Comparative Performance of State-of-the-Art Classifiers in Computer-Aided Detection for CT Colonography

Sang Ho Lee, Janne J. Näppi, and Hiroyuki Yoshida

3D Imaging Research, Department of Radiology,
Massachusetts General Hospital and Harvard Medical School,
25 New Chardon St., Suite 400C, Boston, Massachusetts 02114, USA
{lee.sangho,yoshida.hiro}@mgh.harvard.edu, jnappi@partners.org

Abstract. Several effective machine learning and pattern recognition schemes have been developed for medical imaging. Although many classifiers have been used with computer-aided detection (CAD) for computed tomographic colonography (CTC), little is known about their relative performance. This pilot study compares the performance of several state-of-the-art classifiers and feature selection methods in the classification of lesion candidates detected by CAD in CTC. There were four classifiers: linear discriminant analysis (LDA), radial basis function support vector machine (RBF-SVM), random forests (RF), and gradient boosting machine (GBM). There were five feature selection methods: sequential forward inclusion (SFI) of principal components (PCs), univariate filtering (UF), UF of PCs, recursive feature elimination (RFE), and RFE of PCs. A strategy of using all available features was tested also. For evaluation, 232,211 detections by a CAD system on 1,211 patients were subsampled randomly to create 10 different populations of 500 true-positive (TP) and 500 false-positive (FP) detections. The classifier performance was evaluated by use of the area under the receiver operating characteristic curve of 3 repeated 10-fold cross-validations. According to the result, the discrimination performance of the RBF-SVM classifier with feature selection by the RFE of PCs compared favorably with other methods, although no single classifier outperformed other classifiers under all conditions and feature selection schemes.

Keywords: Classification, feature selection, comparative performance, machine learning, virtual colonoscopy.

1 Introduction

Computed tomographic colonography (CTC) is a promising alternative to traditional invasive colonoscopy methods used in the detection and removal of polyps of the colon [1-3]. Computer-aided detection (CAD) systems for CTC typically make use of a classifier to discriminate between true-positive (TP) and false-positive (FP) findings generated by a polyp candidate detection system based on a set of features extracted

H. Yoshida, D. Hawkes, M.W. Vannier (Eds.): Abdominal Imaging 2012, LNCS 7601, pp. 78–87, 2012.

from the candidates [4-6]. However, CAD systems for CTC still display large numbers of FP detections [7]. Consequently, improving the detection specificity of CAD remains a challenging task in CTC, and a powerful classification engine is needed to deal with this difficult classification problem [8, 9].

The aim of a classification system is to classify an input pattern presented to the system to a correct category based on a feature vector of the input pattern. The complexity of the classification problem relies on the variability of the feature values for patterns in the same class relative to the difference between feature values for patterns in different classes. As a result, the optimality of a classifier depends on a specific dataset [10]. Thus, the goal of achieving the optimal performance for a pattern recognition system may be inconsistent with obtaining the best performance for a single classifier, which may also be associated with different feature selection schemes. This pilot study compared several state-of-the-art classifiers and feature selection schemes by using a large database in the classification task for CAD in CTC.

2 Method

2.1 Feature Selection

The goal of feature selection is to select a subset of relevant features for building robust classifiers by removing irrelevant and redundant features from input data. This is expected to improve the speed of construction and the accuracy of the final classifier.

From a theoretical perspective, it can be shown that optimal feature selection for supervised learning problems requires an exhaustive search of all possible subsets of features. However, for a large number of features or samples, an exhaustive search for an optimal feature set is impractical. Therefore, instead of an optimal set, in practice a supervised learning algorithm searches for a satisfactory approximation of the optimal set of features for a particular classifier.

In this study, three principal state-of-the-art methods were considered for feature selection, including 1) principal component analysis (PCA) [11], 2) univariate filtering (UF) [12], and 3) recursive feature elimination (RFE) [13].

