Hybrid Metaheuristic and Machine Learning Technique for Prognosis of Heart Disease

Rajatdeep Kaur[1*], Kamaljit Kaur[1]
[1] Department of Computer Engineering and Technology, Guru Nanak Dev University, Amritsar, India

Abstract:

Aim: Enhancing the accuracy of prediction of Heart Disease by using Random Forest and Genetic Algorithm.

Background: Cardiovascular Disease is the most prevalent cause of mortality in this world. It is more important to assess the risk of Heart Disease in early stages. In the initial stages, it can be diagnosed by many heart specialists with their experience and knowledge. But sometimes a minor mistake can cost death of the patient. That's why it is necessary to apply some methods such as Machine Learning Algorithms for prediction of Heart disease, so that it will be easy for doctors to understand the patient's condition and able to give appropriate care to them.

Method: Hybrid Genetic Algorithm and Random Forest are used to predict Heart Disease. Genetic Algorithm is applied to Random Forest to tune its hyper parameters and enhance its performance. The Model is evaluated using Italian Dataset, which is collected from the Institute of Clinical Physiology - Reggio Calabria Unit, Laboratory of Bioinformatics, National Research Council, Italy.

Result: TPOT classifier is used to select the optimal parameters of the Random Forest classifier and increase the accuracy of the proposed model by 3% and F1-score by 2% on Italian Dataset.

Keywords: Heart Disease, Data Mining, Machine learning, Random Forest, Genetic Algorithm, TPOT Classifier.

1. INTRODUCTION:

In today's era the volume of data increases, which cannot be extracted with human involvement like previously. Formerly, Manpower is used to analyze the data, although this method is disorganized yet huge patterns and data remains hidden [1].

Computational learning theory and Pattern Recognition are the roots of Machine Learning. In the field of Data analytics use these methods are used to predict entity by composing some models and algorithms. Through these models analysts, scientist, researchers and engineers can generate reliable and valid results. Various hidden patterns or features can also find using past learning and trends in data [1]. Automation of model building is performed using machine learning methods. In machine learning we have to generate models on the basis of training data and then test the model using testing data set. We don't need to instruct the machines, its iterative behavior permits the machine to conform its methods and produce outputs according to new situations and data [2]. Figure 1 shows the different categories of machine learning.

Supervised Learning: Supervised Learning algorithms permits to gather information and produced a yield from past experiences. This algorithm is good to use when data is labeled [3]. In the proposed system, supervised learning algorithm is used because data used for training the model is labeled and the model predicts the outcome of testing data based on that training.

- Support Vector Machine
- Decision Tree

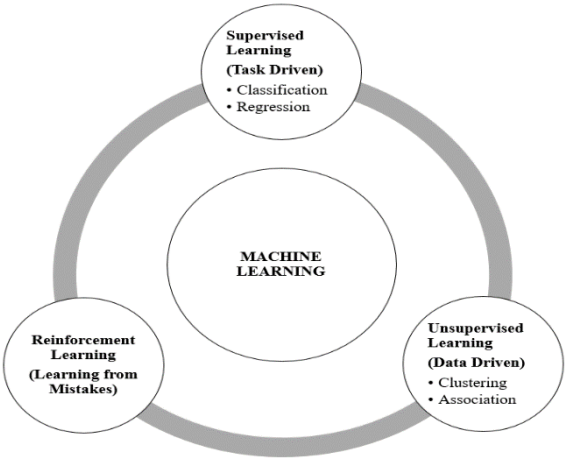

Figure 1: Categories of Machine Learning

- Random Forest
- Naïve Bayes
- Artificial Neural Network

In his paper *Random Forest* is used for training the model because it has three major advantages over other Classifiers. Firstly, usage of multiple decision trees decreases the risk of overfitting. Secondly, it runs efficiently in large databases and predict the highly accurate predictions and also there is no need for pre-processing, if the data is missing as it has ability to estimate the missing data [4]. Random Forest has three chief parameters, (1) Number of decision Trees' n(trees), (2) Minimum depth of trees, and (3) Number of features required for splitting in each tree [5]. In the proposed model, Maximum no. of trees and depth of trees have more impact on the accuracy of the model.

Unsupervised Learning: This algorithm is used to deal with unlabeled data and able to find unidentified patterns by arranging the data in clusters [6].

Reinforcement Learning: It involves controlling the system in order to enhance the numeral performance metrics. In this a machine is trained in a way to maximize the rewards of its actions. That means, the machine is bound to learn from its experience [7].

Sometimes machine learning algorithms alone are not sufficient to get accurate and efficient results, some methods and techniques need to embed with these algorithms. Such as hybridization of two or more machine learning algorithms OR using some metaheuristic techniques to optimize the result produced by machine learning algorithms

Metaheuristic Technique: Metaheuristic technique is the HL-problem independent that presents a set of rules to elaborate Heuristic Optimization Algorithms. A standard idea of all metaheuristics is the coordination among Exploration and Exploitation [8]. The Global search behavior of metaheuristics enables it to use it with other machine learning algorithms to optimize the result [9]. Metaheuristic Algorithms are shown in the figure 2.

