Title: Support Vector Machine Based Decision Tree to Classify Voice Pathologies Using High-Speed Videoendoscopy

Authors: Yue Gao[1], Alicia J. Sprecher[2], Jack J. Jiang M.D., Ph.D.[2,3]

[1]Deparment of Computer Science, University of Wisconsin, Madison, WI, 53706
[2]Department of Surgery, Division of Otolaryngology - Head and Neck Surgery, University of Wisconsin School of Medicine and Public Health, Madison, WI, 53792
[3]Shanghai EENT Hospital of Fudan University, Shanghai, China

Running Title: Computer Analysis of HSV
Article/Category: Original Research
Section/Category: Voice

Corresponding author: Jack J. Jiang, M.D., Ph.D.
Address: 1300 University Ave
5745 Medical Sciences Center
Madison, WI 53706
Telephone number: 608-265-9854
Fax number: 608-265-2139
E-mail address: jiang@surgery.wisc.edu

ABSTRACT

Objective/Hypothesis: Little research has explored the potential of computer assisted decision making applied to high-speed videoendoscopy. In this paper, we propose a computer based method for differentiating normal and pathological larynges on the basis of HSV.

Methods: HSV recordings were collected from 101 patients with normal larynges, leukoplakia, nodules or polyps. After pre-processing, samples were analyzed for the number of glottal regions present during the open phase, the symmetry of the glottal area, the convex nature of the vocal folds and the ratio of the minimal to maximal glottal area. A decision tree based method with support vector machines at the tree nodes was used to separate samples.

Results: Normal samples were differentiated from pathological samples with a sensitivity of 91.1% and a specificity of 81.8%. When samples were divided into normal, nodule, polyp and leukoplakia groups, samples were correctly separated 70.3% of the time.

Conclusions: The combination of SVM and decision tree improves the differentiating capabilities of the parameters employed. While our approach was successful in separating normal from abnormal samples, the classification of unique pathologies requires the development stronger individual parameters.

Keywords: High-speed videoendoscopy; Decision Tree, Support Vector Machine, Polyp, Nodule, Leukoplakia

I. INTRODUCTION

The diagnosis of laryngeal pathologies remains a complex and highly subjective process. The development of an automatic analysis programs to act as support systems during diagnosis could prove extremely useful, particularly for less experienced clinicians. Most commonly diagnosis of laryngeal diseases is based on visualization of the larynx, allowing the physician to evaluate the color, shape, geometry and smoothness of the vocal folds.[1] Stroboscopy generates a simulated slow motion video feedback; however, it cannot be used with aperiodic vocal fold vibrations.[2-4] High-speed videoendoscopy (HSV) overcomes this limitation by providing a frame by frame visualization of the glottal cycle that is not disrupted by aperiodicities.[5] Furthermore, HSV images are approaching the same quality as those generated by stroboscopy.[2]

Computer aided classification of laryngeal images has been a topic of recent interest. In a 2003 study, Ilgner et al. differentiated normal from diseased with 81.4% accuracy based on color texture.[6] Verikas and his team classified laryngeal images into normal, nodular pathology and diffuse pathology groups, using parameters based on laryngeal color, texture and geometry. In three separate studies Verikas and his team classified samples with 87%, 92% and 94% accuracy.[1,7,8] Despite considerable success in developing computer systems applied to still images, no study has applied this technology to HSV data. Physicians choosing to use HSV need a program to organize the ample information generated during a single recoding.[2] Additionally, automated systems could provide assistance during the highly subjective process of diagnosing laryngeal diseases.

In analyzing parameters collected from laryngeal visualization techniques, support vector machines (SVMs) are becoming increasingly popular. SVMs determine a separation hyperplane that achieves the largest margin of separation between two groups.[9,10] Although SVM is a powerful tool for differentiating groups, its application becomes time consuming and its success is reduced when sample groups become large or complexly distributed.[9] Additionally, SVM only differentiates between two groups, making it inadequate for classifying individual pathologies. Combining SVM with a decision tree overcomes this limitation.[10,11]

In this paper we present an automated method for classifying laryngeal pathologies using HSV. By calculating geometric and textural features previously determined to be important in differentiating larynges and adding the temporal features that may also be defined from a HSV recording,[1,7,8,12] we differentiated normal larynges from those displaying nodules, polyps or leukoplakias. After collection, videos underwent initial processing: smooth filtering, threshold histogram, noise filtering and glottal area detection. Next parameters concerned with the concavity, symmetry, number of glottal regions and the ratio of minimal to maximal glottal areas were computed. Finally, samples were classified using a decision tree based on support vector machines.

