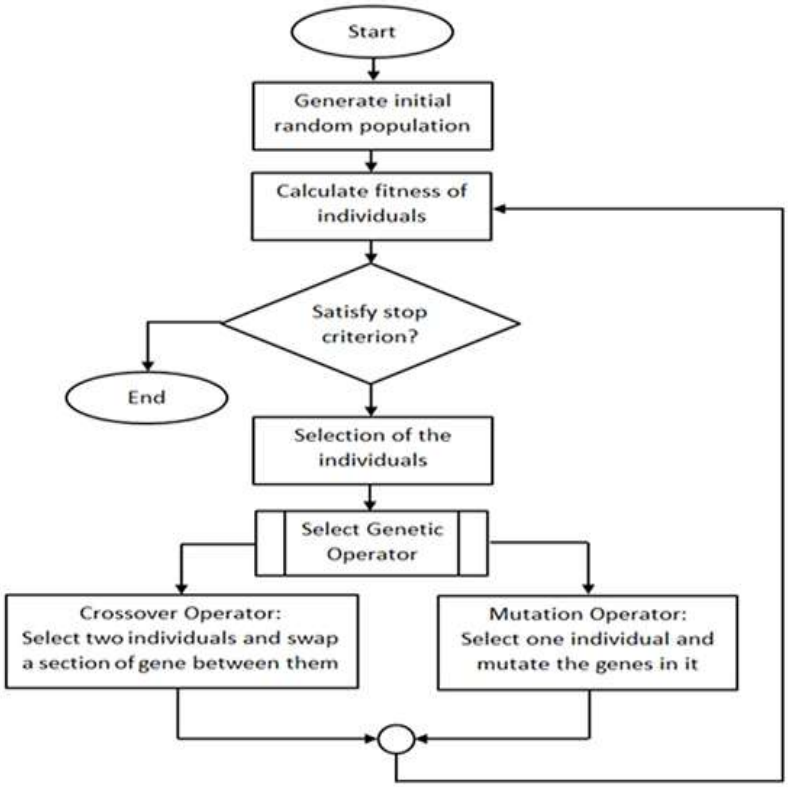

Figure 3: A Flow chart view of Genetic Algorithm

I worked on heart disease dataset obtained from a repository of UCI dataset from University of California (Irvine) was obtained for heart diseases and only 10 attributes from this datasets are used like ag, sx, cp, rbp, fbs, cp, chol, ecg, mhr,sl, mv,th, cl, op, and ei having 304 instances as target described in Table I. First of all dataset be cleaned and then processing takes place by using preprocessing methods as Data reduction, Data transformation, Data cleaning, and Data Integration using the tool named as pandas. A complete architecture design for machine learning process is as shown in the Figure 4.

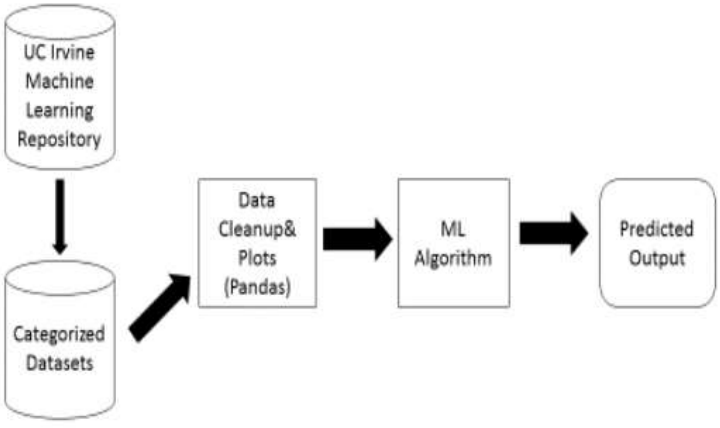

Figure 4: Architectural Design of Machine Learning Process

ML architecture passes through various layers like data acquisition, data processing, data modeling, execution and deployment and having some major steps for transformation of raw data into training data sets capable for enabling the decision making of a system.

