Optimizing the Support Vector Machines (SVM) Committee Configuration in a Colonic Polyp CAD System

Jianhua Yao[1], Ronald M. Summers[1], Amy Hara[2]
1. Clinical Center, National Institutes of Health
2. Mayo Clinic, Scottsdale

ABSTRACT

This paper presents a method to optimize the SVM committee used in a colonic polyp CAD system to achieve high detection performance and efficiency. In our CAD system, characteristic features of polyp candidates are fed into a committee of SVMs to determine if one detection is a true polyp. The committee consists of *M* different SVMs, and each of them is established by an *N*-feature vector. A progressive feature vector selection scheme was proposed to select a population of feature vectors, in which *N*-feature vectors are composed progressively in *N* stages. To optimize the SVM committee configuration, two-way ANOVA is performed to analyze the effect of committee-member-number (*M*) and feature-vector-length (*N*). The area under the ROC curve (AUC) in a ten-fold cross validation is used as the performance metric. Pairwise Tukey's tests are performed to reveal if the performance differences between two configurations are statistically significant. The experiments were tested on 29 patients with 53 polyps. The committee configuration in comparison are *N*=1 to 7 and *M*=1, 3, 5, 7, or 9. ANOVA showed that $N = 3$ has statistically significant performance improvement over *N*=1 and 2, but is statistically equivalent with $N = 4$ to 7. It also showed that there is statistical improvement from $M = 1$ to 7, while $M = 7$ and 9 are statistically equivalent. Based on the result, we chose a committee configuration with $N = 3$ and $M = 7$ since it is the most efficient committee with statistically best performance.

Keywords: Computer Aided Detection, committee of Support Vector Machines, CT colonography, two-way ANOVA

1. INTRODUCTION

With the advancement in digital imaging modalities in recent years, computer aided detection (CAD) is becoming an important field in medical imaging research. Some commercial CAD systems have already emerged, such as breast cancer screening using mammography [1], and more recently, detection of lung nodules in 3D CT acquisitions [2]. Most CAD systems are still in research stage, such as CAD in detecting colon polyps [3] and bone abnormalities [4].

Developing a typical CAD system has two phases: training phase and application phase. A CAD system needs to be trained using training data before its application to new data. Training data is usually manually or semi-automatically preprocessed to identify lesions and structures of interest. Then a computer program (a CAD system) is developed and optimized to detect as many true lesions with few false detections. The trained CAD system generally has a segmentation program, a feature extraction program and a classifier. In the application phase, previously unseen data first undergoes the segmentation and feature extraction process. Then the features of potential detections are fed to the classifier to determine whether they are true lesions or false ones.

Medical Imaging 2005: Physiology, Function, and Structure from Medical Images, edited by Amir A. Amini, Armando Manduca, Proceedings of SPIE Vol. 5746 (SPIE, Bellingham, WA, 2005) · 1605-7422/05/$15 · doi: 10.1117/12.594547

