OPEN ACCESS

International
Journal
Of Modern Engineering Research (IJMER)

A Threshold Fuzzy Entropy Based Feature Selection: Comparative Study

Miss. K.Barani[1], Mr. R.Ramakrishnan[2]
[1, 2] (PG Student, Associate Professor, Sri Manakula Vinayagar Engineering College, Pondicherry-605106)

Abstract: *Feature selection is one of the most common and critical tasks in database classification. It reduces the computational cost by removing insignificant and unwanted features. Consequently, this makes the diagnosis process accurate and comprehensible. This paper presents the measurement of feature relevance based on fuzzy entropy, tested with Radial Basis Classifier (RBF) network, Bagging(Bootstrap Aggregating), Boosting and stacking for various fields of datasets. Twenty benchmarked datasets which are available in UCI Machine Learning Repository and KDD have been used for this work. The accuracy obtained from these classification process shows that the proposed method is capable of producing good and accurate results with fewer features than the original datasets.*
Keywords: *Fuzzy entropy, Feature selection, RBF network, Bagging, Boosting, Stacking, Fuzzy C-means clustering algorithm.*

I. Introduction

Data mining is the process of efficient discovery of non-obvious valuable patterns from a large collection of data. It has been discussed widely and applied successfully in the field of medical research, scientific analysis and business applications. Feature selection has many advantages such as shortening the number of measurements, reducing the execution time and improving transparency and compactness of the suggested diagnosis.

Feature selection is the process of selecting a subset of'd' features of the set D, such that d ≤ D. the primary purpose of the feature selection is to reduce the computational cost and improve the performance of the learning algorithm. Feature selection deals with different evaluation criteria and generally, are classified into filter and Wrapper models. The filter model evaluates the general characteristics of the training data to select the feature subset without relation to any other learning algorithms, thus, it is computationally economical. Nevertheless, it carries the risk of selecting subset of features that may not be relevant. The wrapper models which requires a pre-determined induction algorithm, which assesses the performance of the features that are chosen. The selected features are related significantly to the choice of the classifier and do not generalize to other classifiers. However, this tends to be computationally expensive. Therefore, the filter and wrapper model would complement each other; wrapper models provide better accuracy, whereas filter models search the feature space efficiently.

This paper proposes a filter-based feature subset selection based on fuzzy entropy measures and presents the different selection strategies for handling the datasets. The proposed method is evaluated using RBF network, Bagging, Boosting and stacking for the given benchmarked datasets.

II. Literature Review

Recently, a number of researchers have focused on several feature selection methods and most of them have reported their good performance in database classification. Battiti [7] proposes a method called Mutual-Information-based Feature Selection (MIFS), in which the selection criterion is based on maximizing the mutual information between candidate features and the class variables, and minimizing the redundancy between candidate features and the selected features. Hanchuan et al. [8] follow a similar technique to MIFS, which has been called the minimal-redundancy-maximal-relevance (mRMR) criterion. It eliminates the manually tuned parameter with cardinality of the features already selected. Pablo et al. [9] present a Normalized Mutual Information Feature Selection algorithm. The mutual information among features should be divided by the minimum value of their entropies in order to produce a normalized value, which is to be measured by the redundant term. Yu and Liu [10] developed a correction-based method for relevance and redundancy analysis and then removed redundant features using the Markov Blanket method.

In addition, feature selection methods are analyzed by a number of techniques. Abdel-Aal [1] developed a novel technique for feature ranking and selection with the group method of data handling. Feature reduction of more than 50% could be achieved and improved in the classification performance. Sahanetal [11] built a new hybrid machine learning method for a fuzzy-artificial immune system with a k-nearest neighbour algorithm to solve medical diagnosis problems, which demonstrated good results. Jaganathan etal. [12] Applied a new improved quick reduct algorithm, which is a variant of quick reduct for feature selection and tested it on a classification algorithm called AntMiner. Sivagami Nathan et al. [13] proposed a hybrid method combining Ant Colony Optimization and Artificial Neural Networks (ANNs) to deal with feature selection, which produced promising results. Lin et al. [14] proposed a Simulated Annealing approach for parameter setting in Support Vector Machines, which is compared with a grid search parameter setting and was found to produce higher classification accuracy.

Lin et al. [15] applied a Particle-Swarm-Optimization-based approach to search for appropriate parameter values for a back- propagation network to select the most valuable subset of features to improve classification accuracy. Unler et al [16] developed a modified discrete particle swarm optimization algorithm for the feature selection problem and compared it with tabu and scatter search algorithms to demonstrate its effectiveness. Chang et al [17] introduced a hybrid model for integrating a case-based reasoning approach with a particle swarm optimization model for feature subset selection in medical database classification. Salamo et al [18] evaluated a number of measures for estimating feature relevance based on rough set theory and also proposed three strategies for feature selection in a Case Based Reasoning classifier. Qasem et al [19] applied a time variant multi- objective particle swarm optimization to an RBF Network for diagnosing medical diseases.

This paper describes in detail how to combine the relevance measures and feature subset selection strategies.

III. Fuzzy Entropy-Based Relevance Measure

In information theory, the Shannon entropy measure is generally used to characterize the impurity of a collection of samples. Assuming X as a discrete random variable with a finite set of n elements, where X={x1, x2, x3,…,xn}, then if an element x_i occurs with probability $p(x_i)$, the entropy H(X) of X is defined as follows:

$$H(X)=-\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) \quad (1)$$

Where n denotes the number of elements.

An extension of Shannon entropy with fuzzy sets, which is used to support the evaluation of entropies, is called fuzzy entropy. It was introduced in 1972, after which a number of modifications were introduced to the original fuzzy entropy method.

The proposed fuzzy entropy method is based on the utilization of the Fuzzy C-Means Clustering algorithm (FCM), which is used to construct the membership function of all features. The data may belong to two or more clusters simultaneously and the belonging of a data point to the clusters is governed by the membership values. Similar data points are placed in the same cluster and dissimilar data points normally belong to different clusters. The membership values of the data points are reorganized iteratively to reduce the dissimilarity. The Euclidean distance is used to measure the dissimilarity of two data points.

The FCM algorithm is explained as follows.

Step1: assume the number of clusters(C), where $2 \leq C \leq N$, C – number of clusters and N – number of data points.

Step2: calculate the j^{th} cluster center C_j using the following expression

$$C_J=\left(\sum_{i=1}^{N} \mu_{ij}^{g}\ x_{ij}\right)/\left(\sum_{i=1}^{N} \mu_{ij}^{g}\right) \quad (2)$$

where $g \geq 1$ is the fuzziness coefficient and μ_{ij} is the degree of membership for the i^{th} data point x_i in cluster j.

