Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 6 July 2021: 5465 – 5476

Research Article

Genetic Algorithm based Data Analysis for Fuzzy Extreme Learning Machine

Archana Kale
Associate Professor
Modern Education Society's College of Engineering Pune
India
archana.mahantakale@mescoepunr.org

ABSTRACT

The crucial objective of this paper is to focus on the research work which aims is to design a hybrid model of the genetic algorithm for fuzzy extreme learning machine classifier (GA-FELM), which selects an optimal feature subset by using the multilevel parameter optimization technique. Data analysis is an important task in pattern classification and knowledge discovery problems. The generalization performance of the system is not only depending on optimal features but also dependent upon the classifier (learning algorithm). Therefore, it is an important task to select a fast and efficient classifier. Research efforts have affirmed that extreme learning machine (ELM) has superior and accurate classification ability. However, ELM failed to handle the uncertain data and weighted classification problem. One of the alternative solutions is fuzzy – ELM, which combines the advantages of fuzzy logic and ELM.

GA-FELM is able to handle curse of dimensionality problem, optimization problem and weighted classification problem with maximizing classification accuracy by minimizing the number of features. In order to validate the performance of GA-FELM, the comparison is made with three approaches viz. 1. ELM and GA-ELM 2. GA-ELM and GA-FELM 3. GA-FELM and GA-Existing classifier. The comparative analysis shows that classification accuracy is improved with 9% by reducing 62% features.

KEYWORDS

Pattern Classification Problem, Data analysis, Extreme Learning Machine, Fuzzy Extreme Learning Machine

1 INTRODUCTION

Data analysis is an intricate process in the fields of knowledge discovery and data mining. The key objective of Data analysis or feature subset selection (FSS) is to provide the same or improved classification accuracy with a minimum number of relevant and non-redundant features instead of using all features. It is very intricate to decide the importance of and hence requirement of features without any prior information [1]. Hence, a large number of features are usually included in the input dataset, which contains all types of features like relevant, irrelevant, bad and redundant etc. Perhaps, only relevant and non-redundant features are required for classification and also improve the system generalization performance in the absence of irrelevant and redundant features. However, in many real-time applications, it may be possible that the redundant or irrelevant features may become relevant while functioning jointly with other features, which makes it one of the most critical tasks to appropriately discriminate these features [2].

An optimal feature subset [3] has the ability to collect corresponding important features. One of the most important optimization techniques is the genetic algorithm (GA) which has exploitation and exploration search characteristics. GA is specifically used to calculate the analytical solution for multi-criteria optimization problem [4]. GA is able to handle the big search spaces effectively [5] and has maximum chances of a global optimal solution. However, it faces several difficulties in using this approach in practice. The main reason is the use of conventional neural network classifiers that have local minima problem, over fitting problem etc. When the number of neurons is more than the required then the network faces with overfitting problem. And in the opposite case, if the number of neurons in the neural network is less than required, then the classifier will unable to find the target classification function which leads to poor generalization performance. Though the best optimal feature subset is used, the system degrades its performance due to the use of the poor performance classifier. Therefore, in this paper ELM is used which has established a very good performance in terms of training time, compact network size and simplification.

The main contribution presented in this paper is to design a hybrid model to select an optimal feature subset by using GA for FELM classifier. As per the literature survey, this is the unique attempt which proposes an integrated approach of GA and F-ELM with improved classification accuracy by using multilevel optimal feature subset in which only relevant and non-redundant features (minimum number of features) are present.

The various sections of the paper are organized as follows. Section 2 defines F-ELM with ELM concepts. Section 3 explains FSS algorithms and related approaches with various search strategies. Section 4 discusses the methodology of the proposed hybrid model. Experimental results with comparative performance are described in section 5. Finally, conclusion and future scope are listed in Section 6.

2 FUZZY EXTREME LEARNING MACHINE

ELM is one of fastest learning algorithm which is used for SLFNs [14]. ELM has various advantages over Back Propagation algorithm and SVM in terms of speed, reliability and generalization.

