Original Article

ECG Data Classification Using Machine Learning Approaches

Dr.S.J Saritha[1*], Shaik Ghouhar Taj[2], K. Chandrasekhar[3]

[1*]Assistant Professor, CSE, JNTUA College of Engineering Pulivendula. Email: sjsaritha.cse@jntua.ac.in
[2,3]Assistant Professor(Adhoc),CSE, JNTUA College of Engineering Pulivendula. Email: sgtaj786@gmail.com[2],chandra507shiva@gmail.com[3]

***Corresponding author:** Dr.S.J Saritha

[*]Assistant Professor, CSE, JNTUA College of Engineering Pulivendula. Email: sjsaritha.cse@jntua.ac.in
Doi: 10.47750/pnr.2022.13. S05.215

Abstract

Electric heartbeat was monitored with an electrocardiogram. It was an easy and fun way to diagnose heart problems but also to keep a healthy record of our bodies. Heart disease often manifests itself as chest tightness, sickness, light head edness, severe instability, chest tightness, rapid heartbeat, shortness of breath, fatigue, drowsiness, or even decreased exercise rate. The same basic principle would be for machines to be straight within it, but without the intervention of users, and also to change their habits accordingly. Distinguishing is among the machine-readable learning strategies they use to classify data sets effectively. This paper introduces a multidisciplinary survey in the separation of ECG data using machine learning methods found at the beginning of the literature. It is very useful in both computer and medical research in the field of automatic diagnosis of heart disease. Class dividers improve physician accuracy and early detection for better treatment of affected patients.

Keywords: ECG Information,Classification, Machine Learning,Heart disease Attributes.

INTRODUCTION:

An electrocardiogram (ECG) is a diagnostic test that produces a clear heart rhythm to detect cardiovascular problems (heart). This electrocardiogram (ecg is simply a tool that can detect the pulse of a client. This echocardiography produces electrical activity of the muscle but instead shows this monitoring or writing just as a trace. A health care professional usually processes the information.

ECG:

There is no need to fast or reduce your diet before an ECG test. Before you do an ECG, tell your doctor what drugs you are taking and if you have any allergies to adhesive tapes that may be used to attach electrodes. You will need to take off your outer garments for an ECG test so that the electrodes are attached to your chest and extremities. When paired with pants or a skirt, a separate top allows for easy access to the chest.

When are ECGs needed?

In some cases, getting this test is important. If you have risk factors for increased volume , such as high blood pressure, other evidence from heart disease, including chest tightness, shortness of breath, rapid heartbeat, especially rapid heartbeat, then Electrocardiogram may definitely be needed. If you have ever had a personal and family history of heart problems, high blood pressure, or other risks and you really want to exercise regularly, you will need to undergo tests or job obligations. Electrodes (nerves) and adhesive glue were applied to the chest, limbs, and ankles. An electrocardiograph measures and traces the electrical electrodes formed mainly by the heartbeat, which is detected by these electrodes.

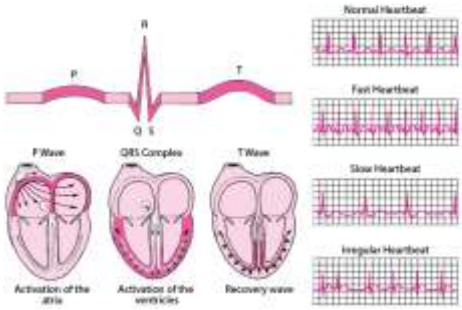


Figure 1: Reading ECG Waves

P-wave:
• This is the first rising wave to emerge.
• Refers to the deterioration of the atrial (systole), where excitement also arises from the SA node increases throughout the atrium.
• The Atria contract is approximately 0.1 seconds after the start of the P-wave. The Atrial systole is therefore expressed by the P-wave.

QRS wave:
• The QRS wave is indeed a second wave, which starts as a low-frequency wave and progresses to large dry triangular waves before ending as a vertical wave.
• Describes ventricular (systole) degeneration.
• Ventricles begin to deteriorate soon after the onset of the QRS wave. The Ventricular systole was therefore largely represented by the QRS wave.

T-wave:

• A T-wave is a small third wave with a dome-shaped curve.
• Indicates the onset of ventricular contraction and ventricular repolarization (diastole).
• COMBINATION OF ATRIAL DIASTOLE AND QRS-WAVEP-R interval:
The P-R interval is indeed the length of time it takes for an impulse to do that from the atria and ventricles it via AV node and bundles of character.