Step3: calculate the Euclidean distance between the i^{th} data point and the j^{th} cluster center as follows:

$$d_{ij}=\left|C_{ij}-x_i\right| \quad (3)$$

Step4: update the fuzzy membership values according to d_{ij}. If $d_{ij} \geq 0$, then

$$\mu = 1/\left(\sum_{m=1}^{C} (d_{ij}/d_{im})^{2/(g-1)}\right) \quad (4)$$

If d=0, then the data point coincides with the j^{th} cluster center (C) and it will have the full membership value, i.e., $\mu_{ij}=1.0$

Step5: repeat Steps 2–4 until the changes in $[\mu]$ are less than some pre-specified values.

The FCM algorithm computes the membership of each sample in all clusters and then normalizes it. This procedure is applied for each feature. The summation of membership of feature 'x' in class 'c', divided by the membership of feature 'x' in all 'C' classes, is termed the class degree CDc($\check{A}$), which is given as:

$$CD_c(\check{A})= \sum_{x\epsilon c} \mu_{\check{A}}(x)/\sum_{x\epsilon C} \mu_{\check{A}}(x) \quad (5)$$

Where $\mu_{\check{A}}$ denotes the membership function of the fuzzy set and $\mu_{\check{A}}(x_i)$ denotes the membership grade of x belonging to the fuzzy set $\check{A}$.

The fuzzy entropy FEc ($\check{A}$) ofclass 'c' is defined as

$FE_c(\breve{A})=-CD_c(\breve{A})\log_2 CD_c(\breve{A})$ (6)

The fuzzy entropy $FE(\breve{A})$ of a fuzzy set X is defined as follows:

$FE(\breve{A})=\sum_{c\epsilon C} FE_c(\breve{A})$ (7)

The probability $p(x_i)$ of Shannon's entropy is measured by the number of occurring elements. In contrast, the class degree $CDc(\breve{A})$ in fuzzy entropy is measured by the membership values of the occurring elements and the highest fuzzy entropy value of the feature is regarded as the most informative one.

IV. Feature Selection Strategies

This section explains three different criteria for the feature selection process. The features are regulated with respect to decreasing values of the fuzzy entropy. A feature in the first position is the most relevant and the one in the last position is the least relevant in the resulting rank vector. The framework of feature selection is depicted in Fig. 1.

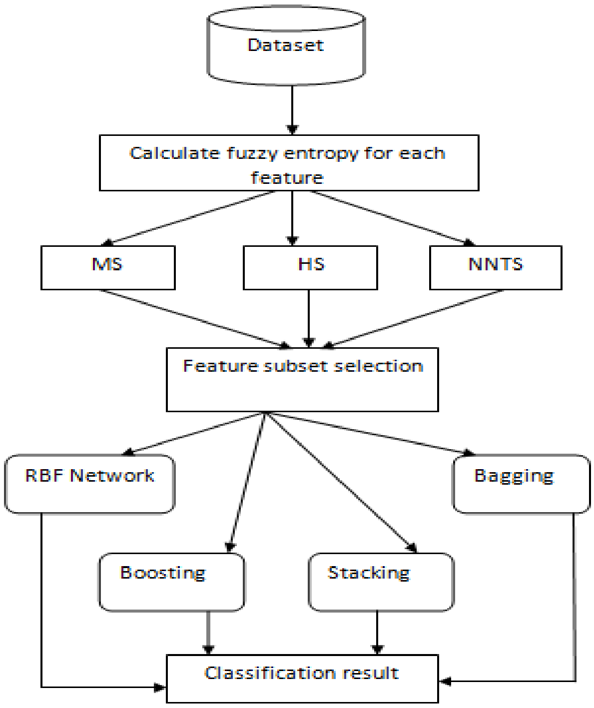

Fig 1

Mean Selection (MS) Strategy: A feature f ϵ F is selected if it satisfies the following condition:

$$\sigma(f) \geq \sum_{f\epsilon F} \sigma(f)/|F|$$

where $\sigma(f)$ is the relevance value of the features, which is selected if it is greater than or equivalent to the mean of the relevant values. This strategy will be useful in examining the suitability of the fuzzy entropy relevance measure.

Half Selection (HS) Strategy: The half selection strategy aims to reduce feature dimensionality to select approximately 50% of the features in the domain. The feature f ϵ F is selected if it satisfies the following condition:

$P_a \geq |F|/2$

Where Pa is the position of the feature in the rank vector. It represents the selected features having a relevance value higher than a given threshold, which is calculated as $|F|/2$. This strategy does produce great reductions, close to 50%. At the same time, some of the selected features are irrelevant despite them passing the threshold. Similarly, some of the omitted features may also be relevant despite them not being selected. This suggests that a new feature selection strategy must be based on the relevance value of each feature instead of a predefined number of features that are to be reduced. The last feature selection strategy described below has a relatively smaller number of features but at the same time, it retains the most relevant.

Neural Network for Threshold Selection (NNTS): An ANN is one of the well-known machine learning techniques and it can be used in a variety of applications in data mining. The ANN provides a variety of feed forward networks that are generally called back propagation networks. It possesses a number of inter- connected layers that consist of an input layer, a hidden layer and an output layer. The fuzzy entropy value of each feature is an initial value for each node in the input layer. The value from the input layer to the output layer is achieved by hidden layers using weights and activation functions. A sigmoid function is used as an activation function and a learning rate coefficient determines the size of weight adjustments made at the each iteration. An output layer is used to represent an output value. The output value can be considered as a threshold value of the given fuzzy entropy.

V. Methodology Description

There are four methodologies used for calculating an accuracy after the features are selected using the above three strategies.

RBF Network:

An RBF network is a type of ANN, which is simpler network structure with better approximation capabilities. It is an artificial neural network that uses the radial basis function as the artificial network. Radial basis function is the real-valued function whose values depends on distance from origin or any other point called as C.RBF can be used as kernel in support vector classification.RBF network trains the hidden layer in unsupervised manner.

Bagging:

Bagging (Bootstrap Aggregating) is a machine learning ensemble meta algorithm that is used to find the stability and accuracy of the training data. This method creates separate samples of the training dataset and classifier for each sample. The result of multiple classifiers is combined to find accuracy.

Bagging leads to improvements in unstable procedures. It helps to reduce variance avoids over fitting. This is the special case of model averaging approach.

Boosting:

Boosting is an ensemble method that starts out with a base classifier that is prepared on the training data. A second classifier is then created behind it to focus on the instances in the training data that the first classifier goes wrong. The process continues to add classifiers until a limit is reached in the number of models or accuracy. It helps to remove noisy data and removes outliers.

Stacking:

Stacking also called Blending or Stacked generalization. It is an ensemble method where multiple different algorithms are prepared on the training data and a Meta classifier is prepared that learns how to take the predictions of each classifier and make accurate predictions on unseen data.

It involves training learning algorithms to combine predictions of several other learning algorithms. First, all of the other algorithms are trained using the available data, then a combiner algorithm is trained to make a final prediction using all the predictions of the other algorithms as additional inputs. It combines algorithms like ID3 and J48.