Hill Climbing and Tabu search are local search optimization techniques. In Hill climbing, new Candidate solution of current candidate has been tested and adopted if the generated solution is optimized [10]. However, In Tabu search some situations are generated using a Tabu List data structure which helps to evade being trapped in Local Optima [11]. *Simulated Annealing* Method is grounded on the correlation among the simulation of the annealing of metals and the conundrum of resolving wide combinatorial optimization issues [12]. *Evolutionary Algorithm* is an arbitrary heuristic search Algorithm that can be worked in a variety of tasks [13]. That are managed by several parameters which are pivotal for effectiveness of search [14]. In evolutionary algorithms the members with more fitness value will survive whereas unfit members will not contribute in any further evaluation [15]. Genetic Algorithm and Particle Swarm Optimization are most widely used Evolutionary Algorithms. The Proposed system uses a Genetic Algorithm for hyper parameter tuning.

Genetic Algorithm: Genetic Algorithm (GA) represent an optimization algorithm in binary space. Successor proposition (hypothesis) is produced by the GA by mutation and recollection of the finest presently known proposition [15]. A genetic algorithm is utilized to locate an ideal arrangement by acquiring the most appropriate genes from every generation to upcoming generation, this most suitable algorithm for hyper parameter optimization than other methods [16]. Crossover, Mutation, and Selection of most fit are general genetic operators that are applied to pick the solutions on the basis of the fitness [17]. *Particle Swarm Optimization* follows the actions performed by a group of various organisms; for instance, Bird and Fish schooling. This method imitates the behavior of organisms that shows how different members interacts to pass information to each other [18]. The most optimal solutions can be searched through representative, commonly known as a particle. In this

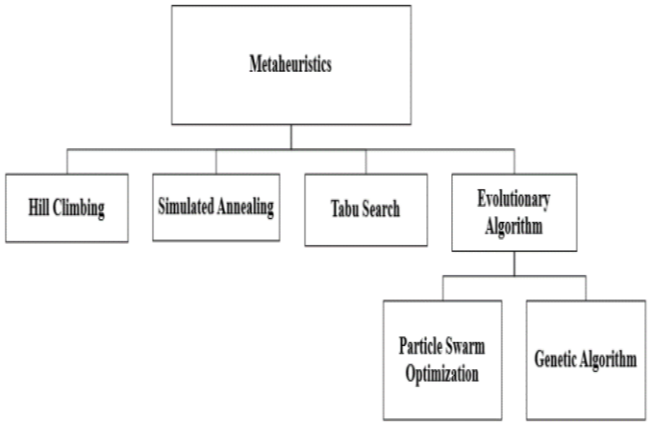

Figure 2: Types of Metaheuristic Techniques

algorithm the domination of each particle is defined by the best position acquired by it [19].

1.1. Machine Learning and Heart Disease

Various Medical domains use machine learning methods to identify and predict problems or disease, e.g. early prediction of cancer or heart disease. Whereas Machine learning with metaheuristic optimization performs well with better accuracy and reduced execution Time. Clinical parameters or data can be obtained from various medical centers and hospitals.

One of the Major reason for demise in Human Beings is Heart illness. It is estimated that, annually 17.5 million individuals are passing away ascribable to heart disease [20]. The heart is a necessary part of our body and our life is dependent upon the functioning of heart. Problem in heart can affect the whole functioning of heart [21]. Various other organs of the body, such as Brain can be affected with the reduction in Blood flow due to heart disease [22]. General features required for prediction of heart disease are Patient's age, Gender, Blood Pressure, Type of chest pain, ECG, Cholesterol level, Blood sugar Status, Smoking Status and Body Mass Index [23]. Common symptoms of Heart Disease are Feeling of tiredness, difficulty in breathing, irregularities in heartbeat, swear pain in chest and dizziness [24]. Due to the rapid growth of the digital technologies, medical center saves chunks of heart related information in their data repositories that is very critical to examine. Data mining procedure and Machine Learning methods play huge contribution in the dissection of non-identical content in medical facilities. These procedures and methods can be applied to sets of data for generating models and get out a conclusion from it [25].

1.2. Related work

Jabbar et al. [26], used Random Forest and Chi Square method for estimating heart disease. The proposed model was evaluated by using three different databases, which are some Hospitals data from Hyderabad, Cleveland and Stalog dataset. The Accuracy and other performance measures were used to make a comparison among various Machine Learning (ML) algorithms.

Yeshvendra et al. [27], applied Random Forest to predict Heart Disease. The efficiency of intended method was determined using Cleveland Heart Disease Dataset. K-fold cross-validation was performed (where K =10) for the authentication of accuracy that was 85.81%.

Amin UI et al. [28], developed a system for Heart illness diagnosis using some Attribute selection techniques and Machine learning algorithms. Some common performance measures like accuracy, specificity, sensitivity and Mathew's' correlation coefficient were used to check the efficiency of the model. Performance on the full featured dataset and reduced feature dataset was compared in propounded system and analyzed that data with reduced feature resulted arise in accuracy and decline in execution time.

Akrivos,E et al.[29], performed a consolidated model of sequential contrast patterns using MHMM (Multichannel Hidden Markov Model). Connection between exposure and Control group was presented in the proposed model. Mean sensitivity of 78% with peak specificity was offered to examine high risk patients.

Gurpreet et al. [30], proposed a system that validates thev diagnostic performance as well as the prognostic implication of anatomic and physiologic finding. It applied Machine Learning algorithms to extract data from angiography and Cardiac CT scans with performance optimization.

Nair et al. [31], applied ML Model on Big Data for developing an RT remote system for predicting the health status of patients. In the proposed system health related attributes get tweeted by users and real time application received that input for further extracting appropriate attributes and applies the ML model to the prognosis health status of respective users for alerting user to take appropriate action.

Jingshu et al. [32], propounded a multi task framework that was the combination of free-text medical notes and structured information for the prediction of disease. The performance of various deep learning architectures was compared to standard text-based models. The proposed method analyzed that models that are using text can perform well than models that used structured data only. Furthermore, different visualization methods for medical professionals to portray model predictions were also compared.

