International Journal of Scientific and Technical Advancements
IJSTA
ISSN: 2454-1532

Feature Selection Using Genetic Algorithm on Breast Cancer Dataset

Pooja Devi[1], Namrata Karlupia[2], Pawanesh Abrol[3]
[1, 2,3]Department of Computer Science and IT
University of Jammu, Jammu
J&K, India-180006
[1]bajgotra.pooja@gmail.com,
[2] namrataphonsa@gmail.com,
[3] pawanesh.abrol@gmail.com

***Abstract*—** Breast cancer is the most life dangerous disease among women. In Breast cancer the cells in the breast grow out of control. Cancer detection is still a major research challenge. In this work, a wrapper-based approach using genetic algorithm has been used for selecting relevant features to detect breast cancer. The aim of the proposed work is to select the relevant features using genetic algorithm and apply ANN for classification to obtained best accuracy. To evaluate the performance of features selected by Genetic Algorithm (GA) artificial neural network (ANN) classifier and breast cancer dataset from UCI machine learning database has been used. The classification error of ANN has been used to design the objective function. An experiment has been performed to analyze the effect of iterations and population size on the accuracy of ANN and best features have been selected. The best accuracy achieved by using ANN for Breast cancer detection for optimized feature subset is 90.5%.

***Keywords*—** Artificial neural network, Breast cancer, Genetic Algorithm, Wrapper method

I. INTRODUCTION

Optimization is the process of finding the best value for the variables of a problem from all the possible solution [1]. The conventional technique can only find local optimum and trouble in solving discrete optimization problems. Recently, Nature inspired algorithms have been attracting considerable attention for their good performance. These are the algorithm which inspired by nature i.e. how nature accommodates the challenging circumstances. Nature-inspired algorithms are a set of novel problem-solving approaches and can be used to produce a nearly optimal solution in very less time [2]. Some of the important nature-inspired algorithms which are used for optimization are genetic algorithm (GA), cuckoo search(CS), particle swarm optimization(PSO), artificial neural network(ANN), ant colony optimization, Artificial bee colony optimization, Fish swarm algorithm etc. Genetic algorithm is an evolutionary algorithm. It is based on the Principle of Charles Darwin's theory of survival of the fitness [3]. Particle Swarm optimization is based on swarm intelligence and it is inspired by the swarm behavior of birds searching their foods. A few real life problems are being solved using nature-inspired algorithms. Feature selection is an optimization problem of finding the best solutions from a given set of features [4]. It is a procedure to reduce the irrelevant, useless and redundant features that don't contribute in increasing the accuracy of the predictive model. Feature selection methods can be categorized as: Filter Method, Wrapper Method and Embedded Method. In filter method the selection of features is independent of the classifier used. The Wrapper Method is that method in which selection is reliant on the classifier used. Embedded Method include variable selection as a part of the training process without a split the data into training and testing sets. Nowadays, nature-inspired algorithms like GA, PSO, ACO and CSA are being used to select best feature subset. Nature-inspired algorithms are highly efficient and try to find the best solution in a large search space. These algorithms are relevant for difficult optimization problems [5].Genetic algorithm has been attesting to find a better solution in a great diversity around optimum search problems. GA can be easily parallelized in computer clusters [6]. For optimization of features using GA a proper objective function is needed to be design. The main aim of the proposed research is to select the most relevant feature and to apply ANN classifier for classification. To select the features genetic algorithm has been applied and for classification ANN classifier has been used.

The paper is divided in VI different sections. Section II provides a detailed literature review for feature selection methods. Section III provides methods used for feature selection. Section IV discuss the proposed work. Experimental results are presented in Section V and Section VI concludes this paper finally.

