DOI 10.3233/BME-151392
IOS Press

Tuning to optimize SVM approach for assisting ovarian cancer diagnosis with photoacoustic imaging

Rui Wang[a] , Rui Li[a*,] , Yanyan Lei[a] and Quing Zhu[b]
[a]*Key Laboratory of Precision Opto-Mechatronics Technology, Ministry of Education, The School of Instrumentation Science and Opto-Electronics Engineering, Beihang University, Beijing, 100191, China*
[b]*Department of Electrical and Computer Engineering, University of Connecticut, 371 Fairfield Way, U-2157, 06269 CT, Storrs, USA*

Abstract. Support vector machine (SVM) is one of the most effective classification methods for cancer detection. The efficiency and quality of a SVM classifier depends strongly on several important features and a set of proper parameters. Here, a series of classification analyses, with one set of photoacoustic data from ovarian tissues ex vivo and a widely used breast cancer dataset- the Wisconsin Diagnostic Breast Cancer (WDBC), revealed the different accuracy of a SVM classification in terms of the number of features used and the parameters selected. A pattern recognition system is proposed by means of SVM-Recursive Feature Elimination (RFE) with the Radial Basis Function (RBF) kernel. To improve the effectiveness and robustness of the system, an optimized tuning ensemble algorithm called as SVM-RFE(C) with correlation filter was implemented to quantify feature and parameter information based on cross validation. The proposed algorithm is first demonstrated outperforming SVM-RFE on WDBC. Then the best accuracy of 94.643% and sensitivity of 94.595% were achieved when using SVM-RFE(C) to test 57 new PAT data from 19 patients. The experiment results show that the classifier constructed with SVM-RFE(C) algorithm is able to learn additional information from new data and has significant potential in ovarian cancer diagnosis.

Keywords: Support vector machines, ovarian cancer detection, feature selection, correlation filter, SVM-RFE

1. Introduction

Ovarian cancer is the fifth leading cause of cancer-related death among women, and has the highest mortality of gynecologic cancer due to the lack of precancerous symptom in early stage [1]. Efficient technologies to detect and diagnose ovarian cancer at a pre-cancer level would produce a considerable impact on improving the cure rate. Photoacoustic tomography (PAT), an emerging imaging modality characterized by both excellent optical contrast and good ultrasound resolution, provides a huge potential in ovarian cancer diagnosis noninvasively via transvaginal approach [2]. The principle and the construction of the PAT imaging system have been reported in previous publications [3]. However,

[*] Address for correspondence: Rui Li, The School of Instrumentation Science and Opto-Electronics Engineering, Beihang University, Beijing, 100191, China. Tel.: +86 010-82338896; Fax: +86 010-82338896; E-mail: lirui_buaa@163.com.

0959-2989/15/$35.00 © 2015 – IOS Press and the authors.

distinguishing a malignant tumor from a benign one is still a challenging task due to the multitude of variants and the lack of visual identifiable vascular distribution features between the two.

It is necessary and urgent to extract describable features from ovarian PAT images for cancer diagnosis with computer aided detection (CAD) systems, which have been demonstrated a positive result on early cancer detection [4, 5]. Unfortunately, literature is scarce referring to utilization of CAD in ovarian cancer detection. Concentrated on the imaging patterns, power distribution over the spatial frequency, and spatial statistical properties, Umar Alqasemi etc. [1] concluded that SVM with a polynomial kernel, which was able to exclusively separate cancerous cases from non-cancerous, outperformed other classification methods such as the generalized linear model and neural network for ovarian cancer detection. Nevertheless, they did not investigate the design and feature selection for SVM classifier, which might improve contribute to the classification accuracy.

Hence, with a consensus that the SVM is a distribution-free algorithm that can overcome poor statistical estimation [6, 7], and the radial basis function (RBF) is an universal kernel function that can be applied to any of the distribution of the samples through the choice of parameters [6, 7], we focus on an automatic design, which couples feature selection and optimizing parameters based on SVM with RBF, to construct an improved better SVM classifier system for ovarian cancer detection. A new feature selection approach is proposed with the combination of the Support Vector Machines Recursive Feature Elimination (SVM-RFE) [8] and a linear correlation filter to obtain an optimal feature subset. Moreover, the optimal values of SVM parameters are derived using grid searching based on 10-fold cross validation (10-CV). Five metrics are adopted to evaluate the performance of SVM classifiers, while Receiver Operating Characteristic (ROC) curve and Area Under ROC Curve (AUC) are additionally calculated to proof the robustness of our proposed approach.