CVD analysis and findings are done as experiment in this section. Experimental platform is actually a critical part for data setup and setting of its infrastructure. So UCI data related with CVD used for experimental purpose on high end machine and having a high speed network system with high primary memory with huge data storage equipment. For this a quad-core i5 processor with some co-processors taken having RAM size minimum 8 GB having more than 2 TB secondary storage media with some tools like Ipytho, pandas, Matplotlib SciPy and StatsModels using an environment Jupyter as web application. Experiment done in two stages, first cleaning of datasets takes place with tool pandas and then at second stage classification of CVD tidy data using ML classifiers for better prediction and get the accuracy level.
Different classifiers having different accuracy is as shown in the Table 2 and an Accuracy comparisons showed in Figure 5.

Table 2: Showing a comparison report of various Classifiers having different Accuracy level.

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

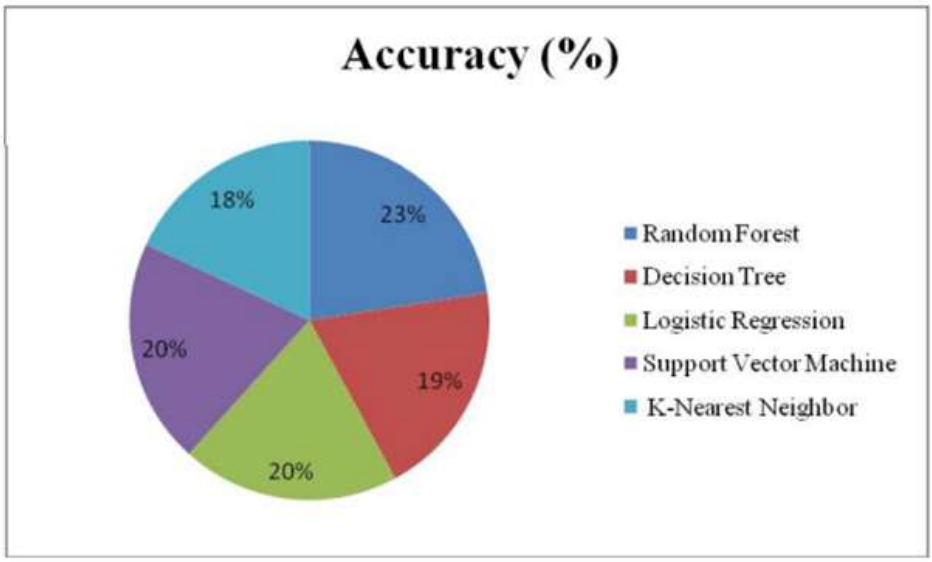

Figure 5: Classifiers Accuracy Analysis

Objective is this mechanism to do a classification of CVD datasets with optimization techniques and some machine learning methods. WEKA environment used for classification experiment and also some cross validations were done on some selected features. For optimal accuracy comparisons in classification of selected featured data, every experiment was done 8 times. Effective evaluation of all classifiers using some aspects like accuracy computation, incorrectly classified instances, correctly classified instances and building model done in three phases as:

Phase1: Classifiers without optimization

Phase2. Classifiers optimized by FCBF

Phase3. Classifiers optimized by FCBF, GA and PSO

Outcomes without optimization showed in Table 3, optimized by FCBF in Table 4 and optimized by FCBF, GA and PSO in Table 5. Also simulation error considered for improving the classifier performance measurements and for this effectiveness and prediction was evaluated 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). Outcomes are shown in Figures 5, 6 and 7.

Efficiency of predictive model was checked by comparing the accuracy for SVM, MLP, RF, K-NN and NB optimized by FCBF, GA and PSO.

