International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-3S, February 2019

Accurate and Fast Diagnosis of Heart Disease using Hybrid Differential Neural Network Algorithm

O. Bhaskaru, M.Sree Devi

ABSTRACT--- Heart disease is the most threatened issues which is more dangerous than the other kind of diseases. The most of the people in world is causes by heart disease which increases the death rate of humans considerably. The detection of heart disease is the more trivial task of medical researchers which cannot be done more accurately. And also manual prediction of heart disease is most complex task, thus it is required to implement the automated system by considering the recent computer technologies which can help medical researchers to diagnosis heart disease fastly and accurately. This is done in this research method by introducing the new method namely Hybrid Differential Evolution based Fuzzy Neural Network (HDEFNN) for heart disease diagnosis. In this method DE is hybridized with the FNN algorithm to ensure the better performance of heart disease diagnosis. The proposed method guarantees the accurate and reliable identification of heart disease with the help of neural network based learning. Here the performance of neural network is enhanced by introducing the genetic algorithm which will update the initial weight values hidden layers, thus the learning prediction accuracy can be improvised. Here the genetic algorithm ensures the 10% performance improvement of neural network. The simulation evaluation of the proposed method HDEFNN is carried out on dataset namely Cleveland heart disease dataset which is collected from the University of California at Irvine (UCI) machine learning repository. The performance evaluation of the proposed method is done in the matlab simulation environment under k-fold cross validation procedure which proves that the proposed method shows better performance in terms of accurate disease diagnosis within less execution time.

Keywords: Heart Disease, Differential Evolution, Fuzzy Neural Network, Genetic Algorithm, Accuracy.

I. INTRODUCTION

Data mining is the most popular field in real world which allows users to find the knowledge from available large volume of data [1]. Data mining includes various fields to extract the useful information. Some of the fields are machine learning [2], neural networks [3], statistics and pattern recognition [4]. The data mining has been significantly utilized in the medicinal fields for early determination of illnesses by the treatment of huge datasets. It is likewise used in numerous different fields for building up the compelling promoting systems. Cardiovascular disorder is a noteworthy reason of inauspiciousness and mortality in the present living style [5]. Today heart sicknesses are most basic ailment in our general public. Around 60% of our population is experiencing coronary illness in the advanced everyday life [6]. The general impacts of this infection result in ailment, handicap and passing. Data mining based coronary illness prediction framework can help in deciding the coronary illness in beginning periods. The forecast framework lessens the high danger of coronary illness and furthermore helps a non-particular specialist to settle on right choice about the coronary illness hazard level [7]. The different data mining procedure utilized for this examination are Classification, Clustering, Fuzzy framework and Association Rules to build the exactness for the diagnosis procedure.

Cardiovascular is the most widely recognized ailment which is steadily expanding the death rate in the people in the present life [8]. Coronary Artery Disease (CAD) is the most significant sort of cardiovascular ailment in which the veins get limited and keeps the supply of oxygen-rich blood to the heart which results in a stroke [9]. The different finding procedures likely Angiography, Electrocardiogram (ECG), blood test are engaged with the identification of CAD. Among these, Angiography is the broadly utilized diagnostic method. Be that as it may, it is of surprising expense and may prompt huge complexities like disease. So to dispense with those dangers, detection model has been proposed dependent on artificial neural networks and genetic algorithms. Along these lines this model guides to give the better outcomes.

The rest of the paper is organized as follows. Section 2 discusses the literature review. Section 3 overviews the proposed technique for disease prediction. Experimental results of the proposed scheme are presented in Section 4. Concluding remarks with future work are covered in Section 5.

II. LITERATURE SURVEY

The authors of [10] provided the model for diagnosis of heart disease by the data mining technique. The data is retrieved from Z- AlizadehSani dataset which contains a record of 303 patients out of which 216 were suffered from CAD. The feature process consists of four ranking methods namely Gini index, weight by SVM, information gain and Principal component analysis.