Principal Component Analysis. The PCA is a well-established method for feature extraction and dimensionality reduction. It is based on the assumption that most information about features is contained in the directions along which the variation of the features is largest. The most common derivation of PCA is a standardized linear projection, which maximizes the variance in the projected space.

Univariate Filtering. UF is a feature selection method that reviews the features by using univariate statistical methods, such as the *t*-test or ANOVA models, to assess the efficacy of each individual feature in class prediction. UF is relatively dominantly used because of its simplicity and efficiency. However, it does not take into account feature-feature interactions, possibly leading to less accurate classifiers. UF is based on including the highest-ranked individual features depending on a chosen association measure. Since UF applies independent evaluation criteria without the process of discovering patterns in data, it does not inherit any bias of a learning algorithm and it is also computationally efficient. UF is preferred in applications where application of data mining algorithms would be too costly or unnecessary in dealing with high dimensional features.

Recursive Feature Elimination. RFE is a multivariate approach based on the information content of a group of features, which uses successive elimination of individual features ranked lowest according to a criterion, aimed at keeping the discrimination ability as high as possible. It attaches a weight to each available feature. Based on the assumption that the features with the smallest weights are least informative in a feature set, a predefined number of features is removed iteratively from the set of available features. RFE involves combinatorial searches through the space of feature subsets, guided by the prediction ability of a specific classification model. Since grouping and predictive analysis of multidimensional features are used to control the selection of feature subsets, RFE tends to give superior performance as feature subsets found are better suited to the predetermined learning algorithm. Consequently, it is more computationally expensive than the UF.

In this study, a total of five feature selection methods derived from PCA, UF, and RFE were considered: 1) sequential forward inclusion (SFI) of the principal components (PCs) of PCA, 2) UF, 3) UF of PCs, 4) RFE, and 5) RFE of PCs. For comparison, also a strategy of using all available features without explicit feature selection was considered.

2.2 Classification

The goal of classification is to identify the correct category of an input pattern. The classification is typically based on an initial training set of samples whose category is known.

In this study, the following state-of-the-art classifiers were considered: 1) linear discriminant analysis (LDA) [11], 2) radial basis function (RBF) support vector machine (RBF-SVM) [14], 3) random forest (RF) [15, 16], and 4) gradient boosting machine (GBM) [17, 18]. Each classifier, except for LDA, evaluated the effect of its model tuning parameters by using resampling. Optimal tuning parameters were chosen across those parameters. Finally, the classification performance was estimated from a training set.

Linear Discriminant Analysis. LDA is a robust and fundamental classifier. It is used for finding an optimal transformation that maps input data into a lower dimensional space to minimize the within-class distance and simultaneously to maximize the between-class distance, thus achieving maximum discrimination. LDA is closely related to PCA in that both look for linear combinations of features which best explain the data. LDA attempts to model the difference between classes of data explicitly, whereas the PCA does not consider differences between classes.

Support Vector Machine. The SVM is based on the concept of decision planes that define boundaries. A decision plane is one that separates between a set of features having different class memberships. The classification is based on separating hyperplanes that distinguish between objects of different class memberships in a multi-dimensional space. The basic idea behind the SVM is to create nonlinear boundaries by generating linear boundaries on a higher-dimensional space, where the original features are rearranged by use of a set of mathematical functions known as kernels. There are a number of kernels that can be used in SVM models, including linear, polynomial, sigmoid, and RBF kernels. The RBF-SVM is the most popular choice among the kernel types used in the SVM.

Random Forest. The RF classifier is an ensemble of decision trees, which combines the predictions of many classification trees to obtain more accurate classifications. Many samples of the same size as the original dataset, called bootstrap samples, are drawn from the dataset with replacement. In each bootstrap sample, approximately 68% of the observations in the original dataset occur one or more times. The observations in the original dataset that do not occur in the bootstrap sample are said to be out-of-bag for that bootstrap sample. For each bootstrap sample, a decision tree is built. At each step of the building process, only a small number of variables are available for construction of the decision tree. There is no pruning of the decision trees of a RF classifier. The trees of the RF are then used for constructing predictions for all out-of-bag observations of bootstrap samples. The predicted class of an input sample is acquired by voting for the predicted class among all the trees.