Sarabant et al. [33], proposed hybridization technique using DT and ANN classifiers using WEKA. Accuracy, Sensitivity and Specification of Classifiers is analyzed individually as well as using Hybrid Technique.

Narayan et al. [34], applied machine learning technique with Fourier transform to presage cardiovascular disease accurately. Proposed system used bagging model and three classifiers say ANN, Naïve Bayes and SVM for evaluation.

Abderrahmane et al. [35], presented a system for Real time Heart Disease prognosis that was in accordance with Apache Spark. The model contains a couple of major segments such as Streaming processing and, Data Storage and visualization. Streaming processing was used Spark's machine learning Library with spark streaming and Data Storage and Visualization used Apache Cassandra for accruing the huge volume of data generated.

Rachhpal Singh [36], proposed a model that uses hybrid Metaheuristic, Genetic Algorithm and Variable neighborhood search to maintain stabilization of Local and Non-local searching for processing of the evolutionary period to get best efficient results.

Gokulnath et al. [37], proposed an optimization method with the use of SVM (Support Vector Machine) and GA (Genetic Algorithm used for feature selection). Furthermore, the outcome of these algorithms (GA-SVM) is contrasted with various ML Algorithms for Feature Selection. The Receiver Operator Characteristic examination is conducted.

Abdar et al. [38], presented a new system N2-Genetic Optimizer using the combination of SVM with GA or PSO. In this paper different ML Algorithms were used to evaluate heart disease dataset of Z-Alizadeh Saini from which best performing method was SVM and selected for further experimentation. Preprocessing of data was done using Normalization whereas, GA or PSO was used for feature selection and reduction of redundancy. As a result of this system GA along with SVM performed better than PSO with SVM.

Eskandari et al. [39], propounded an advanced hybrid method of Whale Optimization and Dragonfly algorithm by applying Machine Learning Algorithm. The Evaluation of this model was done using Cleveland Heart disease dataset and conclude that SVM has better accuracy of 88.89% by selecting nine attributes from data.

Akgul et al. [40], uses Artificial Neural Network for recognition of Heart Problem. Thereafter, GA (Genetic Algorithm) a Metaheuristic algorithm and ANN (Artificial Neural Network) a Machine learning algorithm used together, hybridized approach was assessed using a heart disease dataset of Cleveland University and better accuracy is experienced than ANN and other ML algorithms.

2. METHODS

In literature review different methods are used to predict heart disease with different measuring parameters and every method have its pros and cons. The proposed method uses Hybrid Genetic Algorithm and Random Forest for prognosis of Heart Disease. Italian Dataset obtained from *Institute of Clinical Physiology – Reggio Calabria Unit, Laboratory of Bioinformatics, National Research Council, Italy* [41], is used for evaluation of the proposed model. The dataset consist of 522 entries with 14 features (ID, Gender, Age, Smoking, and Diabetes, CV comorbidities, Education, Glucose, cholesterol, CRP, Glycated HB, BMI, Hypertensive drugs and Death) and 1 class label (CV Death). Out of 522 entities, 391 entities are used for Training dataset and 131 entities used for testing dataset.

The dataset used to check the efficiency of the model is represented in the form of histogram as shown in figure 3.