Manuscript published on 28 February 2019.
* Correspondence Author (s)
O.Bhaskaru, Research scholar, CSE Department, Koneru Lakshmaiah Education Foundation, Green Fields, Guntur District, Vaddeswaram, Andhra Pradesh, India. 522502, (bha_jesus@yahoo.co.in)
Dr.M.Sree Devi, Professor, CSE Department, Koneru Lakshmaiah Education Foundation, Green Fields, Guntur District, vaddeswaram, Andhra Pradesh, India. 522502, (msreedevi_27@kluniversity.in)

Thus the model is based on artificial neural network and genetic algorithm to diagnose the heart disease by clinical means there by eliminating the angiography.

K.Geethaet al., introduced the system that deals with Thyroid diseases [11]. The data has been collected from the UCI repository and undergo pre-processing. The data which had pre-processed is multivariate in nature. Using Hybrid Differential Evolution based algorithm the Dimensionality has been followed to optimize 21 to 10 number of attributes. The subset has been provided to Support Vector Machine classifier algorithm. Then in order to stabilize the errors this process has been carried iteratively and data is classified. Thus accuracy of classification has been observed.

The authors of [12] proposed a system that analyzes the fuzzy neural networks uses and implementations based on FPGA in order to predict various physiological diseases. This Strategy works with path physiological data that has been collected. Then Fuzzy Rule are prepared as per the guidelines of doctors. Finally using Neural Network map these rule in order to predict the specific disease. This process has been carried out based on predefined parameters. FPGA had implemented using Fuzzy-Neural algorithm which is used as an expert system in the field of disease prediction.

Isaac Trigueroet al., proposed a methodology for optimize the position of prototypes relay on differential evolution approach [13]. They made various study regarding performance of recent things in differential evolution. This process has been carried out by optimizing the position which is prior to nearest neighbor classification. Thus results have been identified.

The authors of [14] brought out the system for diagnosis the heart disease which is based on the ReliefF and Rough Set method by sing hybrid classification system. This system contains RFRS feature selection and a classification along with an ensemble classifier. These include data discretization, feature extraction and feature reduction. To deal with these authors had developed the algorithm called heuristic Rough Set reduction. On the other hand, an ensemble classifier has been proposed depends on C4.5 classifier. The dataset on heartis collected from UCI database. Finally system achieved the maximum classification accuracy of about 92.59%.

The authors of [15] used the Bat Imperialist Competitive Algorithm which is based on an evolutionary algorithm on the development of human's socio-political. In this initial population has been considered with m features and n cardiovascular heart disease observations. This approach helps to minimize number of features that indirectly reduce the diagnosis tests mandatory for patients. This system acquires 94.4% accuracy and provides better accuracy for diagnosis cardiovascular heart disease.

In [16] authors proposes a method use Differential Evolution (DE) for hybrid classification algorithm and Least Squares Support Vector Machine which is used for classification. DE algorithm is using the parameters that guarantee the effectiveness. This algorithm evaluates Breast Cancer dataset collected from UCI Repository. This algorithm pays a way for comparing with different classifier based algorithms on same database. Thus it achieves an accuracy of 99.75%.

Alsalamahet al., aims to develop an informatics system for classification of heart diseases with the help of data mining techniques for Radial Basis functions and emerging Neural Networks approach [17]. The author deals with various processes such as classification system using Radial Basis neural networks for Coronary Artery Disease. Then by deep learning approach different types of heart diseases has been identified. Thus it provides novel technique by using Neural Networks techniques, and it evaluates the accuracy and performance by compare the results of different classification models.

In [18] the authors proposed a technique for artificial neural system learning by differential transformative calculation. DE with global and local neighborhood-based change (DEGL) calculation scans for synaptic weight coefficients and to limit the learning mistake. Both global and local neighborhood-based mutation operator have joins to shape contributor vector. This strategy works for order of certifiable information and along these lines results drew out the proficiency and viability.