II. METHODS AND MATERIALS

Sample Collection. The following experiment is based on HSV recordings from 101 subjects, collected between January 2007 and January 2008 at the Ear, Eye, Nose and Throat Hospital of Fudan University. Samples were obtained in accordance with the ethics committee of Fudan University. After examination by an attending physician, subjects were classified as displaying nodules (21), polyps (37), leukoplakias (21) or normal larynges (22). Summary statistics are shown in table 1. For HSV recordings, subjects were asked to phonate the vowel /a/ at a conformable pitch. Images were recorded with a digital high-speed camera (Kay Elemetrics, Lincoln Park, NJ) with a sampling rate of 2000 frames/second and a resolution of 128 x 256 pixels.

Preprocessing. The preprocessing routine focused primarily on video smoothing and video size reduction. In order to remove error, a smoothing filter took a weighted average of the brightness of each pixel. The intensity of each pixel was averaged with the intensities of the pixels to the right and left of it:

$$\mathrm{Im}\,(i) = \frac{\mathrm{Im}\,(i-1) + 2 * \mathrm{Im}\,(i) + \mathrm{Im}\,(i+1)}{4}$$

Im represents each frame, while *i* is the column number in each frame. Im(*i*-1) is the column to the left of *i*, and Im(*i*+1) is the column to the right of column *i*.[4] In order to produce results within the boundary of the image, the value of *i* is between two and the total number of columns minus one. After processing the raw videos, a representative frame was manually cropped into a more explicit view of the glottal area. This cropping was applied to all frames of the video reducing the non-glottal area and increasing computing accuracy.

Image Segmentation. After the preprocessing phase a threshold was selected to isolate the glottal area. The threshold was determined using a computer generated histogram comprised of two normally distributed peaks representing the object intensities and the background intensities, with a valley containing intermediate gray levels. The threshold was selected as the lowest point within the valley, creating a binary image separating the object from the background.[4] Sometimes due to light reflection on the surface of the vocal folds, object regions are misclassified as background. By inputting a minimum value, the program can scan through the image and erase regions with a total number of pixels below the set minimum. After setting a threshold and eliminating reflective noise, an image of the glottis and threshold during the open phase was saved for further analysis (Figure 1).

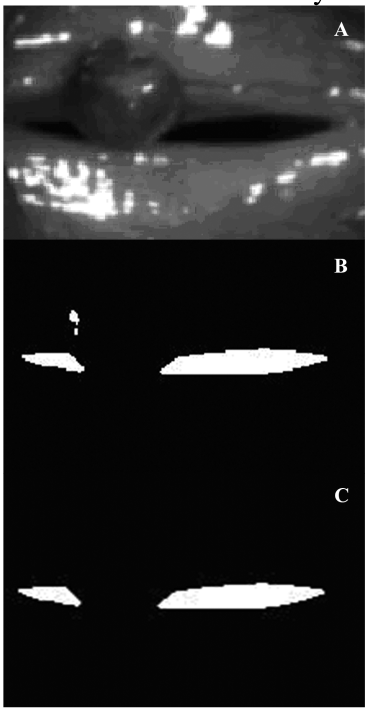


Figure 1: Glottal image and threshold image generated from a subject with vocal fold polyps before selecting a minimum pixel value. The glottal image is cropped to isolate the glottal opening. Laryngeal tissue appears black in the threshold image while the glottal opening is white. Additional areas of white indicate reflective interference. The lower panel is the version after applying a minimum value.

Temporal Feature Calculation. Using the binary threshold image, the glottal area can be easily identified as the white, background region, while the laryngeal tissue is the black, object region. The glottal area was computed as the total number of white pixels in a given frame:

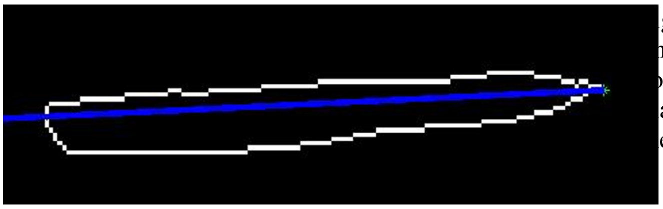
White pixel $\Rightarrow$ Image(i,j) =1 Black pixel $\Rightarrow$ Image(i,j) =0

$$\textit{Glottal}\ \text{Area} = \sum_{i=1}^{R}\sum_{j=1}^{C}\text{Image(i,j)}$$

R and C are the number of rows and columns in a single frame. We collected local information on 30 consecutive frames. The ratio of the minimal and the maximal glottal areas during these 30 frames was recorded.

