Old Dominion University

ODU Digital Commons

Hybrid Committee Classifier for a Computerized Colonic Polyp Detection System

Jiang Li
National Institutes of Health, jli@odu.edu

Jianhua Yao
National Institutes of Health

Nicholas Petrick
U.S. Food and Drug Administration

Ronald M. Summers
U.S. Food and Drug Administration

Amy K. Hara
Mayo Clinic Scottsdale

See next page for additional authors

Follow this and additional works at: https://digitalcommons.odu.edu/ece_fac_pubs

Part of the Artificial Intelligence and Robotics Commons, Biomedical Commons, Digestive System Commons, Investigative Techniques Commons, and the Theory and Algorithms Commons

Original Publication Citation

Li, J., Yao, J., Petrick, N., Summers, R. M., & Hara, A. K. (2006) Hybrid committee classifier for a computerized colonic polyp detection system. In J.M. Reinhardt & Josien P.W. Pluim (Eds.), *Medical Imaging 2006: Image Processing, Proceedings of SPIE Volume 6144* (61445A). SPIE. https://doi.org/10.1117/12.652724

Authors

Jiang Li, Jianhua Yao, Nicholas Petrick, Ronald M. Summers, Amy K. Hara, Joseph M. Reinhardt (Ed.), and Josien P.W. Pluim (Ed.)

Hybrid Committee Classifier for a Computerized Colonic Polyp Detection System

Jiang Li[a], Jianhua Yao[a], Nicholas Petrick[b], Ronald M. Summers[a], and Amy K. Hara[c]

[a]Diagnostic Radiology Department, Warren G. Magnuson Clinical Center, NIH
[b]NIBIB/CDRH, Joint Laboratory for the Assessment of Medical Imaging System, FDA
[c]Mayo Clinic Scottsdale, Scottsdale, AZ, USA

ABSTRACT

We present a hybrid committee classifier for computer-aided detection (CAD) of colonic polyps in CT colonography (CTC). The classifier involved an ensemble of support vector machines (SVM) and neural networks (NN) for classification, a progressive search algorithm for selecting a set of features used by the SVMs and a floating search algorithm for selecting features used by the NNs. A total of 102 quantitative features were calculated for each polyp candidate found by a prototype CAD system. 3 features were selected for each of 7 SVM classifiers which were then combined to form a committee of SVMs classifier. Similarly, features (numbers varied from 10-20) were selected for 11 NN classifiers which were again combined to form a NN committee classifier. Finally, a hybrid committee classifier was defined by combining the outputs of both the SVM and NN committees. The method was tested on CTC scans (supine and prone views) of 29 patients, in terms of the partial area under a free response receiving operation characteristic (FROC) curve (AUC). Our results showed that the hybrid committee classifier performed the best for the prone scans and was comparable to other classifiers for the supine scans.

Keywords: Computer-Aided Detection, Pattern Recognition, Statistical Methods, Classifier Committee, Neural Network, Support Vector Machine

1. INTRODUCTION

Colon cancer is the second leading cause of cancer deaths in the US. Research for the development of computer aided procedures for screening patients for colonic carcinoma has grown as a result of recognized disadvantages that accompany the current standard procedure, colonoscopy. CAD combined with CT colonography is an alternative. There is an ongoing effort by several institutions to develop classification schema that optimize the performance of CAD methods for colon polyp detection. Summers et al. describe recent work on a version of computer automated polyp detection that uses geometric and volumetric features, acquired from the CT data, as the basis for polyp identification.[1,2] Our software first segments the colon using a region growing algorithm, regions-of-interest along the colon wall are then identified. A total of 102 quantitative features are calculated for each polyp candidate. Finally, a decision of true polyp or false positive is made by a classifier based on features that are proved clinical relevant. In this paper, we investigated several committee classifiers used in the decision making stage.