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

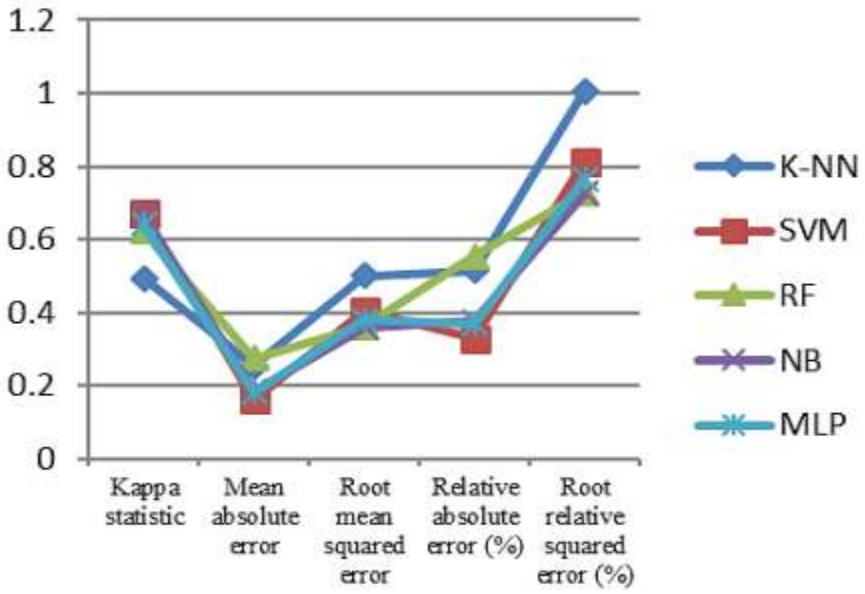

Figure 5: Simulation Error without Optimization

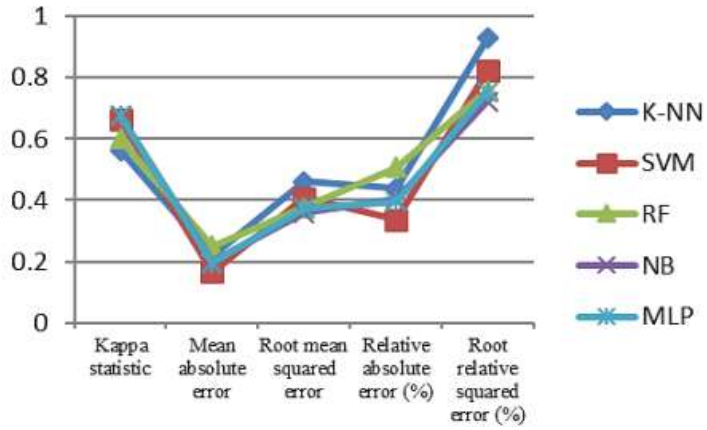

Figure 6: Simulation Error with FCBF Optimization

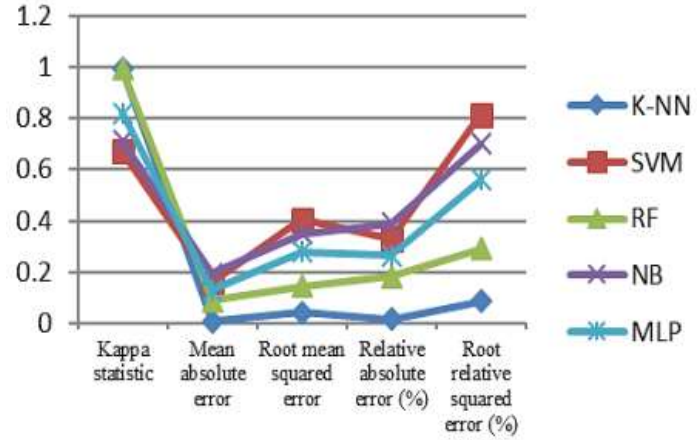

Figure 7: Simulation Error with FCBF, GA and PSO optimization

Every algorithm outperformed than other algorithms. RF performed better than SVM and similarly comparisons elaborated better here in terms of accuracy. At end, a comparison done with various techniques with hybrid proposed optimized algorithm using FCBF, GA and PSO and found K-NN as best having accuracy 99.65 % as compared to RF having accuracy 99.6 %.

So proposed hybrid classifiers models optimized using FCBF, GA and PSO as compared to other methods for classification of CVD problems and a complete comparison was shown in Table 6 by comparing proposed methodology with previous research outcomes and some more classifier models.

Table 6: Various Research Techniques performances

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

Objective in this research paper is to make a comparison in various models having various performance measurements with machine learning algorithms. Here given datasets first pre-processed and then a test prediction applied on it. Every technique worked better in some cases and worst in some other cases. ANN, NB MLP, RF and K-NN worked best on assigned datasets. Experimental outputs showed the increase in accuracy level for prediction using the proposed optimized hybrid technique. A Comparison of proposed approach was done with supervised algorithms having some existing approximate classification accuracy and given datasets for performance evaluation. During analysis, effectiveness of proposed hybrid GA and PSO model was demonstrated for diagnosing the disease as compared to some existing techniques. Hybrid proposed optimized approach using FCBF, GA and PSO achieved 99.65% accuracy with KNN and 99.6% with RF. This research first will deal with learning the way to diagnosis heart problems with automatic learning methods and for future research work that can be extended.