II. LITERATURE REVIEW

The literature related to the proposed topic has been studied and a general review of the same is presented as under:

Dr. S. Mary *et al.* [7] features of Brain MRI Scan Images were selected using a genetic algorithm and classification with the help of a random forest classifier has been done. Babatunde

Pooja Devi, Namrata Karlupia and Pawanesh Abrol, "Feature selection using genetic algorithm on breast cancer dataset," *International Journal of Scientific and Technical Advancements*, Volume 7, Issue 2, pp. 1-7, 2021.

Oluleye *et al*. [8] implemented a binary genetic algorithm on the Flavia dataset for selecting the various feature and also applied WEKA software for the comparison. Dongkoo Shon *et al.* [9] worked on EEG signal by using GA and PCA for feature selection with a KNN classifier to recognize whether each EEG data point represents stress state or not has been presented in [9]. Four steps were used for stress classification on public EEG data. Three different experiments for stress analysis were carried out i.e K-NN without feature selection; K-NN with GA- based feature selection and K-NN with PCA - based feature selection. The GA-based feature selection shows better results as compared to others. Tariq Ali *et al.* [10] implemented a Genetic Algorithm with a K-NN classifier on two datasets for eye state classification i.e. EEG eyes, Eye data from epileptic seizure recognition. EEG Eyes with 14 attributes 14977 instances and 2 classes. Eye data from epileptic seizure recognition with 178 attributes 4601 instances and 2 classes. WEKA software was used for comparison. He compared the accuracy results without feature selection and the accuracy obtained with feature selection. The GA based shows better results as compared to WEKA and without feature selection. In his work [11] an experiment was conducted for feature extraction using a 3D transform-based method and for feature selection using a genetic algorithm with different classifiers on CUAVE and TULIPS database. The results were compared with WEKA Software and obtained an accuracy of 74%.Hivi I. Dino *et al*. [12] disscussed the Facial Expression Recognition dataset with three various feature selection methods like correlation feature selection, gain ratio, information gain was used for resolving the most recognized features of face images using a multi-classification algorithm. R. Kuppuchamy *et al.* [13] was applied information gain, correlation-based feature selection on Pima Indians diabetes dataset with C4.5 classifier and achieved better accuracy.

Statistical feature selection approach for classification of emotions from the speech was proposed on surrey audiovisual expressed emotion dataset with 480 utterances. Three models were designed based on the normality test, non-normality test and PCA approach with SVM, K-NN, RF, NN classifiers [14]. Dorsewamy *et al.* [15] applied K-NN, LR, SVM, RF and NB classifiers to predict household food insecurity with wrapper, filter and embedded based feature selection methods. With these five classifiers, the entire dataset, the best ten, seven best and five best were evaluated, and then results were compared with each other. The seven features and random forest were generated the best results. Paper [16] presented a comparison of six machine learning algorithms on the Wisconsin diagnostic breast cancer dataset to predict and classify cancer as benign or malignant. Mithal Doshi *et al.* [17] conducted an experiment to predict the performance of students with 380 instances and 32 attributes from Mumbai College. Chi-square, Info Gain, Gain Ratio feature selection attribute algorithms were applied with NB tree, Multilayer perception, NB and instance-based K-NN to predict the performance of the student. Karthik Sekaram *et al.* [18] worked on Wisconsin diagnostic breast cancer dataset with a deep neural network for classification and RFE for the feature selection. Wisconsin diagnostic breast cancer was taken from the UCI repository with 699 instances and 9 attributes and 2 class labels called benign and malignant. Accuracy, sensitivity, specificity, precision and recall were measured on the performance of the system. Muhammad Aqeel Aslam *et al.* [19] applied a Deep convolutional neural network for the detection of breast cancer on two WDBC datasets. The first dataset contains 699 samples, 11 attributes, 16 missing values and these missing values were discarded from samples automatically the actual number of samples for the second dataset was 683 taken from the UCI repository. Sukhchain Singh *et al.* [20] developed a deep neural network for the classification to detect fungal disease in the grapes. GA was used for image segmentation and obtained an accuracy of 97.7%. S. Poorani *et al.* [21] presented different feature selection methods with different classifiers like logistic regression, NB, SVM, K-NN were applied on breast cancer dataset with 80 features to classify cancer as benign or malignant. Comparing these algorithms with each other and obtained accuracy. Logistic regression has the highest accuracy with 96.9%.