Recent work showed successful applications of committee classifiers in medical imaging researches, such as breast cancer screening,[3,4] bone abnormalities detection[5] and colonic polyp identification.[6,7] The reason for using committee classifiers is that a committee classifier can often achieve a better performance than that of its committee members. This is true for the case if the errors of the individual committee members cancel out to some extent, i.e., the committee members have a diversity property so that they are unlikely to make errors in the same feature space. In order to obtain the diversity property, committee members are usually trained by bagging, boosting[8] if training data size is fixed, or by independent data sets if they are available. In this paper,

---

Further author information: (Send correspondence to Ronald M. Summers.)
Ronald M. Summers: E-mail: rms@nih.gov, Tel. 1 301 402-5486
Website: http://www.cc.nih.gov/drd/summers.html
Jiang Li: E-mail: lij3@cc.nih.gov, Tel. 1 301 451-8363

we made each committee member use a different set of features. We hope such training method will reduce correlations amongst committee members to some extent.

There exist many ways to combine the outputs of committee members. If only the label output is available for each member, a majority vote scheme is often used. In some situations where continuous outputs like posteriori probabilities are supplied by the members, average or other linear combination methods have been investigated. Theoretic analysis showed that the sum rule (simple average) outperforms other combination schemes in practice though it is based on the most restrictive assumptions.[9] The sum rule is superior because it is most resilient to estimation errors.

We present a committee classifier which involved SVMs and NNs for colonic polyp detection, where SVMs and NNs used different feature sets to increase diversity of committee members. Each SVM and NN produced a posterior probability and the final decision was made by combining the individual output using the sum rule. The features used by SVM or NN members were selected by two feature selection algorithms designed specifically for SVM and NN. In the following sections, we first review SVM and NN classifiers. We then describe two feature selection algorithms. Finally, results for colonic polyp classifications by three committee classifiers are presented.

2. TWO CLASSIFIERS

In this section, we give brief reviews for SVM and NN classifiers.

2.1. Support Vector Machine

Given a set of data pairs $\{\mathbf{x}_p, i_p\}_{p=1}^{N_v}$, where $\mathbf{x}_p \in \mathcal{R}^N$ is the feature vector extracted from a polyp candidate, and $i_p \in \{+1, -1\}$ is a class label (true polyp, true negative) associated with $\mathbf{x}_p$, an SVM defines a hyperplane

$$f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b = 0 \tag{1}$$

to separate the data points. Here $\mathbf{w}$ and b are the plane parameters, and $\phi(\mathbf{x})$ is a function mapping the vector $\mathbf{x}$ to a higher dimensional space. The hyperplane (1) is determined using the concept of *Structural Risk Minimization*[10] by solving the following optimization problem,

$$\min_{\mathbf{w}, b, \xi} \left(\frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{p=1}^{N_v} \xi_p \right) \tag{2}$$

subject to

$$i_p(\mathbf{w}^T \phi(\mathbf{x}_p) + b) \geq 1 - \xi_p, \xi_p \geq 0 \tag{3}$$

here C is a penalty parameter and ξ_p a slack factor. After the hyperplane is determined, a polyp is declared if $f(\mathbf{x}_p) > 0$. Otherwise, an non-polyp is decided. In order to combine the outputs by NN and SVM classifiers we utilized a method suggested by Platt[11] to transfer the SVM output, $f(\mathbf{x}_p)$, to a posterior probability by fitting a sigmoid,

$$p(i_p = 1 | f(\mathbf{x}_p)) = \frac{1}{1 + exp(Af + B)} \tag{4}$$

The parameters A and B were fit using a maximum likelihood estimation from a training set $\{f(\mathbf{x}_p), i_p\}$ by minimizing a cross-entropy error function.[11]