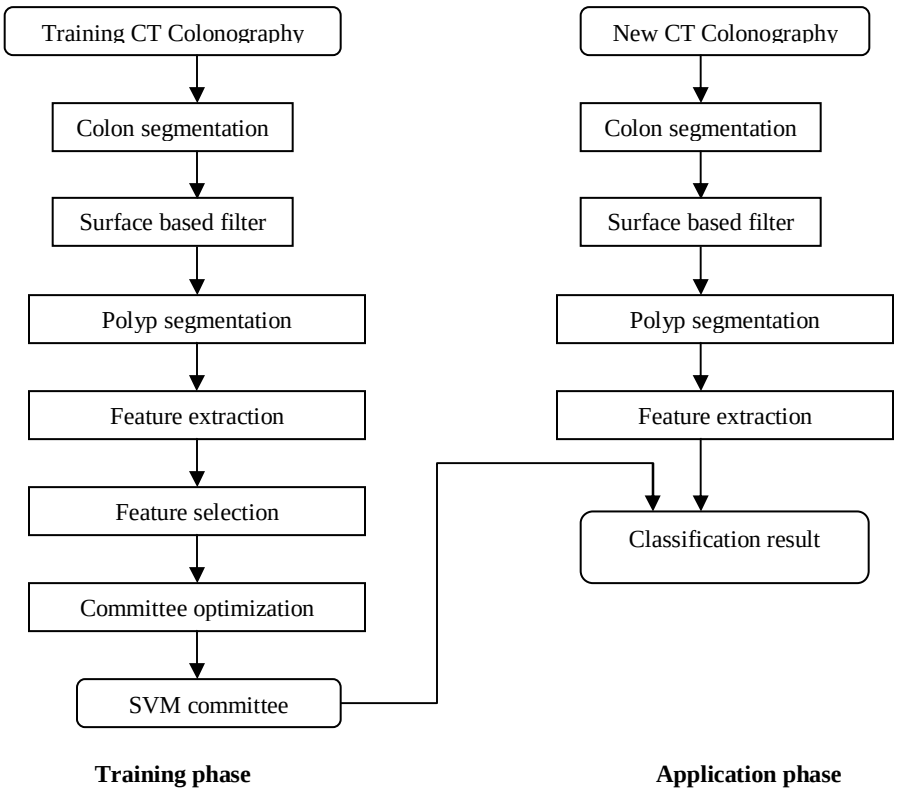

Figure 1. Flow chart of CTC CAD system

A Classifier is an essential component of a CAD system. Classifiers such as neural networks (NN) [5] and support vector machines (SVM) [6] have been widely used in CAD systems. In order to obtain an effective classifier, a subset of features needs to be selected from the entire feature space based on their individual or joint performances. Feature selection methods such as forward stepwise search (FFS) and genetic algorithm (GA) [7] have been proposed. However, the number of features in a classifier needs to be determined. If too few features are used, the classifier may not have enough power to distinguish true detections from false ones. On the other hand, if too many features are selected, the classifier may become too complex and over-trained, and cannot be generalized to data outside the training set. One solution is to use a committee of classifiers, and each classifier only includes a small number of features [5]. The majority vote of the committee is one possible decision function for the classifier. The goal of this investigation is to optimize the configuration of the committee, i.e. how many classifiers in the committee and how many features in each classifier.

This paper will focus on a CAD system that detects colon polyps on computed tomographic colonography (CTC). Colon cancer is the third leading cause of cancer deaths in the US. Compared to colonoscopy, CTC is less invasive. CAD for CTC has been investigated in several institutes over the past few years [3]. In this paper, section 2 will describe our methods, especially the classifier and the optimization of the committee configuration. Section 3 will present results, and section 4 will provide the discussions and conclusions.

2. METHODS

The flow chart of our CAD system is illustrated in Figure 1. In the training phase, the training CT colonographies are segmented and the colon surfaces are extracted. Surface based filters are then applied to

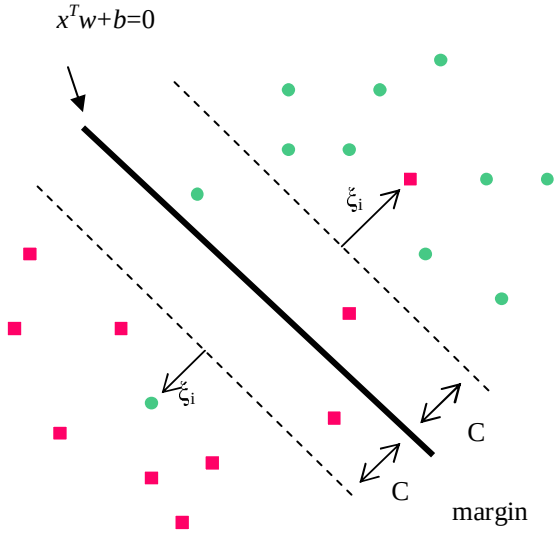

Figure2. Illustration of SVM

identify suspected polyps. After that, polyp detections are segmented to extract potential polyp regions. Quantitative features are then computed for each detection. A progressive feature selection [8] is run to select pertinent features. Each selected feature vector establishes an SVM. Then a committee optimization process is performed to form a committee of SVMs. The configuration of the SVM committee is determined by two-way ANOVA analysis [9].