Gradient Boosting Machine. Boosting is a process that combines many separate prediction rules, some of which may be quite weak on their own, to produce a more powerful combined classifier. The GBM is another procedure that, like the RF, fits many trees to a single dataset. The GBM differs from the RF in that the trees are built sequentially, with observation weights updated according to whether the observations are correctly or incorrectly classified. Boosting iteratively adds basis functions in a greedy fashion such that each additional basis function further reduces the selected loss function. The GBM is one of the more novel classifiers that, to date, has rarely been applied in the analysis of medical images.

2.3 Materials and Evaluation

The empirical data for this study included potential lesion candidates detected by a CAD system [19] from a large clinical CTC screening population of 1,211 patients at 20 medical centers [20]. The patients were prepared cathartically for the CTC examination. Orally administered fecal tagging was used for 37% of the patients. The CTC data were acquired by use of 11 CT scanners with an average slice thickness of 2.35 mm (range, 1.0 – 5.0 mm) and average current of 156 mA (range, 50 – 408 mA). Approximately 18% of the patients had clinically significant colonoscopy-confirmed lesions. There were 317 lesions $\geq$6 mm: 40% of the lesions measured $\geq$10 mm and

60% measured 6 – 9 mm in the largest diameter. Approximately 84% of the lesions had polypoid morphology and 16% had flat morphology.

There were 232,211 CAD detections, including 929 TP detections and 231,282 FP detections. Because some of the lesions were detected multiple times, the number of CAD detections is higher than that of confirmed true lesions in the patients. The detections were sampled randomly without replacement for construction of 10 population samples for an unbiased evaluation of classifier performance under various conditions. Each subsample contained 500 TP and 500 FP CAD detections. Each detection was characterized by a total of 67 shape and texture features.

The classifier performance was evaluated by use of three repeated 10-fold cross-validations, where the performance was measured by use of the area under the receiver operating characteristic curve (A_z). The A_z was assessed for each population sample by use of the four different classifiers with each of the five different feature selection schemes. Fig. 1 illustrates the study design.

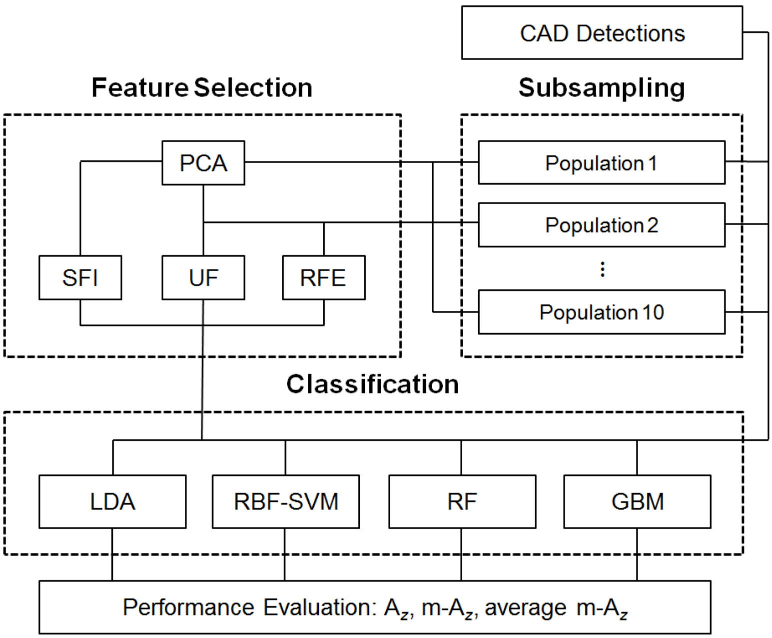

Fig. 1. Diagram illustrating the experiments of the study. The CAD detections were sampled randomly to construct 10 population samples. For each population, feature selection was performed using one of 5 methods (see Section 2.1) and classification was performed with or without feature selection using one of 4 methods (see Section 2.2). Performance evaluated was based on the area under the receiver operating characteristic curve for each population (A_z), average of A_z over the 10 populations (m-A_z), and average m-A_z over the feature selection methods or over the classifiers and each feature selection method (average m-A_z).