TABLE I: Review of various nature inspired algorithms used for feature selection

S. No.	NATURE INSPIRED ALGORITHM	CLASSIFIER	REMARKS
1.	A Binary Grasshopper Optimization Algorithm[22]	Decision Tree	Presented a binary grasshopper optimization algorithm for feature selection on diabetes dataset taken from UCI repository with decision tree classifier for classification.
2.	Genetic Algorithm [23]	Naïve Bayes, Decision Tree	Proposed GA for feature selection to predict the performance of heart disease with NB, decision tree classifier and obtained accuracy. The best accuracy obtained by using these two classifiers was 99.05%
3.	Genetic Algorithm [24]	KNN, SVM, Random Forest	Suggested a model which detects the disease of plants from the leaves. SVM was used on the basis of accuracy, precision, Recall
4.	Particle Swarm Optimization [25]	Support Vector Machine	Particle swarm optimization used for feature search which was applied to find optimal feature space, which can enhance the performance of text classification
5.	Genetic Algorithm [26]	KNN , Naïve Bayes	Applied GA based wrapper feature selection technique on medical dataset with NB, K-NN and J48 supervised learning algorithms to calculate the accuracy in the iteration manner.

Pooja Devi, Namrata Karlupia and Pawanesh Abrol, "Feature selection using genetic algorithm on breast cancer dataset," *International Journal of Scientific and Technical Advancements*, Volume 7, Issue 2, pp. 1-7, 2021.

ISSN: 2454-1532

A comprehensive review has been given in which different techniques for classification and selection of features have been discussed. Thus, the literature review shows that several feature selection techniques are there but nowadays nature-Inspired algorithms like PSO, CSA, ACO, GA etc. are widely used by researchers and showing better results. So, for proposed work GA has been selected.

III. METHODS

A. *Genetic algorithm*

Genetic Algorithm (GA) is a search-based optimization method based on the principles of 'Genetics and Natural Selection. It is used to find out such values of input so that we get the best output values of results. A Genetic algorithm is a method of natural selection where the fittest persons are selected for reproduction to generate offspring for the next generation. It is basically used as a problem-solving technique to provide an optimal solution.

Steps:

1. Choose a coding to indicate downside parameters, a selection operator, a crossover operator, mutation operator, population size, crossover chance and mutation chance.
2. Indiscriminately initialize population of strings of size l, t_{max} set t = 0.
3. Calculate all string in population.
4. If $t > t_{max}$ or alternate execution criterion is consummated, terminate
5. Perform replica on the population.
6. Execute crossover on hit and miss combine of string
7. Perform mutation on each string.
8. Calculate strings within the novel population. Set t = t+1 and held to step three.

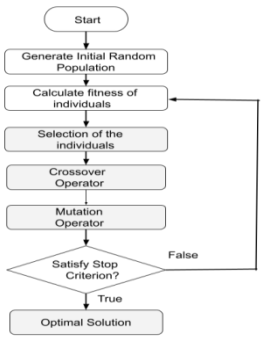


Fig.1. Flowchart of genetic algorithm

Algorithm 1 shows the various steps and operation performed in GA.

Algorithm1:

Step1*: Randomly generate the initial population of n strings (chromosomes)*

Step2*: Evaluate the fitness of each string in the population*

Step3: *Repeat the following steps until the next generation of n individual string produced*

- *a. Select pair of parent chromosomes from the current population according to their fitness i.e., higher fitness individuals are selected more often*
- *b. Apply crossover*
- ***c.*** *Apply mutation*

Step4: *Go to step 2 if termination condition met*
Else end

A. *Artifical neural network*

The artificial neural network is an information processing model that is inspired by the way of natural neurons system.