In the application phase, given a new CTC exam, the same segmentation and feature computation process is carried out to obtain potential polyp detections and their features. The detections and features are then fed into the SVM committee formed in the training phase to determine whether they are true polyps or false detections.

Section 2.1 and 2.2 will introduce the SVM committee. Section 2.3 will describe our progressive feature selection method. Section 2.4 will present the two-way ANOVA analysis to optimize the committee configuration. Section 2.5 will describe our data and experiments.

2.1 Support Vector Machines (SVM)

SVM is a relatively new technique for data classification. It uses hyperplanes in a high dimensional feature space to separate data in different classes [6]. SVM is trained with a learning system derived from statistical learning theory, and is generalizable to unknown data.

In the training phase, detections are given a class label (polyp, non-polyp) to form feature-class pairs (x, y). Given a training set of S detections (x_1, y_1), (x_2, y_2), …, (x_s, y_s), for p-dimensional feature space $x_i \in \Re^p$ and $y_i \in \{+1,-1\}$, we first define a hyperplane:

$$f(x) = w^T \phi(x) + b = 0 \qquad (1)$$

here w and b are plane parameters, and $\phi(x)$ is a function to map vector x into a higher dimensional space. $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ is called the kernel function. We are using radial basis functions as the kernel function, i.e.

$$K(x_i, x_j) = \exp\left(-\left\|x_i - x_j\right\|^2\right). \qquad (2)$$

To separate two training classes, SVM is employed to solve the following optimization problem:

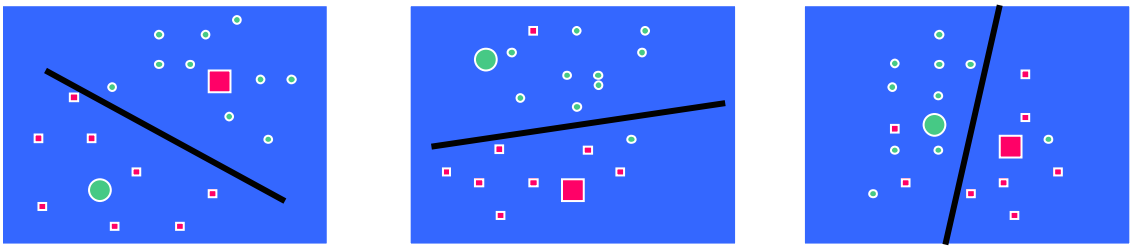

Figure 3. SVM committee

$$\min_{w,b,\xi}\left(\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i\right) \qquad (3)$$

$$\text{subject to} \quad y_i(w^T\phi(x_i)+b) \geq 1-\xi_i,\ \xi_i \geq 0$$

here C is the penalty parameter. The mechanism of SVM is illustrated in figure 2, where a hyperplane is fit to separate two groups of dots. SVM allows a soft margin on each side of the hyperplane. For each data point, the distance to the margin of hyperplane is computed. If the point is on the correct side of the plane, the distance is 0. The optimization process is to minimize the total distance of all training points. After the hyperplane is determined, the decision function for classification rule can be written as

$$h(x) = sign(f(x)) \qquad (4)$$

For a new detection x, it is declared a polyp if $h(x)>0$, or a non-polyp if $h(x)<0$. The feature values in SVM are normalized to the range of [-1, +1]. The normalization factor is obtained from the training data and applied to the testing data.

An SVM in higher dimensional space (more features) can lead to more accurate classification. However, SVM in very high dimensional space may increase the complexity of the model, over-train the data and decrease the generality of the model. Investigation in [10] also suggested that SVM does not work well in very high dimensional space. In order to address this problem, we adopt a scheme to use a committee of SVM, which will be detailed in next section.

