International Journal of Pure and Applied Mathematics
Volume 119 No. 14 2018, 1671-1681
ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

A NEW PRECEPTION OF GREY WOLF OPTIMIZATION IN CLOUD CLASSIFICATION: AN CLOUD BASED APPLICATION FOR SOVING MEDICINE ORIENTED REAL WORLD COMPLEX PROBLEMS

Balaji. S[1], Mandy, Sonio Newcastle[2]
Research Schalor[1],Assistant Prof[2], Dept of IT, Bihar University

ABSTRACT

Bio-Inspired Computational Algorithms and its Applications are used. It integrates various techniques such as genetic algorithms, Ant colony algorithm, Bee algorithm, particle swarm optimization and many hybrid models are used for solving many real-world problems. It is very easy and interesting to learn animal behavior that can be implemented for useful purpose. In this paper existing system are optimization is done using various constraints like alpha, beta and delta and sine cosine algorithm but still they have not extended their concept up to classification. In our project, a new approach to proposed system is Known as Grey Wolf Optimizer (GWO) – Sine Cosine Algorithm (SCA) with five bio-medical dataset and one sine dataset problems are used. The clinical data is a complex data model that correlates the patient's disease, treatment, reaction and historical medical records. The data availability of all the attributes of this complex clinical data model is a serious challenges. The approach should be taken based on the availability of the data. The GWO technique is used to analyze the attributes or symptoms by predicting the diseases. This work enables the prediction of the diseases with improved accuracy. In the second attempt, The clinical information collected from the different dataset is uncertain in nature. Then the Rough set is used to classify the data into best case, worst case and finally classify the rate of the disease. The rules are extracted from the attributes. Finally, the data are classified as best case, worst case. In the proposed work the critical attribute are identified and the irrelevant attributes are removed. The reduction is done without compromising the efficiency and accuracy of the diagnosis process.

Keywords: Bio-inspired computing, Grey Wolf Optimization (GWO), Sine Cosine transformation, Machine Learning, Medical Dataset.

INTRODUCTION

A Bio Inspired Computing

Bio-inspired computing, is one of the biological inspired computing. In this field of study that easily knits together sub fields related to the concepts of connection, social animal behavior and reptiles behavior and then small insects are performed in this domain. It is full of closely related to the field of artificial intelligence, cloud computing, data mining etc as many of its pursuits can be linked to machine learning. It combination subjects are on the fields of biology, computer science and mathematics. Briefly put, it is the use of computers to model the living things, organization of biological structures, genetic concepts are also performed and simultaneously the study of life to improve the usage of computers. Biological inspired computing is a major subset of natural computation. It is one of the techniques of used genetic algorithms, artificial intelligence systems [30-35], particle swarm optimization, ant colony algorithm, cuckoos optimization, bat algorithm and hybrid models to solve many real-world problems [36- 43]. It is very interesting and easily understandable concept of bio inspired domain. And it is one of the part of cloud computing is used and having the small concept of animal behavior of things.

Bio-inspired computational algorithms are usually shortly called as bio-inspired computation. As a result, the bio-inspired computation is generally considered as a total field of investigation are mentioned in subfields together. Moreover, the bio-inspired computational algorithms are tightly related with the community of artificial intelligence, especially the machine learning research field. Both of the bio-inspired computational algorithms and machine learning are based on some concepts and mechanisms of biology, mathematic theories and models, and computer science. To make it more specific, the bio-inspired computational algorithms are motivated from living phenomena.

B Machine learning

Machine learning is the learning process used to extract the knowledge with empirical data. It is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly. Machine learning enables analysis of massive quantities of data. While it generally delivers faster, more accurate results in order to identify profitable opportunities or dangerous risks, it may also require additional time and resources to train it properly. Combining machine learning with AI and cognitive technologies can make it even more effective in processing large volumes of information.

C Types of machine learning:

1) Supervised Learning
2) Unsupervised Learning
3) Semi-supervised Learning
4) Reinforcement Learning

a) Supervised Learning

The classification of the data using the labeled class labels are known as Supervised Learning shown in Fig. 1.1. Both the training data and the testing data are labeled with the classes. Prediction and classification are the data mining technique which comes under supervised learning.