The authors of [19] gave writing on different optimization algorithms, machine learning algorithms and applications in smart healthcare. This paper additionally talked about with difficulties, security, pilot studies and genuine venture and correspondence between medicinal staffs and information investigation which are basic for savvy human services. This proposition clears an approach to guarantee the wellbeing administrations accessible in future that are earth supportable and monetarily reasonable.

In [20] the author utilizes the last populace of heterogeneous flexible neural trees (HFNTs) or, in other words Pareto-based multi-objective genetic programming (MOGP). The parameter tuning by differential development draws out the astounding gathering frameworks. Enhancement of exactness, decent variety and intricacy gives answer for auxiliary multifaceted nature. In this manner chose hopefuls helps as great outfit framework. These methodology backings to better execution over the calculations that are gathered from correlation. Consequently this paper demonstrates that HFNT is viably utilized for data analysis and modelling.

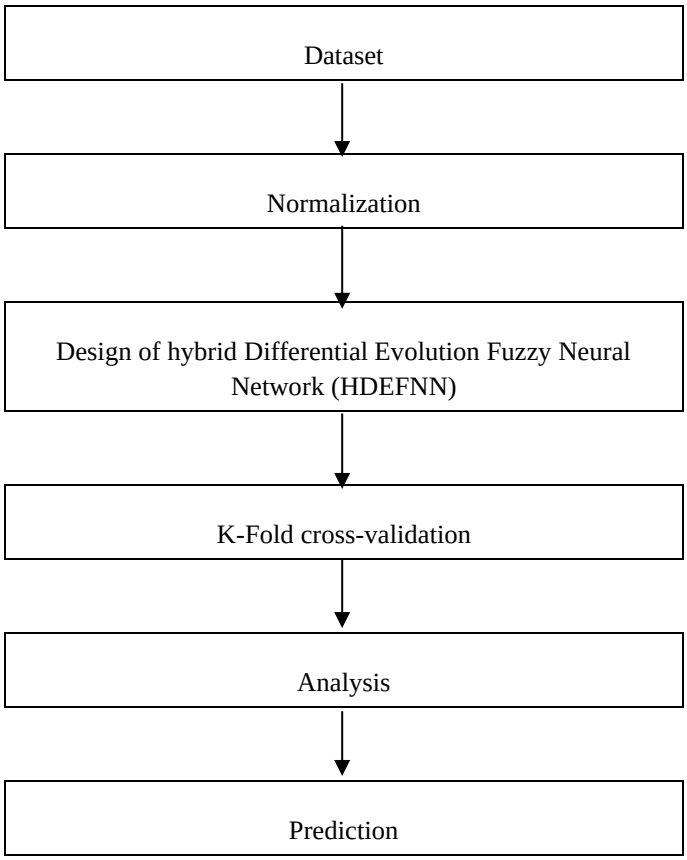
III. PROPOSED SYSTEM

Heart disease diagnosis plays a more important role in real world where most of people including both men and women of all ages affected lot. Heart disease would cause more serious threat to the people which needs to be diagnosed more efficiently. It is required to diagnosis the heart disease early in order to ensure the early treatment procedure thus the unwanted health illness can be avoided. The main goal of this research work is to introduce the system which can predict the heart disease in the automated manner as soon as possible with increased accuracy level. This is achieved in this research work by introducing the Hybrid Differential Evolution Fuzzy Neural Network (HDEFNN) which can predict the heart disease occurrence fastly and accurately.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The performance of proposed method in terms of accurate diagnosis of heart disease is attained by improving the initial weight updation of neural network which is done by introducing the genetic algorithm. The genetic algorithm can select the most optimal weight values for the hidden layers of neural network. The processing flow of the proposed research method is shown in the following figure 1. The obtained dataset is first normalized before applying algorithms. Then designed Differential Evolution Fuzzy Neural Network (HDEFNN) algorithm is applied to the normalized dataset in k-fold cross-validation. Theperformance of the proposed algorithm is analyzed.