The remainder of this paper is organized as follows: Section 2 presents a review of the principle of SVM and Section 3 discusses the proposed automatic design of SVM. Section 4 describes the datasets used and gives the experimental results. Discussion and conclusion have been given in Section 5.

2. Support vector machine

Support vector machine (SVM) [9] is a pattern classification technique developed by Vapnik, et al. based on statistical learning theory. It targets on minimizing the structural risk and uses kernel function to tackle nonlinearly separable problem [6]. For a binary classification, let $T=(x_1,x_2,\ldots,x_L)^{\mathrm{T}}$ denotes a training set with L samples, in which each sample $x_i \in \mathrm{R}^N$ (i=1,2,…,L) is a N-dimension vector, also called feature vector. Let $y_i \in \{-1,+1\}$ denotes the target class label of x_i, then SVM attempts to find an optimal hyperplane maximizing the margin between it and input data. Then, the optimization problem can be solved by its dual form with Lagrange multipliers $\alpha=(\alpha_1,\alpha_2,\ldots,\alpha_L)^{\mathrm{T}}$ [9]:

$$\min\{\frac{1}{2}\sum_{i=1}^{L}\sum_{j=1}^{L}\alpha_i\alpha_j y_i y_j K(x_i \cdot x_j) - \sum_{i=1}^{L}\alpha_i\} \tag{1}$$

$$s.t.\ \sum_{i=1}^{L}\alpha_i y_i = 0,\ 0 \le \alpha_i \le 1, i = 1,2,\cdots,L\ , \tag{2}$$

where $K(x_i \cdot x_j)$ =exp(-0.5$\|x_i$-$x_j\|^2/\sigma^2$) is the RBF kernel function that maps a feature vector into a high dimension space and C is a regularization constant penalizing the training errors. In this paper, we focus on constructing SVM based on RBF kernel due to its flexibility and accuracy [9].

3. Proposed method

The quality of SVM depends strongly on both its kernel parameters and input features. The parameters determines the distribution of training data and controls the tradeoff between confidence range and experiential risk [9]. Also, an efficient feature selection is important in improving classification precision and computing complexity [10]. Thus we propose hereon a methodology on SVM that can with automatically selecting feature subsets and then tuning SVM parameters under the selected features. The scheme of the ensemble SVM design is presented in Figure 1. It illustrates the procedure about how the feature selection (FS) module produces an optimal feature subset under which the Parameter Estimation (PE) module can estimate appropriate SVM parameters.

3.1. Feature selection

Methods of feature selection can be approximately categorized into two types: Filter and Wrapper [10]. The former measures the performance metric of attributes only based on intrinsic characteristics of data, while the latter attempts to seek features yielding the best classification accuracy [10]. Compared with the Filter method, therefore, Wrapper approach might be less computationally efficient but available for better accuracy [10]. A common wrapper-based method is Support Vector Machine Recursive Feature Elimination (SVM-RFE), proposed by Guyon, et al. [8], for selecting genes and now it is widely used for feature selection. It starts with the full feature set, and eliminates recursively the least important feature at each step. Usually, SVM-RFE could generate a good feature subset for classification. However, it does not take the redundancy among features into account.