2.2. Neural Network

Most commonly used neural network classifier is the multilayer perceptron (MLP).[12] The MLP approximates the general Bayes discriminant.[13, 14] MLPs are designed by minimizing the standard training error,

$$E = \sum_{i=1}^{N_c} E(i) \tag{5}$$

where N_c is the number of classes and $E(i)$, the mean-squared error for the ith output, is

$$E(i) = \frac{1}{N_v} \sum_{p=1}^{N_v} [t_p(i) - y_p(i)]^2 \tag{6}$$

Here $t_p(i)$ denotes the ith desired output for the pth feature vector $\mathbf{x}_p$, $y_p(i)$ denotes the ith observed output for $\mathbf{x}_p$. In this paper we have 2 outputs ($N_c = 2$), one is for polyps and another is for non-polyps. We let $t_p(1) = 1$ and $t_p(2) = 0$ if $i_p = 1$ (polyp). Similarly, if $i_p = -1$, we let $t_p(1) = 0$ and $t_p(2) = 1$. We normalized the observed output $y_p(i)$ to approximate the posterior probability,

$$p(i_p = 1|\mathbf{x}_p) = \frac{\text{sigmoid}(y_p(1))}{\sum_{i=1}^{N_c} \text{sigmoid}(y_p(i))} \tag{7}$$

where $\text{sigmoid}(x) = 1/(1 + exp(-x))$.

There are many well-developed algorithms, including the Back Propagation (BP), Conjugate Gradient (CG) and the Levenberg-Marquardt (LM) algorithm[15, 16] for training the network. In this paper we are interested in an algorithm called output weights optimization-hidden weights optimization-output reset (OWO-HWO-OR),[17–19] in which hidden weights and output weights were iteratively trained to reduce the training error. A three layer, fully connected MLP networks with sigmoid activation function for the hidden layer was used. For the pth feature vector, the jth hidden unit net and activation functions are

$$net_p(j) = \sum_{k=1}^{N+1} w(j, k) \cdot x_p(k) \tag{8}$$

where $w(j, k)$ is the weight connecting the jth hidden unit with the kth input, $x_p(N + 1) = 1$ is used to handle the thresholds for hidden units and outputs.

$$O_p(j) = \text{sigmoid}(net_p(j)) \tag{9}$$

the ith observed output is

$$y_p(i) = \sum_{k=1}^{N+1} w_{oi}(i, k) \cdot x_p(k) + \sum_{j=1}^{N_h} w_{oh}(i, j) \cdot O_p(j) \tag{10}$$

where N_h is the number of hidden units, $w_{oi}(i, k)$ and $w_{oh}(i, j)$ are weights connecting to the ith output unit from the kth input and the jth hidden unit respectively. The output weights $w_{oi}(i, k)$ and $w_{oh}(i, j)$ were found using the OWO method. In the HWO procedure, the hidden weighs $w(j, k)$ were updated by minimizing a separate error function for each hidden unit.[17–19]

3. FEATURE SELECTION ALGORITHMS

We currently calculated 102 quantitative features for each polyp candidate. However, many of these features were based on heuristics and are not eventually useful. Irrelevant or redundant features can lead to several problems including unnecessarily demanding computational resources and memory allocation, the *curse of dimensionality* and convergence difficulties for training a large network. The goal of feature selection is to obtain a compact feature set with high classification accuracy.

3.1. Feature Selection Algorithm for SVM

We used a SVM-based wrapper method, progressive feature selection, for SVM committee member classifiers.[20] An SVM classifier was designed to evaluate each feature subset candidate, where the fitness criterion was defined as the average of sensitivity and specificity of the involved SVM. Sensitivity denotes classification accuracy for polyp candidates whereas specificity represents classification accuracy for non-polyp candidates in the data.

In the progressive feature selection algorithm, N-feature vectors were formed progressively in N stages. In the first stage, 1-feature vectors were ranked based on their fitness values, and the top 1000 vectors were passed to the next stage. In the Nth stage, N-feature vectors were formed by adding one feature to the $(N-1)$-feature vectors. N was determined as 3 in the experiment. For our polyp classification task we used a committee of 7 SVMs classifier, where each SVM took 3 features as inputs. This configuration was chosen based on statistical analysis to achieve the best classification performance with the least complexity.[21]

3.2. Feature Selection Algorithm for NN

Training an MLP classifier usually takes a longer time than that of training an SVM classifier. This makes designing an MLP-based wrapper feature selection algorithm is difficult when data size or feature size is large. We utilized a fast feature selection algorithm, PLNFS,[22] for NN classifier in this paper.