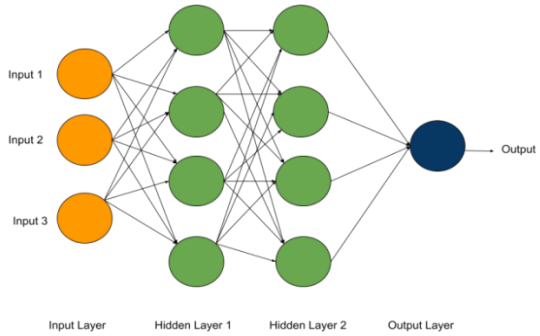


Fig.2. Artifical neural network

ANN is similar to the human brain. It can perform tasks like classification, prediction, pattern recognition. Every neuron is linked with another neuron through a connection link as shown in Fig. 2. There is a weight that is connected with a connection link that has information about the input signals. Weight is the most important information for neurons to solve a particular problem. Every neuron has an input layer, hidden layers, and one output layer. The neurons can take input data and performs some operations on that data and then the results of that data go to the output layer in the form of output Hidden layers are present between the input layer and the output layer and that layer performs all the operations to find hidden features or patterns.

Pooja Devi, Namrata Karlupia and Pawanesh Abrol, "Feature selection using genetic algorithm on breast cancer dataset," *International Journal of Scientific and Technical Advancements*, Volume 7, Issue 2, pp. 1-7, 2021.

IV. PROPOSED METHODOLOGY

The proposed system aims to select the features which give the best accuracy and reduced that features that are not relevant or not important for research. First, step is to generate a random initial population for GA that consists of string of 0's and 1's. In each generation of GA, the selection, crossover and mutation steps are carried out. The fitness function for each chromosome is calculated by using ANN classifier. The chromosomes with best fitness function value are kept for next generation. The process is carried repeatedly till the desired results or an evaluation criteria does not met.

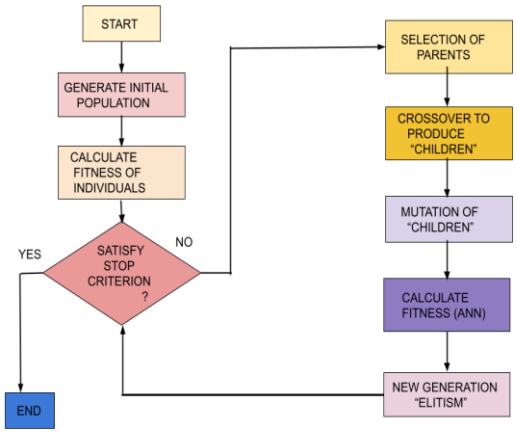


Fig.3. Proposed methodology for feature selection

The Breast cancer Coimbra-dataset used in the proposed system was taken from the UCI library [27]. The Breast cancer Coimbra dataset has 10 features and 116 instances. Among 116 instances 52 instances are about healthy patients and the remaining 64 are about breast cancer patients. Among ten features nine are autonomous variables i.e. age, BMI, Glucose, Insulin, HOMA, Leptin, Adiponectin, Resistin, MCP-1, and the 10th variable is the dependent variable which has the label 2express the presence of disease, and label 1 express the absence of the disease. The detail description of the dataset is shown in TABLE II.

TABLE II. Dataset description

Attribute Characteristics	Integer
Number of Instances	116
Number of Attributes	10
Number of classes	2
Missing value	N/A

To evaluate the performance of GA an objective function is also designed based on ANN classifier and also parameters like specificity, sensitivity and accuracy are computed. So, the objective is to minimize the error generated by ANN. The equation 1 shows the objective function designed for the propo sed work.

$$err = \frac{80}{100} Train_e + \frac{20}{100} Test_e \quad \text{.....(1)}$$

Where, $Train_e$represents the training error and $Test_e$represents the testing error generated by ANN. Therefore, total error generated by testing and training of ANN is taken as objective function. These metrics can be calculated using the formula given in equations (2),(3) ,(4) and (5).