P-R interval:
• PR-R time is really the length of time it takes for the pressure to do that from the atria and ventricles through the AV node and the bulk of the letter.

S-T Section:
- The S-T component is the visual separation between the last Wavefront moment and the start of a T-wave object.
- Indicates the point at which the ventricular fibers are fully depolarized.

Related Work:

Previous studies have assessed user states, especially the flow state, using data from different physiological and psychological technologies like galvanic skin response (GSR), electroencephalography (EEG), electrocardiogram (ECG), electromyography (EMG), and electrodermal activity (EDA) (Berta et al, 2013; Rissler et al, 2018). There are other approaches such as self-reported questionnaires and interviews that are based on the users' recall of the experience (Bhattacherjee, 2012). Recent developments in information systems (IS) have offered more ways to analyze user states. They include more objective measures that combine EEG signals and machine learning techniques to classify the user states.
Machine learning techniques provide a systematic approach for classifying multi-channel EEG signals (Garrett et al, 2003). Recent studies have used machine leaning to optimize players' gaming experience (Hair, 2007), where players are segregated based on their experience in gaming and their momentary scores. Analyzing variables such as scores and responses to situational changes in the computer-based gaming environment helps designers and developers understand both their target population and design dynamics to optimize gaming experience (Hair, 2007). The SVM model is considered as a state-of-the-art machine learning technique for classifying brain activity obtained from EEG (Berta et al., 2013).
Berta et al (2013) focused on building a machine learning classifier that can distinguish three user states, namely, boredom, frustration/anxiety, and flow. They trained the SVM model with radial basis function kernel (RBF) in two different conditions:1) user-dependent with a classification accuracy of 50.1%, and 2) user independent with an accuracy of classification of 66.4%. Berta et al (2013) also 6 implemented a feature selection method to extract important EEG components and then analyzed these components using SVM for reduced computational times and better classification accuracies. After comparing the models with and without feature selection variables, they found that the model with all the components from the data collected have higher performance than any other models. Another study by Chatterjee et al. (2016) also applied machine learning models to identify cognitive flow. They implemented the Bayesian network to detect cognitive flow during gaming and derived an accuracy of 62.2 % based on data from the EEG and GSR technologies. Another research has used the SVM model to classify emotions into boredom, engagement, and anxiety while playing the Tetris game and obtained an accuracy of 53.33 % (Chanel et al., 2008). Chanel et al. used EEG and GSR data to classify the above-mentioned emotions using the SVM (Radial Basis Function kernel) model.
Plotnikov et al. (2012) used a gaussian kernel SVM model to assess flow in games based on EEG data and obtained an average accuracy of 57%. A study by Rissler et al. (2018) implemented SVM and random forests models to classify low flow and high flow in gaming using physiological data that include electrocardiography (ECG), blood volume pressure (BVP), and electrodermal activity (EDA). The result shows that cardiac features play an important role in categorizing the flow state, with random forests being a more accurate model (72.3%) than SVM (Rissler et al., 2018).
Lin et al. (2008) implemented the SVM – RBF model to classify 32 channel EEG data into four states – joy, arousal, sadness, and pleasure – based on emotions triggered by music. To classify emotions, the EEG data was divided into the following frequency bands: delta (1-3 Hz), theta (4-7 Hz), alpha (8-13 Hz), beta (14-30 Hz), and gamma (31-50 Hz). The study resulted in successful classifications of the emotions with 7 a maximum accuracy of 92.73% that used all the frequency bands combinations. Another study with the same context of listening to music utilized the multilayer perceptron classifier to classify the EEG data into joy, angry, sadness, and pleasure and obtained an accuracy of 69.69 % using a sample size of five (Lin et al., 2007).
Similarly, another study by Wang et al. (2011) used machine learning algorithms to classify user states in the context of movie elicitation. The time domain features and frequency domain features of EEG data were compared to assess which features classify emotions more correctly. They used the SVM-RBF model, k-NN model, and multilayer perceptron model to classify user states into joy, sad, relax, and fear. The SVM-RBF model achieved higher accuracy (66.51%) than other models with frequency domain EEG features as input. A similar study was conducted by Wang et al (2014) that compared three different EEG features, specifically power spectrum, wavelet, and nonlinear dynamical analysis, to understand the relationship between emotion and EEG data in the context of movie elicitation. The emotional state classification was done using the different kernels (RBF, polynomial, linear) of the SVM model across all the combinations of frequency bands (delta, beta, alpha, theta, and gamma). The results indicate that the power spectrum plays an important role in classifying the emotions with the linear kernel SVM (87.53%) model achieving the

highest classification accuracy using a combination of all bands (Wang et al., 2014).