Hence, a novel two-stage feature selection approach, called SVM-RFE(C), is proposed in FS module, shown in Figure 1. In the first stage, all features are ranked in terms of importance for classification by SVM-RFE [8], and then, a linear correlation filter is applied to identify and discard high redundant features. Specially, the support vector count [7], denoted by $Sr{=}N_{\mathrm{SV}}/L$ (where N_{SV} and L is the number of support vectors and the total input samples respectively), is employed as the feature ranking criterion in SVM-RFE. It bounds the possible error of removing a feature (e.g., *fi*) from the current set S and is given by the average of 10-fold cross validation (10-CV). And, the Pearson correlation coefficient is used to measure the correlation between two features f_i, f_j:

$$r(f_i,f_j)=\frac{Cov(f_i,f_j)}{\sigma_{f_i}\sigma_{f_j}}=\frac{E[(f_i-\mu_{f_i})(f_j-\mu_{f_j})]}{\sqrt{E(f_i^2)-E^2(f_i)}\sqrt{E(f_j^2)-E^2(f_j)}}\ , \tag{3}$$

where μ_{f_i}, $E(f_i)$ and $\mathrm{E}(f_i^2)$ are the mean, expectation and square expectation of f_i.

3.2. SVM parameter estimation

Two parameters need to be estimated in RBF-based SVM: the width of RBF σ and regularization constant C. Ideally we would like to choose σ and C minimizing the true risk of the SVM classifier. Unfortunately, since this quantity is not accessible, one has to build estimates or bounds for them [7]. Here, a kind of grid-search [7, 11] is implemented to obtain the best (σ, C). Giving a range of different combinations of (σ, C), a grid is obtained. And the task is to pick the best one maximizing the accuracy of SVM [11] on the grid. To locate the best (σ, C) as possible, the seeking is conducted in two steps. First a coarse grid is exploited with a larger searching step to generate a suboptimal (σ, C).

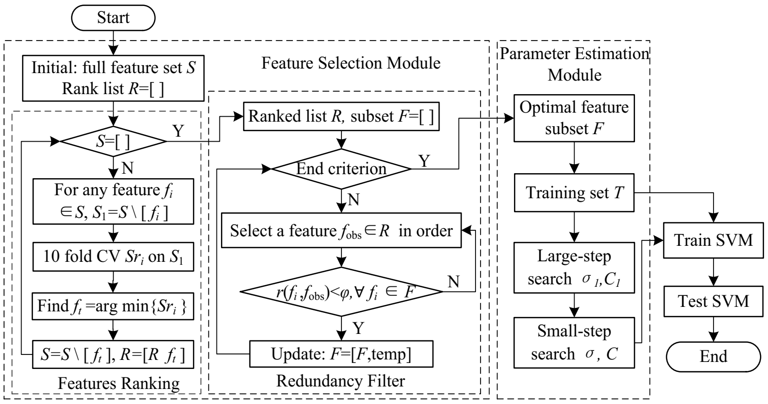
Start
Initial: full feature set S
Rank list R=[ ]
Feature Selection Module
S=[ ]
Y
N
For any feature f_i $\in S$, S_1=$S \setminus [f_i]$
10 fold CV Sr_i on S_1
Find f_t =arg min{Sr_i }
S=$S \setminus [f_t]$, R=[R f_t]
Features Ranking
Ranked list R, subset F=[ ]
End criterion
Y
N
Select a feature $f_{\text{obs}} \in R$ in order
$r(f_i, f_{\text{obs}})<\varphi, \forall f_i \in F$
N
Y
Update: F=[F,temp]
Redundancy Filter
Parameter Estimation Module
Optimal feature subset F
Training set T
Large-step search σ_1,C_1
Small-step search σ, C
Train SVM
Test SVM
End

Fig. 1. The general flowchart of the proposed SVM scheme.

Then, the best values of (σ, C) would be derived on a finer grid with a reduced step.

Once an optimal feature subset is achieved, the corresponding SVM parameters could be estimated with the grid-search on the basis of 10-CV. Thus an optimal SVM classifier would be trained on the training set and could be evaluated by some metrics on the testing set.

4. Experiments and results

4.1. Datasets and evaluation metrics

In order to evaluate our methods, this paper first conducted comparative experiments between SVM-RFE(C) and SVM-RFE on UCI benchmark database, Wisconsin Diagnostic Breast Cancer (called as WDBC) [12].This dataset consists of 569 pathological samples and 30 features.

Then, a series of experiments were carried out via a dataset (called OvaryPAT) including 169 PAT data from 19 ex vivo ovaries , which was obtained with Photoacoustic imaging system reported in [1]. 27 features, which were summarized in Table 1, were extracted with the principle similar to that described in [1, 13]. In this data set, 112 are from malignant ovarian tissues and 57 from benign ones.

