A Comparative Study of Improved F-Score with Support Vector Machine and RBF Network for Breast Cancer Classification

P. Jaganathan, N. Rajkumar, and R. Kuppuchamy

***Abstract*—Feature selection is an important issue in classification of cancer diagnosis. In this paper, a new feature selection method, named improved F-Score is applied for breast cancer diagnosis. First, the improved F-Score values of all the features are calculated using improved F-Score formula. Then the mean value is computed for the calculated improved F-Score values. The improved F-Score values which are greater than the mean improved F-Score are selected. Wisconsin breast cancer dataset (WBCD) is used in this study. As classification algorithms, Support Vector Machine and RBF Network are sued. The results obtained from improved F-Score with Support Vector Machines have produced efficient results compared to improved F-Score with RBF Network. Therefore we show that improved F-Score combined with promising than improved F-Score with RBF Network.**

***Index Terms*—Breast cancer, feature selection, improved F-Score, RBF network, SVM.**

I. Introduction

Cancer is a major health problem for the people worldwide and breast cancer is the most common cause of cancer deaths among women than any other type. It is reported that the incidence of breast cancer is rising in every country of the world. A most effective method for early diagnosis is very much essential than the existing methods to eliminate this deadly disease [6].

Feature selection is an important process in data pre-processing which is essential in classification task. All the features are not always relevant for the classification. Some features are irrelevant and redundant. These features ultimately affect the performance of the classification algorithms in terms of time and cost [4]. A good subset of features will always yield better results.

There are different kinds of feature selection methods namely Filter and Wrapper methods. Filter methods have low computational cost and there is no reliability classification task, where as wrapper methods have high computational cost with high reliability. In our work, the improved F-Score is used as an evaluation criterion for diagnosis of breast cancer. Support Vector Machine and RBF Network are used as the classification algorithms. It is observed that, the method improved F-Score and Support Vector Machines applied on WBCD shows improved classification accuracies than improved F-Score and RBF Network for various training-test partitions.

This paper is organized as follows. Section 2 summarizes the earlier researches on diagnosis of breast cancer. Section 3 explains the feature selection method. Section 4 describes about Support Vector Machine. Section 5 brings out the experimental observations. Section 6 analyzes the experimental results. Section 7 arrives at the conclusion.

II. Related Work

Support Vector Machine is an effective statistical method used in medical diagnosis for pattern recognition, machine learning and data mining (cortes and vapnik 1995) [15]. In the literature, there are some works related to breast cancer diagnosis. Among these, Abonyi and Szeifer (2003) using supervised fuzzy clustering techniques produced a classification accuracy of 95.57% [1]. Goodman, D.E., Boggess, L., & Watkins, A. produced three different results with three different methods such as Optimized-LVQ, Big LVQ, AIRS and accuracies 96.70%, 96.80%, 97.20% respectively [11]. Mehmet Faith Akay has proposed a feature selection method with F-score and support vector machines reaching a classification accuracy of 99.51% [8]. Nauck, D., & Kruse, R., proposed 95.06% of classification [10]. Logarithmic simulated annealing and perception algorithm applied by Albrecht obtained 98.80% [2]. Hamilton et al. (1996) using RIAC method obtained 95.50% classification accuracy [7]. With LDA techniques Ster and Dobnikar (1996) produced a classification accuracy of 96.80% [13]. Pena-Reyes and Sipper (1999) obtained classification accuracy of 97.36% using Fuzzy-GAI method [11]. Setiono (2000) using Neuro-rule 2a technique obtained classification accuracy of 98.10% [12]. With AR and NN Murat Karabatak & M.Cevdet Ince (2009) produced classification accuracy of 97.40% [9]. T. S. Subashini, V. Ramalingam, and S. Palanivel (2009) obtained 97.33% classification accuracy using RBFNN and SVM techniques [14].

III. Feature Selection

A. Introduction to Feature Selection

Feature selection is an optimization technique for reducing dimensionality of data in machine learning and pattern recognition. The main idea of feature selection is to select an optimal subset of input variables by removing features with little or no predictive information. In this paper, a new feature selection method, named improved F-Score is applied for breast cancer diagnosis. First, the improved F-Score values of all the features are calculated using improved F-Score

Manuscript received August 9, revised September 7, 2012.

The authors are with Department of Computer Applications, PSNA College of Engineering & Technology, Dindigul, Tamilnadu, India (email:rknpsna@gmail.com)

10.7763/IJMLC.2012.V2.227

Formula. Then the mean value is computed for the calculated improved F-Score values. The improved F-Score values which are greater than the mean improved F-Score are selected. The F-Score method and the improved F-Score methods are described below.