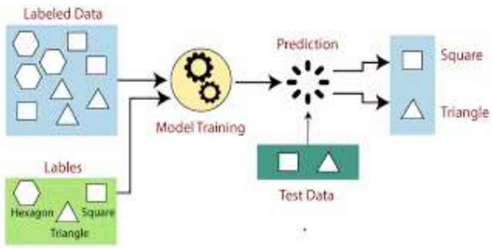
Several studies in the medical field studied the classification of EEG signals based on machine learning techniques, where the SVM model was frequently used. Lotte et al. (2007) reviewed the performance of all machine learning algorithms available for the purpose of classification from EEG to BCI systems. The SVM model is the most efficient for synchronous BCI due to its regularization property, simplicity, 8 and robustness. Vladimir et al. (2015) investigated the performance of the SVM model for seizure prediction using EEG signals. The SVM – RBF kernel model was used in the classification of EEG signals into seizure and non-seizure signals with an accuracy of 95.33 % (Joshi et al., 2014). Another study classified EEG signals into epileptic seizure or not using the SVM model with an accuracy of 98.75 %, where principal component analysis (PCA), linear discriminant analysis (LDA), and independent component analysis (ICA) were used for the feature reduction process (Subasi et al., 2010).

Liang et al. (2006) evaluated the performance of backward propagation neural networks and SVM models for mental task classification based on EEG signals. Other models like k-NN and decision trees were used to classify the sleep stages, with k-NN achieving higher classification accuracy than decision tree (Güneş, Polat, & Yosunkaya., 2010). Alkan et al (2005) proposed an automatic seizure detection model using EEG, logistic regression, and neural networks models, with neural networks achieving higher accuracy (92%).

From the previous studies in the literature, we see that the SVM model has been implemented to categorize user states based on EEG data. There are only a few studies on classification of user states based on frequency bands, especially for the flow state. Hence, in this study, we explore different machine learning models to classify the user states into boredom, flow, and anxiety with different combinations of the frequency bands. Also, we are interested to identify the best performing machine learning model to distinguish the flow state from all the other states. Table 2.1 provides a brief overview 9 of previous studies that have applied various machine learning models in classifications of user states.

Machine Learning:

Another machine learning challenge that can be addressed using machine learning methods involves fragmentation. Nearby, Naive Bayes, Decision Trees, Divorce Law Students, Neural Networks, and Vector Support Machine were prominent surveillance learning strategies. Figure 1 shows the various actions required to find solutions for using a machine learning system.



There is a lot of research done on the separation of ECG data using a variety of machine learning methods. This paper is further developed into Phase II, which provides information on ECG data representation, section III provides an overview of machine learning methods, Phase IV provides an ECG data classification survey and finally section V provides a discussion and conclusion.

Machine Learning Techniques:

To identify various heart problems, ECG devices employ the algorithms and methodologies shown below.

1. Theory of Fuzzy Sets
2. Artificial Neural Networks (ANNs)
3. Vector Machine Support
4. Algorithms for Genetics
5. Hybrid Methodologies
6. Markov model and rough set theory

1. Theory of Fuzzy Sets

Fuzzy set theory provides effective tools to handle and manipulate imprecise and noisy data and to make decisions based on such data. Fuzzy logic offers a linguistic approach that represents an excellent approximation to medical texts. In addition fuzzy logic provides reasoning methods capable of making approximate references. These facts suggest that fuzzy set theory might be suitable for the development of computerized diagnosis system [1][4].. In the diagnosis problems.we need to observe and classify symptoms caused by a disease. These symptoms are not well defined in many cases. Many research techniques have been devised based on fuzzy set theory, they are Compositional rule Inference, Linguistic variables, Fuzzy patterns recognition, Fuzzy classification, Fuzzy clustering and Fuzzy clustering analysis image processing and Fuzzy inverse problem [11 1. Adlassnig et a [I] did a complete diagnosis system based on fuzzy theory. They constructed fuzzy relations from the frequency of occurrence of symptoms and diseases. The fuzzy relations are represents in two dimensional arrays. Diagnosis was performed using these fuzzy relations and the compositional rule of interference [9][IO]. This paper discusses a fuzzy classification problem approach towards epilepsy risk level from EEG signals.