REFERENCES

[1]. https://www.who.int/news-room/fact-sheets/detail/cardiovasculardiseases-(cvd)

[2]. Kelly, B. B., & Fuster, V. (Eds.). (2010), “Promoting cardiovascular health in the developing world: a critical challenge to achieve global health”, National Academies Press.

[3]. Poirier, Paul, et al. (2006), "Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss: an update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease from the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism.", Circulation 113.6: 898-918.

[4]. Bhatnagar, Prachi, et al. (2016), "Trends in the epidemiology of cardiovascular disease in the UK.", Heart 102.24: 1945-1952

[5]. Beunza, Juan-Jose, et al. (2019), "Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease).", Journal of biomedical informatics 97: 103257.

[6]. Zhao, Lina, et al. (2019), "Enhancing Detection Accuracy for Clinical Heart Failure Utilizing Pulse Transit Time Variability and Machine Learning.", IEEE Access 7: 17716-17724.

[7]. Gandhi, Monika, and Shailendra Narayan Singh (2015), "Predictions in heart disease using techniques of data mining.", International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), pp. 520-525. IEEE.

[8]. Nabil Alshurafa, Costas Sideris, Mohammad Pourhomayoun, Haik Kalantarian, Majid Sarrafzadeh (2016), "Remote Health Monitoring Outcome Success Prediction using Baseline and First Month Intervention Data", in IEEE Journal of Biomedical and Health Informatics, 21: 507-514.

[9]. Ponrathi Athilingam, Bradlee Jenkins, Marcia Johansson, Miguel Labrador (2017), "A Mobile Health Intervention to Improve Self-Care in Patients With Heart Failure: Pilot Randomized Control Trial", JMIR Cardio 2017, vol. 1, issue 2, pg no:1 .

[10]. DhafarHamed, Jwan K. Alwan, Mohamed Ibrahim, Mohammad B. Naeem (2017), "The Utilisation of Machine Learning Approaches for Medical Data Classification", Annual Conference on New Trends in Information & Communications Technology Applications, pp. 213-218.

[11]. Mai Shouman, Tim Turner, and Rob Stocker (2012), "Applying k-Nearest Neighbour in Diagnosing Heart Disease Patients", International Journal of Information and Education Technology, Vol. 2, No. 3, pp.220-223.

[12]. Borkar, Sneha, and M. N. Annadate (2018), "Supervised Machine Learning Algorithm for Detection of Cardiac Disorders", Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), IEEE, 2018, pp.1-4.

[13]. Omar Boursalie, Reza Samavi, Thomas E. Doyle (2015), "M4CVD: Mobile Machine Learning Model for Monitoring Cardiovascular Disease.", Procedia Computer Science, 63 : 384-391.

[14]. Chen, Rui, et al. (2019), "Using Machine Learning to Predict One-year Cardiovascular Events in Patients with Severe Dilated Cardiomyopathy.", European Journal of Radiology, 117 : 178-183.

[15]. Dhar, Sanchayita, et al. (2018), "A Hybrid Machine Learning Approach for Prediction of Heart Diseases.", 4th International Conference on Computing Communication and Automation (ICCCA). IEEE, 2018, pp.1-6.

[16]. Dinesh, Kumar G., et al. (2018), "Prediction of Cardiovascular Disease Using Machine Learning Algorithms.", International Conference on Current Trends towards Converging Technologies (ICCTCT). IEEE, 2018, pp.1-7.

[17]. Mezzatesta, Sabrina, et al. (2019), "A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis.", Computer Methods and Programs in Biomedicine 177 : 9-15.

[18]. Terrada, Oumaima, et al. (2019), "Classification and Prediction of atherosclerosis diseases using machine learning algorithms.", 5th International Conference on Optimization and Applications (ICOA). IEEE, 2019, pp.1-7.

[19]. Alić, Berina, Lejla Gurbeta, and Almir Badnjević (2017), "Machine learning techniques for classification of diabetes and cardiovascular diseases.", 6th Mediterranean Conference on Embedded Computing (MECO). IEEE, 2017, pp.1-4.