ID3: Generates decision tree from the dataset and is used in machine learning and natural language processing domains. It begins with original set S at the root node and iterates through unused attribute of the set. It is calculated using Entropy and information gain value.

J48: It is the extension of ID3 algorithm. It is used to generate decision tree that can be used for classification and so it is called as statistical classifier.

VI. Dataset Description

The performance of the proposed method is evaluated using several benchmarked datasets.

DATASET	NO OF FEATURES	NO OF INSTANCES
Diabetes	768	8
Hepatitis	155	19
Heart-Statlog	270	13
Wisconsin breast cancer	699	9
Grub damage	155	8
White clover	63	31
Squash unstored	52	23
Squash stored	50	23
Tic-tac-toe	51	9
Chess	42	6

Dermatology	105	34
Car	1117	6
Liver disorder	187	6
Hypothyroid	312	29
Pasture	36	22
Eggs	48	3
Fiber	48	4
Ionosphere	351	34
Balance	17	3
Cleveland heart disease	302	13

1. Wisconsin breast cancer:

The dataset was collected by Dr.William H.Wolberg (1989– 1991) at the University of Wisconsin–Madison Hospitals. It contains 699 instances characterized by nine features: (1) Clump Thickness, (2) Uniformity of Cell Size, (3) Uniformity of Cell Shape, (4) Marginal Adhesion, (5) Single Epithelial Cell Size, (6) Bare Nuclei, (7) Bland Chromatin, (8)Normal Nucleoli and (9)Mitoses, which are used to predict benign or malignant growths. In this dataset, 241(34.5%) instances are malignant and 458(65.5%) instances are benign.

2. Pima Indians diabetes:

The dataset is available at the National Institute of Diabetes and Digestive and Kidney Diseases. It contains 768 instances described by eight features used to predict the presence or absence of diabetes. The features are as follows: (1) number of pregnancies, (2) plasma glucose concentration, (3) diastolic blood pressure, (4) triceps skin fold thickness,(5) serum insulin, (6)body mass index, (7)diabetes pedigree function and (8)age in years.

3. Heart-Statlog:

The dataset is based on data from the Cleveland Clinic Foundation and it contains 270 instances belonging to two classes: the presence or absence of heart disease. It is described by13 features (age, sex, chest, resting blood pressure, serum cholesterol, fasting blood sugar, resting electro cardiographic, maximum heart rate, exercise induced angina, old peak, slope, number of major vessels and thal).

4. Hepatitis:

The dataset is obtained from the Carnegie–Mellon University and it contains 155 instances belonging to two classes: live or die. There are 19 features (age, sex, steroid, antivirals, fatigue, malaise, anorexia, liver big, liver film, spleen palpable, spiders, ascites, varices, bilirubin, alk phosphate, SGOT, albumin, protime and histology).

5. Cleveland heart disease:

The dataset was collected from the Cleveland Clinic Foundation and contains about 296 instances, each having 13 features, which are used to infer the presence or absence of heart disease. The features are (1) age, (2)sex, (3)chest pain type, (4)resting blood pressure, (5) cholesterol, (6)fasting blood sugar, (7)resting electro cardio- graphic results, (8)maximum heart rate, (9)exercise induced angina, (10)depression induced by exercise relative to segment, (11)slope of peak exercise, (12) number of major vessels and (13)thal.

6. Chess:

The dataset consist of 6 attributes namely:(1)White_king_file,(2) White_king_rank (3)White_rook_file, (4)White_rook_rank, (5) Black_king_file ,(6)Black_king_rank and two classes like win or lose for 42 instances.

7. Grub_damage:

The dataset consists of 158 instances consisting of attributes like year-zone, year, strip, pdk, damage-rankRJT, damage-rankALL, dry_or_irr and zone with two classes: low or high.

8. Pasture:

This dataset contains two classes like low or high with 22 attributes like fertilizer, slope, aspect_dev_NW, OLsenP, MinN, TS, Ca-Mg, LOM, NFIX, Eworms-main-3, Eworms-No-Species, KUnset, OM,Air-Perm, Porosity, HFRG-pct-mean, jan-mar-mean-TDR, Annual-mean-Runoff, root-surface-area and Leaf-p.

9. Squash-stored:

The dataset containing two class and 50 instances with 24 attributes like site, daf, fruit, weight, storewt, lene, solids, brix, a*, egdd, fgdd, ground slot a*, glucose, fructose, sucrose, total, glucose+fructose, starch, sweetness, flavor, dry/moist, fiber, heat inlut emerg and heat inlut flower.

10. Squash-Unstored:

The dataset containing two class and 52 instances with 23 attributes like site, daf, fruit, weight, lene, solids, brix, a*, egdd, fgdd, groundslot_a*, glucose, fructose, sucrose, total, glucose+fructose, starch, sweetness, flavor, dry/moist, fiber, heat_inlut emerg and heat inlut flower.

11. Tic-tac-toe:

This dataset contains 51 instances with 9 attributes like top-left-square, top-middle-square, middle-left-square, middle-middle-square, middle-right-square, bottom-left-square, bottom-middle-square and bottom-right-square with two classes.

12. White-clover:

The dataset contains 63 instances with two classes and with 31 attributes like strata, plot, paddock, whiteclover-91, bareground-91, cocksfoot-91, othergrasses-91, otherlegumes-91, RyeGrass-91, Weeds-91, whiteclover-92, bareground-92, cocksfoot-92, othergrasses-92, otherlegumes-92, RyeGrass-92, weeds-92, whiteclover-93, bareground-93, cocksfoot-93, othergrasses-93, otherlegumes-93, RyeGrass-93, weeds-93, whiteclover-94, bareground-94, cocksfoot-94, othergrasses-94, otherlegumes-94, RyeGrass-94, weeds-94 and strata combined. The classes may be either yes or no.

13. Balance:

The dataset contains two classes with 17 instances and 3 attributes like Subject no, forward-backward and side-side.

14. Car:

This contains 1117 instances with 6 attributes like buying, maint, doors, persons, lug boot and safety with two classes.

15. Dermatology:

This dataset contains 105 instances with 34 attributes and two classes. The attributes are like erythema, scaling, definite borders, itching, koebner phenomenon, polygonal papulus, follicular papulus, oral mucosal involvement, knee and elbow involvement, Scalp involvement, family history, melanin incontinence, eosinophils in the infiltrate, PNL infiltrate, fibrosis of the papillary dermis, exocytosis, acanthosis, hyperkeratosis, parakeratosis, clubbing of the rete ridges, thinning of the suprapapillary epidermis, spongiform pastule, munro microabcess, focal hypergranulosis, disappearance of the granular layer, vacuolisatio and damage of basal layer, spongiosis, saw-tooth appearance of the retes, follicular horn plug, perifollicular parakeratosis, inflammatory monoluclear inflitrate, band-like infiltrate and age. The class defines either present or absent.