2. Artificial Neural Networks (ANNs)

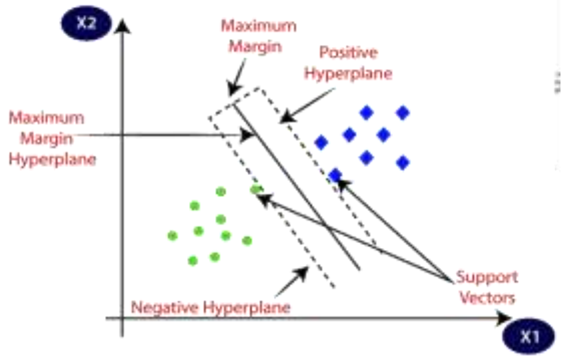
Artificial neural network (ANN) model involves computations and mathematics, which simulate the human–brain processes. Many of the recently achieved advancements are related to the artificial intelligence research area such as image and voice recognition, robotics, and using ANNs. The ANN models have the specific architecture format, which is inspired by a biological nervous system. Like the structure of the human brain, the ANN models consist of neurons in a complex and nonlinear form. The neurons are connected to each other by weighted links. All the processes in ANN models, such as data collection and analysis, network structure design, number of hidden layers, network simulation, and weights/bias trade-off, are computed through learning and training methods. The ANN applications solving problems of the real world include a wide range of scientific fields from finance to hydrology and fall into the three categories: (i) pattern classification, (ii) prediction, and (iii) control and optimization. Clustering, classification, and simulation in the ANN models are performed with different kinds of structures related to the ANN models. The ANN models are categorized into three groups: (i) static ANN, (ii) dynamic ANN, and (iii) statistical ANN. Static ANN model is known as a multilayer perceptron neural network model, dynamic neural network models such as tapped delay lines and recurrent neural network models, and statistical neural network models such as radial basis function model, and generalized regression neural network model. It may also be possible to combine the ANN model with other optimization techniques, for example, adaptive neuro-fuzzy inference system for the better prediction purposes. One of the most important implications of the ANN is in hydrological systems, and this chapter provides an insight into the concept, background, theory, classification, and application of the ANN models in hydrology and water resources management.

3. Vector Machine Support

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:



4.Algorithms for Genetics

The *genetic algorithm* (GA), developed by John Holland and his collaborators in the 1960s and 1970s (Holland, 1975; De Jong, 1975), is a model or abstraction of biological evolution based on Charles Darwin's theory of natural selection. Holland was probably the first to use the crossover and recombination, mutation, and selection in the study of adaptive and artificial systems. These genetic operators form the essential part of the genetic algorithm as a problem-solving strategy. Since then, many variants of genetic algorithms have been developed and applied to a wide range of optimization problems, from graph coloring to pattern recognition, from discrete systems (such as the travelling salesman problem) to continuous systems (e.g., the efficient design of airfoil in aerospace engineering), and from financial markets to multi-objective engineering optimization.

There are many advantages of genetic algorithms over traditional optimization algorithms. Two most notable are: the ability of dealing with complex problems and parallelism. Genetic algorithms can deal with various types of optimization, whether the objective (fitness) function is stationary or non-stationary (change with time), linear or nonlinear, continuous or discontinuous, or with random noise. Because multiple offsprings in a population act like independent agents, the population (or any subgroup) can explore the search space in many directions simultaneously. This feature makes it ideal to parallelize the algorithms for implementation. Different parameters and even different groups of encoded strings can be manipulated at the same time.

However, genetic algorithms also have some disadvantages. The formulation of fitness function, the use of population size, the choice of the important parameters such as the rate of mutation and crossover, and the selection criteria of the new population should be carried out carefully. Any inappropriate choice will make it difficult for the algorithm to converge or it will simply produce meaningless results. Despite these drawbacks, genetic algorithms remain one of the most widely used optimization algorithms in modern nonlinear optimization.