The random search [15] between the input and hidden layer is computationally efficient in SLFN as no further tuning is required. However, ELM [16-21] is unable to handle the uncertain dataset and the weighted classification problems. The conventional ELM lacks the ability to resolve those problems. One of the alternative solutions to solve those problems is F-ELM.

F-ELM is the hybrid approach of fuzzy logic and extreme learning machine. F-ELM [22] is able to handle the uncertainty data and also handle the weighted classification problem [23]. As in FNN, the whole network is needed to be tuned but in F-ELM the weights between input layer neurons to hidden layer neurons are randomly assigned and the neurons from hidden layer to output layer are analytically tuned. Hence, it is a fast learning algorithm as compared to the FNN.

3 FEATURE SUBSET SELECTION (FSS)

The performances of the FSS algorithm mainly depend upon the approach and hence search strategy to be used. The various approaches like filter, wrapper and hybrid are available with different search strategies like heuristic search, exhaustive search and randomized search. Filter approach is totally independent of the learning algorithm or classifier. It uses information, dependency, distance and consistency criteria to rank the features [24]. And according to the feature rank order, the features are selected as a feature subset.

Wrapper approach depends on the classifier and its evaluated predictive accuracy. Filter approach specifically used for data mining where wrapper approach mainly used for machine learning. A Hybrid approach is one which combines the advantages of filter and wrapper approach. For wrapper and hybrid approach, the search strategy is required for selecting feature subset. Focus algorithm (Almuallim and Dietterich, 1994) is an example of the exhaustive search. Sequential Forward Search (SFS) or Sequential Backward Search (SBS) are the examples of the heuristic search. In SFS algorithm the subset is created by inserting one by one feature from the feature rank order. Accordingly, the total number feature subset which is equal to the number of features. However, it is difficult to use the heuristic search method for the high dimensional dataset. The random search strategy is suitable for such large attributed datasets. One of the examples of the random search strategy is GA which returns an optimal feature subset without creating the order of features which has proved as principally smart approach for multi-criteria optimization [25].

4 METHODOLOGY

In this section, the working of a multilevel hybrid model for GA-FELM is explained. Figure 2 shows the information flow architecture of multilevel hybrid model for GA-FELM. The architecture is divided mainly into four parts – preprocessing, FSS by using GA, fuzzificaion and classification. which are described in detail below.