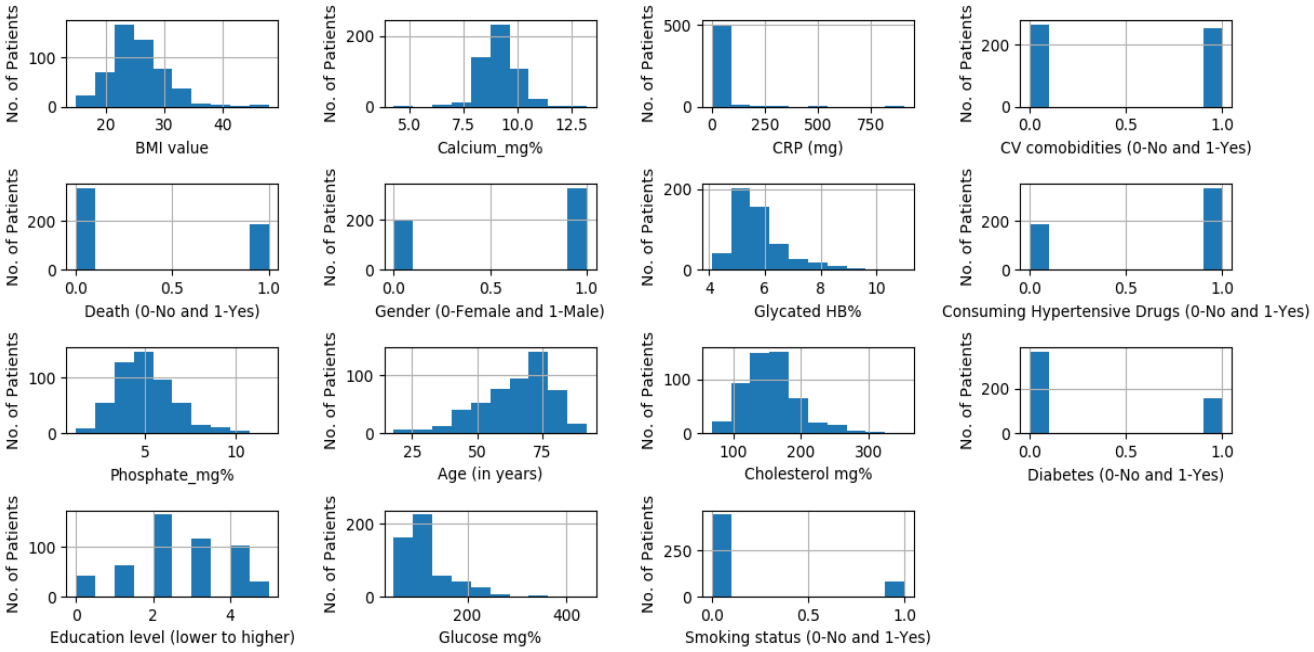

Figure 3: Histogram of Dataset

In our proposed system HGARF, Genetic Algorithm is used for tuning hyper parameters of Random Forest and that improves the accuracy of predicting Deaths caused by Heart Disease. Confusion Matrix is used to detect the actual no. of correct and incorrect predictions of the model as shown in table II.

Table-II: Confusion Matrix

		Predicted	
Observed		Presence of disease	Absence of Disease
	Presence of disease	True Positive	False Positive
	Absence of disease	False Negative	True Negative

- True Positive (TP): Positive observation is predicted as positive.
- True Negative (TN): Negative observation is predicted as negative.
- False Positive (FP): Negative observation is predicted as positive.
- False Negative (FN): Positive observation is predicted as negative.

Following Performance measures are applied to evaluate the performance of our proposed model:

- Accuracy: (TP + TN)/(TP + TN + FP + FN)
- Sensitivity: (TP)/(TP + FN)
- Specificity: (TN)/(FP +FN)
- Precision: (TP)/(TP +FP)
- False Positive Rate: (FP)/(TN +FP)

1.1. Algorithm of Random Forest:

Step 1: Select a new bootstrap instance from the training set.

Step 2: Grow a un-pruned tree on the selected bootstrap sample.

Step 3: Randomly select the number of attributes needed to sample from each tree in each internal node and analyze the best partition.

Step 4: Check whether each tree is fully developed? If yes, do not prune more.

Step 5: Check the majority vote from all the trees and that vote is the output for estimation.

1.2. Algorithm of proposed system:

Step 1: Import Heart Disease Italian Dataset.

Step 2: Perform Normalization of the Data by scaling to range [0, 1]

Step 3: Split Dataset into two sets, Training set and testing set in 3:1 Ratio.

Step 4: Import the Classifier to be used that is Random Forest classifier.

Step 5: Define the range of parameters of Random Forest to be optimized.

Step 6: Import TPOT classifier for tuning the Hyper parameters of Random Forest using Genetic Algorithm.

Step 7: Define the parametric values of TPOT classifier.

Step 8: Fit the model on Training set of Heart Disease Dataset.

Step9: Compute the Accuracy of the model on Testing Dataset and Generate Confusion Matrix to get Performance measures.