S.NO	NAME OF RESEARCHERS	MACHINE LEARNNING APPROACHES	ECG FEATURES	TARGET CLASSES	DATASET	ACCURACY
1	AmitaMalav, Kalyani Kadam et. al. 2018	ANN algorithm, K-means algorithm	Which is more reliable to diagnosis	Normal,Noise/artifact, VT, AF, Bigeming, PVC	Cleveland dataset	89.01%.
2	B Sankara Babu, A Suneetha et. al. 2018	Grey Wolf optimization algorithm, RNN algorithm	It improves accuracy, sensitivity and specificity	Normal,Noise/artifact, VT, AF, Bigeming, PVC	standard medical UCI dataset	97.46%
3	Mohd Faisal Ansari, Bhavya AlankarKaur Et. al. 2021	modified algorithm using logistic regression	Various classification problems	Normal,Noise/artifact, VT, AF, Bigeming, PVC	repository dataset	87%
4	Shrinivas D. Desai, ShantalaGiraddi Et. al. 2019	Logistic regression and back-propagation neural networks (BPNN) (LR)	It is essence of neural network training	Normal,Noise/artifact, VT, AF, Bigeming, PVC	Cleveland heart disease dataset	BPNN -85.074% LR-92.58%
5	FadouaKhennou, Charif Fahim et. al. 2019	naive Bayesian algorithm, SVM algorithm	It supports both classification or regression challenges	Normal,Noise/artifact, VT, AF, Bigeming, PVC	Hungarian dataset ,Switzerland datasets	Naïve bayes-86% Svm-87%
6	Mr. ChalaBeyene, Prof. Pooja Kamat 2018	Decision tree, Naive Bayes, K-Nearest Neighbour algorithm	Various Classification problems	Normal,Noise/artifact, VT, AF, Bigeming, PVC	StatLog heart disease dataset	Navie baye-83% KNN-80% Decision tree-77%
7	SP Rammohana, K.karunya et. al. 2018	Feature selection, Multi-Layer perceptron	To reduce the number of input variables	Normal,Noise/artifact, VT, AF, Bigeming, PVC	Cleveland heart disease dataset	88.46%
8	Qasem Al-Tashi, Helmi Rais et. al. 2018	ant colony optimization algorithm	It improves accuracy, sensitivity, and specificity	Normal,Noise/artifact, VT, AF, Bigeming, PVC	Cleveland heart disease dataset	…………..
9	Cincy Raju, PhilipsyE et. al. 2018	Naïve Bayesian classifiers, Association Rule,SVM	To predict disease by using datamining techniques	Normal,Noise/artifact, VT, AF, Bigeming, PVC	Cleveland heart disease dataset	Navie bayes-80%

SURVEY ON ECG DATA CLASSIFICATION

A lot of research has been done in the past and now about the separation of ECG data. The mechanical reading methods are mostly used rather than the few textbooks available. Many learning classification systems, including those Vector support (SVM), Neural Networks (NN), Neuro-fuzzy algorithms, Adaptive neuro-fuzzy algorithms, and Multimodal determinant learning, have been used. in a previous study. examines the ECG signal and data classification process primarily which includes pre-processing, filtering, feature extraction, feature reduction, classification, and performance measurement. Table I clearly summarizes previous research on ECG classification.