16. Hypothyroid:

The dataset contains 29 attributes for 312 instances. The attributes are as: age, sex, on thyroxine, query on thyroxine, on antithyroid medication, sick, pregnant, thyroid surgery,I131 treatment, query hypothyroid, query hyperthyroid, lithium, goiter, tumor, hypopituitary, psych, TSH measured, TSH, T3 measured, T3, TT4 measured, TT4,,T4U measured, T4U,FTI measured, FTI, TBG measured, TBG, referral source with class negative or positive.

17. Eggs:

The dataset contains 3 attributes like Gat_content, Lab, and Technician with two classes G and H for 48 instances.

18. Fiber:

This dataset contains two classes yes or no with 4 attributes. The attributes considered are cracker, diet, and subject and digested.

19. Ionosphere:

This dataset contains 34 attributes from a01to a34 for 351 instances and consists of two classes 1 or 2.

20. Liver-disorder:

The dataset contains mcz, alkphos, sgpt, sgot, gammagt, drinks as attributes for 187 instances with two classes 1 or 2.

VII. Result

The selected features from the three strategies are tested with RBF network, Bagging, Boosting and stacking to calculate the accuracy.

7.1 Wisconsin Breast Cancer:

In fig 2, the report is clearly depicted and is found that Bagging and Boosting yields the highest accuracy of 99.57 in mean selection strategy.

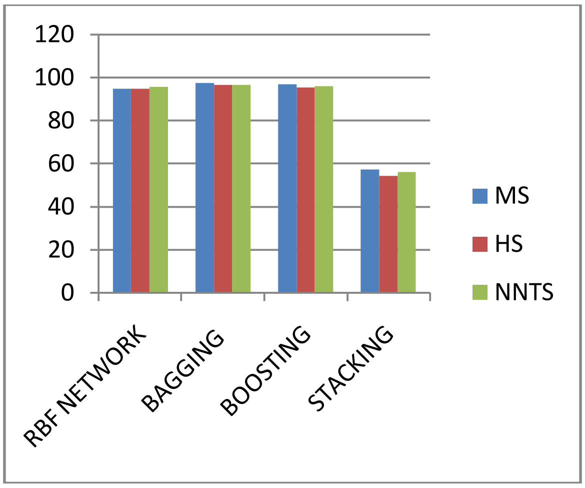

Fig 2

7.2 Pima Indian Diabetes:

In fig 3, the report is clearly depicted and is found that Bagging yields the highest accuracy of 100.0 in mean selection strategy.

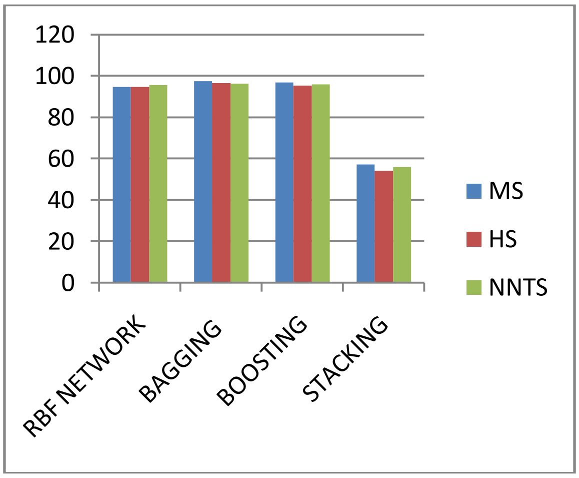

Fig 3

7.3 Heart- Statlog:

In fig 4, the report is clearly depicted and is found that Bagging and Boosting yields the highest accuracy of 99.62 in mean selection strategy.

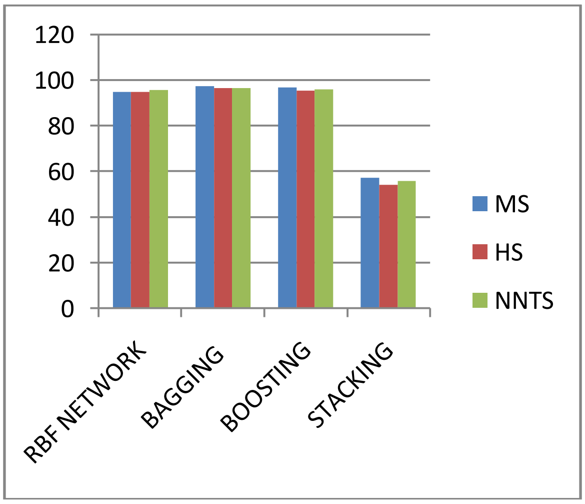

Fig 4

7.4 Hepatitis:

In fig 5, the report is clearly depicted and is found that RBF network yields the highest accuracy of 90.32 in half selection strategy.

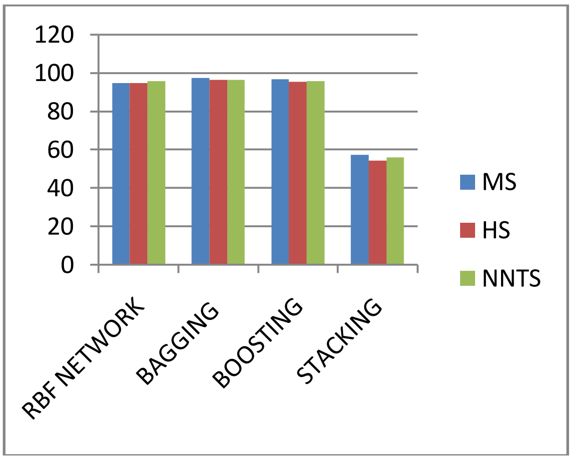

Fig 5

7.5 Cleveland heart Disease:

In fig 6, the report is clearly depicted and is found that RBF network yields the highest accuracy of 99.00 in neural network for threshold selection strategy.

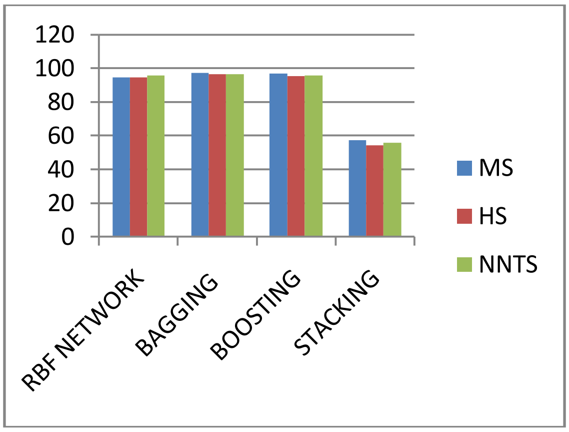

Fig 6

7.6 Chess:

In fig 7, the report is clearly depicted and is found that Bagging and Boosting yields the highest accuracy of 100 in mean selection strategy.