2.2 SVM committee

Instead of using one SVM in a very high dimensional feature space, we break the feature space into subsets of low dimension feature spaces (feature vectors). Each feature vector established one SVM, and all SVMs form a committee. We allow overlap of features between different feature subsets. This scheme combines the advantages of using a large number of features and keeping the feature space small for single SVM in the committee. Each member in the committee has one vote for the classification, i.e., if the decision function of the SVM is greater than 0, the vote is 'yes', otherwise the vote is 'no'. The majority vote is used as the decision function of the committee. The committee approach generally produces improved results, provided that the error rate for each member is less than 50%. Figure 3 demonstrates how the SVM committee works. This is a committee of three SVMs. In the first SVM, there are two misclassified data (big square and big circle), but in the second and third SVM, they are correctly classified. By a majority vote, a correct classification is reached.

In order for the committee to achieve optimal performance, SVM members should be able to compensate each other. If only a few top feature vectors are selected, they usually tend to overlap each other and putting them together won't enhance the power of the committee. Therefore, a large pool of feature vectors should be available for committee member selection. We developed a progressive feature vector selection method for this purpose [8]. The details will be described in section 2.3.

The configuration of a committee is essential for its performance. An SVM committee consists of M SVM classifiers and each SVM is in an N-dimensional feature space. We propose a statistical approach based on two-way ANOVA analysis to determine the optimal committee configuration (section 2.4).

2.3 Feature vector selection

The goal of feature vector selection is to generate a large pool of feature vectors to be used as candidates for committee members. The task is to select *K* feature vectors with best performance, and each vector has *N* features. Here *K* is a large number, and *N* is a relatively small number.

There are several commonly used feature selection schemes, including exhaustive search, forward stepwise search, and genetic algorithm. Exhaustive search can be very time consuming if hundreds of features are available as candidates. Forward stepwise search is easily trapped in local minimum, and genetic algorithm is sensitive to the initial population.

We proposed a progressive search method to efficiently select a group of *K* best *N*-feature vectors. In this method, *N*-feature vectors are formed progressively in *N* stages. In each stage, one more feature is added to the vectors selected from the previous stage. Those new feature vectors are ranked by their performance and only the *K* top feature vectors are passed to the next stage. The rationale behind this scheme is that feature vectors with the worst performance in *N-1*st stage are unlikely to be in the top group in the stage *N* after one more feature is added. Essentially, this method combines the benefit of exhaustive search and forward stepwise search. Only a limited number of vectors are exhaustively examined in each stage, and the performances of feature vectors improve from stage to stage.

2.4 Two-way ANOVA analysis for optimizing committee configuration

The configuration of the SVM committee has two variables: the number of committee members (*M*) and the feature vector length of single SVM member (*N*). These two variables need to be determined and optimized before the formation of a committee. Generally, bigger *M* and *N* can lead to better performing committees in the training phase. However, bigger *M* and *N* also make the committee more complex and may over-train the data which may eventually decrease the performance in the testing phase. Furthermore, the improvement may not be statistically significant after *M* and *N* reach a certain level. The goal of the configuration optimization process is to find a configuration with the smallest *M* and *N* and statistically equivalent best performance.

Analysis of variance (ANOVA) is a statistical technique used to compare the means of two or more groups of observations (committee configuration in our case) [9]. The ANOVA model in our application can be written as

$$Y_{MN} = \mu + \alpha_M + \beta_N + \alpha\beta_{MN} + \varepsilon_{MN} \quad (5)$$

Here Y_{MN} is the observed committee performance, μ is the overall mean of the performance, α_M is the effect of committee member number, β_N is the effect of feature vector length, $\alpha\beta_{MN}$ is the joint effect, and ε_{MN} is the residual error for unaccounted variation. The effect α_M is defined as the difference between the population mean of configuration with *M* committee members and the overall mean μ. Similarly, β_N and $\alpha\beta_{MN}$ can be defined. ANOVA assumes that the distribution of each effect is normal. The null hypothesis is that the means of all configurations are equal. Several tests, such as F-test and Tukey's test, are designed to test if the hypothesis can be rejected. Since two independent variables are involved, two-way ANOVA analysis is applied.