Accuracy = (TP+TN)/(TP +TN + FP +FN)…(2)
Precision = TP / (TP + FP) …(3)
Sensitivity = TP /(TP +FN) …(4)
Specificity =TN /(TN + FP) …(5)

TABLE III shows the GA parameters like population size, mutation rate, type of selection etc. and their initialization.

TABLE III. GA Parameter initialization

GA Parameter	Value
Population size	10
Maximum number of Iterations	50
Crossover Percentage	0.7
Mutation Percentage	0.1

V. RESULTS AND DISCUSSION

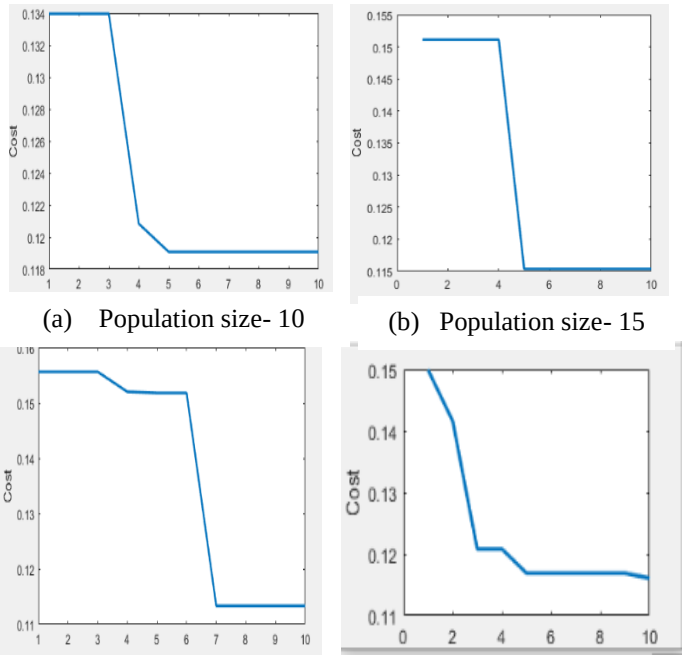
For the analysis, two experiments have been performed. In first experiment the effect of population size on objective function is analyzed. TABLE IV shows s the variation of objective function at different population size. The accuracy achieved when Population size varies from 10 to 30.

TABLE IV. Effect of change in Population size on objective function

Iteration	Population Size	Accuracy	Objective function value
10	10	88.0%	0.092
10	15	88.4%	0.063
10	20	88.6%	0.068
10	30	88.3%	0.095

Pooja Devi, Namrata Karlupia and Pawanesh Abrol, "Feature selection using genetic algorithm on breast cancer dataset," *International Journal of Scientific and Technical Advancements*, Volume 7, Issue 2, pp. 1-7, 2021.

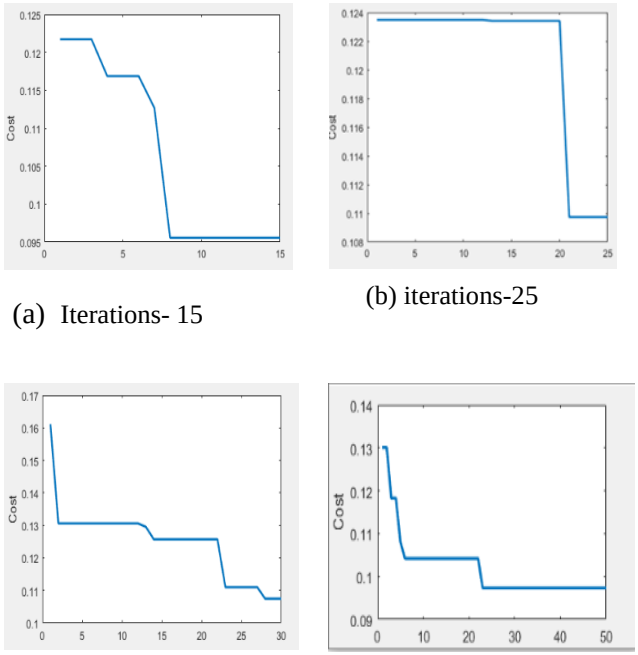
International Journal of Scientific and Technical Advancements
ISSN: 2454-1532