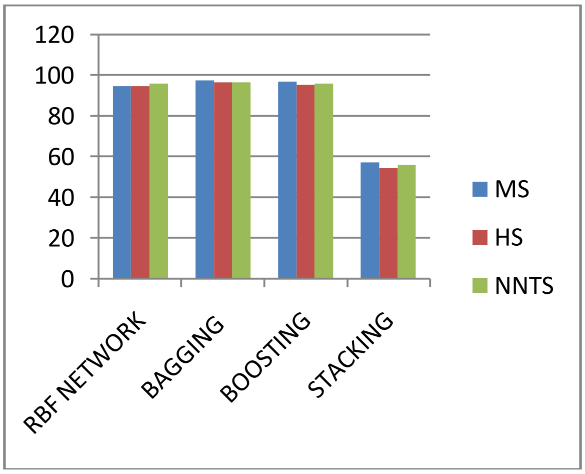

Fig 7

7.7 Grub Damage:

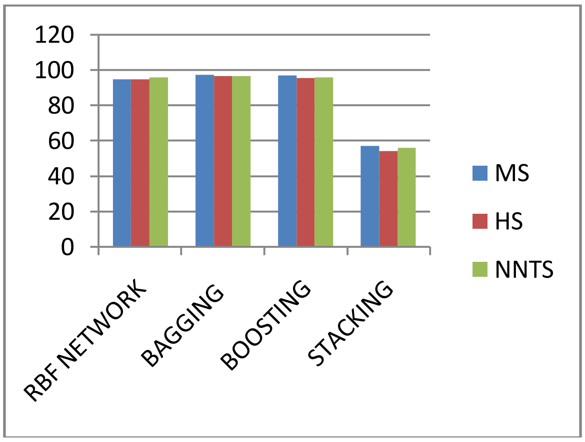

Fig 8

In fig 8, the report is clearly depicted and is found that Bagging and Boosting yields the highest accuracy of 98.06 in mean selection strategy.

7.8 Pasture:

In fig 9, the report is clearly depicted and is found that RBF network yields the highest accuracy of 100.0 in mean selection strategy.

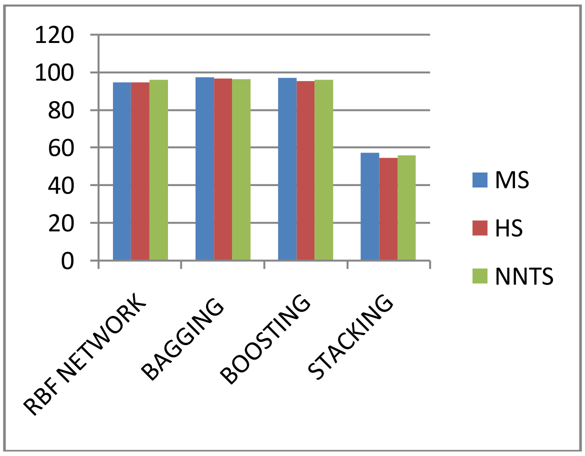

Fig 9

7.9 Squash-Stored:

In fig 10, the report is clearly depicted and is found that Boosting yields the highest accuracy of 98.0 in mean selection strategy.

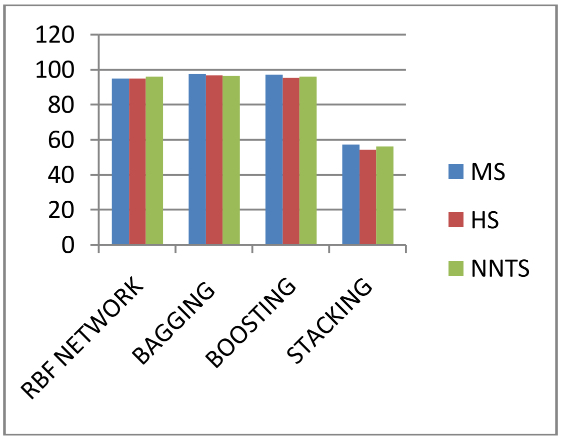

Fig 10

7.10 Squash-Unstored:

In fig 11, the report is clearly depicted and is found that Bagging yields the highest accuracy of 100.0 in mean selection strategy.

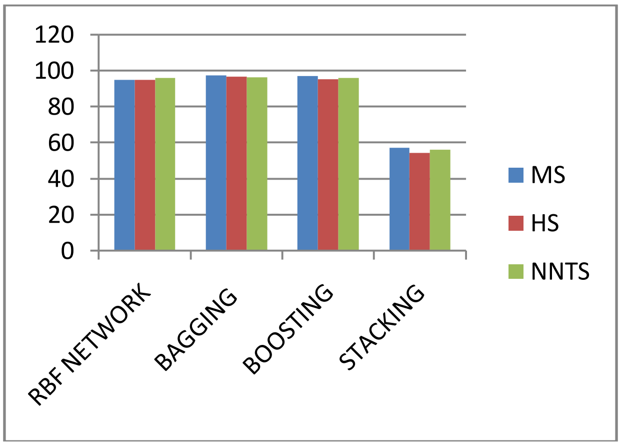

Fig 11

7.11 Tic-tac-toe:

In fig 12, the report is clearly depicted and is found that Bagging and Boosting yields the highest accuracy of 100.0 in mean selection strategy.

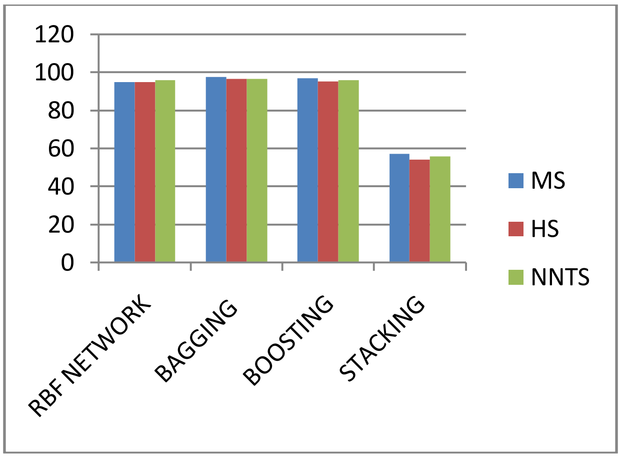

Fig 12

7.12 White- Clover:

In fig 13, the report is clearly depicted and is found that Bagging and Boosting yields the highest accuracy of 96.82 in mean selection strategy.