Tukey's test is employed to do all pair-wise comparisons. It is a common technique for *post-hoc* comparisons (exploring the data to uncover large differences, without limiting investigation by a priori theory). The test assigns Tukey group letters to all configurations at 95% confidence level. If the same letter is assigned to any two configurations, they are not significantly different from each other; otherwise, the difference is statistically significant. Compared to other tests, Tukey's test is generally less conservative and tends to find smaller statistical difference.

We used randomized block method to design the experiment. That is, for each configuration, we randomly generate the same number of samples. Given a configuration with specific *M* and *N*, the SVM committee samples are formed by randomly choosing *M* vectors from the candidate pool of *N*-feature vectors. The ANOVA analysis is conducted using SAS 9.1 and SAS Enterprise Guide 2.1.

2.5 Data and experiments

The CTC data in our experiment was obtained from 29 patients (each patient had a supine study and a prone study). CT scans were done on a Siemens Volume Zoom (Sensation 4 and Sensation 16, 4 or 16 detector rows) multi-detector helical CT scanner. CT scanning parameters were 140 kVp, routine (non-reduced) tube current 183 mAs (mean; range 140 to 201), field-of-view to fit (29 - 49 cm), 3 mm collimation, 1.5 mm reconstruction interval. Based on colonoscopic examination and CTC of the same patient, 53 polyps were identified.

The image processing and filtering procedure in the polyp CAD system detected 1235 potential polyp detections. Among those, 214 were true positive detections (multiple detections on the same polyp). 110 quantitative features were computed for each detection. The committee configurations in comparison were those with 1, 3, 5, 7, or 9 members (M=1,3,5,7, or 9), and feature vector length from 1 to 7 (N=1 to 7) in each SVM member. The progressive feature selection generated the top 1000 feature vectors in each vector category. In the randomized block experiment, we randomly generated 100 committee samples for each configuration. The performance of the committee was evaluated using ten-fold cross validation. That is, the committee were trained on 90% of the training data and tested on 10% of the data to evaluate its performance. Receiver operating characteristic (ROC) curve was plotted for the test. Area under the ROC curve (AUC) was used as the performance metric.

Two-way ANOVA analysis was performed on the committee samples using AUC as dependent variable and committee_member_number (M) and feature_vector_length (N) as effects. Tukey's test was employed to do all pair-wise comparisons.

3. RESULTS

Figure 4 shows the means plot of AUC by committee_member_number and feature_vector_length. Figure 5 shows the means plot of AUC by committee_member_number alone. Figure 6 shows the means plot of AUC by feature_vector_length alone. These plots show the mean and variation of AUC for each committee configuration. Table 1 shows the Tukey's test for the effect of committee_member_number (M). Since M=7 and M=9 has the same Tukey grouping letter, it indicates that they are not statistically different. Table 2 shows the p-values for the effect of committee_member_number, which confirms that the null hypothesis (M=7 and M=9 are equivalent) cannot be rejected. Table 3 shows the Tukey's test for the effect of feature_vector_length (N). Tukey's test shows that N=3 is statistically different from N=1 and N=2, and that there are no significant differences between N=3 and N=4, 5, nor between N=5 and N=6, 7. Table 4 shows the p-values for the effect of feature_vector_length and confirms the finding.

4. DISCUSSION and CONCLUSIONS

Based on the observation from the ANOVA analysis, we chose M=7 and N=3 as the committee configuration. This configuration constructs the most efficient committee while maintaining the statistically best performance.

Designing an effective classifier for CAD systems involves several issues. One issue is the number of features to be included in the classifier. If too few features are included, the classifier may not have enough classification power; on the other hand, if too many features are used, the classifier may be over-trained. The committee of classifiers with small number of features is a feasible solution. The configuration of the committee, i.e. the number of committee members and the feature vector length of each member, needs to be optimized to obtain an efficient committee. This paper proposed a statistical method based on two-way ANOVA to analyze the performance of different configurations. The analysis can help establish the smallest committee with statistically best performance.