For the set of converted data pairs $\{\mathbf{x}_p, \mathbf{t}_p\}_{p=1}^{N_v}$, where $\mathbf{x}_p \in \mathbf{R}^N$ and $\mathbf{t}_p \in \mathbf{R}^{N_c}$. PLNFS first utilized a piecewise linear network (PLN) to approximate the nonlinear mapping from $\mathbf{x} \to \mathbf{t}$. The PLN employed a clustering method to partition the feature space into a hierarchy of regions (or clusters), where simple hyperplanes were fit to the local data. Thus local linear models constructed local approximations to the global nonlinear mapping. For each cluster PLNFS used a linear regression model to represent the mapping from input to output,

$$\mathbf{t}_{N_v \times N_c} = \mathbf{x}^{(q)}_{N_v \times N} \mathbf{w}^{(q)}_{N \times N_c} + \mathbf{\Xi}^{(q)}_{N_v \times N_c} \tag{11}$$

where the superscript denotes when the feature data belongs to the qth cluster, the weight $\mathbf{w}^{(q)}$ and error $\mathbf{\Xi}^{(q)}$ become valid. Using the Modified Schmidt procedure[23] to each of the clusters yields,

$$\mathbf{x}^{(q)}_{N_v \times N} = \mathbf{\Theta}^{(q)}_{N_v \times N} \mathbf{A}^{(q)}_{N \times N} \tag{12}$$

here

$$\mathbf{\Theta}^{(q)} = [\mathbf{\Theta}^{(q)}_1, \mathbf{\Theta}^{(q)}_2, \cdots, \mathbf{\Theta}^{(q)}_N] \tag{13}$$

with orthonormal columns satisfying $\mathbf{\Theta}^{(q)T}_i \mathbf{\Theta}^{(q)}_j = \mathbf{I}$, where $\mathbf{I}$ denotes identity matrix, and $\mathbf{A}^{(q)}$ is an upper triangle matrix. Substituting (12) into (11) yields a piecewise linear orthonormal (PLO) system,

$$\mathbf{t} = \mathbf{\Theta}^{(q)} \mathbf{A}^{(q)} \mathbf{w}^{(q)} + \mathbf{\Xi}^{(q)} = \mathbf{\Theta}^{(q)} \mathbf{w}^{(q)}_o + \mathbf{\Xi}^{(q)}, \tag{14}$$

where $\mathbf{w}^{(q)}_o$ are weights for the orthonormal system.

Let $\mathbf{X}(d) = \{\mathbf{x}(i) : 1 \leq i \leq d, \mathbf{x}(i) \in \mathbf{Z}\}$ be a set of d features from the set $\mathbf{Z} = \{\mathbf{z}(i) : 1 \leq i \leq N\}$ of N available features. The fitness value for the feature set, $\mathbf{X}(d)$, involved in system (14) was defined as the output variance explained by $\mathbf{X}(d)$,

$$J(\mathbf{X}(d)) = \sum_{i=1}^{N_c} \sum_{k=1}^{d} \sum_{q=1}^{N_{cluster}} (w^{(q)}_o(i,k))^2 \tag{15}$$

where $N_{cluster}$ is the number of clusters the feature space was partitioned.

PLNFS selected one best feature from the remaining feature pool at each step based on the fitness value calculated using (15), it then determined if there were some features in the selected feature set that became less important after the new feature was added. If yes, the least useful one will be excluded from the selected pool. These steps were repeated until either the preset number of features were selected or adding new features did not significantly improve fitness value. Finally, PLNFS gave a series feature sets, i.e., the best sets of size 1, size 2,..., etc. Note that the best N feature set did not necessarily contain all the best $N-1$ features. See[22] for algorithm details.

4. HYBRID COMMITTEE CLASSIFIERS

We describe the committee diversity and outline the proposed hybrid classifier in this section.