Figure 1: Proposed System – Process Flow



3.1. Differential Evolution

Differential Evolution (DE) is known to be more simpler and effective technique for searching process to solve the continuous optimization problems. DE adapts the direction aware search procedure to find the reliable solution among various solutions. DE tends to store their candidate solution in the vector format. The updation of stored candidate solutions is performed by following procedures namely mutation, crossover and selection. The major difference found between the conventional procedures and the DE method is adaptation of mutation operator which generates the new candidate solution from the available population of solutions. The processing steps of DE are shown in the following algorithm 1.

Algorithm 1:

1: Initialize and configure the algorithm parameters
2: Repeat till ending criteria met
3: For each and every candidate solution stored in vector vi do
4: Apply mutation operator and generate new candidate solution and stored in new vector ui
5: Apply cross over operator between the candidate solutions stored in vectors ui and vi and generate the test vector ti
6: end for
7: For each candidate solution in vector vi do
8: if f(ti) ≤ f(vi) then
9: vi ← ti
10: end if
11: end for
12: end while

3.2. Fuzzy Neural Network

Feed forward network is also defined as the multi-layer perceptron which differs from the conventional methods with the utilization of fuzzy signals [21]. The difference between the conventional and the neuro fuzzy system can be identified based on the definition of NFS system which is represented as

$$NFS = (\mathbf{F}_n, \mathbf{F}_m, \tilde{\mathbf{I}}, \tilde{\mathbf{O}}, \mathbf{A}, \mathbf{L}, f) \quad 1$$

Where F_n represents the input field which is sorted vector of neurons with size n×1,

F_m represents the output field which is sorted vector of neurons with size m×1

$\tilde{\mathbf{I}}$ represents fuzzy input vector $(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n) \in \Re^n$, where $\tilde{x}_i$ represents the input to neuron i in the input field, F_x,

$\tilde{\mathbf{O}}$ represents fuzzy output vector, $(\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_m) \in \Re^m$, where $\tilde{y}_i$ represents output from neuron i in the output field, F_y,

A: $\Re^n \rightarrow \Re^m$ reprents association function,

$\mathbf{L} \subseteq \mathbf{A}$ represents set of learned associations, and

f : $\Re^k \rightarrow \Re$ represents neurons' activation function.

3.3. Genetic Algorithm

Genetic Algorithm is a population based iterative pursuit calculation which seeks starting with one population then onto the next, concentrating on the territory of the best candidate solution up until now, while persistently looking through the solution space age shrewd. To accomplish best solution, the GA applies stochastic administrators, for example, selection, crossover and mutation. The significant strides of GA include: encoding, instatement of the population, fitness assessment, selection, crossover and mutation. In this work we utilize genuine esteemed chromosome rather than binary representation of the weights.

Pseudocode

Set i=0
Initialize the candidate solution population
Calculate the fitness solution for each candidate solution present initial population
Repeat until end criteria met
{

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Apply selection operator on candidate solution
Apply cross over operator candidate solutions resultant from selection operator
Apply mutation operator on resultant solution obtained from the cross over operator
Evaluate the fitness value of new candidate solution obtained from mutation operator
Increment i value i = i + 1;
}
end

3.4. Hybrid Differential Evolution Fuzzy Neural Network (HDEFNN)

In this work FNN technique is hybridized with the DE algorithm to ensure the accurate diagnosis outcome of heart disease. Like genetic algorithm, DE is also known to the population based optimal search space seeking algorithm which can be able find the most optimal search solutions from the entire search space. Here DE is hybridized with the FNN algorithm to optimize the decision making of FNN algorithm. The proposed heart disease diagnosis method HDEFNN following two processing steps namely parameter initialization and the FNN training. The processing flow of proposed HDSFNN method is shown in the following pseudo code.