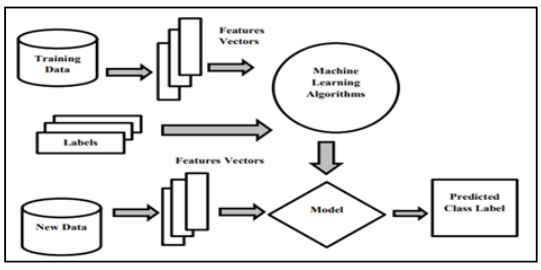


Fig.1.1 (Supervised learning)

b) Unsupervised Learning:

The classification of the data using the unlabeled class labels are known as Unsupervised Learning shown in Fig. 1.2. The system is used establish the class label with the available data. Clustering, Association mining, Dimension reductions are the data mining technique which comes under unsupervised learning.

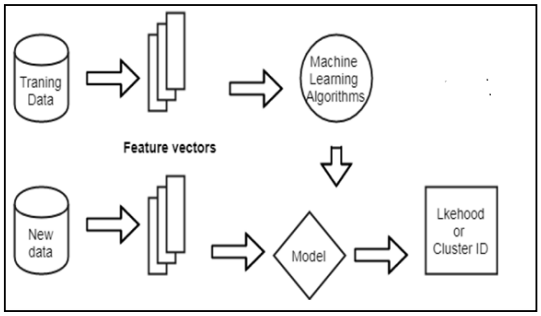


Fig.1.2(Unsupervised learning)

c) Semi-supervised Learning:

Both labeled and unlabeled examples are used in creating the function or classifier.

d) Reinforcement Learning:

It is another type of machine learning used in the field of artificial intelligence. It expects an additional behavior of the machine or agent to perform actions automatically within a specific context with increased performance.

1. RELATED WORK

Aijun Zhu [1],et.al., proposed the A new meta-heuristic method is proposed to enhance current metaheuristic methods for global optimization and test scheduling for three-dimensional (3D) stacked system-on-chip (SoC) by hybridizing grey wolf optimization with differential evolution (HGWO). Because basic grey wolf optimization (GWO) is easy to fall into stagnation when it carries out the operation of at-tacking prey, and differential evolution (DE) is integrated into GWO to update the previous best position of grey wolf Alpha, Beta and Delta, in order to force GWO to jump out of the stagnation with DE's strong searching ability. The proposed algorithm can accelerate the convergence speed of GWO and improve its performance. Twenty three well-known benchmark functions and an NP hard

problem of test scheduling for 3D SOC are employed to verify the performance of the proposed algorithm. Experimental results show the superior performance of the proposed algorithm for exploiting the optimum and it has advantages in terms of exploration.

Zhang Dan [2] et. al., was proposed an artificial neural network (ANN) prediction model that incorporates with Constructive Cost Model (COCOMO) which is improved by applying particle swarm optimization (PSO), PSO-ANN-COCOMO II, to provide a method which can estimate the software develop effort accurately. The modified model increases the convergence speed of artificial neural network and solves the problem of artificial neural network's learning ability that has a high dependency of the network initial weights. This model improves the learning ability of the original model and keeps the advantages of COCOMO model. Using two data sets (COCOMO I and NASA93) to verify the modified model, the result comes out that PSO-ANN-COCOMO II has an improvement of 3.27% in software effort estimation accuracy than the original artificial neural network Constructive Cost Model (ANN-COCOMO II).

Duangjai Jitkongchuen [3] et. al.,proposed a hybrid differential evolution algorithm with grey wolf optimizer for solving continuous global optimization problems. The proposed algorithm introduces a new improved mutation schemes. In this algorithm, the control parameters are self-adapted by learning from previous evolutionary search. Beside, the grey wolf optimizer algorithm is used to enhance the crossover strategy. The performance of the proposed algorithm was evaluated on nine well-known benchmark functions and it was compared to particle swarm optimization, the traditional differential evolution algorithm and the self-adaptive differential evolution algorithm (jDE). The experimental results suggested that the proposed algorithm performed effectively to solving complex optimization problems.

Nampetch Sinsupan, Uthen Leeton and Thanatchai Kulworawanichpong [4] presented the use of the harmony search method for solving optimal power flow problems. The harmony search method mimics a jazz improvisation process by musicians in order to seek a fantastic state of harmony. To assess the searching performance of the proposed method, a 6-bus test power system was challenged. Satisfactory results obtained from the proposed method were compared to those obtained by genetic algorithms, particle swarm optimization, and the quasi-Newton with BFGS formula. This paper described the use of harmony search algorithm to find optimal power flow solutions. This work was conducted by 30 trials. The test also applied the BFGS, genetic algorithms and particle swarm optimization of 30 trials each for comparison. The results showed that the harmony search can be the fastest among the intelligent search methods and also give the smallest S.D. of the solution for the 30 trials.