(a) Population size- 10 (b) Population size- 15

(c) Population size- 20 (d) Population size- 30

Fig.4. Effect of change in population size on objective function



(a) Iterations- 15 (b) iterations-25

(c) Iterations- 30 (d) Iterations- 50

Fig.5. Effect of change in iterations on objective function

By changing the population size, it is analyzed that accuracy does not increase after 30 population size and the highest accuracy attained is 88.6%. Another experiment has been conducted in which effect of change in iterations on objective function is studied.

TABLE V . Effect of change in Iterations on objective function

Iteration	Population size	Accuracy	Objective function Value
15	10	88.%	0.092
25	10	89.%	0.027
30	10	89.%	0.056
50	10	90.5%	0.066

TABLE V shows the accuracy achieved when number of iterations are increased from 15 to 50. It is observed that graph starts converging after about 30 iterations.

The most relevant 5 features are selected out of 9 features. The selected features are Age, BMI, Glucose, Adiponection and Resistin. Fig.6 shows the confusion matrix which is used to measure the performance of an algorithm. TP are those cases which correctly get classified as true and are false. TN are those cases which correctly get classified as false and are false.

FP describes those cases which wrongly get true but are false. FN are those cases which wrongly get classified as false but are true.

Accuracy: It can be defined as the number of properly classified patterns to the total number of samples. The Formula for accuracy measurement as

$$A = \frac{TP + TN}{TP+TN+FP+FN}$$

Precision: The number of positive samples taking out from the total number of samples acknowledge positive by the classification model.

$$P = \frac{TP}{TP+FP}$$

Recall: The number of true positive patterns taking out from the total number of positive declared patterns.

Pooja Devi, Namrata Karlupia and Pawanesh Abrol, "Feature selection using genetic algorithm on breast cancer dataset," *International Journal of Scientific and Technical Advancements*, Volume 7, Issue 2, pp. 1-7, 2021.

Sensitivity: It is the ability of a diagnosis test to identify true positive cases. It measures the performance of correctly identified positives.

Specificity: It is the ability of a diagnostic test to identify true negative cases. Specificity measures the proportion of negatives that are correctly identified.

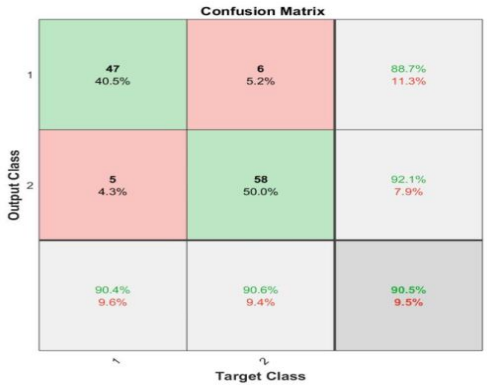


Fig.6. Confusion matrix

TABLE VI shows the final results attained using selected feature set in terms of accuracy, sensitivity, precision is 90.5%, 92.1%, 90.4% respectively

TABLE VI. Performance evaluation results

Selected features	Accuracy	Sensitivity	specificity	Precision
5	90.5	88.7	92.1	90.4

Experimental analysis shows that for population size 10 and for number of iterations 30, the objective function starts converging and gives best accuracy 90.5%.