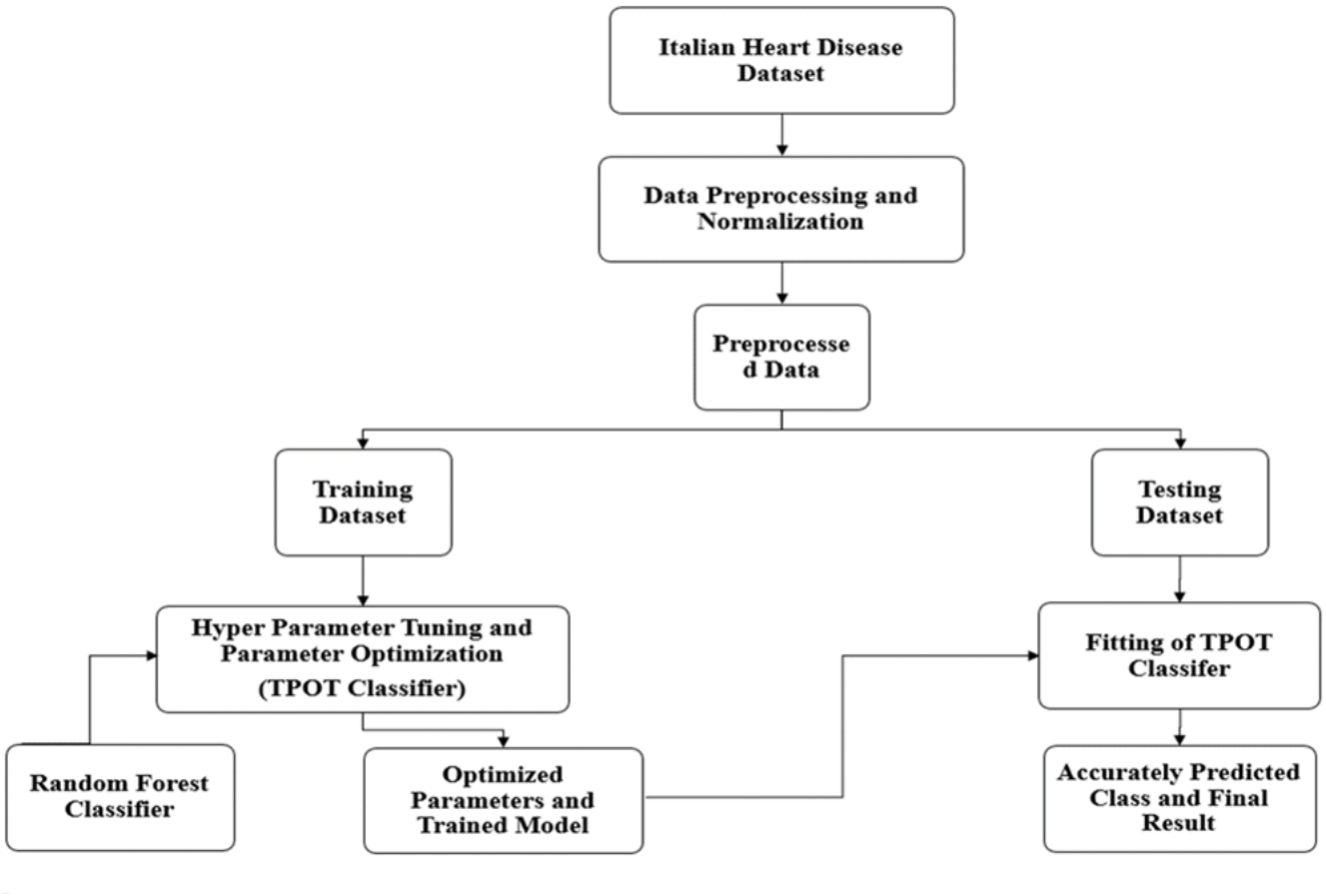

Figure 4: Working Flow of Proposed System

3. RESULTS AND DISCUSSION

For our Experimentation, we are splitting the dataset in 3:1 ratio. For training the model we use 75% of a dataset and the rest of the 25% data is used for Testing the Model. In order to evaluate the performance of our model we are using Hybrid Random Forest and Genetic Algorithm (HGARF). Genetic Algorithm is used to tune the hyper parameters of Random Forest. For the implementation of GA we are using TPOT Classifier that helps to optimize ML Pipeline using Genetic Algorithm. Here is the Best Pipeline obtained on Italian Dataset:

Best Pipeline:- RandomForestClassifier(Combine DFs (input_matrix, RandomForestClassifier (input_matrix, criterion = entropy, max_depth = 935, max_features = sqrt, min_samples_leaf = 12, min_samples_split = 10, n_estimators =384)), criterion = entropy, max_depth = 671, max_features = log2, min_depth = 671, ma_features = log2, min_samples_leaf = 4, min_samples_split = 5, n_estimators = 500)

Best Parameters of GA used in the implementation are:

1. Population size = 2
2. Generations = 7
3. Crossover rate = 0.1
4. Mutation rate = 0.9
5. Offspring_size = 12

Whereas f1-score and accuracy is used as the Fitness function. As per our model, accuracy is defined as the ability of the model to accurately predict the occurrence of death from cardiovascular disease. Figures 5 and 6 shows the variation in the accuracy of the model with variations in values of some Random forest Parameters such as n_estimators (No. of Trees) and maximum depth. Figure 5 depicts the highest accuracy when the number of Tress are 500. However, figure 6 shows the highest accuracy at 935 maximum depth. In both the figures the value of accuracy is fluctuating with changes in values.

Our model is validated on different datasets that is Italian Heart disease Dataset and Cleveland Dataset (obtained from UCI Machine Learning Repository).

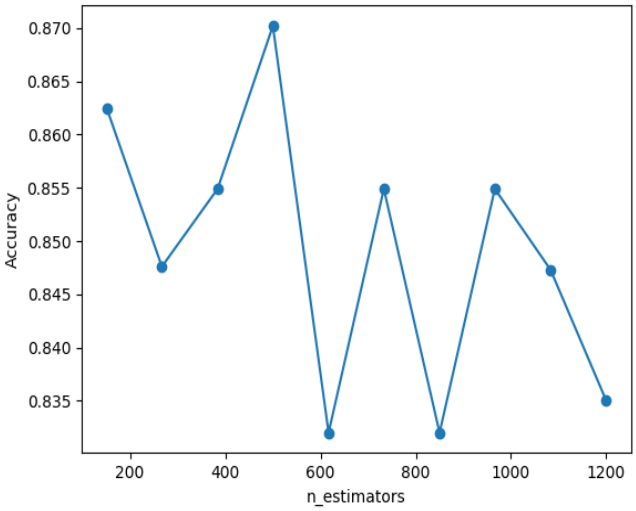

Figure 5: Change in Accuracy with respect to no. of trees

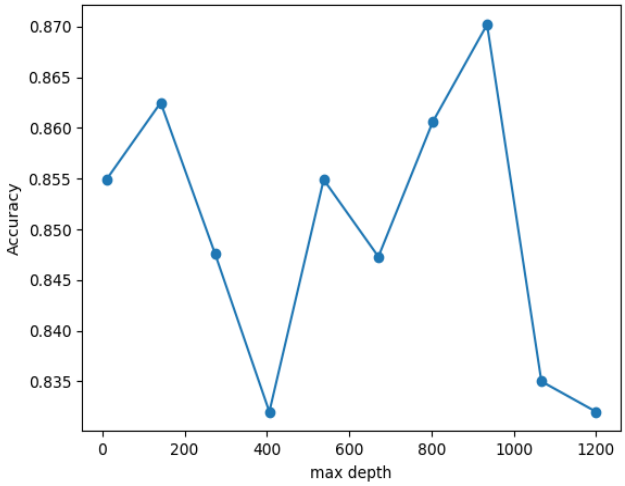

Figure 6: Change in Accuracy with respect to maximum depth of trees.

The accuracy of our model is compared with random forest using Italian dataset and Cleveland Dataset in figure 7 and 8 respectively.