The two datasets (WDBC and OvaryPAT) were divided randomly into training set (about 2/3 of all used samples) and testing set. Furthermore, as mentioned in section 1, five common metrics [14, 15]: accuracy (Acc), sensitivity (Se), specificity (Sp), positive predict value (PPV) and negative predict value (NPV) were used to evaluate the performance of different SVM classifiers. Moreover, ROC

Table 1

Distribution and ranking of 27 features on OvaryPAT (the selected features by SVM-RFE(C) at φ=0.6 were in bold)

Feature category	Feature index	Ranking in descending order of significance
(I) Statistical parameters of PAT image and its envelope data	1-7	**10**, 9, **13**, **23**,12, **26**, **24**, **8**, **7,** 25, **11**, 21, 18, 20, **27**, 19, **4**, 17, 3, 16, 14, **2**, 15, **6**, 1, 22, 5
(II) Peak outputs of 3spatial filters (for malignant template, normal template resp.)	8-13	
(III) Texture features of PAT image	14-23	
(IV) Spectral parameters of PAT beams	24-27	

curves and AUC [15] were calculated to further quantify the proposed algorithm.

4.2. Feature selection and parameter estimation

To verify the effectiveness and feasibility of the proposed algorithm, all features are ranked firstly by SVM-RFE, and then the relationship between the support vector count and the number of selected features is observed to determine the best number of features for SVM-RFE. Finally a suitable value of φ is decided for SVM-RFE(C) to acquire a similar size feature subset.

The procedure above would be conducted on both WDBC and OvaryPAT. Table 1 listed the ranking results of all features on OvaryPAT using SVM-RFE. Then the curves of support vector count with respect to feature number on both datasets were generated respectively, shown in Figure 2. And the first minimums of the two curves were also marked with purple '*'. Evidently, the curves decline initially and then fluctuate repeatedly or rise slightly along with the increase of features. As illustrated in Section 3.1, a lower support vector count means a smaller classification error, thus one could conclude the best feature number from the curves. Consequently, the top 14 features on WDBC and 8 features on OvaryPAT were obtained by SVM-RFE respectively.

Furthermore, the number of features that might be identified as redundancy by SVM-RFE(C) was collected in Figure 3. On WDBC, the number of 14, exactly coincides with the numbers at asterisk in Figure 2, is generated for the non-redundant features at φ =0.8. Thus the 14 were selected as the number for uncorrelated features. On OvaryPAT, however, the number of non-redundant features is always larger than 8 when φ varied from 0.4 to 0.9. Furthermore, all of the 11 selected non-redundant features at φ =0.6 were listed in Table 1. Obviously, some of them seemed to be at the back in the ranking list of significance. Thus, a tradeoff between significance and redundancy was decided. With the golden section ratio 0.618 for φ, the top 8 non-redundant features were finnally selected.

Under the feature subsets obtained by the two algorithms, the corresponding SVM parameters were estimated. And for the two subsets obtained by SVM-RFE(C), the best (σ, C) were logσ =0.5, log$C$=1.5 on WDBC and logσ =0, log$C$=1 on OvaryPAT, repspectively.

4.3. Comparison and analysis

From section 4.2, different SVM classifiers obtained by RFE and RFE(C) respectively were built on two datasets. The ROC curves of SVMs were illustrated in Figure 4, and Table 2 listed the test performance as well as the AUC value of different SVMs trained with feature subsets over each dataset. The performance of SVM built on every full feature set was also derived for comparison.

In Table 2, SVM-RFE(C) outperforms SVM-RFE in kinds of aspects. Firstly, the SVMs generated

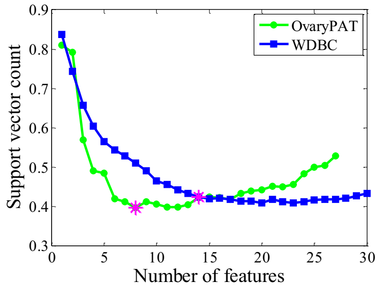

Fig. 2. The support vector count versus the number of features after ranking.