Spatial Feature Calculation. Using glottal and threshold images, three spatial parameters were calculated. Once the threshold image was selected, the program calculated the number of separate regions within the glottis. This value, the threshold value, was recorded and used as the initial partition criteria in the classification phase.

Next the previously saved glottal and threshold images were used to calculate a measure of symmetry. A reference line was drawn between the anterior and posterior ends of the glottal gap, such that it connected the furthest points of the glottal gap or in cases where one of the ends was defined by a straight edge; bisected this straight edge. The program calculated the area on either side of the reference line and produced an error value representing the difference in the two areas. Due to slight differences in image orientation we allowed the reference line to deviate 10 degrees from the initial line constructed by the two end points, determining the minimal error value achieved within the rotation range (Figure 2). Finally, a measure of vocal fold concavity was calculated. Taking the threshold image, the program added to the glottal area until the edge was a smooth, concave curve. The measure of concavity was calculated as the total pixels added to the image, with samples more close to concave requiring less additional area.



gure 2: Threshold image with symmetry reference ie. The program calculated the line such that it onnected the two ends of the glottal gap with no more an a 10 degree deviation from horizontal to maximize e symmetry reading.

Decision Tree Combined with Support Vector Machine. SVM determines the maximum separation between two groups (Figure 3).[9,10]

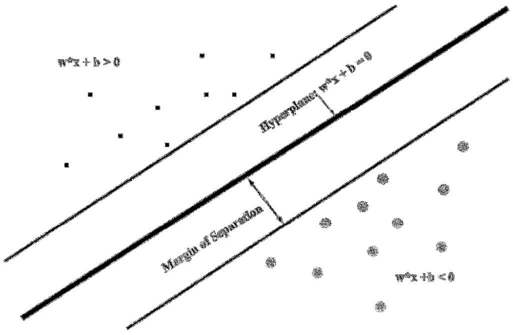


Figure 3: Visual representation of SVM. The dots represent two separable groups. The thick line between them is a hyperplane. The goal of SVM is to maximize the distance of the two groups from this hyperplane as represented by the arrows.

It differs from other classification algorithms because it provides a large margin of separation with relatively simple computations. In our SVM, data from each parameter was first fitted to a normal curve, such that $-1 \leq X \leq 1$, where a negative X value would suggest a more abnormal sample. Each sample was also labeled with a Y value of plus one if normal or negative one if abnormal as classified by the attending physician. Using a training sample, we defined the separation hyperplane with slope "w" and y-intercept "b". Both w and b were initially equal to zero. Each sample (x_i,y_i) was entered into the hyperplane function of $y_i(w*x_i + b)$. If both the sample data and physician classification were in agreement, then the function generated a positive number and no adjustments were necessary. If the function returned a number less than zero, the following adjustments were made to the slope and y-intercept: w' = w + y_ix_i and b' = b + y_i. This process was repeated until $y_i(w*x_i + b)$ was greater than zero for all data, or in the case of outliers that could not be resolved, the program ran for a predetermined duration (1000 integrations) before accepting the hyperplane as maximizing the margin shown in figure 3.[13] The calculations can be summarized as follows:

$$if\ y_i(w\cdot x_i+b)\leq 0\ \ then\ \ \ w'=w+y_ix_i,\ b'=b+y_i$$

$$untill\ y_i(w\cdot x_i+b)>0 for\ all\ i$$

We used the SVM-light developed by Joachims which allows the parameters involved to be adjusted to address the trade-off between sensitivity and specificity.[14]. The preceding explanation reflects SVM with two parameters; the system can evaluate more parameters simultaneously using an increasing number of dimensions for the separation hyperplane and also "kernel tricks" to reduce the dimensions.[13]

SVM only differentiates between two groups. Combining SVM with a decision tree overcomes this limitation. The decision tree is the most commonly used tool for classification in many fields.[15] In our approach, SVMs were trained to separate classes at each level of the tree. After determining the differentiating capabilities of each parameter, we arranged the parameters into a decision tree to separate each pathology cases as shown in figure 4.