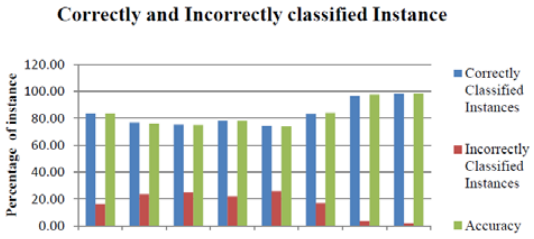
Mahmad Raphiyoddin S. Malik E. Rasul Mohideen Layak Ali [5] have presented the Grey Wolf Optimizer (GWO) is one of the recently introduced Swarm Intelligence (SI) algorithms, developed by the inspiration from grey wolves prey search characteristics. The GWO algorithm imitate the hierarchical leadership and hunting mechanism of grey wolves in nature. The GWO simulates the major steps of grey wolves like hunting, searching for prey, encircling and attacking. The GWO move the wolves pack toward prey by updating location vector, which is an average of best locations of the pack. This paper addresses the GWO issues and proposes the Weighted distance Grey Wolf Optimizer (wdGWO). In proposed wdGWO algorithm, the location update strategy is modified and weighted sum of best locations is used instead of just a simple average. The proposed algorithm is well tested over set of complex benchmark problems and the performance is comprehensively compared with SI algorithms counterpart. The dimensions of the problems are varied from 10 to 50 for fair comparison among basic state of-the-art. Simulation results supports the superior performance of the proposed algorithm.

Additionally, the solutions are compared relying on the metaheuristics reported in the review of literature. This paper discussing the Grey Wolf Optimizer (Algorithm) and Sine Cosine Algorithm (SCA). The HGWOSCA approach mathematically model are used. In this paper presenting the Classification rate, Numerical experiments and parameter settings. It aims at discussing the experimental results, bio-medical real life and medical dataset problems. Finally, the conclusion of the work is summarized at end of the text. This article is organized with different works and subsections to simplify the presentation of work.

2. EXISTING SYSTEM

In Existing System , optimization is done using various constraints like alpha, beta and delta and sine cosine algorithm but still they have not extended their concept up to classification.

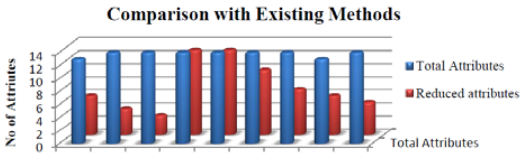
Performance analysis



Methods

Methods is called as a diseases like Heart diseases, AF Heart Disease and Diabetics etc.

Comparison of existing system



3. PROPOSED SYSTEM

3.1 Grey wolf optimization (gwo)

Mirjalili et al. [14] developed a new population based nature inspired algorithm called Grey Wolf Optimization (GWO). This approach mimics the hunting behavior and social leadership of grey wolves in nature. Four types of grey wolves such as alpha, beta, delta, and omega are employed for simulating the leadership hierarchy. The first three best position (fittest) wolves are indicated as a; b and d who guide the other wolves (x) of the groups toward promising areas of the search space. The position of each wolf of the group is updated using the following mathematical equations:

The encircling behavior of each agent of the crowd is calculated by the following mathematical equations:

$$D = /C.Xp(t)-A.x(t))$$

$$X(t+1)=Xp(t)-A.D$$

where t is the current Iteration A &C are coefficient vectors of the prey & X indicates the position vectors a grey wolf.

Where components of a are linearly decreased from 2 to 0 over course of iterations & r1, r2 random vector in (0,1)

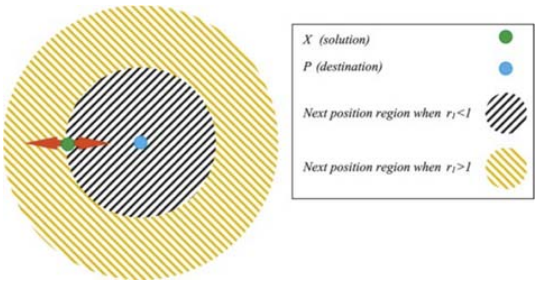


Fig. 2.1. Illustrating next step towards targeted optimum solution.