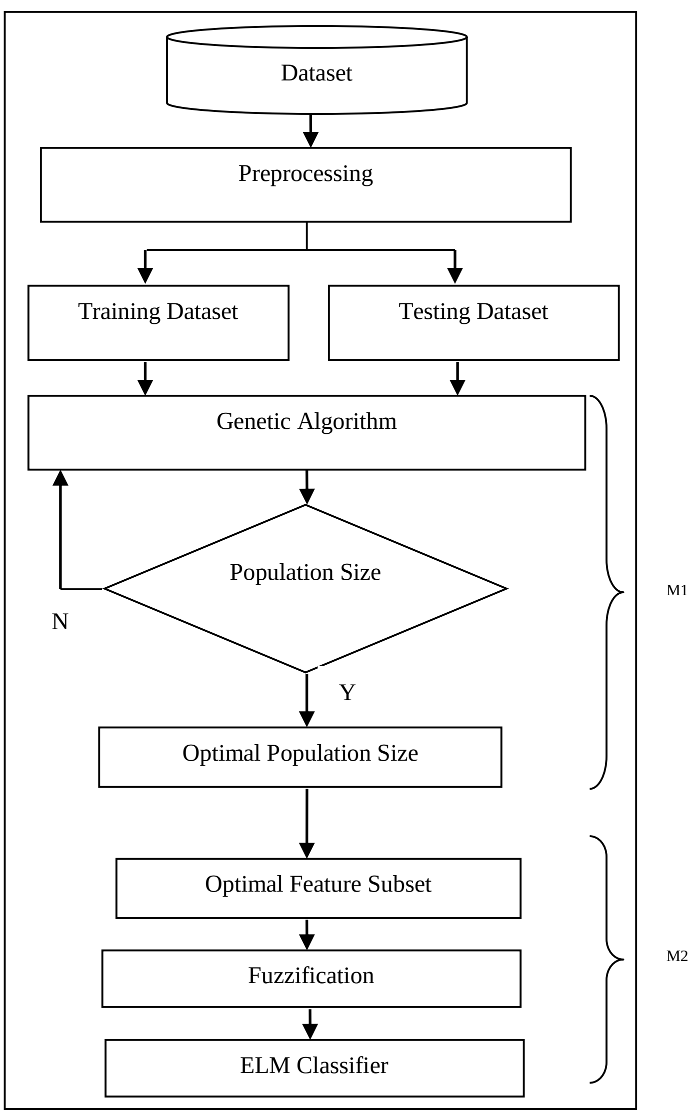

Figure 2: Multilevel hybrid model Architecture of GA-FELM

4.1 Datasets

In a binary classification problem, the input features are assigned to two classes like either +1 or -1. Therefore for experimentation, the datasets with only two classes are considered. Total six datasets from UCI Repository [26] are considered like Pima Indian Diabetes (PID), Heart-Statlog (SHD), Ionosphere, Brast cancer, Australian, German. Table 2 shows the detail description of these datasets like the number of features, the number of instances present in the dataset and the class distribution.

4.2 Preprocessing

Preprocessing is one of the important tasks for building any efficient model. The handling of missing values and selecting feature subset are used as preprocessed methods. For any model, the wrong input definitely degrades the quality of the system. Hence, it becomes very important to provide an accurate input. FSS is one of the preprocessing methods which selects perfect input (required features) and also reduces the learning time.

Table 2 Dataset Information

Dataset [26]	Features	Instances	Class 1	Class 2
PID	8	768	500	268
SHD	13	270	150	120
Ionosphere	34	351	126	225
BC	10	699	458	241
Australian	14	690	383	307
German	20	915	644	271

4.3 Feature subset selection using GA

A GA is one of the important techniques for FSS which returns optimal feature subset. GA work iteratively with a set of candidate solutions to the problem which are also known as a population [27]. In each iterative step, three processes are executed like evaluation, selection and recombination process with the help of genetic operators - selection, crossover and mutation. The iteration is repeated till it reaches some termination condition.

GA calculates the fitness value which is depending on the quality of solutions for each individual by using fitness function. Fitness function (function of the problem) is also used in evaluation to determine which of the candidate solutions are better. Selection is used to choose the strings for next generation, which have the super probability that is based on the fitness comparative to that of n other strings. It also removes those points that have low fitness value from the population. Mutation and crossover create new solutions, for exploration. Mutation is restored lost generic material where Crossover is allowed information exchange between points by protecting the fittest value of all individuals without introducing a new value [28].

However, GA has faced the optimal population size problem. As the population size changes, the feature subset is also changed. Many researchers initialize the population size as 50 or 70, but how to define an optimal population size is missing in the vast literature survey. To solve this problem, a multimodal architecture (M1+M2) of GA-FELM is proposed, in which GA results are calculated with varying the population size from 10:10:90 (from 10 to 90 with the incremental size of 10). Hence, total 9 feature subsets are evaluated with varying population size. Using all these subsets, one subset is chosen as an optimal feature subset which has maximum occurrences. Thus in this paper, population size is used as an additional parameter for optimization of feature subset which is evaluated in M1 level function.