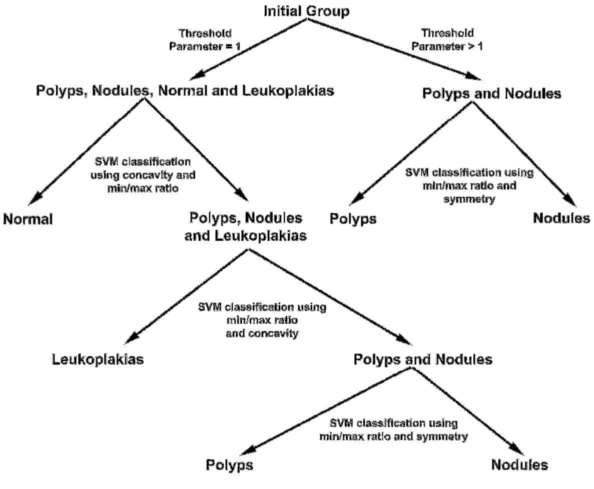

Initial Group
Threshold Parameter = 1
Threshold Parameter > 1
Polyps, Nodules, Normal and Leukoplakias
Polyps and Nodules
SVM classification using concavity and min/max ratio
SVM classification using min/max ratio and symmetry
Normal
Polyps, Nodules and Leukoplakias
Polyps
Nodules
SVM classification using min/max ratio and concavity
Leukoplakias
Polyps and Nodules
SVM classification using min/max ratio and symmetry
Polyps
Nodules

Figure 4: Schematic of the decision tree used to classify individual samples. SVM was used at each node to separate groups.

Statistical Analysis

A Mann Whitney rank sum test was employed to detect significant differences between normal and abnormal samples for each parameter. Next a one way ANOVA on ranks determined the ability of each parameter to differentiate between normal, nodule, polyp and leukoplakia groups. Samples were analyzed with a SVM based decision tree twice. First, the sensitivity and specificity were calculated for the separation of normal from abnormal samples.

Next the percent of samples correctly classified was calculated after data was divided into normal, nodule, polyp and leukoplakia groups.

III. RESULTS

Parameter Evaluation: Before combining all four parameters in the classification phase, we evaluated each parameter on the basis of its ability to differentiate normal from pathological samples. Shown in table 2 are the results of Mann Whitney rank sum tests for each parameter. Measures of threshold, glottal area ratio and concavity all were significantly different for normal and abnormal data (p=0.007, p=0.013 and p<0.001, respectively). Our symmetry parameter did not change significantly with pathology (p=0.294).

Next we evaluated parameters for their ability to classify individual pathologies. The results of a one way ANOVA on ranks revealed significance in all four parameters. Using Dunn's method for pairwise comparison, the threshold measure best differentiated polyps while glottal area ratio found more significant differences with nodules. Concavity was different for normal samples when compared to all three pathologies and the symmetry measure detected a difference between polyps and nodules (Table 2).

SVM Based Decision Tree. Using a decision tree combined with SVM we correctly classified 89.1% of samples as either normal or abnormal. The specificity was 81.9%, while the sensitivity was 91.1%. Table 3 is a confusion matrix reporting these results. Attempts to separate samples into specific disease classes yielded less compelling results. Normal and leukoplakia samples were classified best with accurate classification rates of 81.8% and 76.2%, respectively. Polyps were difficult to discern from nodules resulting in lower correct classification rates of 64.9% and 61.9% respectively. In the nodule group 28.6% of samples were misclassified as polyps. Likewise, 24.3% of polyps were classified as nodules. Overall, samples were correctly classified 70.3% of the time. The confusion matrix for these results is shown in table 4.

IV. DISCUSSION

In this study we were able to successfully apply a computer based classification system to HSV data. In separating normal from abnormal our system correctly identified 89.1% of samples. Especially promising is high sensitivity, missing only 8.9% of abnormal samples. Despite strong results in the detection of laryngeal pathology, our system struggled to differentiate specific conditions. Overall 70.3% of samples were correctly classified as normal, leukoplakia, nodule or polyp. Nodules and polyps were particularly hard to separate.

Recently, much research has focused on the application of computer analysis to various forms of laryngeal imaging.[1,6-8,16,17] Currently, studies by Verikas provide the most exciting results with accuracies as high as 94% when classifying normal, nodular and diffuse pathology groups.[1,7,8]. While our current investigation does not obtain results as strong as these studies, our research represents a novel investigation into the application of these technologies to HSV data. Despite its advantages over stroboscopy, HSV remains relatively uncommon in the clinical setting. We believe that computer systems for the analysis of HSV data are a necessary prerequisite for more widespread implementation. Our aim was to develop a program to be used with HSV for initial screening. Although not yet worthy of clinical implementation; this type of classification system has several clinical advantages. First, the measures detect physically observable parameters with specificity unobtainable by a subjective observer. Additionally, these parameters require minimal computer processing. In contrast, past research has focused on more sophisticated computer processing, requiring longer run times and more user training.[1,6-8] In a clinical setting, a physician may quickly and easily utilize this type of program to generate immediate feedback.