[20]. Awan, Shahid Mehmood, Muhammad Usama Riaz, and Abdul Ghaffar Khan (2018), "Prediction of heart disease using artificial neural network." VFAST Transactions on Software Engineering 13.3: 102-112.

[21]. Fathalla, Karma M., et al. (2016), "Cardiovascular risk prediction based on Retinal Vessel Analysis using machine learning.", IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2016.

[22]. Gjoreski, Martin, et al. (2017), "Chronic Heart Failure Detection from Heart Sounds Using a Stack of Machine-Learning Classifiers.", International Conference on Intelligent Environments (IE). IEEE, 2017, pp.1-14.

[23]. Metsker, Oleg, et al. (2018), "Dynamic mortality prediction using machine learning techniques for acute cardiovascular cases.", Procedia Computer Science 136 : 351-358.

[24]. Balasubramanian, Vineeth Nallure, et al. (2009), "Support vector machine based conformal predictors for risk of complications following a coronary drug eluting stent procedure.", 36th Annual Computers in Cardiology Conference (CinC). IEEE, 2009, pp. 5-8.

[25]. Groselj, C., et al. (1997), "Machine learning improves the accuracy of coronary artery disease diagnostic methods.", Computers in Cardiology. IEEE, 1997, pp. 57-60.

[26]. Zhou, Yijiang, et al. (2019), "Machine Learning-Based Cardiovascular Event Prediction For Percutaneous Coronary Intervention.", Journal of the American College of Cardiology 73.9 Supplement, pp. 1-127.

[27]. Singh, Manpreet, et al. (2016), "Building a cardiovascular disease predictive model using structural equation model & fuzzy cognitive map.", IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). IEEE, 2016, pp.1377-1382.

[28]. Alaa, Ahmed M., et al. (2019), "Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants.", PloS one 14.5 : e0213653.

[29]. Farooq, Kamran, and Amir Hussain (2016), "A novel ontology and machine learning driven hybrid cardiovascular clinical prognosis as a complex adaptive clinical system.", Complex Adaptive Systems Modeling 4.1 : 12.

[30]. Abdar, Moloud, et al. (2019), "A new machine learning technique for an accurate diagnosis of coronary artery disease.", Computer methods and programs in biomedicine 179 : 104992.

[31]. Chen, Rui, et al. (2019), "Using Machine Learning to Predict One-year Cardiovascular Events in Patients with Severe Dilated Cardiomyopathy.", European Journal of Radiology, 177: pp.178-187.

[32]. Amma, NG Bhuvaneswari (2012), "Cardiovascular disease prediction system using genetic algorithm and neural network.", International Conference on Computing, Communication and Applications. IEEE, 2012, pp. 1-5.

[33]. Tang, Xiaochen, et al. (2019), "A Real-time Arrhythmia Heartbeats Classification Algorithm using Parallel Delta Modulations and Rotated Linear-Kernel Support Vector Machines.", IEEE Transactions on Biomedical Engineering vol 67, no.4, pp. 978-986.

[34]. Squire, Kenneth J., et al. (2019), "Photonic crystal-enhanced fluorescence imaging immunoassay for cardiovascular disease biomarker screening with machine learning analysis.", Sensors and Actuators B: Chemical 290 : 118-124.

[35]. Alizadehsani, Roohallah, et al. (2019), "Machine learning-based coronary artery disease diagnosis: A comprehensive review.", Computers in biology and medicine, 111 : 103346.

[36]. Kumar, N. Komal, and D. Vigneswari (2019), "Hepatitis-Infectious Disease Prediction using Classification Algorithms." , Research Journal of Pharmacy and Technology 12.8 : 3720-3725.

[37]. Kumar, N. Komal, et al. (2018), "Predicting Non-Small Cell Lung Cancer: A Machine Learning Paradigm.", Journal of Computational and Theoretical Nanoscience 15.6-7: 2055-2058.

[38]. Vigneswari, D., et al. (2019), "Machine Learning Tree Classifiers in Predicting Diabetes Mellitus.", 5th International Conference on Advanced Computing & Communication Systems (ICACCS). IEEE, 2019, pp. 84-87.

[39]. Kumar, N. Komal, et al. (2019), "An Optimized Random Forest Classifier for Diabetes Mellitus.", Emerging Technologies in Data Mining and Information Security. Springer, Singapore: 765-773.