3.2 Classification

Classification is termed as data analysis process that is used for extracting hidden patterns in the form of models to describe the classes and to predict the future data. The classification system is provided with the class label known as exemplars and these are said to be labels. The data mining task is to compute a model from the labeled objects and to use this model to predict the classes of unlabeled objects. A model or a classifier is built to predict the attributes that are labeled categorical. This classification process consists of two steps namely building the classifier or model and using the classifier for classifications.

During this phase, the model stored on the disk is loaded into the main memory and used to classify new unlabeled instances. The classification rules are used for classification shown in Fig. 1.4. These rules are applied on the test data. To improve the accuracy of the classifier, test data is used.

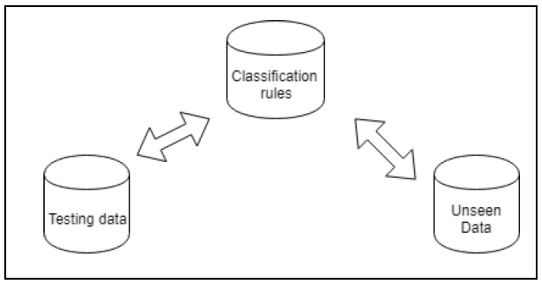


Fig.1.3(Classification)

3.2.1 Classification Model

Given a set of n labeled instances x1…xn represented by vectors v1...vn, and corresponding classes c1...cn the classification problem is to determine a function *f* such that *f(xi)=Ci*. It is also called as supervised learning process hence the classes are labeled. The classification process is a two-phase process.

3.2.2 Model construction

During this phase, the labeled examples are analyzed, a model is built and stored on disk. The learning algorithm is used to build the classifier. Using the class labels present in the training data the classifier is built shown in Fig. 1.3. The data present in the data set are also referred as instance or tuples.

3.3 Training and testing data

The classifier model is tested with the error rate which measures the performance of the classifier with the number of incorrectly classified instances. The training data which consists of the known classes is used to build the classifier and testing data which consists of unknown classes are used to test the performance of the classifier built. Validation set is also another dataset considered in the machine learning used for performance testing. The universal set is split into training and testing set.

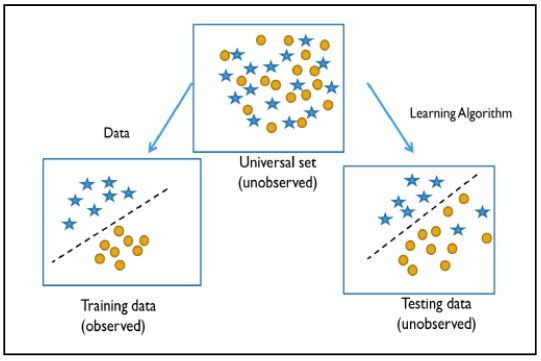


Fig.1.4(Training and Testing data)

3.4 Data set

The sample data set is considered from Machine Learning repository. It consists of data which are used by the machine learning researchers for their empirical analysis. Heart disease test data comprises of a number of tests required for the disease diagnosis process. The heart disease can be diagnosed on a patient with the result of 14 test attributes [6] [8] [36]. The attributes are real and categorical. The attributes that acts on decision are called as decision attribute. The shows the list of attributes for the Heart disease diagnosis.

3.5 Parameter setting

Parameter Values
Search Agents 500
Max. number of iterations 5–50

Fig.1.5 (Parameter Setting)

3.6 Rough set theory :

It has been applied in health care industry in the recent years. It is the useful way of finding knowledge from dataset to estimate the imprecise concept. The rough set theory is different from, since it provides the vagueness due to the lack of information on attributes present in the data set. It also acts as a mathematical tool.

The Rough set is the formal approximation of the crisp set. The set can be represented as the best case, the worst case and boundary region. Universe is called the data set.

The worst case consists of the set that are certainly contained in the set from the universal set. The best case contains the data that are possibly contained in the observed features. The boundary region consists of those objects that cannot be classified from the data set. The relevant attribute is identified by evaluating the attribute set called as reducts. It is the technique used to remove the redundant attributes which provides the effective way for prediction of diseases.