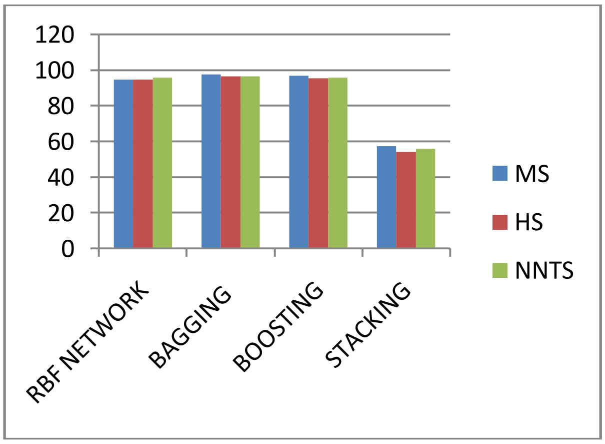

Fig 13

7.13 Balance:

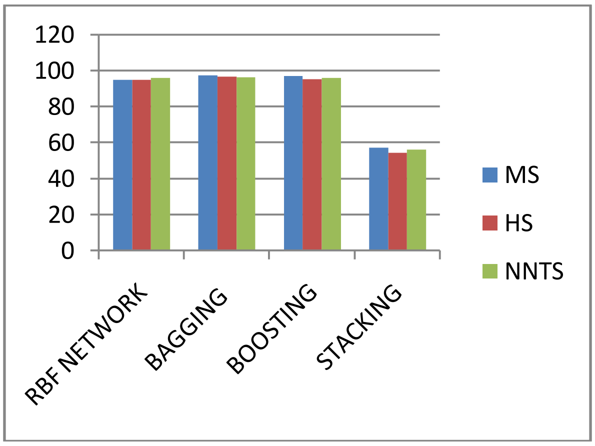

Fig 14

In fig 14, the report is clearly depicted and is found that Bagging and Boosting yields the highest accuracy of 100.0 in mean selection strategy and the same accuracy in RBF network using half selection.

7.14 Car:

In fig 15, the report is clearly depicted and is found that Bagging yields the highest accuracy of 98.06 in neural network for threshold selection strategy.

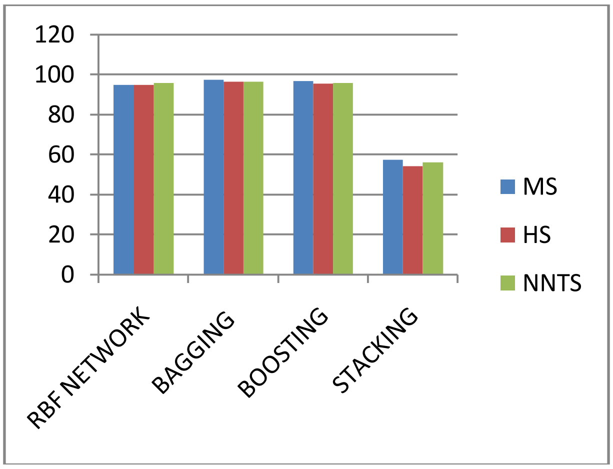

Fig 15

7.15 Dermatology:

In fig 16, the report is clearly depicted and is found that Bagging and Boosting yields the highest accuracy of 99.04 in mean selection strategy.

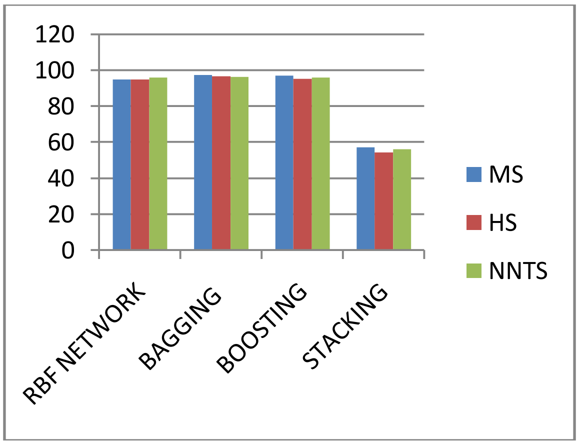

Fig 16

7.16 Hypothyroid:

In fig 17, the report is clearly depicted and is found that all the four methodologies yields the highest accuracy of 94.23 in mean selection strategy.

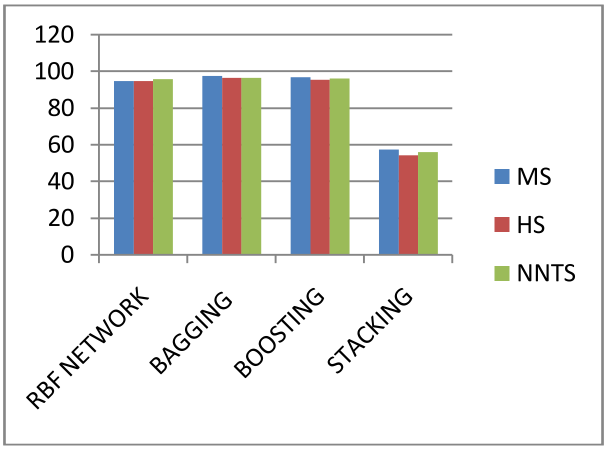

Fig 17

7.17 Eggs:

In fig 18, the report is clearly depicted and is found that RBF network yields the highest accuracy of 100.0 in mean selection strategy.

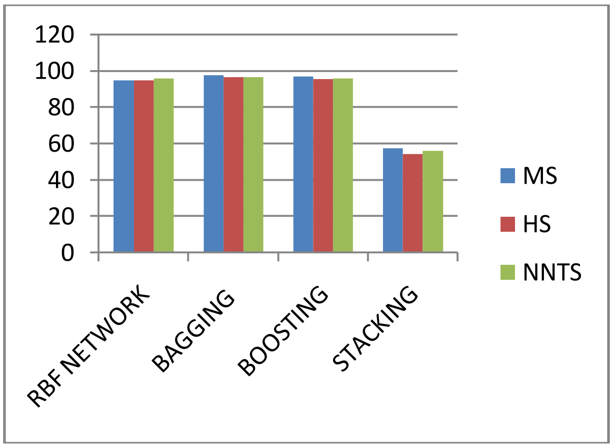

Fig 18

7.18 Fiber:

In fig 19, the report is clearly depicted and is found that Bagging and Boosting yields the highest accuracy of 97.91 in mean selection strategy.

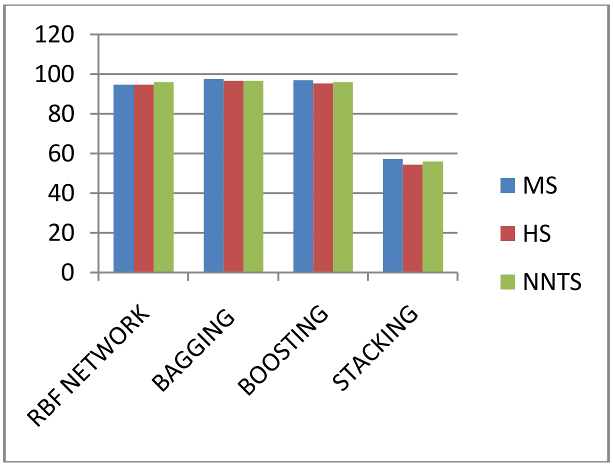

Fig 19

7.19 Ionosphere:

In fig 20, the report is clearly depicted and is found that Boosting yields the highest accuracy of 99.43 in mean selection strategy.