Table 1. Tukey's test for effect of committee_member_number (*M*) in two-way ANOVA

committee_member_number	Mean AUC	Tukey Grouping
1	0.816	A
3	0.861	B
5	0.878	C
7	0.890	D
9	0.896	D

Table 2. Pr > |t| for Null hypothesis: Mean(*M*=i)=Mean(*M*=j) for the effect of committee_member_number

M=i \ M=j	1	3	5	7	9
1		<0.0001	<0.0001	<0.0001	<0.0001
3	<0.0001		<0.0001	<0.0001	<0.0001
5	<0.0001	<0.0001		<0.0001	<0.0001
7	<0.0001	<0.0001	<0.0001		0.130
9	<0.0001	<0.0001	<0.0001	0.130	

Table 3. Tukey's test for effect of feature_vector_length (*N*) in two-way ANOVA

Feature_vector_length	Mean AUC	Tukey Grouping
1	0.739	A
2	0.794	B
3	0.902	C
4	0.900	C
5	0.907	C D
6	0.914	D
7	0.919	D

Table 4. Pr > |t| for Null hypothesis: Mean(*N*=i)=Mean(*N*=j) for the effect of feature_vector_length

N=i \ N=j	1	2	3	4	5	6	7
1		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
2	<0.0001		<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
3	<0.0001	<0.0001		0.955	0.329	0.032	0.028
4	<0.0001	<0.0001	0.955		0.289	0.024	0.019
5	<0.0001	<0.0001	0.329	0.289		0.227	0.076
6	<0.0001	<0.0001	0.032	0.024	0.227		0.405
7	<0.0001	<0.0001	0.028	0.019	0.076	0.405	