We believe the goal of specific disease identification is not beyond the reach of this type of computerized classification. Using SVM and decision tree algorithms we can separate samples into groups defined by a disease specific combination of parameters. In our study, the parameters of threshold, concavity and the ratio of glottal area differentiated samples well; however, polyps and nodules produce similar values in these categories and thus were not distinguishable. As in past studies, our symmetry parameter struggled to produce consistent results.[12] Tissue surrounding the vocal folds often appeared in the HSV image, obscuring the actual symmetry of the vocal folds, and a proper reference line was difficult to select. Further research could endeavor to determine a better way to quantify symmetry and to explore other parameters that may be valuable in disease classification.

SVM has been used in several previous laryngeal-imaging studies and is a popular analysis tool in computer science and engineering.[18] SVM analysis has a fast training and learning speed even when applied to relatively large sample sizes and may be used to evaluate many parameters at once.[15] While SVM analysis can only separate samples into two groups; by using it at each level of decision tree structure, we can separate data into many more specific groups.[10,11] Moreover, SVM combined with a decision tree can be used to combine individually weak parameters into a robust classification scheme.[11] While our parameters were able to detect differences between specific groups, no one parameter was capable of adequately differentiating all four groups. By combining the parameters in a SVM based decision tree we were able to improve our classification results.

Currently the classification of laryngeal images is limited by the image quality. Higher resolution images make glottal area detection both easier and more accurate. Moreover, measures of vocal fold texture require quality images. Past research has implicated color texture analysis as the most valuable single parameter extracted from vocal fold images.[1,6] Physicians agree that the interpretation of color information contributes heavily to their assessment.[7] Color imaging is a forthcoming development in HSV, which should greatly improve diagnostic potential.[2]

V. CONCLUSION

In this paper, we address the need for an automated support system during the diagnosis of laryngeal pathologies. The concept of applying computerized categorization to HSV data is promising for future work. The use of a decision tree based method in combination with SVM also holds potential. Improvements in HSV image quality and segmentation are needed before a system can generate results strong enough for clinical implementation. Further research could focus on the differentiation of normal and pathological data, as well as the development of parameters to assist in the computerized identification of specific diseases.

Acknowledgement:
The research was supported by NIH grant number 1-R01 DC05522 and DC008850 from the National Institute on Deafness and Other Communication Disorders.

REFERENCES

1. Verikas A, Gelzinis A, Bacauskiene M, Valincius D, Uloza V. A Kernel-based approach to categorizing laryngeal images. Comput Med Imaging Graph. 2007;31: 587-594.
2. Deliyski DD, Petrushev PP, Bonilha HS, Gerlach TT, Martin-Harris B, Hillman RE. Clinical implementation of laryngeal high-speed videoendoscopy: Challenges and evolution. Folia Phoniatr Logop. 2008; 60: 33-44.
3. Bonilha HS, Deliyski DD. Period and glottal width irregularities in vocally normal speakers. J Voice. 2008 [In Press].
4. Zhang Y, Bienging E, Tsui H, Jiang JJ. Efficient and effective extraction of vocal fold vibratory patterns from high-speed digital imaging. J Voice. 2008;[In Press].
5. Heman-Ackah YD. Diagnositc tools in laryngology. Curr Opin Otolaryngol Head Neck Surg. 2004;12(6): 549-552.
6. Ilgner JFR, Palm C, Schütz AG, Spitzer K, Westofen M, Lehmann TM. Colour texture analysis for quantitative laryngoscopy. Acta Otolaryngol. 2003;123: 730-734.
7. Verikas A, Gelzinis A, Bacauskiene M, Uloza V. Towards a computer-aided diagnosis system for vocal cord diseases. Artif Intell Med. 2006; 36: 71-84.
8. Verikas A, Gelzinis A, Bacauskiene M, Uloza V. Integrating global and local analysis of color texture and geometric information for categorizing laryngeal images. Int J Pattern Recogn 2006;20(8): 1187-1205.
9. Zhong W, Chow R, Stolz R, He J, Dowell M. Heirarchical clustering support vector machines for classifying type-2 diabetes patients. In: Măndoiu I, Sunderraman R, Zelikovsky A, Editors. Lecture Notes in Computer Science: Bioinformatics Research and Applications. Springer Berlin 2008:379-389.
10. Takahashi F, Abe S. Decision-tree-based multiclass support vector machines. Proceedings of the 9th International Conference on Neural Information Processing. 2002;3:1418-1422.
11. Nguyen T, Li M, Bass I, Sethi IK. Investigation of combining SVM and decision tree for emotion classification. Proceedings of the Seventh IEEE International Symposium on Multimedia. 2005;3:540-544.
12. Shohet JA, Courey MS, Scott MA, Ossoff RH. Value of videostroboscopic parameters in differentiating true vocal fold cysts from polyps. Laryngoscope. 1996;106(1):19-26.
13. Russell S, Norvig P. Artificial Intelligence: A modern Approach. 2nd Edition. Englewood Cliffs, NJ: Prentice Hall; 2003.
14. Joachims T. Making large-Scale SVM Learning Practical. In: Schölkopf B, Burges C, Smola A, Ed. **Advances in Kernel Methods - Support Vector Learning.** MIT-Press, 1999: 41-56.
15. Bennett KP, Blue JA. A support vector machine approach to decision trees. In Rensselaer Polytechnic Institute Department of Methematical Sciences, Math Report No. 97-100. Troy: NY, 1997:2396-2401.
16. Godino-Llorente JI, Gómez-Vilda P, Sáenz-Lechón N, Blanco-Velasco M, Cruz-Roldán F, Ferrer-Ballester MA. Support vector machines applied to the detection of voice disorders. In: Amir Hussain, Marcos Faundez-Zanuy, Gernot Kubin authors. Lecture