4.1. Committee Diversity

In a regression context, Tresp[24] rigorously defined why and how the diversity amongst member predictors contributes toward overall ensemble accuracy, using "bias-variance-covariance" decomposition. In a classification context, there is no such neat theory. However, if member classifiers' outputs are estimates of the posterior probabilities and a simple average rule (sum rule) is used to combine the committee members' outputs, Tumer and Ghosh[25, 26] provided a theoretical framework for analyzing the committee diversity for classification tasks. Suppose $P(\mathbf{x})$ is the probability of $\mathbf{x}$ to be a polyp, $1 - P(\mathbf{x})$ the probability to be a non-polyp, and we have M committee classifiers to form a posterior probability prediction for $\mathbf{x}$ using the sum rule,

$$\hat{P}(\mathbf{x}) = \frac{1}{M} \sum_{i=1}^{M} \hat{p}_i \qquad (16)$$

where $\hat{p}_i$ is the estimated posterior probability of $\mathbf{x}$ to be a polyp by the ith classifier. Assume each classifier has the same expected error and identical correlation coefficient between classifiers, Tumer and Ghosh showed that the expected error for the committee classifier is

$$E\{\hat{P} - P\} = m \left(\frac{1 + \rho(M-1)}{M} \right), \qquad (17)$$

where $m = E\{P - p_i\}, \forall i$, is the expected error of the ith committee classifier and is assumed identical for all members, and ρ is the identical correlation coefficient for all members. For extreme situation, $\rho = 0$, all classifiers are uncorrelated, we have $E\{\hat{P} - P\} = \frac{1}{M} m$. For the case of $\rho = 1$, all classifiers are perfectly correlated, the committee error will be unchanged. Therefore, appropriate committee members should be those that have low errors and less correlations between each other. As a guidance, we used NN and SVM classifiers as committee members and each member used a different set of features with a goal to achieve the committee diversity.

4.2. Permuting Imbalanced Data

The term "Imbalanced data" describes the situation where there exist more samples in one class than that in the other classes. For our case, the polyp detections have many more false positives than true positives. Traditional classifiers like neural networks usually produce poor predictive results for the minority class because the NN training is dominated by the majority class. There have been several methods for this bias learning, including over-sampling the minority class, under-sampling the majority class,[27] and utilizing boosting for balancing training.[28] We used a condense algorithm[29] to filter out the false positives which are internal samples using a nearest neighbor rule. Internal samples are those that do not form decision boundaries. All true positives were kept even though they were identified as internal samples by the nearest rule.

4.3. Proposed Hybrid Classifier

Given the labelled data set, we are now ready to describe our hybrid system as follows.

1. Select feature set for SVM using the algorithm described in section 3.1, the feature set contains 7 feature vectors and each with a length of 3.

2. Train the SVM committee classifier using the selected feature set. The imbalanced learning was dealt with using the under-sampling method.

3. Select feature set for NN using the algorithm described in section 3.2. The selected feature set contains 11 feature vectors with length from 10 to 20.

4. Train NN committee classifier using the selected feature set. The imbalanced learning was dealt with by the method described in section 4.2.

5. Combine the outputs of NN and SVM committees using the sum rule to form the hybrid committee classifier and apply it to the test data set.

The configuration of the NN committee is based on experiments that show it can provide a good performance.

4.4. Data Acquisition and Experiments

CTC procedure was performed on 29 patients with a high suspicion of colonic polyps or masses. There were 19 males and 10 females. The mean age was 69 years (st. dev. 11 years; range 41 to 86 years). All patients had at least one polyp and 27 of them had at least one polyp or mass 1 cm or larger. These patients were chosen from a larger cohort who underwent contrast-enhanced CTC. Selection criteria included that patients had at least 1 polyp > 5 mm, a majority of which were identified on both the prone and supine views. The software first segmented the colon using a region growing algorithm, regions-of-interest along the colon wall were then identified. A total of 102 quantitative features were calculated for each polyp candidate based on a 3-D segmentation algorithms.[30] Finally, the software produced 221 true polyp detections with 1034 false positives for the supine scans, and 148 true polyp detections with 1022 false positives for the prone scans.