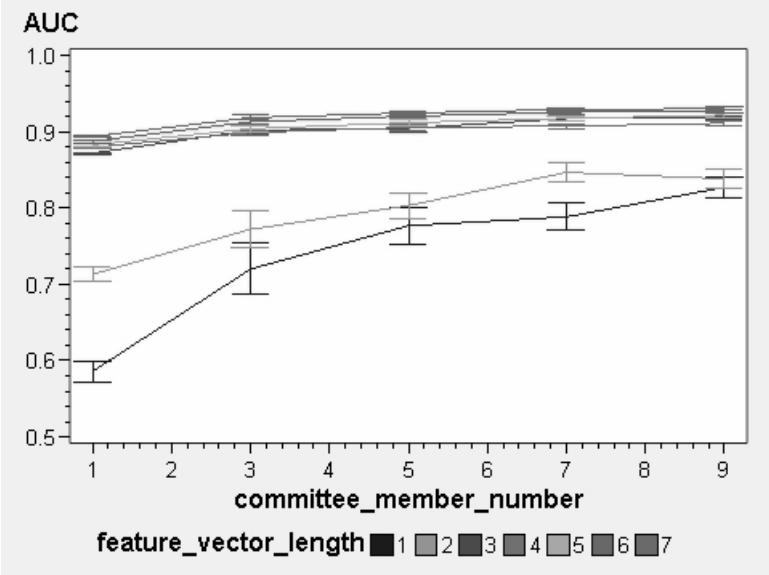

Figure 4. Means plot of AUC by committee_member_number and feature_vector_length

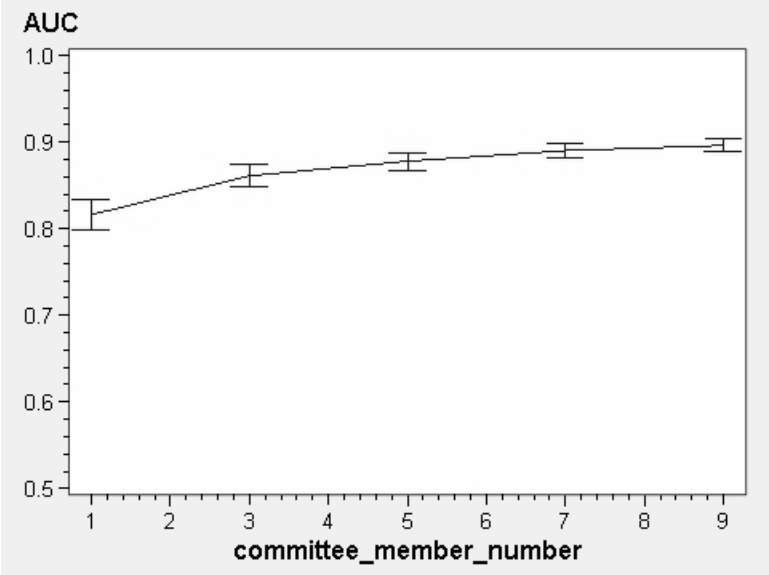

Figure 5. Means plot of AUC by committee_member_number

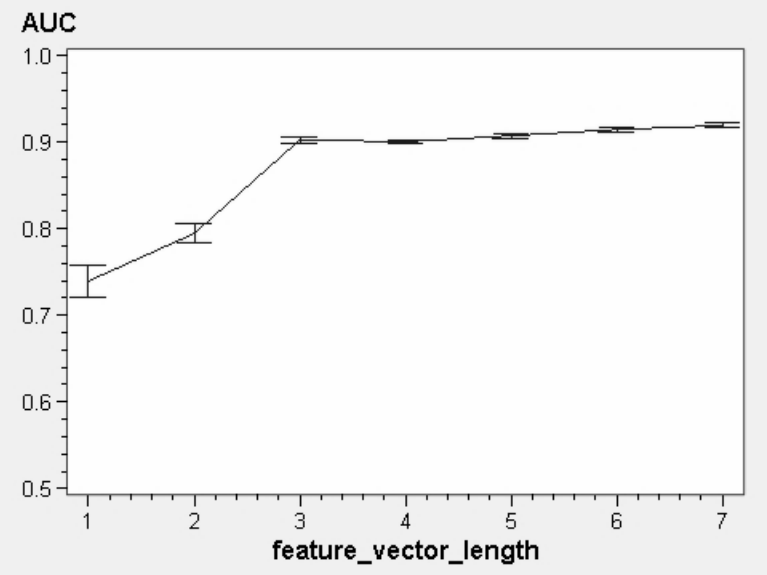

Figure 6. Means plot of AUC by feature_vector_length

REFERENCES

1. Lechner, M., N. Michael, and E. Elvecrog, *Comparison of two commercially available computer aided detection (CAD) systems.* Applied Radiology, 2002: p. 31-35.
2. Digumarthy, S., et al. *Computer-aided Diagnosis (CAD) for Detection of Lung Nodules: Illustration of Common Pitfalls and Advantages*. in *RSNA*. 2004.
3. Summers, R.M. *Current concepts and future directions in computer-aided diagnosis for CT colonography*. in *CARS 2002*. 2002.
4. Yin, T.-K. 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 Trans. Med. Imag., 2004. 23(5): p. 639-654.
5. Jerebko, A.K., et al. *Multi network classification scheme for detection of colonic polyps in CT colonography data sets*. in *SPIE Medical Imaging*. 2002.
6. Cristianini, N. and J.S. Taylor, *An Introduction to Support Vector Machines.* 2000: Cambridge University Press.
7. Miller, M., et al. *Feature Selection for Computer-Aided Polyp Detection using Genetic Algorithm*. in *SPIE Medical Imaging*. 2003.
8. Yao, J., et al. *Progressive feature vector selection scheme for computer aided colonic polyp detection*. in *RSNA 2004*. 2004. Chicago, IL.
9. SAS, *SAS Course Notes.* 2004.
10. Hsu, C.-W., C.-C. Chang, and C.-J. Lin, *A Practical Guide to Support Vector Classification*. 2003. p. 12.