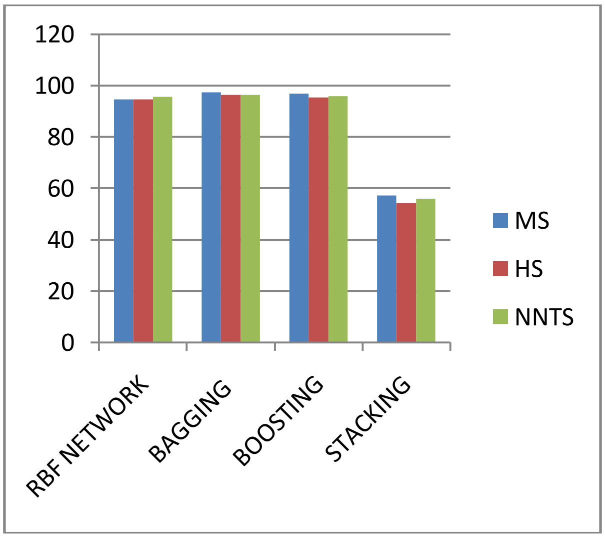

Fig 20

7.20 Liver Disorder:

In fig 21, the report is clearly depicted and is found that Bagging yields the highest accuracy of 97.32 in mean selection strategy.

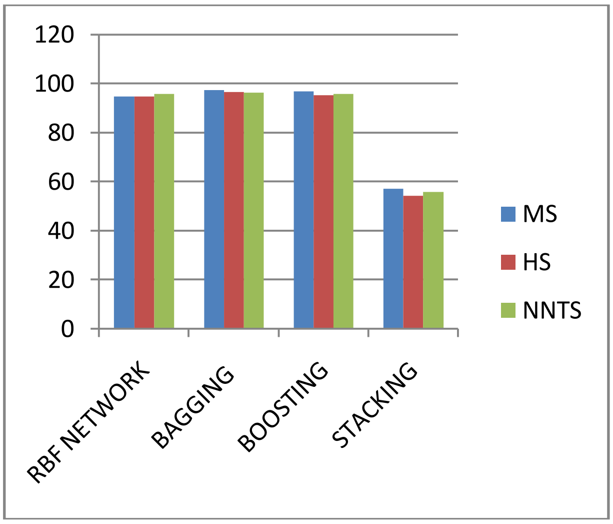

Fig 21

The overall result of the dataset being used is depicted in table 1 and table 2.

Table 1

S.NO	DATASET & ATTRIBUTES	STRATEGY	SELECTED FEATURES	RBF NETWORK	BAGGING	BOOSTING	STACKING
1	Wisconsin Breast Cancer(9)	MS	1,5	99.42	**99.57**	**99.57**	65.52
		HS	2,3,4,6,7,8,9	98.99	98.67	98.07	62.53
		NNTS	All	93.33	98.06	98.04	75.84
2	Pima Indian Diabetes (8)	MS	2,3	98.30	**100**	99.86	65.10
		HS	1,4,5,6,7,8	96.09	99.10	98.36	62.11
		NNTS	2,3	98.30	99.02	98.86	63.80
3	Heart Statlog (13)	MS	1,4,5,8	98.51	**99.62**	**99.62**	55.55
		HS	2,3,6,7,9,10,11,12,13	95.55	98.73	98.12	52.56
		NNTS	1,4,5,8	98.51	98.64	98.62	54.25
4	Hepatitis(19)	MS	1,15,16,18	84.51	89.67	89.67	79.35
		HS	2,3,5,6,7,8,9,10,11,12, 13,14,17,19,4	**90.32**	88.78	88.17	76.36
		NNTS	1,15,16,18	84.51	88.69	88.67	78.05
5	Cleveland Heart Disease(13	MS	4,5,8	98.34	98.67	98.67	3.63
		HS	1,2,3,7,9,10,11,12,13, 6	94.71	97.78	97.17	0.64
		NNTS	1,4,5,8	**99.00**	97.69	97.67	2.33
6	Chess(6)	MS	1,4,6	88.09	**100**	**100**	71.42
		HS	2,3,5	88.09	99.10	98.5	68.44
		NNTS	4,6	88.09	99.02	99.00	70.13
7	Grub Damage(8)	MS	2	96.77	**98.06**	**98.06**	68.38
		HS	1,3,4,5,6,7,8	83.87	97.16	96.56	65.40
		NNTS	2	96.77	97.08	97.06	67.08
8	Pasture(22)	MS	5,6,17,20,22	**100**	88.88	88.88	66.66
		HS	1,2,3,4,7,8,9,10,11,12, 13,14,15,16,18,19,21	80.55	87.99	87.38	63.68
		NNTS	2,3,4,5,6,10,13,16,17, 19,20,21,22	80.55	87.90	87.88	65.36
9	Squash stored(24)	MS	4,5,10,11,19,20,23,24	84.0	90.0	**98.0**	86.0
		HS	1,2,3,6,7,8,9,12,13,14, 15,16,17,18,21,22	86.0	89.10	96.5	83.01
		NNTS	All	86.0	89.02	97.0	84.70
10	Squash Unstored(23)	MS	4,9,10,18,19,22,23	82.69	**100**	98.07	5.76
		HS	1,3,5,6,7,11,12,13,14, 15,16,17,20,21,2	86.53	99.10	96.57	2.78
		NNTS	2,4,5,6,7,8,9,10,11,14, 15,16,17,18,19,20,21, 22,23,	78.84	99.02	97.07	4.47