Table 3
Results of Multimodal GA-FELM for PID dataset binary classification problem

Genetic Algorithm									
Population Size	10	20	30	40	50	60	70	80	90
Feature Subset	{2,4,8}	{2,5,6}	{2,5,6}	{2,5,8}	{2,5,6}	{2,5,6}	{2,5,6}	{2,5,6}	{2,5,6}
M1 - Optimal Feature Subset {2,5,6}									
GA - Wrapper Approach									
Classifier	**BN**	**NB**	**SVM**	**MLP**	**RBF**	**J48**	**RF**	**ELM**	**F-ELM**
GA Algorithm Accuracy	71.22	75.39	65.88	74.47	75.35	75.26	71.35	69.56	77.82
Optimal Feature Subset for each classifier	{2,5,6}	{2,5,6}	{2,5,6}	{2,5,6}	{2,5,6}	{2,5,6}	{2,5,6}	{2,5,6}	{2,5,6}
Efficient Classifier	**M2 F-ELM 77.82%**								

Table 4
Results of Multimodal GA-FELM for SHD dataset binary classification problem

Genetic Algorithm									
Population Size	10	20	30	40	50	60	70	80	90
Feature Subset	{3,8,9, 10,13}	{2,3,10, 12,13}	{1,2,3,7, 12,13}	{1,2,3, 9,12}	{3,8,9, 10,13}	{1,2,3, 9,12}	{1,2,3, 9,12}	{1,2,3, 9,12}	{1,2,3, 9,12}
M1 - Optimal Feature Subset {1,2,3,9,12}									
GA - Wrapper Approach									
Classifier	**BN**	**NB**	**SVM**	**MLP**	**RBF**	**J48**	**RF**	**ELM**	**F-ELM**

GA Algorithm Accuracy	80.19	78.21	75.4	78.54	78.21	79.86	76.23	83.95	87.65
Optimal Feature Subset for each classifier	{1,2,3, 9,12}	{1,2,3, 9,12}	{1,2,3, 9,12}	{1,2,3, 9,12}	{1,2,3, 9,12}	{1,2,3, 9,12}	{1,2,3, 9,12}	{1,2,3, 9,12}	{1,2,3, 9,12}
Efficient Classifier	**M2 F-ELM 87.65%**								

M1 is the level function where an optimal feature subset is finalized by considering all feature subsets from the various population sizes. The M2 level function is used to evaluate the subset with various classifiers and select the efficient classifier. To illustrate, Table 3 and Table 4 show the stepwise evaluation for PID and SHD dataset respectively. For PID dataset, {2,5,6} is an optimal feature subset, finalized from the nine feature subsets which are calculated from the various population sizes. By using the same subset, the classification accuracy is calculated with various existing classifiers like Navie Bays (NB), Multilayer Perceptron (MLP), Radial Basis Function (RBF), Bays net (BN), random forest (RF), J4.8, Support vector machine (SVM), ELM and F- ELM. Though results, it is inferred that
even though the number of features is same for all classifiers, F-ELM classifier provides an increased accuracy as compared to other traditional classifiers. Hence, it is concluded that the system performance does not only depend on the optimal feature subset but also depends on the classifier to be used. Therefore, F-ELM classifier is chosen to improve the generalization performance. The hybrid model ensures the creation of an optimal feature subset that reduces the computational rate without drastically affecting the performance of the classifier.

4.4 Fuzzification

Fuzzification is a process where the input features are converted into the weighted features [29]. The trapezoidal membership function Equation 1 is used for fuzzification of each feature.

$$f(M,a,b,c,d)=\begin{cases} 0 & M<a, M>d \\ \frac{(M-a)}{(b-a)} & a\le M\le b \\ 1 & b\le M\le c \\ \frac{(d-M)}{(d-c)} & c\le M\le d \end{cases} \quad \text{..................(1)}$$

4.5 Classification

Classification is a vital process in machine learning and data mining, which is used to categorize every instance in the input dataset into various classes [30]. As classifier plays an important role in system generalization performance; ELM classifier is used which is one of the fast learning algorithms.