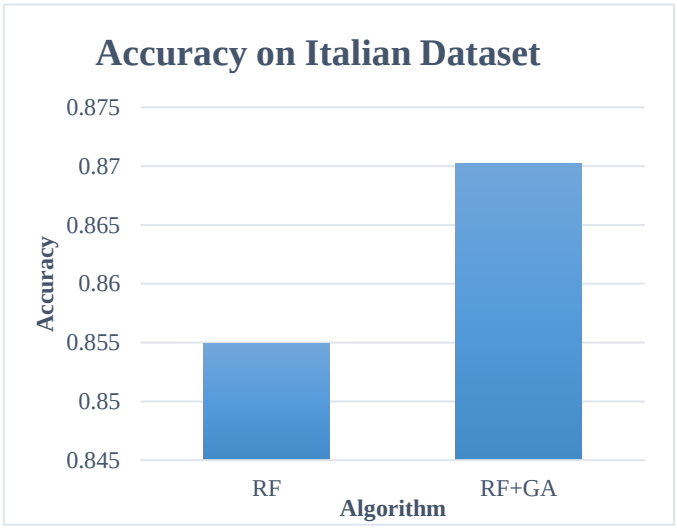

Figure 7: Compare the accuracy of model with RF on Italian Dataset.

Random Forest obtained 85% accuracy which is increased in our model by 2% in Italian Dataset. However, 82.5 % accuracy is obtained by Random Forest in Cleveland Dataset

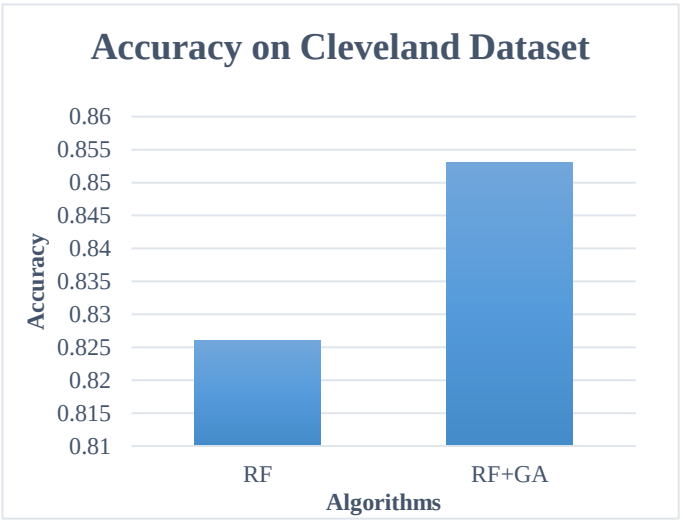

Figure 8: Compare the accuracy of the model with RF on Cleveland Dataset.

and increased in our model (HGARF) by 3%. Figure 9 and 10 shows the performance measures on both datasets.

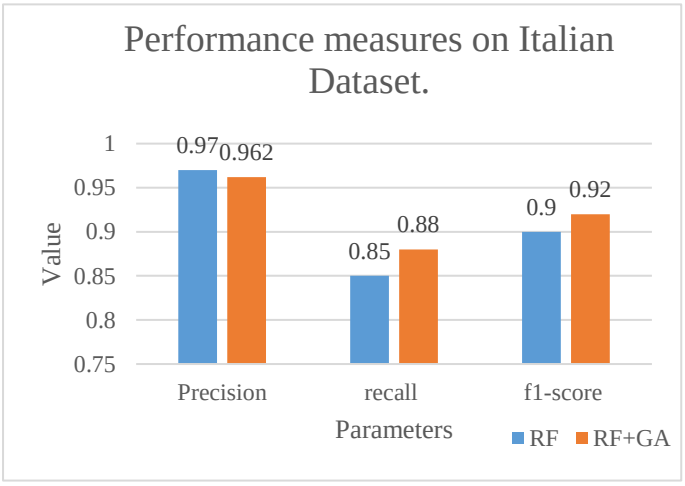

Figure 9: Performance measures on Italian Dataset.

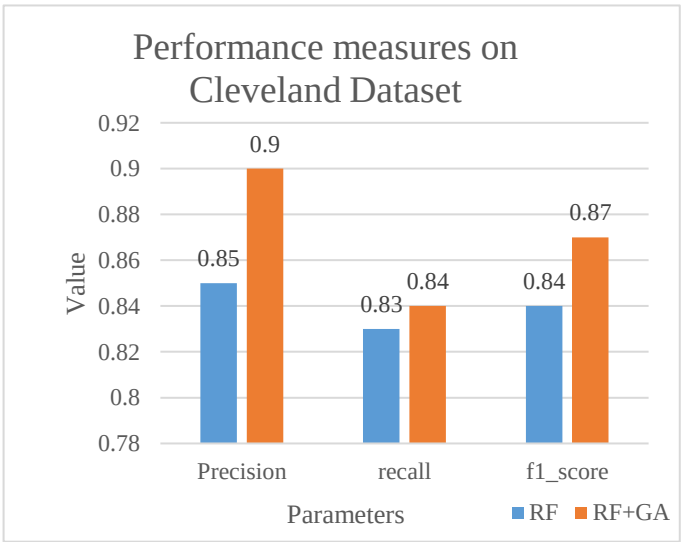

Figure 10: Performance measures on Cleveland Dataset.

Various Performance measures parameters such as Precision, Recall and F1-score shown in figure 9 and 10 for both the databases. Both specificity (precision) and sensitivity (recall)

have different roles in the graphs above. However, Precision of random forest is higher than that of RF+GA in Italian Dataset but f1-score of our proposed Model is higher in both the datasets. Precision and recall individually, not enough to determine an algorithm or model. So the average of both the parameters that is f1-score is considered for evaluation.