[40]. Devi, BAS Roopa (2018), "MSO–MLP diagnostic approach for detecting DENV serotypes.", International Journal of Pure and Applied Mathematics 118.5 : 1-6.

[41]. S. R. W. Garner (2007), “The Waikato environment for knowledge analysis”.

[42]. R. Poli, J. Kennedy, and T. Blackwell (2007), “Particle swarm optimization”, Swarm Intell., Vol.1, No.1, pp. 33–57.

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

[44]. Y. Shi (2004), “Particle swarm optimization”, IEEE Connect., Vol.2, No.1, pp. 8–13.

[45]. Singh, Rachhpal and Rupinder Singh (2021), "Survey of Wheat Crop Diseases using Meta-heuristic and Machine LearningTechniques”, *International Journal of Computer Applications & Information Technology* (IJCAIT), vol 13, issue 1.

[46]. Bhardwaj, Ankur (2021), "Machine learning and optimization techniques in major human fatal disorders: A Short Communication", *International Journal of Computer Applications & Information Technology (IJCAIT)*, Issue: *Machine learning* 13, no. 1.

[47]. Sharma, Anubha, and Nirupma Tivari (2012), "A survey of association rule mining using genetic algorithm.", *International Journal of Computer Applications & Information Technology*, 1, no. 2 (2012): 5-11.

[48]. Sharma, Manik (2013), "Role and working of genetic algorithm in computer science.", *International Journal of Computer Applications and Information Technology (IJCAIT)* 2, no. 1.

[49]. Sharma, Samriti (2017), "Applications of genetic algorithm in software engineering, distributed computing and machine learning.", *International Journal of Computer Applications and Information Technology (IJCAIT)*, vol 9, issue 2, pp: 208-212.

[50]. Singh, Rachhpal, and Rupinder Singh (2019), "Nature Inspired Job Scheduling For E-Health Services In Mobile Cloud Computing.", *International Journal of Computer Applications & Information Technology*, vol 11, issue 2.

[51]. Sharma, Manik, and Gurvinder Singh (2019), "Need and Design of Smart and Secure Energy-Efficient IoT-Based Healthcare Framework.", *Energy Conservation for IoT Devices*. Springer, Singapore, pp.259-281.

[52]. Arora, Sankalap, Manik Sharma, and Priyanka Anand (2020), "A novel chaotic interior search algorithm for global optimization and feature selection.", *Applied Artificial Intelligence*, *vol.* 34, issue..4, pp.: 292-328.

[53]. Monga, Preeti, Manik Sharma, and Sanjeev Kumar Sharma (2021), "A comprehensive meta-analysis of emerging swarm intelligent computing techniques and their research trend.", *Journal of King Saud University-Computer and Information Sciences*.

[54]. Sharma, Manik, Samriti Sharma, and Gurvinder Singh (2020), "Remote monitoring of physical and mental state of 2019-nCoV victims using social internet of things, fog and soft computing techniques.", *Comput Methods Programs Biomed* : 105609-105609.

[55]. Singh, Rachhpal (2012), "Genetic algorithm for parallel process scheduling", *International Journal of Computer Applications and Information Technology (IJCAIT)* 1, no. 2.

[56]. Singh, Rachhpal (2012), "Task scheduling with genetic approach and task duplication technique.", *International Journal of Computer Applications & Information Technology*, 1, no. 1 pp.: 7-12.

[57]. Singh, Rachhpal (2018), "Task Scheduling Techniques Based On Parallel Genetic Algorithm Approaches: A Review.", *International Journal of Computer Applications & Information Technology*, 11, no. 1, pp.: 229-241.

[58]. Kaur, Prableen, Ritu Gautam, and Manik Sharma. "Feature Selection for Bi-objective Stress Classification Using Emerging Swarm Intelligence Metaheuristic Techniques." In *Proceedings of Data Analytics and Management*, pp. 357-365. Springer, Singapore, 2022.

[59]. Sharma, Manik, and Natalia Romero. "Future prospective of soft computing techniques in psychiatric disorder diagnosis." *EAI Endorsed Transactions on Pervasive Health and Technology* 4, no. 15 (2018): e1-e1.

[60] Sharma, Manik, et al. "Analysis of DSS queries using entropy based restricted genetic algorithm." Applied Mathematics & Information Sciences 9.5 (2015): 2599.