3.7 Information system

The information system is represented as a set of 5 variables.

a) BEST CASE
b) WORST CASE
c) FIT VALUE
d) ACTUAL VALUE
e) BOUNDARY VALUE

3.8 Accuracy value

The accuracy of the classifier is based on the number of instance that is correctly classified.

ACCURACY= (Number of predicted value) / (Number of value)

FORMULA:

M=(fit-worst)./(best-worst)

MASS CALCULATION:

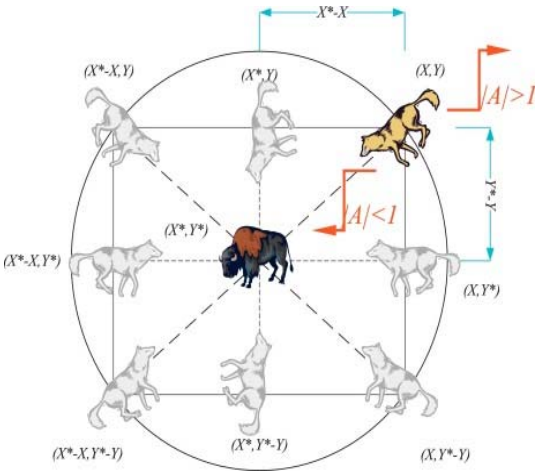
M=M./sum(M);

The third attempt is to reduce the number of attributes used for the effective diagnosis of the disease. In the medical diagnosis process, there is various number of test attributes used for the effective diagnosis of the disease. This work is proposed to reduce the number of attributes by finding the critical attributes. For the experimental purpose the heart disease and diabetics dataset is considered. Initially there are 14 attributes in which only 5 attributes are found as critical attributes. Finally, the performance analysis is compared with 14 and 5 attributes. The results depict that the performance of the classifier improves with the reduced number of attributes.

3.9 Feature selection

Feature selection includes finding the conditional entropy and relevance using the Eqs (5.5) and (5.6). The conditional entropy (CE) is the amount of information needed to describe the outcome of a random variable. The relevance (R) is used to find the relevant and the irrelevant attributes. Feature selection is also useful as part of the data analysis process. Subset selection evaluates a subset of features as a group for suitability.

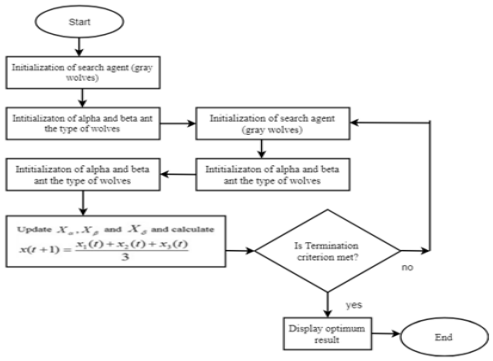
3.10 Proposed diagram & algorithm



CE= − x∈X y∈Y f (x, y) log f (y|x) (5.5) where, f(y/x) =f (y, x)/f(x)

R=ENTROPY(Ai)-CE (Ai,Aj) (5.6)

where, Ai= value of one attribute, Aj= value of another attribute



3.11 Testing Algorithm

Input: Dts-Testing Dataset

Output: Dts (Xi)

Step1: Extract testing dataset and load into database

Step2: Separate the data set with the records and attributes as in training algorithm

Step3: Select the threshold value for different diseases.

4. EXPERIMENTAL RESULTS

The experiments are conducted on the heart disease dataset. the results are compared with the 14 attributes and the 5 attributes.

4.1 Reduced List Of Attributes

The classification approach is used to find the relevant attributes and the irrelevant attributes. The critical attributes are selected based on the critical score computed.

4.2 System architecture:

Supervised learning, assumes training data that has been hand-labelled, and attempts to find prediction in the data that can then be used to determine the correct output value for new data instances.

It has recently been explored is semi-supervised learning, which uses a combination of labeled data. The piece of input data for which an output value is generated is formally termed as instance.

The instance is formally described by a vector of features, which together constitute a description of all known characteristics of the instance. Algorithms for prediction depend on the type of label output, on supervised learning and the algorithm is statistical in nature.