5 EXPERIMENTAL RESULTS

In order to validate, the effectiveness of the introduced GA-FELM approach and as F-ELM is the next version of ELM, the comparative performance is calculated for - 1. ELM and GA-ELM 2. ELM and F-ELM 3. GA-FELM and GA-existing classifier [31]. For experimentation, the dataset is used from UCI

Machine Learning Repository. The data samples are categorized into 70% and 30% for training and testing process. Classification accuracy [32] is calculated by using Equation 2.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \ldots\ldots\ldots\ldots(2)$$

Where, True negative (TN), False negative (FN), True positive (TP) and False positive (FP) indicate correctly classified negative, wrongly classified positive, correctly classified positive and wrongly classified negative cases respectively.

5.1 ELM and GA-ELM

The performance of GA-ELM approach experimented on five real benchmark classification problems. By applying the GA-ELM algorithm for the considered datasets from the repositories, the optimal features are selected and the classification accuracy is calculated. The experimental results of the GA for finding optimal feature subset for six datasets are as shown in Table 6. GA can select optimal feature subset by reducing the classification error and returns the relevant and non-redundant feature subset. However, as the population size changes, the feature subsets are also changed. So, the nine feature subset is evaluated by varying population size 10:10:90. Though all these subsets, one subset is finalized as an optimal subset which has the maximum occurrences is shown in Table 6. By using this optimal feature subset the classification accuracy is calculated for Extreme Learning Machine (GA-ELM). Table 7 shows the comparative performance of ELM classification accuracy by using all features and GA-ELM. With these results, it is inferred that GA-ELM provides the 9% increased accuracy by using only 38% features. The performance comparison of ELM and GA-ELM is as shown in figure 3.

5.2 GA-FELM and GA-ELM

As F-ELM is the next version of ELM, the comparative results between these two classifiers are calculated. Table 8 illustrate that GA-FELM provides 8.2% and 3.7% increased accuracy than GA-ELM for PID and SHD dataset respectively. It also shows that FELM provides 8.7% and 10% increased accuracy than ELM for PID and SHD dataset respectively.

5.3 GA-FELM and GA-Existing classifiers

To summarize that GA-FELM approach provides an improved generalization performance over the various existing traditional classifiers (learning algorithms) like random forest (RF), Radial Basis Function (RBF), J4.8, Navie Bays (NB), Multilayer Perceptron (MLP), Bays net (BN), Support vector machine (SVM) with the result, it is noticed that GA-FELM gives the highest accuracy than other classifiers even though the number of features present in the optimal subset is same for PID and SHD dataset as shown in Table 9.

Table 6

Results of GA-ELM for seven dataset binary classification problem with optimal population size

	Genetic Algorithm									Optimal Subset	ELM Accuracy (%)
Pop	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**		
PID	{2,4,8}	{2,5,6}	{2,5,6}	{2,5,8}	{2,5,6}	{2,5,6}	{2,5,6}	{2,5,6}	{2,5,6}	{2,5,6}	77.82
SHD	{3,8,9,10,13}	{2,3,10,12,13}	{1,2,3,7,12,13}	{1,2,3,9,12}	{3,8,9,10,13}	{1,2,3,9,12}	{1,2,3,9,12}	{1,2,3,9,12}	{1,2,3,9,12}	{1,2,3,9,12}	83.95

Ionosphere	{5,6,12,15,18,21,29,33,34}	{8,13,14,19,24,34}	{7,8,13,14,18,24,27}	{5,16,22,24,25,27}	{3,4,5,8,15,21,27}	{7,18,21,22,24,27,34}	{7,8,21,24,25,27,34}	{5,7,8,14,21,24,27,28}	{1,5,14,16,21,23,24,25,27}	{3,4,5,8,15,21,27}	99.05
Breast Cancer	{2,7,8,9	{2,7,8,9}	{2,7,8,9}	{2,7,8,9}	{2,3,4,5,6,7,10}	{2,3,4,5,6,7,10}	{2,3,4,5,6,7,10}	{2,3,4,5,6,7,10}	{2,3,4,5,6,7,10}	{2,3,4,5,6,7,10}	99.04
Australian	{3,4,8,9,11}	{3,5,8,9}	{5,7,8,10,11}	{3,5,8,9}	{3,5,8,9}	{3,5,8,9}	{3,5,8,9}	{3,5,8,9}	{3,5,8,9}	{3,5,8,9}	88.88
German	{2,5}	{1,6,7,8,9,11,12,14,18}	{1,4,7,10,13,14,17}	{1,7,11,13,15,18}	{1,2,4,7,11,12,14}	{1,4,7,10,13,14,17}	{1,4,7,10,13,14,17}	{1,4,7,10,13,14,17}	{1,3,6,8,11,13,15}	{1,4,7,10,13,14,17}	74.18