B. F-Score

F-score is a simple technique which measures the discrimination of two sets of real numbers. Given training vectors. Given training vectors x_k, k=1,2,….,m, if the number of positive and negative instances are n_+ and n_- respectively, then the F-score of the i^th feature is defined as

$$F_i = \frac{\left(\bar{x}_i^{(+)} - \bar{x}_i\right)^2 + \left(\bar{x}_i^{(-)} - \bar{x}_i\right)^2}{\frac{1}{n_+ - 1\sum_{k=1}^{n+}\left(x_{k,i}^{(+)} - \bar{x}_i^{+}\right)^2} + \frac{1}{n_- - 1}\sum_{k,i}^{n_-}\left(x_{k,i}^{(-)} - \bar{x}_i^{-}\right)^2} \tag{1}$$

where $\bar{x}_i, \bar{x}_i^{(+)}, \bar{x}$ are the average of the i^{th} feature of the whole, positive, and negative datasets, respectively. The numerator indicates the discrimination between the positive and negative sets, and the denominator indicates the one within each of the two sets. The larger the F-score is, the more likely this feature is more discriminative.

C. Improved F-Score

Improved F-Score is a technique which measures the discrimination among more than two datasets. Given training vectors x_k, k=1,2,…,m, and the number of datasets(l>=2), if the number of j^{th} dataset is n_j, j=1,2,…,l, then the improved F-score of the i^{th} feature is defined as

$$F_i = \frac{\sum_{j=1}^{l}\left(\bar{x}_i^{(j)} - \bar{x}_i\right)^2}{\sum_{j=1}^{l}\frac{1}{n_j - 1}\sum_{k=1}^{nj}\left(x_{k,i}^{(j)} - \bar{x}_i^{(j)}\right)^2} \tag{2}$$

where $\bar{x}_i, \bar{x}_i^{(j)}$ are the average of the i^{th} feature of the whole dataset and the j^{th} dataset respectively; $x_{k,i}^{(j)}$ is the i^{th} feature of the k^{th} instance in the j^{th} dataset, l is the dataset. The numerator indicates the discrimination between each dataset, and denominator indicates the one within each of dataset. The larger the improved F-score is, the feature is more likely to be discriminative (Xie & Wang, 2011).

IV. Support Vector Machines

Support vector machine is a technique for learning in pattern classification and non-linear regression, Pioneered by Cortes and Vapnik in 1995, Boser, Guyon, Vapnik in 1992 and modified by Vapnik in 1999.

Consider the problem of separating the set of training vectors belonging to two linearly separable classes,

$$(x_i, y_i), x_i \varepsilon R^n, y_i \varepsilon \{+1, -1\}, i = 1, \ldots\ldots, n, \tag{3}$$

where x_i is a real-valued n-dimensional input vector and y_i is a label that determines the class of x_i. A separating hyper plane is determined by an orthogonal vector w and a bias b, which identifies the points that satisfy

$$w, x + b = 0 \tag{4}$$

The parameters w and b are constrained by

$$\min\left|w, x_i + b\right| \geq 1 \tag{5}$$

A separating hyper plane in canonical form must satisfy the following constraints,

$$y_i\left(w, x_i + b\right) \geq 1, i = 1, 2, \ldots, n \tag{6}$$

The hyper plane that optimally separates the data is the one that minimizes

$$\Phi(w) = \frac{1}{2}(w, w) \tag{7}$$

Relaxing the constraints of (4) by introducing slack variables

$\xi_i \geq 0, i = 1, 2, \ldots\ldots, n$ becomes

$$y_i\left(w, x_i + b\right) \geq 1\xi_i, i = 1, 2, \ldots\ldots, n \tag{8}$$

In this case, the optimization problem becomes

$$\Phi(w, \xi) = \frac{1}{2}(w, w) + C\sum_{i=1}^{n}\xi_i \tag{9}$$

With a user defined positive finite constant C. The solution to the optimization problem in (7), under the constraints of (6), could be obtained in the saddle point of Lagrangian function

$$L(w, b, \alpha, \xi, Y) - \frac{1}{2}(w, w) + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\left|Y_i(w.x_i)\xi_i\right| - \sum_{i=1}^{n}Y_i\xi \tag{10}$$

where a_i>=0, $£_I$>=0, i=1,2,..,n are the Lagrange multipliers. The Lagrangian function has to be minimized with respect to w,b, and $£_I$. Classical Lagrangian duality enables the primal problem, (8), to be transformed into its dual problem, which is easier to solve. The dual problem is given by

$$\max\left[\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\alpha_i\alpha_j\Upsilon_i\Upsilon_j K\left(x_i, x_j\right)\right] \tag{11}$$

with constraints

$$\sum_{i=1}^{n}\alpha_i, \Upsilon_i = 0, 0 \leq \alpha_i \leq C, i = 1, 2, \ldots, n \tag{12}$$