We used five-fold cross validation with 100 bootstraps to compare the hybrid committee, NN committee and SVM committee classifiers by their FROC curves, in terms of the partial area under the FROC curve (AUC). In the five-fold cross validation, the resulting detections from prone and supine scans were randomly partitioned to 5 parts. Each of the 5 parts was held out as a test set with 100 bootstraps to produce average FROC curves, and the remaining four-fifths were used to train the committee classifiers using the procedures described in section 4.3. Note that the selected feature sets for NN and SVM committee classifiers may be different in each fold. The training and testing procedures were repeated 5 times, and we obtained an overall average FROC curve for each committee classifier. The paired t-test was used to test if the AUCs for different committee classifiers were significantly different at the 95% confidence level.

5. RESULTS

Fig. 1 and Fig. 2 show the mean FROC curves of the 100 bootstraps for supine and prone data respectively. For the prone data set, the means and standard deviations (SD) of AUC were 0.902±0.039, 0.905±0.043 and 0.932±0.035 for the SVM, NN and Hybrid committees, respectively. The hybrid committee was found (by the paired t-test) to be superior to the SVM and NN committees. There was no statistical difference between the SVM and NN. The sensitivity for the three committees, means and SDs, at an FPR=3.0 were 65± 3.7%, 75.9±6.1% and 82.4±3.6%, respectively. For the supine case, the means and SDs for AUC were 0.937± 0.026, 0.911±0.035 and 0.933±0.028 for the SVM, NN and hybrid committees. The hybrid and SVM committees were found to be superior to the NN committee while no significant difference was found between the SVM and hybrid committees.

6. DISCUSSION AND CONCLUSION

We developed a hybrid committee classifier for CAD of colonic polyps in CT colonography. The proposed committee classifier was found to be the best amongst SVM and NN committee classifiers for the prone data set, and one of the best for the supine data set.

We observed that for prone scan data set (Fig. 2), the NN and SVM committee classifiers produced quite different FROC curves, and therefore the hybrid committee classifier was able to provide superior performance to both the NN and SVM committees. This is consistent to the diversity analysis in section 4.1, i.e., if committee members performed differently in different feature spaces, the overall committee classifier will achieve good performance in the whole feature space. The diversity analysis also holds for the supine scan (see Fig. 1). The hybrid committee was not superior to the SVM or NN committee because both the SVM and NN committees produced similar performances in the whole feature space.

It is essential to achieve the diversity property amongst committee members when designing a committee classifier. There are several methodologies to manipulate data for obtaining such diversity, including bagging, boosting, cascade correlation, or the method used in this paper that trained committee members with different features. It is also important to keep in mind that each member should not be much inferior to others since the inferior member will have a negative effect on the committee, see Fig. 1 for instance, in some places the hybrid committee was inferior to the SVM due to the worse performance of the NN committee.

If the optimal classifier is achievable for the given data, our experiments showed that the hybrid committee classifier was more likely to approach this goal: it was always one of the best classifiers though it is not statistically

better than the SVM committee classifier for the supine scan data. However, it is not clear at this point if the SVM committee or the hybrid committee were already the achievable optimal classifier for the supine case. The supine scans had intravenous contrast, which tends to increase CT image quality. However, whether it was the influence of intravenous contrast that made the SVM committee alone approach the optimal classifier is still under investigation.