Table 7 Performance comparison of ELM and GA-ELM with reduction rate of features

	ELM (%)	**GA-ELM(%)**	**Improved Accuracy(%)**	**Total Number of features present in the dataset**	**Total Number of features present in the optimal feature subset**	**Reduction in Number of features (%)**	**Use of number of features (%)**
PID	69.56	77.82	8.26	8	3	63	37
SHD	77.77	83.95	6.18	13	5	62	38
Ionosphere	94.33	99.05	4.72	34	7	80	20
BC	85.16	99.52	14.36	10	7	30	70
Australian	73.91	88.88	14.97	14	4	72	28
German	71.63	74.18	2.55	20	7	65	35
Average	**78.72**	**87.23**	**9**	**-**	**5.5**	**62**	**38**

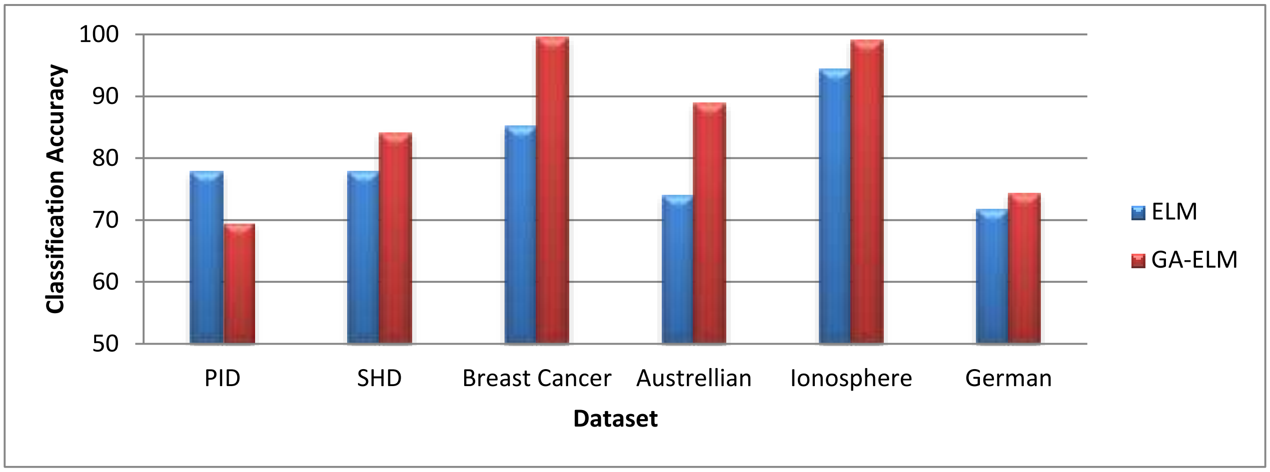

Figure 3 Comparisons of ELM by using all features and GA-ELM by using only 38% features

Table 8 Results of optimal feature subset by using Genetic Algorithm (GA) by using ELM and F-ELM Classifier

Dataset	ELM (%)	FELM (%)
PID with all features	71.73	80.43
PID GA{2,5,6}	69.56	77.82
SHD with all features	77.77	87.65
PID GA{1,2,3,9,12}	83.95	87.65