Notes in Computer Science: Nonlinear Analyses and Algorithms for Speech Processing. Heidelberg: Springer Berlin. 2005: 219-230.

17. Colton RH, Woo P, Brewer DW, Griffin B, Casper J. Stroboscopic signs associated with benign lesions of the vocal folds. J Voice. 1995;9(3): 312-325.
18. Chen W, Peng C, Zhu X, Wan B, Wei D. SVM-based identification of pathological voices. Proceedings of the 29th Annual International Conference of the IEEE EMBS. 2007:3786-3789.

Table 1: Summary statistics for human data used.

Diagnosis	Gender Ratio M:F	Mean Age	Total number
Normal	3:8	45.86±11.62	22
Nodules	1:20	40.52±9.65	21
Polyps	18:19	45.86±11.28	37
Leukoplakia	20:1	52.67±8.69	21

Table 2: Individual parameter evaluation.

Parameter	Mean Normal	Mean Abnormal	Mann Whitney Rank Sum	ANOVA	Significant Comparisons
Threshold	1±0	1.4±0.54	p=0.007	p<0.001	Polyps vs Normal Polyps vs Leukoplakia
Glottal Area Ratio	0.07±0.08	0.15±0.15	p=0.013	p<0.001	Nodule vs Normal Nodule vs Leukoplakia
Symmetry	1071.94±893.60	1428.16±1397.43	p=0.294	p=0.002	Nodule vs Polyp
Concavity	165.86±82.99	708.32±514.50	p<0.001	p<0.001	Polyp vs Normal Nodule vs Normal Leukoplakia vs Normal

Table 3: Confusion matrix for normal vs abnormal classification results

Medical Classification	Classification Result		
		Normal	Abnormal
	Normal	18	4
	Abnormal	7	72

Table 4: Confusion matrix for individual pathology classification results.

	Classification Result				
		Normal	Leukoplakia	Nodule	Polyp
Medical Classification	Normal	18	3	1	0
	Leukoplakia	3	16	1	1
	Nodule	1	1	13	6
	Polyp	1	3	9	24

FIGURES

Figure 1: Glottal image and threshold image generated from a subject with vocal fold polyps before selecting a minimum pixel value. The glottal image is cropped to isolate the glottal opening. Laryngeal tissue appears black in the threshold image while the glottal opening is white. Additional areas of white indicate reflective interference. The lower panel is the version after applying a minimum value.

Figure 2: Threshold image with symmetry reference line. The program calculated the line such that it connected the two ends of the glottal gap with no more than a 10 degree deviation from horizontal to maximize the symmetry reading.

Figure 3: Visual representation of SVM. The dots represent two separable groups. The thick line between them is a hyperplane. The goal of SVM is to maximize the distance of the two groups from this hyperplane as represented by the arrows.

Figure 4: Schematic of the decision tree used to classify individual samples. SVM was used at each node to separate groups.