This is a classic quadratic optimization problem, for which there exists a unique solution. According to the Kuhn-Tucker theorem of optimization theory (Bertsekas, 1995), the optimal solution satisfies

$$\alpha_i\left[\Upsilon_i\left(w, x_i + b\right) - 1\right] = 0, i = 1, 2, \ldots, n \tag{13}$$

has non-zero Lagrange multipliers if and only if the points x_i satisfy

$$\Upsilon_i\left(w, X + b\right) = 1 \tag{14}$$

These points are termed SV. The hyperplane is determined by the SV, which is a small subset of the training vectors.

Hence if α_i^* is the non –zero optimal solution, the classifier function can be expressed as

$$f(x) = sgn\left\{\sum_{i=1}^{n} \alpha_i \Upsilon_i \left(x_i, x+b\right)\right\} \tag{15}$$

When a linear boundary is inappropriate SVM can map the input vector into a high dimensional feature space. By defining a non-linear mapping, the SVM construct an optimal separating hyperplane in this higher dimensional space. usually non-linear mapping is defined as

$$\Phi(.):R^n \rightarrow R^{nh} \tag{16}$$

In this case, optimal function (11) becomes (15) with the same constraints

$$\max\left[\sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{j=1}^{n} \alpha_i \alpha_j \Upsilon_i \Upsilon_j K\left(x_i, x_j\right)\right] \tag{17}$$

where

$$K(x_i x_j) = \{\varphi(x_i).\varphi(x_j)\}$$

is the kernel function performing the non-linear mapping into feature space. The kernel function may be any of the symmetric functions that satisfy the Mercel conditions (Courant &Hilbert, 1953). The most commonly used functions are the Radial Basis Function (RBF)

$$K(x_i, x_j) = exp\{-\gamma|x_i - x_j|^2\} \tag{18}$$

and the polynomial function

$$K(x_i, x_j) = (x_i x_j + 1)^q, q = 1,2,\ldots., \tag{19}$$

where the parameters in (18) and (19) must be preset.

V. Experimental Observations

A. Wisconsin Breast Cancer Dataset

This dataset is taken from the UCI machine learning repository for our experiments. It is part of the collection of databases at the University of California, Irvine collected by Dr.William H. Wolberg (1989-91) at the University of Wisconsin-Madison Hospitals. There are 699 records in this database. Each record in the database has nine attributes. The aim of the dataset is to classify the presence or absence of breast cancer given the result of various medical tests carried out on a patient. This database includes 9 attributes. These features are (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, (9) Mitosis. The nine attributes are represented as an integer value between 1-10 and detailed in Table I. In this database, Two hundred and forty one records (65.5%) are malignant and four hundred and fifty eight records (34.5%) are benign [3].

In order to evaluate the efficiency of the method, performance measures like sensitivity, specificity were considered. The measures were compiled by the following units.

Classification Accuracy (%):

$(TP + TN) / (TP + FP + FN + TN)$

Sensitivity (%) = $TP / TP + FN \times 100$

Specificity (%) = $TN / FP + TN \times 100$

TABLE I: Attribute Description of WBCD

Attribute description	Values of attributes	Mean	Standard deviation
Clump thickness	1-10	4.42	2.82
Uniformity of cell size	1-10	3.13	3.05
Uniformity of cell shape	1-10	3.20	2.97
Marginal adhesion	1-10	2.80	2.86
Single epithelial cell size	1-10	3.21	2.21
Bare nuclei	1-10	3.46	3.64
Bland chromatin	1-10	3.43	2.44
Normal nucleoli	1-10	2.87	3.05
Mitoses	1-10	1.59	1.71

TABLE II: The Obtained Features from Improved F-Score

Method	**The number of original Features in input space**	**The number of reduced features with feature selection**
improved F-Score	9	4

TABLE III: Performance of the Classifier with Different Methods Using Ten-Fold Cross Validation

Method	Sensitivity	Specificity	Classification accuracy
improved F-score + SVM	97.161	92.531	95.565
improved F-score + RBF Network	95.196	95.435	95.278

In Fig. 1, the flowchart shows how the classification accuracy for breast cancer is determined. It demonstrates the computation of improved F-Score values which helps in discriminating the relevant and irrelevant features. First the improved F-Score of each feature is calculated and the mean F-Score value is determined. The features which are above the mean F-Score are selected for classification. With the selected features is passed to SVM classifier with different class validations. The outcome of this procedure has produces efficient results.