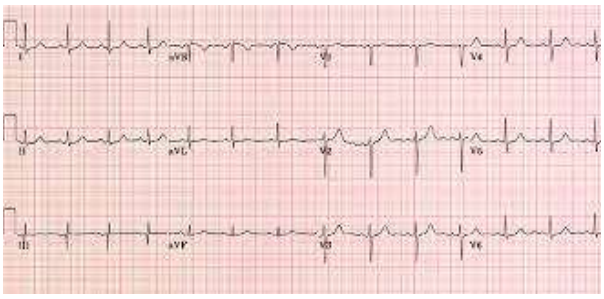


Fig 3- Normal ECG

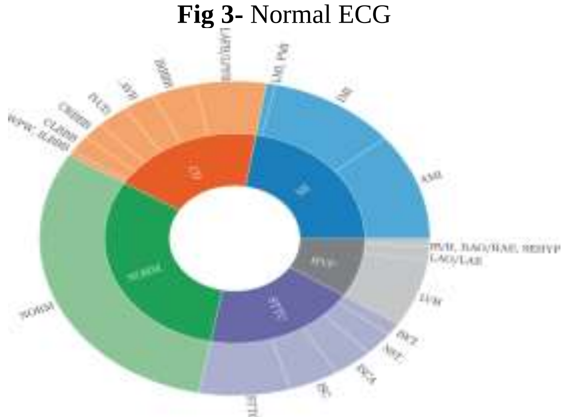


Fig 4- Data Set1

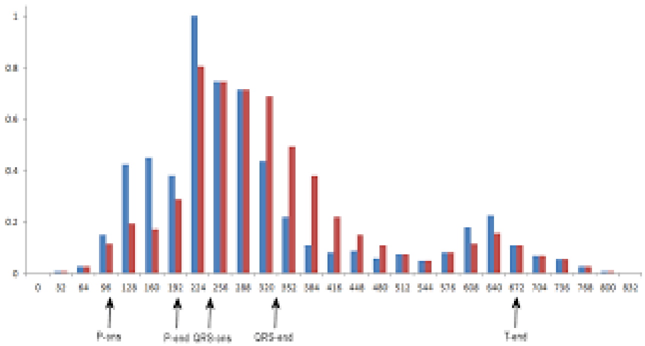


Fig 5- Data Set2

CONCLUSION

Recent research has evaluated different ML algorithms used to predict heart disease using ECG monitoring. Synthetic networks (ANN), vector support systems (SVM), fuzzy logic (FL), pruning trees (DT), Rough set theory (RST), genetics, but also algorithm hybrids are all widely used methods and techniques. In addition to the many previous findings, such a research area requires a thorough investigation into the actual implementation of the classification system. Additionally, this review of the literature reveals that very few books focus on the coronary artery and its STEMI origin. As a response, there is a real need for us to fill this void in the study and make a significant contribution to this field.

REFERENCES:

1. AmitaMalav, Kalyani Kadam "A Hybrid Approach for Heart Disease Prediction Using Artificial Neural Network and K-means"International Journal of Pure and Applied Mathematics Volume 118 No. 8 2018, 103-110.
2. B Sankara Babu, A Suneetha"Medical Disease Prediction using Grey Wolf optimization and Auto Encoder based Recurrent Neural Network"Periodicals of Engineering and Natural Sciences ISSN 2303-4521 Vol.6, No.1, June 2018, pp. 229~240
3. Mohd Faisal Ansari, Bhavya AlankarKaur "A Prediction of Heart Disease Using Machine Learning Algorithms"© The

Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2021 J. I.-Z. Chen et al. (Eds.): ICIPCN 2020, AISC 1200, pp. 497–504, 2021.

4. Shrinivas D. Desai, ShantalaGiraddi "Back-Propagation Neural NetworkVersus Logistic Regression in HeartDisease Classification"© Springer Nature Singapore Pte Ltd. 2019 J. K. Mandal et al. (eds.), Advanced Computing and Communication Technologies, Advances in Intelligent Systems and Computing 702
5. FadouaKhennou, Charif Fahim "A Machine Learning Approach: Using Predictive Analytics to Identify and Analyze High Risks Patients withHeart Disease"International Journal of Machine Learning and Computing, Vol. 9, No. 6, December 2019.
6. Mr. ChalaBeyene, Prof. Pooja Kamat"Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques"International Journal of Pure and Applied Mathematics Volume 118 No. 8 2018, 165-174.
7. SP Rammohana, K.karunya"ANALYSIS OF NEURAL NETWORKS BASED HEART DISEASE PREDICTION SYSTEM"978-1-5386-5025-0/18/$31.00 2018 IEEE
8. Qasem Al-Tashi, Helmi Rais"Hybrid Swarm Intelligence Algorithms with Ensemble Machine Learning for Medical Diagnosis"978-1-5386-4744-8/18/$31.00 ©2018 IEEE
9. Cincy Raju, Philipsy"A Survey on Predicting Heart Disease using Data Mining Techniques"Proc. IEEE Conference on Emerging Devices and Smart Systems (ICEDSS 2018) 2-3 March 2018, Mahendra Engineering College, Tamilnadu, India.
10. Arash Malekian, Nastaran Chitsaz "**Concepts, procedures, and applications of artificial neural network models in streamflow forecasting", Advances in Streamflow Forecasting,**Elsevier,2021, Pages 115-147.
11. Xin-She Yang, " Genetic Algorithms" , Nature Inspired Optimization Algorithms(Second Edition), Academic Press, 2021, Pages 91-100.