Table 9 Comparative performance of GA-FELM with GA+ existing classifier

GA+Existing Classifier	PID (%)	SHD (%)
GA-F-ELM	77.82	87.65
GA-ELM	69.56	83.95
GA-SVM	65.88	75.4
GA-BN	71.22	80.19
GA-RBF	75.35	78.21
GA-MLP	74.47	78.54
GA-NB	75.39	78.21
GA-J48	75.26	79.86
GA-RF	71.35	76.23

6 CONCLUSIONS

In this paper, Genetic Algorithm for F- ELM classifier (GA-FELM) with multimodal parameter optimization is introduced. GA is one of the examples of random search strategy which returns optimal feature subset. The main objective of this research work is to select optimal feature subset by using the genetic algorithm for Fuzzy Extreme Learning Machine classifier (GA-FELM) and also to design a hybrid model with multimodal parameter optimization which selects optimal feature subset by considering the population variance. The hybrid model performs satisfactorily on several datasets with the key advantages of less learning (training) time, high speed, optimal feature selection and better generalization performance. In order to prove, the performance of GA-FELM, the comparative results are calculated by considering ELM, F-ELM, GA-ELM, GA-FELM and GA-existing classifier. The proportional average analysis shows that on an average 9% classification accuracy is increased by using only 38% features. The validation of the proposed multilevel hybrid model over various datasets and results are found to be efficient to a great extent almost for all. This approach can be used in many real-time applications where large numbers of features are present. Presently, the results are evaluated for binary classification problem. In future work, the same work will be extended to multiclass classification and one-class classification problem.

REFERENCES

[1] L. Yu and H. Liu. (2004). Efficient feature selection via analysis of relevance and redundancy, Journal of machine learning research. pp. 1205–1224.

[2] Kittler,J. (1978). Pattern Recognition and Signal Processing, Chapter Feature Set Search Algorithms Sijthoff and Noordhoff, Alphen aan den Rijn, Netherlands, pp. 41–60.

[3] D. Koller and M. Sahami. (1996). Toward optimal feature selection. Tech. rep. Stanford InfoLab.

[4] G. Pahuja and T. Nagabhushan. 2016. A novel ga-elm approach for parkinson's disease detection using brain structural t1-weighted mri data. In: Cognitive Computing and Information Processing (CCIP), Second International Conference on IEEE, pp. 1–6.

[5] L. Mao, L. Zhang, X. Liu, C. Li and H. Yang. 2014. Improved extreme learning machine and its application in image quality assessment, Mathematical Problems in Engineering.

[6] G. Chandrashekar and F. Sahin. 2014. A survey on feature selection methods, Computers & Electrical Engineering 40 (1), pp. 16–28.

[7] G. P. Zhang. 2000. Neural networks for classification: a survey, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 30 (4) pp. 451–462.

[8] V. Rathi and S. Aggarwal. 2014. Comparing the performance of ann with fnn on mammography mass data set. In: Advance Computing Conference (IACC). IEEE International, pp. 1307–1314.

[9] L. Zedeh. 1989, Knowledge representation in fuzzy logic, IEEE Transactionson knowledge and data engineering 1 (1), pp. 89–100.

[10] M. F. Akay. 2009. Support vector machines combined with feature selection for breast cancer diagnosis, Expert systems with applications 36 (2) , pp. 3240–3247.

[11] G.-B. Huang, Q.-Y. Zhu and C.-K. Siew. 2006. Extreme learning machine: theory and applications. Neurocomputing 70 (1). pp. 489–501.

[12] G.-B. Huang, Y.-Q. Chen and H. A. Babri. 2000. Classification ability of single hidden layer feedforward neural networks, IEEE Transactions on Neural Networks, 11 (3), pp. 799–801.