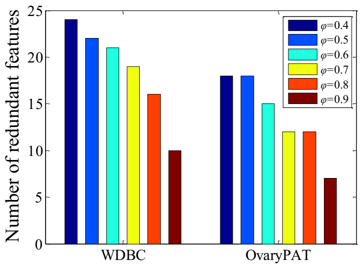

Fig. 3. Number of features might be filtered with different values of φ.

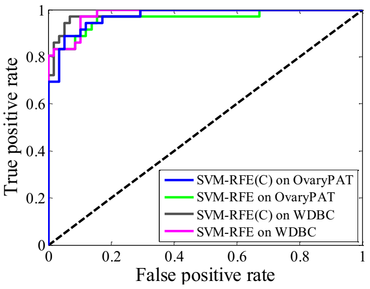

Fig. 4. ROC curves of feature subsets obtained by two methods on datasets.

Table 2

Comparison results of different feature selection methods on WDBC and OvaryPAT

Data	Methods	Size	Acc (%)	Se (%)	Sp (%)	PPV (%)	NPV (%)	AUC
WDBC		30	94.681	88.889	98.276	96.970	93.443	0.98994
	SVM-RFE(C)	14	**97.197**	**94.286**	**100.000**	**100.000**	**96.067**	0.98412
	SVM-RFE	14	92.059	85.758	94.376	92.667	90.063	0.97932
OvaryPAT		27	89.286	91.892	84.211	91.892	84.211	0.97582
	SVM-RFE(C)	8	**94.643**	**94.595**	**94.737**	**97.222**	**90.000**	**0.97318**
	SVM-RFE	8	92.857	91.892	94.737	97.143	85.714	0.96454

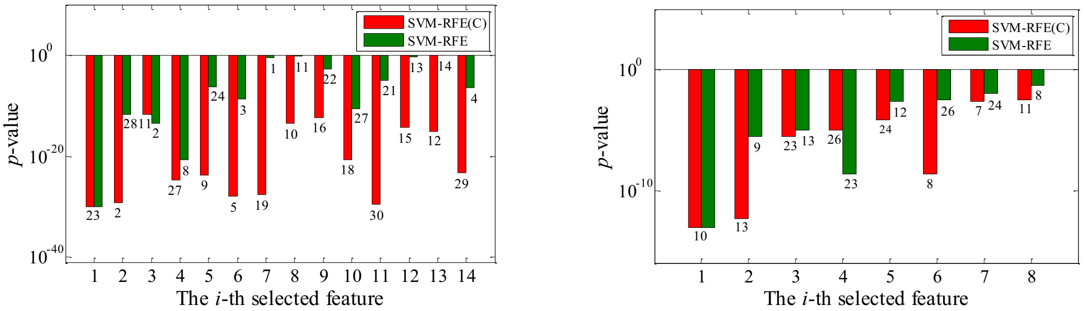

Fig. 5. p-values of the selected feature by SVM-RFE(C) and SVM-RFE on WDBC and OvaryPAT. The index of each selected feature was marked below each bar.

by SVM-RFE(C) achieve the best accuracy of 97.197% and 94.643% on two datasets, respectively, which outstandingly improves the classification performance of the full feature set. Besides, the sensitivity and specificity of SVM classifier constructed by SVM-RFE(C) on OvaryPAT also touched 94.595% and 94.737% respectively, showing a rather high performance. Furthermore, the SVMs constructed using SVM-RFE(C) held larger AUC values than that obtainted using SVM-RFE. This also suggested an improvement of robustness in constructing SVM classifiers with SVM-RFE(C).

The superiority of SVM-RFE(C) was also confirmed by the following statistical analysis. On each dataset, the p-values of features selected by SVM-RFE(C) and SVM-RFE were respectively calculated with t-tests, shown in Figure 5. Generally, most features selected by SVM-RFE(C) achieved relatively high confidence probability to distinguish malignant from benign tissues. On both two datasets, the p-values of all features selected by SVM-RFE(C) were lower than 0.05, showing a remarkable statistical significance for differentiating malignant from benign samples. On the contrary, some of features picked by SVM-RFE produced relative large p-values. These results verified that SVM-RFE(C) is more reliable and practical than SVM-RFE does from a statistical standpoint.