3 Results

The classification performance of the four classifiers is summarized in Tables 1 to 6. Tables 1 through 5 show the classifier performance with each of the five feature selection methods, whereas Table 6 shows the classifier performance without feature selection. Each row shows the result of an indicated population sample, whereas the columns indicate the average and standard deviation of the A_z value over three repeated 10-fold cross-validations of the indicated classifier. Bold numbers indicate the highest average of A_z for a sample. The bottom row shows the average and standard deviation of A_z over the 10 population samples.

If we consider the highest overall classifier performance in terms of the highest mean A_z (m-A_z) over the 10 subsampled populations, the ranking of classifiers varies according to the feature selection method. The performance was highest for the RBF-SVM without feature selection (0.800; Table 6), followed by the RF with RFE (0.799; Table 4), followed by RBF-SVM with the SFI of PCs (0.798; Table 1) and GBM with RFE (0.798; Table 4).

Also, the frequency at which a classifier outperformed the other classifiers in terms of the highest A_z for each of the 10 populations depended on the feature selection method. RBF-SVM outperformed the other classifiers most often with SFI of PCs (Table 1), with UF (Table 2), and without feature selection (Table 6). However, LDA outperformed the other classifiers with UF of PCs (70%; Table 3) and RFE of PCs (40%; Table 5). The RF and GBM classifiers outperformed the other classifiers with the RFE feature selection method (Table 4).

The robustness, or consistency, of a classifier, can be characterized by the average m-A_z value that the classifier yields for the different feature selection methods. In this sense, RBF-SVM yielded the highest performance (0.795), followed by the RF (0.794), GBM (0.793), and LDA (0.786).

Table 1. Performance comparison of classifiers with feature selection by the SFI of PCs. Brackets indicate the standard deviation (SD) of A_z. The numbers in bold indicate the highest value of A_z among classifiers for each population sample.

Population	LDA	RBF-SVM	RF	GBM
1	0.780 [0.040]	**0.796** [0.047]	0.789 [0.040]	0.778 [0.041]
2	0.789 [0.055]	**0.796** [0.051]	0.794 [0.050]	0.780 [0.054]
3	0.810 [0.030]	**0.821** [0.028]	0.814 [0.024]	0.812 [0.032]
4	0.789 [0.039]	**0.808** [0.041]	0.802 [0.045]	0.790 [0.038]
5	**0.794** [0.043]	0.790 [0.046]	0.786 [0.057]	0.789 [0.055]
6	**0.806** [0.035]	0.800 [0.041]	0.793 [0.044]	0.798 [0.041]
7	0.789 [0.035]	**0.804** [0.037]	0.790 [0.034]	0.799 [0.036]
8	0.786 [0.042]	**0.790** [0.041]	0.785 [0.042]	0.783 [0.044]
9	0.776 [0.057]	0.789 [0.064]	**0.792** [0.055]	0.781 [0.057]
10	0.772 [0.052]	0.790 [0.049]	**0.791** [0.043]	0.789 [0.039]
Mean ± SD	0.789 ± 0.012	0.798 ± 0.010	0.793 ± 0.009	0.790 ± 0.011