VI. CONCLUSION

Breast cancer detected at an early stage will help to save the lives of thousands of women. The ANN classifier provides better accuracy for small dataset. In the proposed work, GA has been used for feature selection and ANN classifier implemented to compute the accuracy. Effect of population size and effect of iterations on the accuracy have been analyzed. Confusion matrix obtained shows the best accuracy attained by using GA on breast cancer dataset is 90.5%. In the Breast cancer Coimbra dataset, the 10^{th} variable is the dependent variable which has label 2 to represent the presence of disease and label 1 to represent the absence of the disease.

REFERENCES

[1] N. Karlupia and P. Abrol, "Design and Analysis of Optical Patterns using Bacterial Foraging Algorithm for Optimized Illumination," *Indian Journal of Computer Science and Engineering (IJCSE),* vol. 8, issue 3, 2017.

[2] Vinoothna Manohar Botcha, Gudibandi Monitha, Devi Navya Sri Madala and Bhanu Prakash Kolla, "Analysis of Nature Inspired Algorithms," *Journal of Critical Reviews,* vol. 7, issue 4, pp. 1-3, 2020.

[3] Binitha S and S Siva Sathya, "A Survey of Bio-inspired Optimization Algorithm," *International Journal of Soft Computing and Engineering (IJSCE),* vol. 2, issue 2, pp. 1-15 2012.

[4] Dr. K. James Mathai and Kshiti Agnihotri, "Optimization techniques for feature selection in classification," *International Journal of Engineering Development and Research,* vol. 5, issue 3, pp. 1-4, 2017.

[5] N. Karlupia, P. Sambyal, P. Abrol and P. Lehana, "BFO and GA based Optimization of Illumination Switching Patterns in Large Establishments," *2019 6th International Conference on Computing for Sustainable Global Development (INDIACom),* pp. 349-354, 2019.

[6] Shaohua Wu, Yong Hu, Wei Wang, Xinyong Fengand Wanneng Shu, "Application of Global Optimization Methods for feature selection and Machine learning," *Hindawi Publishing Corporation Mathematical Problems in Engineering*, pp. 1-9, 2013.

[7] Dr. S. Mary Joans and J. Sandhiya, "A Genetic Algorithm Based Feature Selection for Brain MRI Scan Images using Genetic algorithm," International Journal of Advanced Engineering Research and Science (IJAERS), vol. 4, issue 5, pp. 124-130.

[8] Babatunde Oluleye, Armstrong Lesia, Diepeveen Dean and Jinsong Leng, "A Genetic Algorithm-Based Feature Selection," International Journal of Electronics Communication and Computer Engineering, vol. 5, issue 4, pp. 899-905, 2017.

[9] Dongkoo Shon, KichangIm, Jeong-Ho Park, Dong-Sun Lim and Byungtae Jang, "Emotional Stress State Detection Using Genetic Algorithm-Based feature selection on EEG signals*," International Journal of Environmental Research and Public Health*, pp. 3-11, 2018.

[10] Tariq Ali, Asif Nawaz and Hafiza Ayesha Sadia, "Genetic Algorithm Based Feature Selection technique for Electroencephalography data," *Applied Computer Systems*, pp. 199-127, 2019.

[11] Sunil Sudam Morade and Suprava Patnaik, "A Genetic Algorithm-Based 3D Feature Selection for Lip Reading," *International Conference on Pervasive Computing (ICPC)*, 2015.

[12] Hivi I. Dino and Maiwan B. Abdulrazzaq, "A Comparison of Four Classification Algorithms for Facial Expression Recognization," *Polytechnic Journal*, vol. 10, pp. 74-80, 2020.

[13] R. Kuppuchamy, T. Kamalavalli, S. Vinothini, N. Jayalakshmi and N. Vallileka, "Correlation based Ensemble Feature Selection Algorithm for Diagnosis of Diabetics," *International Journal of Innovative Technology and Exploring Engineering (IJITEE),* vol. 9 , pp. 373-377, 2020.