5. Conclusion

This study explored an SVM-based automatic recognition algorithm for ovarian cancer diagnosis using photoacoustic imaging. The proposed approach named SVM-RFE(C) can automatically tune the kernel parameters of a SVM with RBF and the selection of features by adding a correlation filter into SVM-RFE.SVM-RFE. Experiments on WDBC confirmed that SVM-RFE(C) achieved better performance and robustness than that of SVM-RFE. The 8-feature SVM classifier obtained by SVM-RFE(C) exhibited the best accuracy of 94.643% with the highest AUC of 0.97318 on OvaryPAT. The conclusions strongly supported that the architecture can fuse the outputs of different classifiers for a more specificity and comprehensive output. The successful utilization of the constructed SVM system means that it could be used for diagnostic purposes of ovarian tumors using photoacoustic data.

Acknowledgments

This research is supported in part by the National Cancer Institute 1R01CA151570 and sponsored by China Scholarship Council (CSC). Specially, we appreciate Mr. Hai Li for providing great assistance in data and image collection.

References

[1] U. Alqasemi, P. Kumavor, A. Aguirre, et al., Recognition algorithm for assisting ovarian cancer diagnosis from coregistered ultrasound and photoacoustic images: Ex vivo study, Journal of Biomedical Optics **17** (2012), 126003–126003.

[2] A. Aguirre, Y. Ardeshirpour, M.M. Sanders, et al., Potential role of coregistered photoacoustic and ultrasound imaging in ovarian cancer detection and characterization, Translational Oncology **4** (2011), 29–37.

[3] H. Li, P.D. Kumavor, U. Alqasemi, et al., Classification algorithm of ovarian tissue based on co-registered ultrasound and photoacoustic tomography, Photons Plus Ultrasound: Imaging and Sensing, San Francisco, CA, United states, 2014, pp. 894349.

[4] L. Bogoni, P. Cathier, M. Dundar, et al., Computer-aided detection (CAD) for CT colonography: A tool to address a growing need, The British Journal of Radiology **78** (2005), S57–S62.

[5] Y. Rejani and S.T. Selvi, Early detection of breast cancer using SVM classifier technique, International Journal on Computer Science and Engineering **1** (2009), 127–130.

[6] S.S. Keerthi and C.-J. Lin, Asymptotic behaviors of support vector machines with Gaussian kernel, Neural Computation **15** (2003), 1667–1689.

[7] O. Chapelle, V. Vapnik, O. Bousquet, et al., Choosing multiple parameters for support vector machines, Machine Learning **46** (2002), 131–159.

[8] I. Guyon, J. Weston, S. Barnhill, et al., Gene selection for cancer classification using support vector machines, Machine Learning **46** (2002), 389–422.

[9] C. Cortes and V. Vapnik, Support-vector networks, Machine Learning **20** (1995), 273–297.

[10] Y. Saeys, I. Inza and P. Larrañaga, A review of feature selection techniques in bioinformatics, Bioinformatics **23** (2007), 2507–2517.

[11] C.-W. Hsu, C.-C. Chang and C.-J. Lin, A practical guide to support vector classification, Talk at University of Freiburg, July 15, 2003.

[12] W.N. Street, W.H. Wolberg and O.L. Mangasarian, Nuclear feature extraction for breast tumor diagnosis, Biomedical Image Processing and Biomedical Visualization **1905** (1993), 861–870.

[13] M. Hussain, S.K. Wajid, A. Elzaart, et al., A comparison of SVM kernel functions for breast cancer detection, 2011 Eighth International Conference on Computer Graphics, Imaging and Visualization (CGIV), 2011, pp. 145–150.

[14] R. Wang, Z. Zhu and L. Zhang, Improving scale invariant feature transform-based descriptors with shape–color alliance robust feature, Journal of Electronic Imaging **24** (2015), 033002.

[15] H.A. Guvenir and M. Kurtcephe, Ranking instances by maximizing the area under ROC curve, IEEE Transactions on Knowledge and Data Engineering **25** (2013), 2356–2366.