Table 2. Performance comparison of classifiers with feature selection by UF

Population	LDA	RBF-SVM	RF	GBM
1	0.784 [0.040]	0.782 [0.041]	**0.790** [0.038]	0.781 [0.048]
2	0.785 [0.058]	0.793 [0.049]	0.796 [0.054]	**0.798** [0.054]
3	0.809 [0.029]	**0.816** [0.024]	0.815 [0.025]	0.807 [0.032]
4	0.778 [0.047]	**0.799** [0.048]	0.793 [0.044]	0.797 [0.053]
5	0.779 [0.038]	**0.785** [0.050]	0.778 [0.055]	0.783 [0.051]
6	0.790 [0.035]	**0.791** [0.042]	0.776 [0.046]	0.771 [0.042]
7	0.789 [0.035]	**0.811** [0.035]	0.792 [0.031]	0.797 [0.027]
8	0.772 [0.041]	0.791 [0.048]	**0.794** [0.042]	0.790 [0.043]
9	0.764 [0.057]	0.791 [0.053]	**0.802** [0.054]	0.796 [0.052]
10	0.756 [0.056]	**0.800** [0.044]	0.780 [0.043]	0.784 [0.035]
Mean ± SD	0.781 ± 0.015	0.796 ± 0.011	0.792 ± 0.012	0.790 ± 0.010

Table 3. Performance comparison of classifiers with feature selection by UF of PCs

Population	LDA	RBF-SVM	RF	GBM
1	**0.796** [0.040]	**0.796** [0.047]	0.792 [0.047]	0.785 [0.041]
2	**0.805** [0.051]	0.799 [0.051]	0.800 [0.043]	0.791 [0.052]
3	**0.819** [0.024]	0.812 [0.028]	0.816 [0.031]	0.815 [0.034]
4	**0.806** [0.038]	0.798 [0.041]	0.791 [0.040]	0.803 [0.038]
5	**0.797** [0.045]	0.769 [0.046]	0.768 [0.060]	0.779 [0.044]
6	**0.804** [0.039]	0.784 [0.041]	0.783 [0.041]	0.790 [0.042]
7	**0.795** [0.035]	0.786 [0.037]	0.784 [0.034]	**0.795** [0.031]
8	0.789 [0.037]	0.782 [0.041]	**0.792** [0.035]	0.788 [0.037]
9	0.782 [0.053]	0.780 [0.064]	0.775 [0.054]	**0.783** [0.048]
10	0.765 [0.054]	**0.779** [0.049]	0.772 [0.040]	0.776 [0.043]
Mean ± SD	0.796 ± 0.015	0.789 ± 0.013	0.787 ± 0.014	0.790 ± 0.011

Table 4. Performance comparison of classifiers with feature selection by RFE

Population	LDA	RBF-SVM	RF	GBM
1	0.786 [0.041]	0.789 [0.041]	**0.800** [0.040]	0.784 [0.044]
2	0.739 [0.054]	0.798 [0.047]	0.803 [0.050]	**0.810** [0.052]
3	0.805 [0.030]	0.814 [0.025]	**0.819** [0.024]	0.806 [0.028]
4	0.776 [0.047]	**0.800** [0.048]	0.798 [0.045]	**0.800** [0.046]
5	0.769 [0.048]	0.788 [0.050]	0.789 [0.057]	**0.790** [0.053]
6	**0.802** [0.038]	0.789 [0.042]	0.778 [0.044]	0.785 [0.040]
7	0.773 [0.034]	**0.813** [0.038]	0.799 [0.034]	0.810 [0.027]
8	0.765 [0.042]	0.790 [0.047]	**0.798** [0.042]	0.791 [0.041]
9	0.759 [0.063]	0.792 [0.053]	**0.810** [0.055]	0.801 [0.051]
10	0.753 [0.054]	**0.801** [0.044]	0.791 [0.043]	**0.801** [0.039]
Mean ± SD	0.773 ± 0.021	0.797 ± 0.010	0.799 ± 0.011	0.798 ± 0.010

Table 5. Performance comparison of classifiers with feature selection by RFE of PCs