7. ACKNOWLEDGEMENT

REFERENCES

1. R. M. Summers, A. Jerebko, M. Franaszek, J. Malley, and C. Johnson, "Colonic polyps: Complementary role of computer-aided detection in CT colonography," *Radiology* **225**, pp. 391–399, 2002.
2. R. M. Summers, "Challenges for computer-aided diagnosis for CT colongraphy," *Abdominal Imaging* **27**, pp. 268–274, 2002.
3. L. Wei, Y. Yang, R. M. Nishikawa, and Y. Jiang, "A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications," *IEEE Transactions on Medical Imaging* **24**, pp. 371–380, March 2005.
4. S. Joo, Y. S. Yang, W. K. Moon, and H. C. Kim, "Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features," *IEEE Transactions on Medical Imaging* **23**, pp. 1292 – 1300, Oct. 2004.
5. T.-K. Yin and N.-T. Chiu, "A computer-aided diagnosis for locating abnormalities in bone scintigraphy by a fuzzy system with a three-step minimization approach," *IEEE Transactions on Medical Imaging* **23**, pp. 639 – 654, May 2004.
6. A. K. Jerebko, J. D. Malley, M. Franaszek, and R. M. Summers, "Multiple neural network classification scheme for detection of colonic polyps in CT colonography data sets," *Acad. Radiol.* **10**(2), pp. 154–160, 2003.
7. A. K. Jerebko, J. D. Malley, M. Franaszek, and R. M. Summers, "Computer-aided polyp detection in ct colonography using an ensemble of support vector machines," in *CARS*, **1256**, pp. 1019–1024, 2003.
8. L. Breiman, *Combining artificial neural nets*, vol. 31, ch. Combining predictors. Springer, 1999.
9. J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**, pp. 226–239, Mar. 1998.
10. V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
11. J. C. Platt, *Advances in Large Margin Classifiers*, ch. Probabilistic Outputs for Support Vector Machines and Comparisions to Regularized Likelihood Methods. MIT Press, 1999.
12. D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, MIT Press, 1988. Foundations.
13. D. W. Ruck, S. K. Rogers, M. Kabrisky, M. E. Oxley, and B. W. Suter, "The multilayer perceptron as an approximation to a Bayes optimal discriminate function," *IEEE Trans. on Neural Network* **1**, pp. 296–298, Dec. 1990.
14. E. A. Wan, "Neural network classification: A bayesian interpretation," *IEEE Trans. on Neural Network* **1**, pp. 303–305, Dec. 1990.
15. K. Levenberg, "A method for the solution of certain problems in least squares," *Quart. Appl. Math.* **2**, pp. 164–168, 1944.
16. D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *J. Soc. Ind. Appl. Math.* **11**, pp. 431–441, 1963.
17. C. Yu and M. T. Manry, "A modified hidden weight optimization algorithm for feed-forward neural networks," *Thirty-Sixth Asilomar Conference on Signals, Systems & Computers* **1**, pp. 1034–1038, Nov. 2002.
18. J. Li, M. T. Manry, L.-M. Liu, C. Yu, and J. Wei, "Iterative improvement of neural classifiers," in *Proceedings of the Seventeenth International Conference of the Florida AI Research Society*, May 2004.

19. R. G. Gore, J. Li, M. T. Manry, L.-M. Liu, C. Yu, and J. Wei, "Iterative design of neural network classifiers through regression," *International Journal on Artificial Intelligence Tools* , 2005. In Press.

20. J. Yao, R. M. Summers, and A. K. Hara, "Progressive feature vector selection scheme for computer aided polyp detection," in *RSNA*, 2004.

21. J. Yao, R. M. Summers, and A. Hara, "Optimizing the support vector machine (svm) committee configuration in a colonic polyp cad system," in *SPIE Medical Imaging*, 2005.

22. J. Li, J. Yao, R. M. Summers, and A. Hara, "An efficient feature selection algorithm for computer-aided polyp detection," in *FLAIRS*, 2005.

23. F. J. Maldonado and M. T. Manry, "Optimal pruning of feed-forward neural networks based upon the schmidt procedure," in *the 36th Asilomar Conference on Signals, Systems, and Computers*, pp. 1024–1028, 2002.

24. V. Tresp, *Handbook for Neural Network Signal Processing*, ch. Committee Machines. CRC Press, 2001.

25. K. Tumer and J. Ghosh, "Analysis of decision boundaries in linearly combined neural classifiers," *Pattern Recognition* **29**, pp. 341–348, Feb. 1996.