Pseudocode

1: Initialize the configurator parameter values of DE algorithm
2: For each input neurons of FNN algorithm
Initialize the search space in the domain [0, 1] using real values
3: repeat until satisfied New solution <= MaxGen.DO
{

- Measure the fitness values of solution vectors by using Mean Square Error (MSE) metric
- Order the fitness values of candidate solutions in Ascending order

4: If first fitness value <= min. error then

- Choose the optimal solution for Multilayer Perceptron (MLP)

5: Create end vectors for storing solutions obtained from mutation and crossover operators
6: Initialize the FNN parameter values
7: initialize the weight values of MLP in terms selected optimal solution
8: Update weights to minimize error using backpropagation with training data
9: End while
10: End

IV. EXPERIMENTAL DESIGN

Dataset: The dataset is obtained from the University of California at Irvine (UCI) machine learning repository The Cleveland heart disease data set has 303 records. The data set has 13 attributes and 1 class attribute. The class attribute refers to the presence of heart disease in the patient.

Normalization: Normalization is used to convert the data into standard format to make it suitable for algorithm evaluation. Here the data value is transformed within the range of [0,1] by applying the normalization process.

K-fold cross-validation: K-Fold cross validation procedure is used to test the processing results of training phase in order to increase the learning accuracy by adjusting the error value. Here the k value is considered as 10 where the input dataset will be divided into 10 parts where one part will be considered for the testing phase and the remaining part will be considered for the training phase.

Classification accuracy is used to predict the performance outcome of the propsoed methodology in terms of correct diagnosis outcome. The classification accuracy is measured with the help of confusion matrix tool by using the which rows of varying classes in data set can be analysed. The processing outcome of confusion matric is shown in the following figure 2.

Figure 2: Confusion Matrix

	Prediction Outcome		
p'	P	n	total
Actual Value	True Positive	False Negative	P'
n'	False Positive	True Negative	N'
total	P	N	

Accuracy: Accuracy is defined as percentage of the correctly predicted outcome.

Accuracy= (True Positive + True Negative)/ (True Positive + False Negative+
False Positive + True Negative)

TN / True Negative: case was negative and predicted negative

TP / True Positive: case was positive and predicted positive

FN / False Negative: case was positive but predicted negative

FP / False Positive: case was negative but predicted positive

Precision: True Positive/ (True Positive+ False Positive)

Recall: True Positive/ (True Positive+ False Negative)

Results and Discussion

The proposed approach is compared with three algorithms. J48 decision tree, naïve bayesand random forest algorithms are taken into consideration.

Table 1: Accuracy and Execution Time

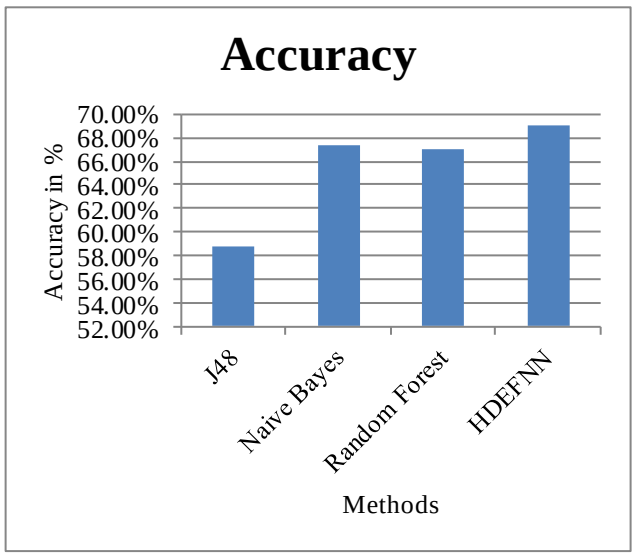
Algorithm	Accuracy	Execution Time (In Seconds)
J48	58.8%	9
Naive Bayes	67.4%	**3**
Random Forest	67.1%	14
HDEFNN	**69.1%**	4

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Table 1 shows the performance comparison for accuracy and execution time. The accuracy of proposed algorithm is 69.1%. The accuracy of other algorithms are less compared to HDEFNN. Similarlyin terms of execution time, NaiveBayees has the better result with 3 seconds.