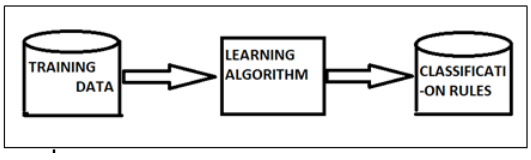


Fig.3.4(Design System)

4.3 Implementaion Detail

Our proposed model to classified the diseases and rate classification implemented by using the **Matlab software**. And also we implement our model by using the GWO (Grey Wolf Optimization) algorithm and obtain the required output

4.4 Matlab Software (R2014a)

MATLAB is a fourth-generation programming language and numerical analysis environment. Uses for **MATLAB** include matrix calculations, developing and running algorithms, creating user interfaces (UI) and data visualization.

A proprietary programming language developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, C#, **Java**, **Fortran** and **Python**.

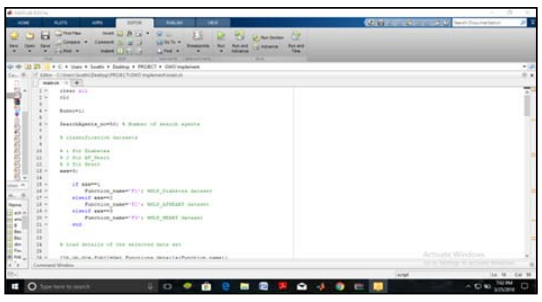
If a user types 2+2 at the matlab command prompt and presses enter, the computer replies with 4, as shown below:

```
> 2+2
[1] 4
```

5. SCREEN SHOT

5.1 Main.m - (Module 1)

Here main function deals with three functions namely F1, F2 and F3.

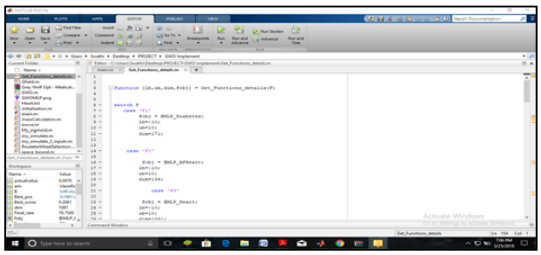


In this step the execution of the loop begins. It determines the position, best fit, worst fit.

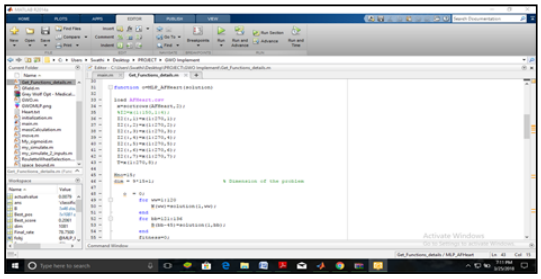
The best position and worst position has some value. Then these are compared with actual value to find the best position.

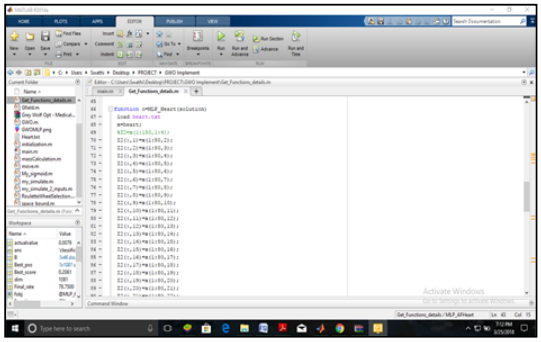
After finding the no.of proper classification we calculate the rate and thereby converting it to %.

5.2 Module 2 -get function

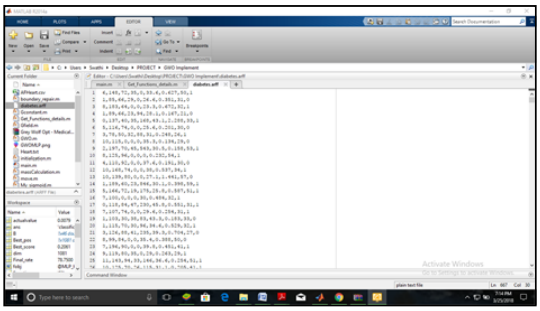


Get function holds the values for various attributes for various test cases.