[12] G.-B. Huang and L. Chen. 2008. Enhanced random search based incremental extreme learning machine, Neurocomputing 71 (16), pp. 3460–3468.
[13] Mehmet Fatih Akay. (2009). Support vector machines combined with feature selection for breast cancer diagnosis. Elsevier journal of Expert Systems with Applications. pp. 3240–3247.
[14] I. Guyon, J. Weston, S. Barnhill and V. Vapnik. 2002. Gene selection for cancer classification using support vector machines, Machine learning 46 (1-3) . pp. 389–422.
[15] G.-B. Huang and L. Chen. (2008). Enhanced random search based incremental extreme learning machine. Neurocomputing. vol. 71. nos. 16-18. pp. 3460–3468.
[16] G.-B. Huang, Q.-Y. Zhu and C.-K. Siew. 2004. Extreme learning machine: a new learning scheme of feedforward neural networks, in: Neural Networks, In proceedings. IEEE International Joint Conference. vol. 2. pp. 985–990.
[17] G.-B. Huang, Q.-Y. Zhu and C.-K. Siew. 2006. Extreme learning machine: theory and applications. Neurocomputing 70 (1). pp. 489–501.
[18] G.-B. Huang, L. Chen, and C.-K. Siew. (2006). Universal approximation using incremental constructive feedforward networks with random hidden node. IEEE Tranactions on. Neural Network. vol. 17. no. 4. pp. 879–892.
[19] G.-B. Huang, H. Zhou, X. Ding and R. Zhang (2012). Extreme learning machine for regression and multiclass classification, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 42 (2). pp. 513–529.
[20] G.-B. Huang. (2014). An insight into extreme learning machines: Random neurons, random features and kernels. Cognit. Computat. vol. 6. no. 3. pp. 376–390.
[21] S. Beyhan and K. Kavaklioglu. (2015). Comprehensive modeling of u-tube steam generators using extreme learning machines. IEEE Transactions on Nuclear Science 62 (5). pp. 2245–2254.
[22] S. Y. Wong, K. S. Yap, H. J. Yap, S. C. Tan and S. W. Chang. (2015). On equivalence of fis and elm for interpretable rule-based knowledge representation. IEEE transactions on neural networks and learning systems 26 (7). pp. 1417–1430.
[23] W. Zhang and H. Ji. 2013. Fuzzy extreme learning machine for classification, Electronics Letters 49 (7). pp. 448–450.
[24] Y. Saeys, I. Inza and P. Larranaga. 2007. A review of feature selection techniques in bioinformatics, bioinformatics 23 (19). pp. 2507–2517.
[25] R. Storn, K. Price. 1997. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces, Journal of global optimization 11 (4) . pp. 341–359.
[26] M. Lichman, 2013. UCI machine learning repository. URL http://archive.ics.uci.edu/ml
[27] L. Zhuo, J. Zheng, X. Li, F. Wang, B. Ai and J. Qian. 2008. A genetic algorithm based wrapper feature selection method for classification of hyperspectral images using support vector machine. In: Geoinformatics and Joint Conference on GIS and Built Environment: Classification of Remote Sensing Images, International Society for Optics and Photonics. pp. 71471J–71471J.
[28] Y. Fei and H. Min. 2016. Simultaneous feature with support vector selection and parameters optimization using ga-based svm solve the binary classification, In: IEEE International Conference on Computer Communication and the Internet (ICCCI). pp. 426–433.
[29] K. B. Nahato, K. H. Nehemiah and A. Kannan. 2016. Hybrid approach using fuzzy sets and extreme learning machine for classifying clinical datasets. Elsevier Journal of Informatics in Medicine Unlocked 2. pp. 1–11.
[30] G.-B. Huang, H. Zhou, X. Ding and R. Zhang 2012. Extreme learning machine for regression and multiclass classification, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 42 (2). pp. 513–529.

[31] H. Temurtas, N. Yumusak and F. Temurtas. 2009. A comparative study on diabetes disease diagnosis using neural networks, Elsevier journal on Expert Systems with applications 36 (4). pp. 8610–8615.
[32] G. Karakaya, S. Galelli, S. D. Ahipa ̧sao ̆glu and R. Taormina, Identifying (quasi) equally informative subsets in feature selection problems for classification: a max-relevance min-redundancy approach, IEEE transactions on cybernetics 46 (6). pp. 1424–1437.