The accuracy comparison of proposed and exisitng method is shown in the following figure 3.

Figure 3. Accuracy comparison evaluation



The execution time comparison is shown in the following figure 4.

Figure 4. Execution time comparison

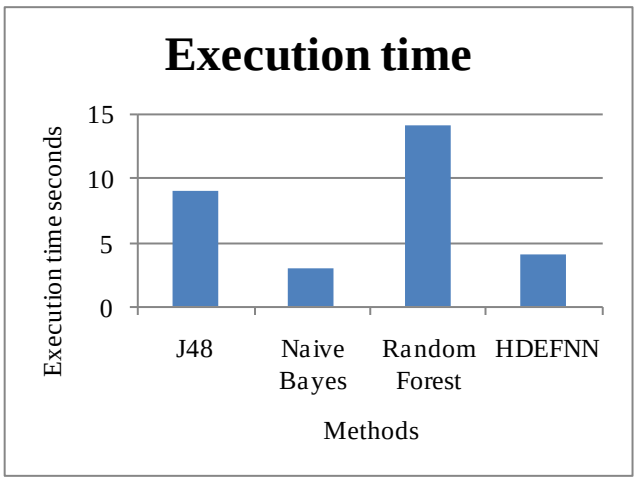


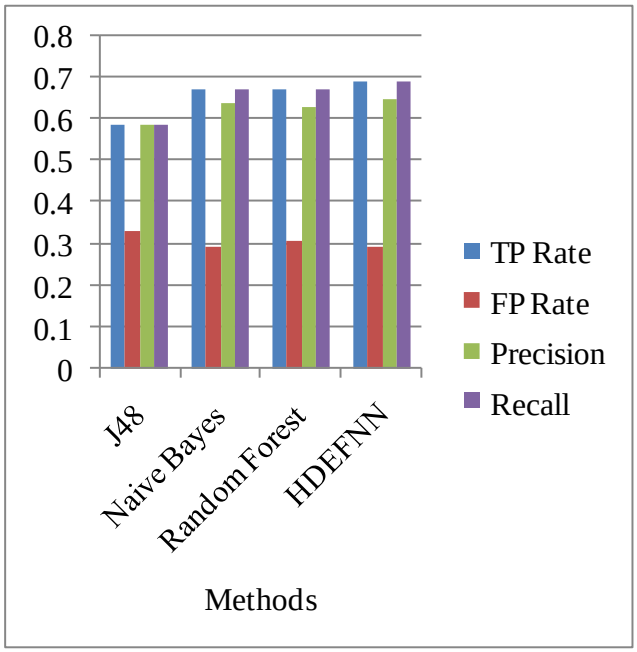
Table 2: Performance Analysis

Algorithm	TP Rate	FP Rate	Precision	Recall
J48	0.588	0.332	0.585	0.588
Naive Bayes	0.674	0.292	0.641	0.674
Random Forest	0.671	0.306	0.631	0.671
HDEFNN	**0.691**	0.293	**0.650**	**0.691**

Table 2 represents the 4 parameters TP Rate, FP Rate, Precision and Recall. Among these 4 performance metrics, the proposed algorithm outperforms other algorithms in TP rate, precision and recall. In the FP rate Naive Bayes slightly performs well compared to other algorithms.

The graphical representation of comparison evaluation between the proposed and existing methods is shown in the following figure 5.

Figure 5. Comparison of proposed and existing methods



V. CONCLUSION

In this work, heart disease diagnosis is performed by introducing the automated research method in order to ensure the faster and accurate prediction of heart disease in early stages. This is performed by introducing the HDEFNN method which can guarantee the accurate identification of the heart disease. In this work, evaluation of the heart disease prediction system is done over the heart disease data set collected from the UCI repository. The performance assessment is done by comparing the proposed system with the various conventional methods. The better after effects of this framework helps the zone experts and even the individual related with the field to prepare for a better decide and give the patient than have early assurance results as it performs sensibly well even without retraining. The proposed calculation gives better characterization results higher preparing speed and lesser estimate mistake.