It is observed from Figure 9 and 10 that average score of performance measures of our model HGARF are higher (92% for Italian Dataset and 87% for Cleveland dataset) than that of Random Forest (90% for Italian dataset and 84% for Cleveland dataset). That means the performance of Random Forest is improved after tuning of hyper parameters.

Many researches have been done for the prediction of Heart disease using different Machine Learning Algorithms such as SVM, Neural Network, Random Forest and Decision Tree. It has been observed that the accuracy obtained from the Random Forest algorithm is not that efficient, which is enhanced in this paper using Genetic Algorithm. This model is applied on two datasets that were obtained from different Data Repositories but this technique is not efficient to work with real-time data. It will be improved in further researches by embedded this model with other technology.

4. CONCLUSION AND FUTURE SCOPE

In this paper, the propounded system model generates the better accuracy for prediction of heart disease than Random Forest. Metaheuristic algorithm plays an essential role in enhancing the accuracy of the model. Hybridization of Genetic Algorithm and Random Forest optimize the results by tuning hyper parameters of Random Forest and with this we get an accuracy of 87% and f1-score of 92% for Italian dataset. TPOT classifier is used to implement GA based Random Forest, which generate best pipeline and select best parametric values of Random Forest. Maximum no. of Trees and Maximum Depth of trees are important parameters to enhance the performance of the model. Better and accurate results can help the area specialists and others related to this field to be ready for upper level determination and give appropriate treatment to patients. So early estimating heart disease helps to give quality service and save the individual's life. In future Internet of things along with proposed work can be used for disease analysis using Real time Data.

REFERENCES:

[1] Simeone, Osvaldo. "A brief introduction to machine learning for engineers," arXiv preprint arxiv: 1709.02840, 2018.

[2] Kubat, Miroslav. "An introduction to machine learning". Vol. 2. Cham, Switzerland: Springer International Publishing, 2017.

[3] Talabis, Mark, Robert McPherson, Inez Miyamoto, and Jason Martin., Information Security Analytics: Finding Security Insights, Patterns, and Anomalies in Big Data. Syngress, 2014.

[4] Liaw, Andy, and Matthew Wiener. "Classification and regression by randomForest," R news, Vol. 2, pp. 18-22, 2002.

[5] Kaur, Jaspreet, Dilbag Singh, and Manjit Kaur. "A Novel Framework for Drug Synergy Prediction using Differential Evolution based Multinomial Random Forest", International Journal of Advanced Computer Science and Applications, Vol. 10, pp. 601-608, 2019.

[6] Aggarwal, Nimish, Shikha Sarkar, and Dilbag Singh. "A BREIF STUDY ON TERRORIST ATTACK PREDICTION USING UNSUPERVISED LEARNING", International Jounral of Information and Computer Science, Vol. 5, pp. 276-287, September 2018.

[7] Montague, P. Read, "Reinforcement Learning: An Introduction, by Sutton, RS and Barto, AG", Trends in cognitive sciences, vol. 3, p. 360, September 1999.

[9] Singh, Rachhpal, "Hybrid Metaheuristic Based Scheduling with Job Duplication for Cloud Data Centers In Harmony Search and Nature Inspired Optimization Algorithms, Advances in intelligent Systems and Computing, vol. 741, pp. 989-997, 2019.

[10] Bandaru, Sunith, and Kalyanmoy Deb, "Metaheuristic techniques", Decision Sciences: Theory and Practice pp. 693-750, 2016.

[11] Luke, Sean. Essentials of metaheuristics, 2009. [E book] Available: http://cs.gmu.edu/~sean/book/metaheuristics/.

[12] Edelkamp, Stefan, and Stefan Schroedl. Heuristic search: theory and applications. Elsevier, 2011.

[13] Van Laarhoven, Peter JM, and Emile HL Aarts. "Simulated annealing", In: Simulated annealing: Theory and applications, Vol. 37, Springer, Dordrecht, 1987, pp. 7-15.

[14] Zhou, Kai, Wei Wan, Xi Chen, Zhijiang Shao, and Lorenz T. Biegler. "A Parallel Method with Hybrid Algorithms for Mixed Integer Nonlinear Programming", In: Computer Aided Chemical Engineering, vol. 32, Elsevier, 2013, pp. 271-276.

[15] Wegener, Stefan Droste Thomas Jansen Ingo. "Evolutionary Algorithms", In: Foundations of Genetic Algorithms 2001, FOGA 6, 2001, p. 275.

[16] Meyer-Baese, Anke, and Volker J. Schmid. Pattern recognition and signal analysis in medical imaging. Elsevier, 2014.

[17] Kaur, Manjit, Dilbag Singh, Kehui Sun, and Umashankar Rawat. "Color image encryption using non-dominated sorting genetic algorithm with local chaotic search based 5D chaotic map", Future Generation Computer Systems, vol. 107, pp. 333-350, 2020.

[18] Yang, Xin-She, Amir Hossein Gandomi, Siamak Talatahari, and Amir Hossein Alavi, eds. Metaheuristics in water, geotechnical and transport engineering. Newnes, 2012.

[19] Singh, Rachhpal, "Hybrid genetic, variable neighbourhood search and particle swarm optimisation-based job scheduling for cloud computing", International Journal of Computational Science and Engineering, Vol.17, pp. 184-191, 2018.

[20] Martinez, Clara Marina, and Dongpu Cao, iHorizon-Enabled Energy management for electrified vehicles. Butterworth-Heinemann, 2018.