VI. Results and Discussion

In this paper, a novel feature selection method called as improved F-Score for breast cancer diagnosis is applied combined with Support Vector Machines and RBF Network. First, the improved F-Score has been computed. The features above the mean improved F-Score are selected for the process. Then the selected features have been used in the classification of benign and malignant cases. Table II shows the obtained reduced number of features after applying the mean improved F-Score selection criteria. We have used two different classification algorithms. i) Support Vector Machines and ii) RBF Network.

Table III shows the performance of the two classifiers. Sensitivity, Specificity, Classification accuracy are presented. Table IV shows the performance comparison of various training-test partitions with two different classifiers. 95.415% for 50-50% training-test partition, 95.357% for 60-40% training-test partition, 95.238% for 70-30% training-test partition, 96.528% for 80-20% training-test partition is obtained for support vector machine classifier. 95.415% for 50-50% training-test partition, 96.071% for 60-40% training-test partition, 95.238% for 70-30% training-test partition, 96.428% for 80-20% training-test partition is obtained for RBF Network classifier. The results here depicts that our method improved F-Score and Support vector machines for diagnosis of breast cancer produces far better result than improved F-Score and RBF Network.

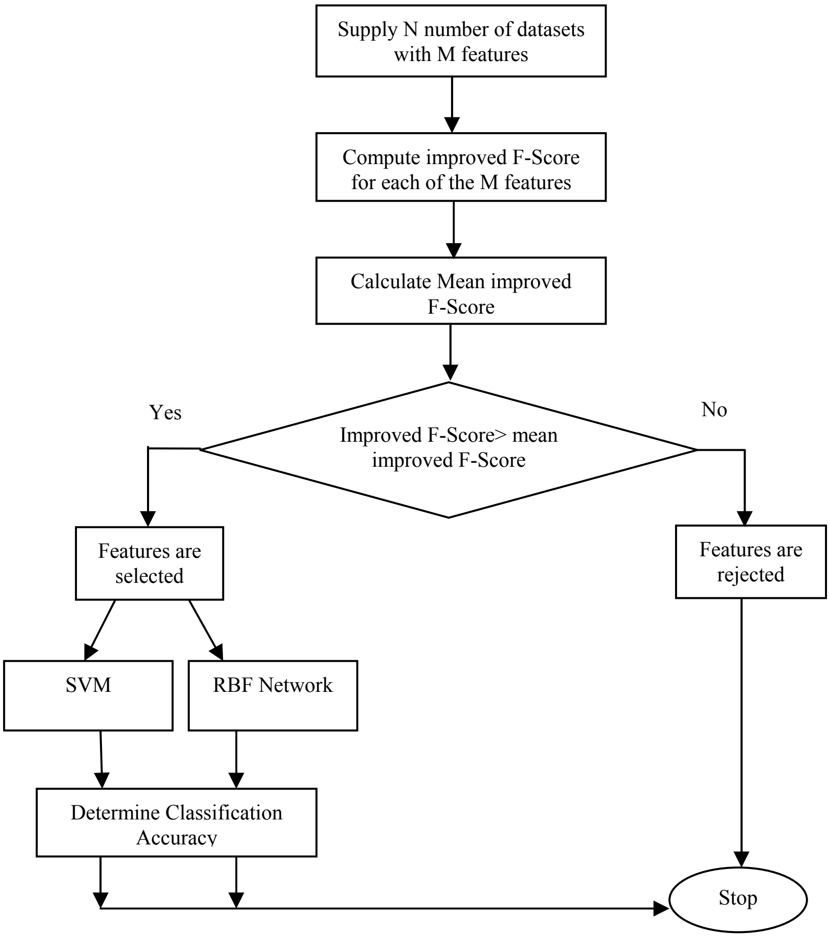
Supply N number of datasets with M features
Compute improved F-Score for each of the M features
Calculate Mean improved F-Score
Yes
No
Improved F-Score> mean improved F-Score
Features are selected
Features are rejected
SVM
RBF Network
Determine Classification Accuracy
Stop