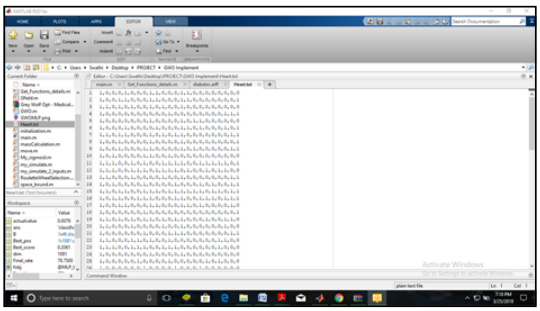




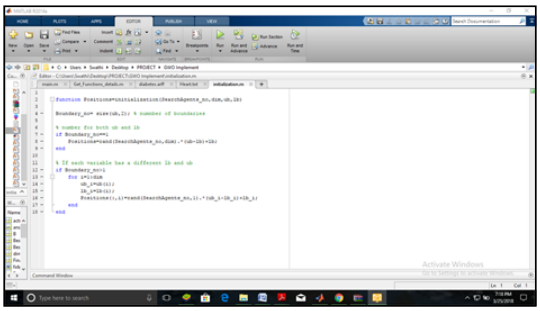
5.3 Module 3 : Dataset for diabetics



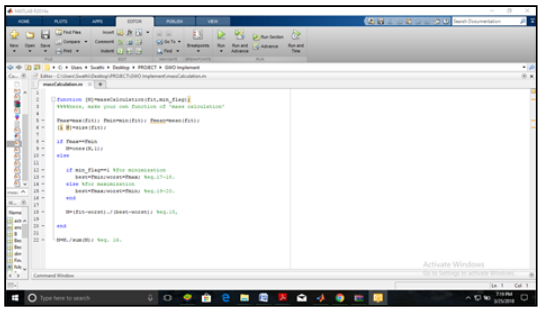
5.4 Module 5 :Dataset for Heart disease



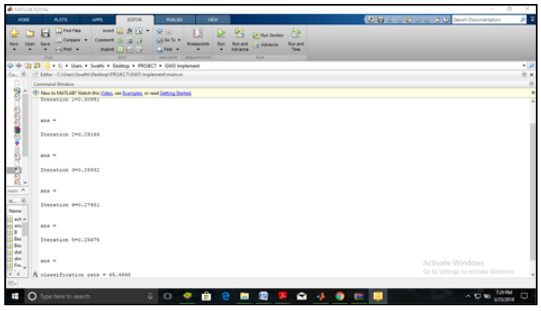
5.5 Module 6 :Initialization



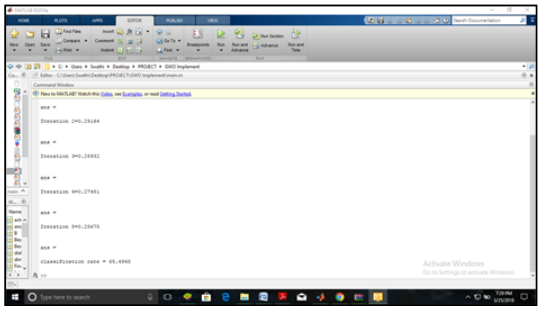
5.6 Module 7 :Mass calculation



5.7 Module 8 :(OUTPUT)



5.6 OUTPUT :



First the final rate is calculated using the formula,

Final_rate(i)=(Rrate/80)*100;

It is the final step it undergoes various iterations and determines the classification rate using the classification rate formula.

6. CONCLUSION

The performance analysis is done by comparing the results of the proposed classifiers with the existing classifiers. The experimental setup was carried with the heart disease dataset from the proposed classifiers in term of performance metrics. It is further required to reduce the number of attributes to decrease the effort and time for the classification. The learning algorithm is developed to reduce the attributes from heart disease dataset to 5 attributes. These five attributes are termed as critical attributes. Also, the comparison analysis is shown with the original and the reduced attributes. The classifiers proposed can be used for the classification and prediction on any dataset. The proposed dimensionality reduction algorithms can also be used for finding the relevant and critical attributes for any kind dataset

7. SUGGESTION FOR FURTHER RESEARCH

The classifier method discussed in this thesis showed improved accuracy compared with the existing algorithms. The classifier considered is a binary classifier where the class labels have only two decisions say yes or no. In the future work the same classification method can be used for multiclass label prediction. The high-volume dataset can be tested to improve the scalability of the classifier performance. Further it can be enhanced to perform data analytics and to develop a intelligent decision support system for any medical dataset consider.