[21] Learning, Machine, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Advances in Computational Sciences and Technology, Vol. 10, pp. 2137-2159, 2017.

[22] Krishnaiah, V., G. Narsimha, and N. Subhash Chandra, "Heart disease prediction system using data mining techniques and intelligent fuzzy approach: a review", International Journal of Computer Applications, Vol. 136, pp. 43-51, 2016.

[23] Bhor, Pooja, Gurjot Singh Sodhi, and Dilbag Singh. "Classification of Heart Variability through Machine Learning", International Journal of Electronics Engineering, Vol. 11, pp. 73-80, 2019

[24] Srinivas, Kavitha, G. Raghavendra Rao, and A. Govardhan, "Survey on prediction of heart morbidity using data mining techniques", International Journal of Data Mining & Knowledge Management Process, Vol. 1, pp. 14-34, 2011.

[25] Wulsin, Lawson R., Jane C. Evans, Ramachandran S. Vasan, Joanne M. Murabito, Margaret Kelly-Hayes, and Emelia J. Benjamin, "Depressive symptoms, coronary heart disease, and overall mortality in the Framingham Heart Study", Journal of Biobehavioral Medicine, Vol. 67, pp. 697-702, 2005.

[26] Kaur, Manjit, Hemant Kumar Gianey, Dilbag Singh, and Munish Sabharwal. "Multi-objective differential evolution based random forest for e-health applications." Modern Physics Letters B 33, no. 05 (2019): 1950022.

[27] Jabbar, M. Akhil, Bulusu Lakshmana Deekshatulu, and Priti Chandra. "Prediction of heart disease using random forest and feature subset selection", In: Innovations in bio-inspired computing and applications, Springer, Cham, 2016, pp. 187-196.

[28] Singh, Yeshvendra K., Nikhil Sinha, and Sanjay K. Singh. "Heart Disease Prediction System Using Random Forest", International Conference on Advances in Computing and Data Sciences, Springer, Singapore, 2016, pp. 613-623.

[29] Haq, Amin Ul, Jian Ping Li, Muhammad Hammad Memon, Shah Nazir, and Ruinan Sun, "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms", Mobile Information Systems, Vol. 2018, pp. 1-21, 2018.

[30] Akrivos, E., V. Papaioannou, N. Maglaveras, and I. Chouvarda, "Prediction of Cardiac Arrest in Intensive Care Patients Through Machine Learning", In: Precision Medicine Powered by pHealth and Connected Health, Springer, Singapore, 2018, pp. 25-29.

[31] Singh, Gurpreet, Subhi J. Al'Aref, Marly Van Assen, Timothy Suyong Kim, Alexander van Rosendael, Kranthi K. Kolli, Aeshita Dwivedi et al. "Machine learning in cardiac CT: basic concepts and contemporary data", Journal of cardiovascular computed tomography, Vol. 12, pp. 192-201, 2018.

[32] Nair, Lekha R., Sujala D. Shetty, and Siddhanth D. Shetty, "Applying spark based machine learning model on streaming big data for health status prediction", Computers & Electrical Engineering, Vol. 65, pp. 393-399, 2018.

[33] Liu, Jingshu, Zachariah Zhang, and Narges Razavian, "Deep EHR: Chronic disease prediction using medical notes", arXiv preprint arXiv:1808.04928, Vol. 1, pp. 1-24, 2018.

[34] Maji, Srabanti, and Srishti Arora, "Decision tree algorithms for prediction of heart disease", Information and Communication Technology for Competitive Strategies, Springer, Singapore, 2019, pp. 447-454.

[35] Narayan, Subhashini, and E. Sathiyamoorthy. "A novel recommender system based on FFT with machine learning for predicting and identifying heart diseases", Neural Computing and Applications, Vol. 31, pp. 93-102, 2019.

[36] Ed-Daoudy, Abderrahmane, and Khalil Maalmi. "Real-time machine learning for early detection of heart disease using big data approach", In: 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems, IEEE, 2019, pp. 1-5.

[37] Singh, Rachhpal. "Genetic-variable neighborhood search with thread replication for mobile cloud computing", International Journal of Parallel, Emergent and Distributed Systems, Vol. 32, pp. 486-501, 2017.

[38] Gokulnath, Chandra Babu, and S. P. Shantharajah, "An optimized feature selection based on genetic approach and support vector machine for heart disease", Cluster Computing, Vol. 22, pp. 14777-14787, 2019.

[39] Abdar, Moloud, Wojciech Książek, U. Rajendra Acharya, Ru-San Tan, Vladimir Makarenkov, and Paweł Pławiak, "A new machine learning technique for an accurate diagnosis of coronary artery disease", Computer methods and programs in biomedicine, Vol. 179, p. 104992, 2019.

[40] Eskandari, Marzieh, and Zeinab Hassani. "Intelligent application for Heart disease detection using Hybrid

Optimization algorithm", Journal of Algorithms and Computation, Vol. 51, pp. 15-27, 2019.

[41] Akgül, Miray, Özlen Erkal Sönmez, and Tuncay Özcan. "Diagnosis of Heart Disease Using an Intelligent Method: A Hybrid ANN–GA Approach", In: International Conference on Intelligent and Fuzzy Systems, Springer, Cham, 2019, pp. 1250-1257.

[42] Mezzatesta, Sabrina, Claudia Torino, Pasquale De Meo, Giacomo Fiumara, and Antonio Vilasi. "A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis", Computer methods and programs in biomedicine, Vol. 177, pp. 9-15, 2019.