Table 2

S. NO	DATASET & ATTRIBUTES	STRATEGY	SELECTED FEATURES	RBF NETWORK	BAGGING	BOOSTING	STACKING
11	Tic-tac-toe (9)	MS	2,4,5,7	96.07	**100**	**100**	56.86
		HS	1,3,6,8,9	88.23	99.103	98.5	53.87
		NNTS	All	87.30	95.84	95.82	59.01
12	White Clover(31)	MS	3,4,7,9,13,16,18, 20,23,26,27,29	87.30	**96.82**	**96.82**	60.31
		HS	1,2,5,6,8,10,11,12, 14,15,17,19,21,22 ,24,25,28,30,31	82.53	95.92	95.32	57.33
		NNTS	3,4,7,9,13,14,16,1 8,20,23,26,27,29	87.30	95.84	95.82	59.01
13	Balance(3)	MS	2,3	94.11	**100**	**100**	5.88
		HS	1	**100**	99.10	98.5	2.89
		NNTS	All	94.11	99.02	99.0	4.58
14	Car(6)	MS	3,4	97.31	96.50	96.50	68.66
		HS	1,2,5,6	94.89	95.61	95.0	65.68
		NNTS	All	93.33	**98.06**	98.04	75.84
15	Dermatology (34)	MS	34	93.33	**99.04**	**99.04**	77.14
		HS	34	95.23	98.15	97.54	74.15
		NNTS	34	93.33	98.06	98.04	75.84
16	Hypothyroid (29)	MS	1,22,26	**94.23**	**94.23**	**94.23**	**94.23**
		HS	All except 1	89.10	93.33	92.73	91.24
		NNTS	22,26	**94.23**	93.25	93.23	92.93
17	Eggs(3)	MS	2	**100**	93.75	95.83	2.08
		HS	1,3	87.5	92.85	94.33	0.92
		NNTS	All	91.66	96.93	96.91	63.28
18	Fiber(4)	MS	4	91.66	**97.91**	**97.91**	64.58
		HS	1,2,3	91.66	97.01	96.41	61.59
		NNTS	3,4	91.66	96.93	96.91	63.28
19	Ionosphere (34)	MS	All except 14	94.01	99.14	**99.43**	64.10
		HS	14,28	97.43	98.24	97.93	61.11
		NNTS	All	91.66	96.93	96.91	63.28
20	Liver Disorder(6)	MS	1,2	94.65	**97.32**	96.79	57.21
		HS	3,4,5,6	94.65	96.42	95.29	54.23
		NNTS	1,2,3,4	95.72	96.34	95.79	55.92

VIII. Conclusion

Feature selection aims to reduce the amount of unnecessary, irrelevant and redundant features. It helps retrieve the most relevant features in datasets and improves the classification accuracy with less computational effort. If the features are not chosen well, even the best classifier performs poorly. In this paper, we describe feature relevance measures based on fuzzy entropy values and devise three feature selection strategies: Mean

Selection, Half Selection and Neural Network Threshold Selection with an RBF Network classifier. The features selected using the above strategies is passed over RBF network, Bagging, Boosting and stacking to predict their accuracy. The intention is to select the correct set of features for classification when datasets contain noisy, redundant and vague information.

Twenty benchmark datasets from the UCI Machine Learning Repository from various fields like medicine, agriculture, sports and others are used for evaluation. The proposed feature selection strategies have produced accuracies that are acceptable or better when compared with the accuracy obtained for the entire feature set without any feature selection. Of all the proponents, the one that maximizes the accuracy is the fuzzy entropy with Mean Selection. It is also found that among the four methodologies used, Bagging yields highest accuracy in most of the cases. Thus, Bagging can be taken a Best case, Boosting and RBF network as Average case and Stacking as Worst case. In future, this can be applied to a wide range of problem domains with hybridization of different feature selection techniques to improve the performance of both the feature selection and the classification.

REFERENCE

[1] R.E. Abdel-Aal, GMDH based feature ranking and selection for improved classification of medical data, J.Biomed.Inform.38 (6)(2005)456–468.

[2] M.F.Akay, Support vector machines combined with feature selection for breast cancer diagnosis, Int.J.Expert Syst.Appl.36 (2)(2009)3240–3247.

[3] Chin-Yuan Fan,Pei-Chann Chang, Jyun-Jie Lin,J.C.Hsieh, A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification, Int.J.Appl.Soft Comput.11(1)(2011)632–644.

[4] Huan Liu, LeiYu, Toward integrating feature selection algorithms for classification and clustering, IEEE Trans.Knowl.Data Eng.17 (4)(2005)491–502.

[5] R. Kohavi, George H.John, Wrappers for feature selection subset selection, Artif. Intell.97 (1–2) (1997)273–324.

[6] II-Seok Oh, Jin-Seon Lee, Byung-RoMoon, Hybrid genetic algorithm for feature selection, IEEE Trans.Pattern Anal.Mach.Intell.26 (11)(2004)1424–1437.

[7] R. Battiti, Using mutual information for selecting features in supervised neural net learning, IEEE Trans.Neural Netw.5(4)(1994)537–550.

[8] Hanchuan Peng, Fuhui Long, Chris Ding, Feature selection based on mutual information: criterion of max-dependency, max-relevance and min- redundancy, IEEE Trans.Pattern Anal.Mach.Intell.27 (8) (2005)1226–1238.

[9] Pablo A.Extevez, Michel Tesmer,Claudio A.Perez, JacekM.Zurada, Normalized mutual information feature selection, IEEE Trans. Neural Netw.20 (2) (2009)189–201. [10] Lei Yu, Huan Liu, Efficient feature selection via analysis of relevance and redundancy, J.Mach.Learn.Res.5 (2004)1205–1224.

[11] Seral Sahan, Kemal Polat, Halife Kodaz, Salih Gunes, A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis, Int.J.Comput.Biol.Med.37(3)(2007)415–423.

[12] P.Jaganathan,K.Thangavel, A.Pethalakshmi, M.Karnan, Classification rule discovery with ant colony optimization and improved quick reduct algorithm,IAENGInt.J.Comput.Sci.33(1)(2007)50–55.

[13] Rahul Karthik Sivagaminathan, Sreeram Rama krishnan, A hybrid approach for feature subset selection using neural networks and ant colony optimization, Int. J.Expert Syst.Appl.33(1)(2007)49–60.

[14] Shih-Wei Lin, Zne-Jung Lee,Shih-Chieh Chen, Tsung-YuanTseng, Parameter determination of support vector machine and feature selection using simulated annealing approach,Int.J.Appl.Soft Comput.8(4)(2008)1505–1512.

[15] Shih-WeiLin, Shih-Chieh Chen, Wen-Jie Wu, Chih-Hsien Chen, Parameter determination and feature selection for back-propagation network by particle swarm optimization, Int.J.Knowl.Inf.Syst.21(2)(2009)249–266.

[16] Alper Unler, Alper Murat, A discrete particle swarm optimization method for feature selection in binary classification problems, Eur.J.Oper.Res.206(3) (2010)528–539.

[17] Pei-Chann Chang, Jyun-JieLin, Chen-Hao Liu, An attribute weight assignment and particle swarm optimization algorithm for medical database classification, Int. J.Comput.Methods Progr.Biomed.107(3)(2012)382–392.

[18] Maria Salamo, MaiteLopez-Sanchez, Rough set based approaches to feature selection for case-based reasoning classifiers, Int.J.Pattern Recognit.Lett.32 (2) (2011)280–292. [19] Sultan Noman Qasem, Siti Mariyam Shamsuddin, Radial basis function network based on time variant multi-objective particle swarm optimization for medical diseases diagnosis, Int.J.Appl. Soft Comput.11(1) (2011)) 1427–1438.