Population	LDA	RBF-SVM	RF	GBM
1	**0.800** [0.042]	0.799 [0.044]	0.799 [0.041]	0.787 [0.044]
2	0.800 [0.055]	0.780 [0.058]	**0.802** [0.046]	0.773 [0.051]
3	**0.817** [0.029]	0.809 [0.029]	0.815 [0.029]	0.814 [0.042]
4	0.796 [0.040]	0.796 [0.036]	**0.805** [0.043]	0.804 [0.042]
5	**0.807** [0.044]	0.796 [0.045]	0.788 [0.054]	0.785 [0.047]
6	0.811 [0.037]	0.794 [0.034]	0.796 [0.039]	**0.817** [0.035]
7	**0.810** [0.035]	0.792 [0.040]	0.773 [0.044]	0.798 [0.034]
8	0.797 [0.037]	0.759 [0.044]	**0.804** [0.034]	0.797 [0.029]
9	0.772 [0.059]	0.782 [0.059]	0.789 [0.057]	**0.792** [0.051]
10	0.752 [0.045]	0.793 [0.046]	0.801 [0.039]	**0.804** [0.038]
Mean ± SD	0.796 ± 0.020	0.790 ± 0.014	0.797 ± 0.012	0.797 ± 0.013

Table 6. Performance comparison of classifier without feature selection

Population	LDA	RBF-SVM	RF	GBM
1	0.785 [0.041]	0.792 [0.043]	**0.797** [0.040]	0.783 [0.040]
2	0.785 [0.054]	0.800 [0.049]	**0.801** [0.050]	0.794 [0.057]
3	0.804 [0.032]	**0.815** [0.025]	**0.815** [0.023]	0.805 [0.033]
4	0.782 [0.043]	**0.809** [0.044]	0.798 [0.041]	0.799 [0.043]
5	0.784 [0.047]	**0.792** [0.050]	0.784 [0.056]	0.788 [0.051]
6	0.798 [0.038]	**0.799** [0.042]	0.776 [0.043]	0.780 [0.038]
7	0.784 [0.035]	**0.813** [0.038]	0.799 [0.031]	0.809 [0.029]
8	0.776 [0.043]	0.791 [0.044]	**0.794** [0.042]	0.790 [0.038]
9	0.772 [0.057]	0.794 [0.055]	**0.803** [0.054]	0.799 [0.052]
10	0.763 [0.048]	**0.798** [0.046]	0.791 [0.043]	**0.798** [0.038]
Mean ± SD	0.783 ± 0.012	0.800 ± 0.009	0.796 ± 0.011	0.794 ± 0.009

Similarly, we can also characterize the robustness of a feature selection method by calculation of the average m-A_z value of the classifiers yielded by each selection method. In this sense, RFE of PCs yielded the highest performance (average of m-A_z = 0.795; Table 5), followed by the strategy without feature selection (0.793; Table 6), SFI of PCs (0.793; Table 1), RFE (0.792; Table 4), UF of PCs (0.791; Table 3), and UF (0.790; Table 2).

4 Discussion

The preliminary results of this pilot study indicate that the RBF-SVM classifier compares favorably with other state-of-the-art classifiers in the discrimination of TP and FP CAD detections in CTC. The feature selection method of RFE of PCs compares favorably with other feature selection methods. However, no single classifier could be considered optimal under all conditions, including the use of different population samples or different feature selection methods.

In this pilot study, we used balanced sets of TP and FP samples. In practice, CAD systems produce unbalanced samples with a large number of FP samples and relatively few TP samples. However, the use of balanced data sets for the purposes of constructing classifiers and for estimating classification accuracy would often be more convenient and faster. Further work is needed for establishing the effect of using balanced and unbalanced sets on the projected classification accuracy.

Computational demands can place constraints on the classification problem. Among feature selection methods, RFE can be considered as a rather slow method for calculation, whereas the calculation for PCA is quite fast. The UF method is faster than the RFE, but slower than the PCA method. Among classifiers, the construction of SVM and GBM classifiers is remarkably slower than that of RF and LDA classifiers. The relatively small differences of the performance results suggest that although the use of fast classifiers and feature selection methods may reduce classification accuracy over that of slower methods, the reduction in overall accuracy is not necessarily meaningful in a practical application.

Acknowledgements. This work was supported by grants from NIH/NCI CA095279, CA166816, and CA131718. We also thank Partners Research Computing for providing high-performance computing services.