26. K. Tumer and J. Ghosh, "Error correlation and error reduction in ensemble classifiers," *Connection Science* **8**(3-4), pp. 385–403, 1996.

27. N. Japkowicz, "Learning from imbalanced data sets: A comparison of various strategies," in *Learning from Imbalanced Data Sets: The AAAI Workshop*, A. Press, ed., pp. 10–15, (Menlo Park, CA), 2000. Technical Report WS-00-05.

28. H. Guo and H. L. Viktor, "Learning from imbalanced data sets with boosting and data generation: The databoost-im approach," *SIGKDD Explor. Newsl.* **6**(1), pp. 30–39, 2004.

29. I. Tomek, "Two modifications of cnn," *IEEE Trans. on Syst., Man., and Cybern.* (SMC-6:), pp. 769–772, 1976.

30. J. Yao and R. Summers, "3d colonic polyp segmentation using dynamic deformable surfaces," in *SPIE Medical Imaging*, 2004.
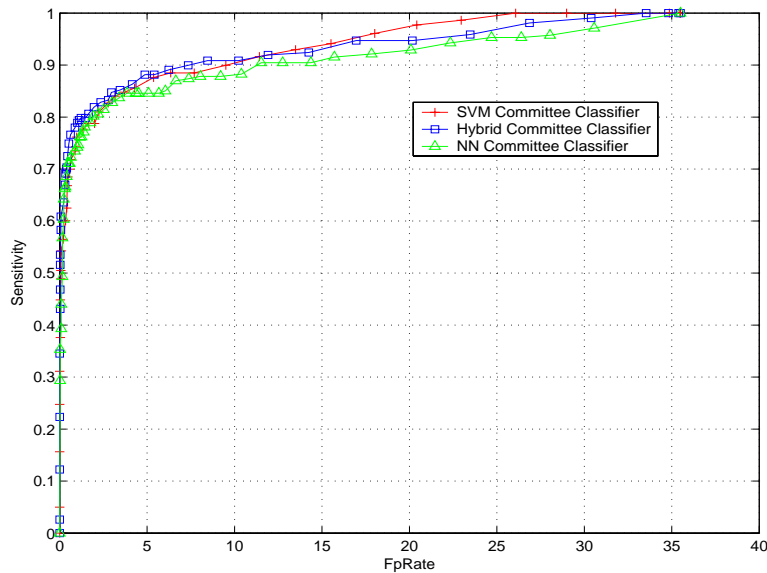
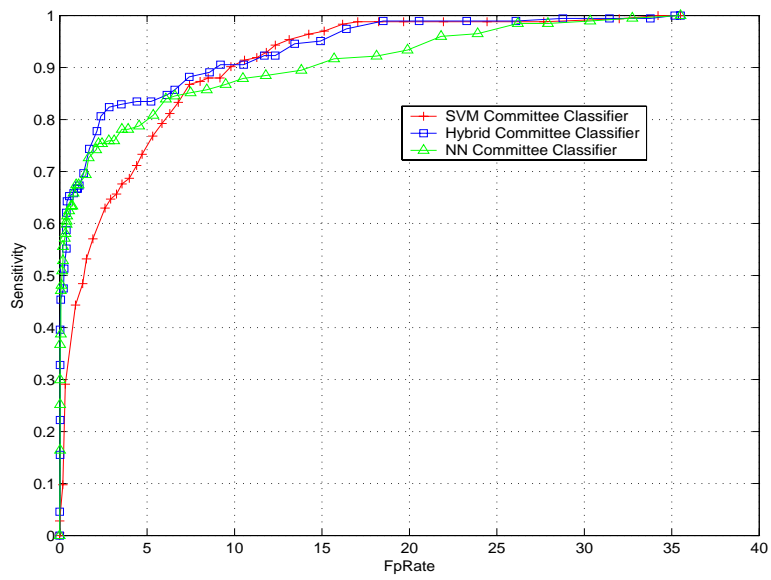**Figure 1.** FROC curves for the supine scan view detections

Figure 2. FROC curves for the prone scan view detections