Fig. 1. Flowchart of improved F-Score

TABLE IV: PERFORMANCE COMPARISON OF VARIOUS TRAINING-TEST PARTITIONS WITH DIFFERENT METHODS

Method	Classification accuracy			
	50-50% Training-test partition	60-40% Training-test partition	70-30% Training-test partition	80-20% Training-test partition
improved F-score + SVM	95.415	95.357	95.238	96.428
improved F-score + RBF Network	95.415	96.071	95.238	96.428

VII. CONCLUSION

Feature selection is an optimization technique for reducing dimensionality of data in machine learning and pattern recognition. The main idea of feature selection is to select an optimal subset of input variables by removing features with little or no predictive information.

In this article, improved F-Score feature selection method is applied with two different classifiers SVM and RBF Network for Wisconsin breast cancer dataset. In this study, improved F-Score and Support vector machines for diagnosis of breast cancer produces far better result than improved F-Score and RBF Network. The performance measurement criteria are classification accuracy, sensitivity– specificity values.

In this way, we have modelled a best expert system on the classification of WBCD datasets. In future, other medical datasets with real values and multiple classes can be used to evaluate this feature selection method. Also other measurement functions such as correlation can be used to measure the distinguishing between two or more classes can be used.

ACKNOWLEDGMENT

This work is supported by All India Council for Technical Education, New Delhi, under Research Promotion Scheme.

REFERENCES

[1] J. Abonyi and F. Szeifert, “Supervised fuzzy clustering for the identification of fuzzy classifiers,” *Pattern Recognition Letters*, vol. 14, no. 24, pp. 2195-2207, 2003.

[2] A. A. Albrecht, G. Lappas, S. A. Vinterbo, C. K. Wong, and L. Ohno-Machado, “Two applications of the LSA machine,” in *proceedings of the 9th international conference on neural information processing*, pp. 184-189, 2002.

[3] C. L. Blake, C. J. Mertz, UCI Repository of machine learning database, Irvine, CA: University of California, [Online]. Available: http://www.ics.uci.edu/mlearn/MLRepository.html.

[4] B. Cao, D. She, J.-T. Sun, Q. Yang, and Z. Chen, “Feature selection in a kernel space,” in *international conference on machine learning (ICML)* Oregon, USA, June 20-24, pp. 121-128, 2007.

[5] Goodman, D. E Boggess L. and A. Watkins, “Artificial immune system classification of multiple class problems,” in *proceeding of the artificial neural networks in engineering,* pp. 179-183, 2002.

[6] Cancer topics, [Online]. Available: http://www.cancer.gov/cancertopics/types/breast

[7] H. J. Hamiton, N. Shan, and Cercone, N, “RIAC: A rule induction algorithm based on approximate classification,” Technical Report CS pp. 96-06, University of Regina, 1996.

[8] M. F. Akay, “Support vector machines combined with feature selection for breast cancer diagnosis,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 3240-3247, 2009.

[9] M. Karabatak, M. Cevdet Ince, ”An expert system for deduction of breast cancer based on association rules and neural networks,” *Expert Systems with Applications*, vol. 36, no. 2, pp.3465-3469, 2009.

[10] D. Nauck, and R. Kruse, “Obtaining interpretable fuzzy classification rules from medical data,” *Artificial Intelligence in Medicine,* vol.17, pp.131-155, 1999.

[11] C. A. Pena-Reyes and M. Sipper, “A fuzzy-genetic approach to breast cancer diagnosis,” *Artificial Intelligence in Medicine*, vol. 17, pp. 131–155, 1999.

[12] R. Setiono, “Generating concise and accurate classification rules for breast cancer diagnosis,” *Artificial Intelligence in Medicine*, vol. 18, no. 3, 205–217, 2000.

[13] B. Ster and A. Dobnikar, “Neural networks in medical diagnosis: Comparison with other methods,” in *Proceedings of the international conference on engineering applications of neural networks,* pp. 427–430, 1996.

[14] T. S. Subashini, V. Ramalingam, and S. Palanivel, “Breast mass classification based on cytological patterns using RBFNN and SVM ,” *Expert Systems with Applications*, vol. 36, no. 3, pp. 5284-5290, 2009.

[15] V. Vapnik, The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995.

[16] J. Xie and C. Wang, “Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases,” *Expert Systems with Applications*, 2011.