[14] Nilima Salankar and Anjali Mishra, "Statistical Feature Selection Approach for Classification of Emotions from speech," *NGCT and University of Petroleum and Energy Studies (UPES),* pp. 1-13, 2019.

Pooja Devi, Namrata Karlupia and Pawanesh Abrol, "Feature selection using genetic algorithm on breast cancer dataset," *International Journal of Scientific and Technical Advancements*, Volume 7, Issue 2, pp. 1-7, 2021.

ISSN: 2454-1532

[15] Dorsewamy and Mersha Nigus, "Feature Selection Methods for Predicting household food insecurity," *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 9, issue 1, pp. 1560-1568, 2020.

[16] Nikita Rane, Rucha Kanade, Jean Sunny and Prof. Sulochana Devi, "Breast Cancer Classification and Prediction using Machine Learning," *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, issue 2, pp. 576-580, 2020.

[17] Mital Doshi and Dr. Setu K Chaturvedi, "Correlation based feture selection technique to predict student performance," *International Journal of Computer Networks & Communications (IJCNC)*, vol. 6, No. 3, pp. 197-206, 2014.

[18] Karthik Sekaram, Srinivasa Perumal and P.V.S.S.R. Chandra Mouli, "Breast Cancer Classification using Deep Neural Networks," *Knowledge Computing and its Application*, pp. 227-241, 2018.

[19] Muhammad Aqeel Aslam, Aslam and Daxing Cui, "Breast Cancer Classification using Deep Convolutional Neural Network," *Journal of Physics*, 2020.

[20] Sukhchain Singh and Er. Rachna Rajput, "Implementation Paper to detect and classification of Fungal Disease in Grapes leaves using the Genetic Algorithm," International journal of advanced research in science engineering, vol. 6, issue 1, pp. 1-16, 2017.

21] S.Poorani and P. Balasubramanie, "Deep Neural Network Classifier in Breast Cancer Prediction," *International journal of Engineering and advanced Technology(IJEAT)*, vol. 9, issue 1, pp. 2106-2109, 2019.

[22] Reyhaneh Yaghobzadeh, Seyyed Reza Kamel, Mojtaba Asgari and Hassan Saadatmand, "A Binary Grasshopper Optimization Algorithm for feature selection," *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, issue 3, pp. 533-540, 2020.

[23] M. ANBARASI, E. ANUPRIYA and N.CH.S.N.IYENGAR, "Enchanced prediction of Heart Disease with feature selectionusing GeneticAlgorithm," *International Journal of Engineering Science and Technology*, pp. 1-7, 2010.

[24] Shubham Gupta, Jasbir Singh, "Implementation of Ensemble Classifier for Plant Disease Detection," *International Research Journal of Engineering and Technology (IRJET)*, pp. 1-7, 2020.

[25] Yong Liu, ShenggenJu, Junfeng Wang and Chong Su, "A New Feature Selection Method for Text Classification Based on Independent Feature Space Search," *Hindawi*, pp. 1-14, 2020.

[26] D. Asir Antony Gnana Singh, E. Jebamalar Leavline, R. Priyanka and P. Padma Priya, "Dimensionality Reduction using Genetic algorithm for improving accuracy in medical Diagnosis," *International journal Intelligent Systems and Application*, pp. 67-73, 2016.

[27] S.Poorani and P. Balasubramanie, "Deep Neural Network Classifier in Breast Cancer Prediction," *International journal of Engineering and advanced Technology(IJEAT)*, vol. 9, issue 1, pp. 2106-2109, 2019.

Pooja Devi, Namrata Karlupia and Pawanesh Abrol, "Feature selection using genetic algorithm on breast cancer dataset," *International Journal of Scientific and Technical Advancements*, Volume 7, Issue 2, pp. 1-7, 2021.