REFERENCES

1. Witten, I. H, Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.
2. Robert, C. (2014). Machine learning, a probabilistic perspective.
3. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural networks, 61, 85-117.
4. Samarasinghe, S. (2016). Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition. Auerbach publications.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

5. Saxena, K., & Sharma, R. (2016). Efficient Heart Disease Prediction System. Procedia Computer Science, 85, 962-969.
6. Price, W. A., & Nguyen, T. (2016). Nutrition and physical degeneration: a comparison of primitive and modern diets and their effects. EnCognitive. com.
7. Sharma, P., & Saxena, K. (2017). Application of fuzzy logic and genetic algorithm in heart disease risk level prediction. International Journal of System Assurance Engineering and Management, 8(2), 1109-1125.
8. Hobbs, F. D. R., Piepoli, M. F., Hoes, A. W., Agewall, S., Albus, C., Brotons, C., ... & Deaton, C. (2016). 2016 European Guidelines on cardiovascular disease prevention in clinical practice. European Heart Journal, 37(29), 2315-2381.
9. Nikpay, M., Goel, A., Won, H. H., Hall, L. M., Willenborg, C., Kanoni, S., ... & Webb, T. R. (2015). A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. Nature genetics, 47(10), 1121.
10. ZeinabArabasadi, RoohallahAlizadehsani, Mohamad Roshanzamir, Hossein Moosaei, Ali AsgharYarifard, "Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm", Computer Methods and Programs in Biomedicine, January 12, 2017.
11. K.Geetha, S.SanthoshBaboo, "Efficient thyroid disease classification using differential evolution with SVM", Journal of Theoretical and Applied Information Technology, Vol.88, June 30,2016.
12. Ketan K. Acharya, R.C.Patel, "Applications of Fuzzy-Neural and FPGA for prediction of various diseased- a survey",IJAEEEIE, Vol.3. December 12, 2014.
13. Isaac Triguero, Salvador Garcia, Francisco Herrera, "Differential evolution for optimizing the positioning of prototypes in nearest neighbour classification", Pattern Recognition, October 24, 2010.
14. Xiao Liu, Xiaoli Wang, Qiang Su, Mo Zhang, Yanhong Zhu, Qiugen Wang, Qian Wang, "A Hybrid classification system for heart disease diagonosis based on the RFRS method", Hindawi Computational and Mathematical Methods in Medicine, 2017.
15. Nithya S, Suresh GnanaDhas C, 'Bat imperialist competitive algorithm(BICA) based feature selection and genetic fuzzy based improved kernel support vector machine(GFIKSVM) classifier for diagnosis of cardiovascular heart disease", Biomedical Research 2018.
16. Omar S.Soliman, EmanAboElHamd, "Classification of Breast Cancer using Differential Evolution and Least Squares Support Vector Machine", IJETTCS, Vol.3, March-April 2014.
17. Alsalamah M, "Artificial Diseases Diagnosis using Artificial neural networks", Conventry University, January 27, 2017.
18. Tapas Si, SimantaHazra, and N.D Jana, "Artificial Neural Network training using Differential Evolutionary Algorithm for Classification", January 2012.
19. Kwok Tai Chui 1, WadeeAlhalabi 2, Sally Shuk Han Pang 3, Patricia Ordóñez de Pablos 4, Ryan Wen Liu 5and Mingbo Zhao 6, "Disease Diagnosis in smart healthcare : Innovation, Technologies and applications", Sustainability 2017.
20. Varun Kumar Ojha, Ajith Abraham, Vaclav Snasel, "Ensemble of heterogeneous flexible neural trees using multi objective genetic programming", Vol.52, May 16, 2017.
21. http://wwwold.ece.utep.edu/research/webfuzzy/docs/Nist98/Nist98.html