References

1. Yoshida, H., Näppi, J.: Three-Dimensional Computer-Aided Diagnosis Scheme for Detection of Colonic Polyps. IEEE Trans. Med. Imaging 20, 1261–1274 (2001)
2. Yoshida, H., Masutani, Y., MacEneaney, P., Rubin, D.T., Dachman, A.H.: Computerized Detection of Colonic Polyps at CT Colonography on the Basis of Volumetric Features: Pilot Study. Radiology 222, 327–336 (2002)
3. Summers, R.M., Jerebko, A.K., Franaszek, M., Malley, J.D., Johnson, C.D.: Colonic Polyps: Complementary Role of Computer-Aided Detection in CT Colonography. Radiology 225, 391–399 (2002)
4. Näppi, J., Yoshida, H.: Automated Detection of Polyps with CT Colonography: Evaluation of Volumetric Features for Reduction of False-Positive Findings. Acad. Radiol. 9, 386–397 (2002)
5. Näppi, J., Yoshida, H.: Feature-Guided Analysis for Reduction of False Positives in CAD of Polyps for Computed Tomographic Colonography. Med. Phys. 30, 1592–1601 (2003)
6. Suzuki, K., Zhang, J., Xu, J.: Massive-Training Artificial Neural Network Coupled with Laplacian-Eigenfunction-Based Dimensionality Reduction for Computer-Aided Detection of Polyps in CT Colonography. IEEE Trans. Med. Imaging 29, 1907–1917 (2010)
7. Näppi, J., Nagata, K.: Sources of False Positives in Computer-Assisted CT Colonography. Abdom. Imaging 36, 153–164 (2011)
8. Yoshida, H., Näppi, J.: CAD in CT Colonography without and with Oral Contrast Agents: Progress and Challenges. Comput. Med. Imaging Graph. 31, 267–284 (2007)
9. Näppi, J.J., Regge, D., Yoshida, H.: Ensemble Detection of Colorectal Lesions for CT Colonography. In: Yoshida, H., Sakas, G., Linguraru, M.G. (eds.) Abdominal Imaging 2011. LNCS, vol. 7029, pp. 60–67. Springer, Heidelberg (2012)
10. Nadler, M., Smith, E.P.: Pattern Recognition Engineering. Wiley, New York (1993)

11. Martinez, A.M., Kak, A.C.: PCA versus LDA. IEEE Trans. Pattern Anal. Mach. Intell. 23, 228–233 (2001)
12. Mankiewicz, R.: The Story of Mathematics, p. 158. Princeton University Press (2001)
13. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene Selection for Cancer Classification Using Support Vector Machines. Mach. Learn. 46, 389–422 (2002)
14. Schölkopf, B., Burges, C.J.C., Smola, A.J. (eds.): Advances in Kernel Methods: Support Vector Learning. MIT Press, Cambridge (1999)
15. Breiman, L.: Random Forests. Mach. Learn. 45, 5–32 (1999)
16. Svetnik, V., Liaw, A., Tong, C., Wang, T.: Application of Breiman's Random Forest to Modeling Structure-Activity Relationships of Pharmaceutical Molecules. In: Roli, F., Kittler, J., Windeatt, T. (eds.) MCS 2004. LNCS, vol. 3077, pp. 334–343. Springer, Heidelberg (2004)
17. Hastie, T., Tibshirani, R., Friedman, J.H.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, Heidelberg (2001)
18. Friedman, J.H.: Greedy Function Approximation: A Gradient Boosting Machine. Ann. Statist. 29, 1189–1232 (2001)
19. Näppi, J., Frimmel, H., Dachman, A., Yoshida, H.: A New High-Performance CAD Scheme for the Detection of Polyps in CT Colonography. In: SPIE Medical Imaging 2004, vol. 5370, pp. 839–848 (2004)
20. Näppi, J., Yoshida, H., Regge, D.: Accuracy of Computer-Aided Detection in Detecting Flat Lesions in CT Colonography. Insights into Imaging 3(S1), 334 (2012)