Using the proposed attribute reduction technique, the 14 test attributes are reduced to 5 test attributes. These attributes are found relevant using the critical score computed. The performance of the classifiers is compared with the original attributes and the reduced attributes. The accuracy of the proposed classifier produced better results compared with the existing classification algorithms. Also, the comparison with other existing classifications methods on the original attributes with the reduced attributes are discussed

REFERENCES

[1] N. Li, G. Li, Z.L. Deng, An improved sine cosine algorithm based on levy flight, in: Proc. SPIE 10420, Ninth International Conference on Digital Image Processing (ICDIP 2017).

[2] O.E. Turgut, Thermal and economical optimization of a shell and tube evaporator using hybrid backtracking search-sine-cosine algorithm, Arabian Journal of Science and Engineering, vol. , no. ,pp. 2017.

[3] Amudhavel, J., Rao, D.N., Sathian, D., Dhavachelvan, P., Raghav, R.S., Prem Kumar, K., "Big data scalability, methods and its implications: A survey of current practice"", (2015) ACM International Conference Proceeding Series, 06-07-March-2015

[4] Amudhavel, J., Prabu, U., Dhavachelvan, P., Moganarangan, N., Ravishankar, V., Baskaran, R., "Non-homogeneous hidden Markov model approach for load balancing in web server farms (NH2M2-WSF)", (2015) Global Conference on Communication Technologies, GCCT 2015

[5] Amudhavel, J., Prabu, U., Dhavachelvan, P., Moganarangan, N., Ravishankar, V., Baskaran, R., "A comprehensive analysis and performance assessment on QoS aware web service selection algorithms", (2015) Global Conference on Communication Technologies, GCCT 2015

[6] Baskaran, R., Basha, M.S.S., Amudhavel, J., Kumar, K.P., Kumar, D.A., Vijayakumar, V., "A bio-inspired artificial bee colony approach for dynamic independent connectivity patterns in VANET", (2015) IEEE International Conference on Circuit, Power and Computing Technologies, ICCPCT 2015

[7] Amudhavel, J., Lakhani, S.H., Rajaguru, D., Vengattaraman, T., Sampath Kumar, S., Prem Kumar, K., "A chaotic Krill Herd optimization approach in VANET for congestion free effective multi hop communication", (2015) ACM

[8] Amudhavel, J., Vigneshwaran, R., Janakiram, A., Jarina, S., Prem Kumar, K., Anantharaj, B., Sathian, D., "An empirical analysis on Quality of Service(QoS) in cloud computing", (2016) Indian Journal of Science and Technology, 9 (22)

[9] Amudhavel, J., Jayabharathi, A., Kumarakrishnan, S., Malarvizhi, M., Gomathy, H., Prem Kumar, K., "An scalable bandwidth reduction and optimization in Smart Phone Ad Hoc Network (SPAN) using Krill Herd Algorithm", (2015) ACM International Conference Proceeding Series,06-07-March-2015

[10] Karthikeyan, P., Sathian, D., Amudhavel, J., Raghav, R.S., Abraham, A., Dhavachelvan, P., "A comprehensive survey on variants and its extensions of BIG DATA in cloud environment", (2015) ACM International Conference Proceeding Series, 06-07-March-2015.

[11] Jaiganesh, S., Jarina, S., Amudhavel, J., Premkumar, K., Sampathkumar, S., Vengattaraman, T., "Performance analysis of collision avoidance frame works in VANETS", (2016) Indian Journal of Science and Technology, 9 (11).

[12] Padmapriya, V., Amudhavel, J., Gowri, V., Lakshmipriya, K., Vinothini, S., PremKumar, K., "Demystifying challenges, opportunities and issues of Big data frameworks", (2015) ACM International Conference Proceeding Series, 06-07-March-2015

[13] [13] Amudhavel, J., Padmashree, D., Kumarakrishnan, S., Harinee, S., Anantharaj, B., Prem Kumar, K., "A novel bio-inspired krill herd optimization in wireless ad-hoc network (WANET) for effective routing", (2015) ACM International Conference Proceeding Series, 06-07-March-2015

[14] Vijayakumar, V., Inbavalli, P., Joseph, K.S., Amudhavel, J., Rajaguru, D., Kumar, S.S., Vengattaraman, T., Premkumar, K., "Research on QoS aware dynamic reconfiguration and performance measures in VANET", (2015) Global Conference on Communication Technologies, GCCT 2015