Breast cancer diagnosis from fluorescence spectroscopy using support vector machine

Jiyoung Choi, Sharad Gupta, Inho Park, Doheon Lee, and Jong Chul Ye[a]

Korea Advanced Institute of Science and Technology (KAIST), 373-1 Guseong-dong Yuseong-gu, Daejon 305-701, Korea

ABSTRACT

A novel support vector machine (SVM) classifier incorporating the *complexity* of fluorescent spectral data is designed to reliably differentiate normal and malignant human breast cancer tissues. Analysis has been carried out with parallel and perpendicularly polarized fluorescence data using 36 normal and 36 cancerous tissue samples. In order to incorporate the complexity of fluorescence spectral profile into a SVM design, the curvature of phase space trajectory is extracted as a useful complexity feature. We found that the fluorescence intensity peaks at 541nm-620nm as well as the complexity features at 621nm-700nm are important discriminating features. By incorporating both features in SVM design, we can improve both sensitivity and specificity of the classifier.

Keywords: fluorescence spectroscopy, breast cancer diagnosis, phase space trajectory, complexity, SVM, random forest

1. INTRODUCTION

Optical spectroscopy provides new ways to characterize physical and chemical changes occurring in tissues and cells, and thereby offers exciting possibilities for novel diagnostic and therapeutic approaches.[1] Optical techniques such as fluorescence, Raman and light scattering can be used to distinguish cancerous from non-cancerous state of a tissue.[2–4] The underlying physical basis is that they are strongly influenced by cellular structure and chemical composition of the tissues. Due to its sensitivity to minute variations, fluorescence spectroscopy can provide quantitative biochemical information about the state of the tissue, which may not be obtained using standard pathology. Over the past 15 years its diagnostic potential has been tested in different organs of the body, including the mouth, breast, esophagus, and bladder, etc.[2–6]

A number of fluorophores ranging from structural proteins to various enzymes and co-enzymes are present in the human tissue and can be excited by ultra- violet and visible light.[1,2] The fluorescence emission can differ significantly in normal and cancerous tissues due to the differences in concentrations of absorbers and scatterer, as also the scatterer sizes. Polarized fluorescence spectroscopy is useful in isolating the characteristic spectral features from the diffuse background since the parallel components of the fluorescence suffers fewer scattering events.[7] In comparison, the intensity of the perpendicular component is affected more by scatterers and contains more information about scattering.

A number of studies conducted so far have established certain broad biochemical changes occurring in tumor tissues, which leave characteristic signatures in the spectral domain. Most of the optical diagnosis of cancer use traditional statistical techniques such as Fisher's linear discriminant analysis,[3,4] principal component analysis (PCA),[8,9] or artificial neural network analysis (ANN).[10,11] Recently, support vector machine (SVM)[12,13] combined with recursive feature elimination (RFE)[14] have been proposed and outperformed the existing classifiers in the literatures.[15] Local features such as the sudden spectral intensity change of the benign tissues around 590nm and the emergence of a minor peak in the malignant tissue around 620nm were found important in a classifier design.

The main contribution of this paper is to develop a more robust diagnostic tools by incorporating the complexity of the fluorescence intensity profile. Recently, nonlinear signal analysis has been applied for analyzing complex signals, like heart beat and brain activities.[16,17] The underlying principle of the nonlinear signal analysis is that some stochastic phenomena often comes from a deterministic dynamic system in a high dimensional

[a] Corresponding Author: Jong Chul Ye, E-mail: jong.ye@kaist.ac.kr, Telephone: 82-42-869-4320.

Optical Tomography and Spectroscopy of Tissue VII
edited by Britton Chance, Robert R. Alfano, Bruce Jason Tromberg, Mamoru Tamura, Eva Marie Sevick-Muraca
Proc. of SPIE Vol. 6434, 64340P, (2007) · 1605-7422/07/$18 · doi: 10.1117/12.700800

phase space. According to physiological findings, there exist morphological changes such as enlargement and hypercromasia of nuclei, overcrowding and irregular cellur arrangement.[18] Hence, we may expect that such morphological changes *deterministically* alters light propagation and scattering properties of tissues and hence affect the complexity of fluorescence spectra.[18]

In order to verify the conjectures, we have designed a support vector machine (SVM) classifier using the fluorescence intensity and complexity features. As a measure of complexity, we use the curvature of the phase space trajectory constructed from the fluorescent intensity profile. Curvature is a scalar valued differential geometric quantity, which is invariant to rotation and translations. Recall that the curvature has been often used to describe the complexity of a signal in image processing and computer vision.[19] Analysis suggests that the curvature features are important at 621-700nm region especially from perpendicular polarized channel, whereas intensity features are most dominating features at 541nm-620nm. The SVM classification results also show that classifier using intensity features exhibits higher sensitivity of cancerous tissues at 541nm-620nm, whereas curvature based SVM shows higher specificity in 621-700nm wavelength regions. Furthermore, we can improve both sensitivity and specificity over 90% by using both signal simultaneously. To our knowledge, this *complexity based* classifier design is novel and has not been reported. Furthermore, we believe that this complexity feature can effectively identify the *deterministic* light scattering changes due to morphological changes of cellular structures.

This paper is organized as following. In Section 2, the theory of nonlinear signal complexity analysis is reviewed. Section 3 describes the methods of data collection and SVM design based on intensity and complexity features. Experimental results are given in Section 4, which is followed by discussions and conclusion in Section 5.

2. THEORY

2.1. Phase Space Trajectory

The nonlinear signal analysis suggests that a noise-like signal may come from a complex deterministic dynamic system in a high dimensional embedding space.[20,21] In fact, it has been shown that if the signal is measured over a sufficiently long interval with sufficient accuracy, it is possible to make the quantitative interference about the dynamics through the "method of delays".[20,21]

In order to apply the method of delay, assume that $x(t)$ is the original signal (fore example, fluorescence intensity profile indexed by the wavelength t). Then, the signal can be represented as a point at t in an m-dimensional space:

$$y_t(m) = [x(t), x(t-\tau), \cdots, x(t-(m-1)\tau)], \tag{1}$$

where the number m is called the *embedding dimension*. There are several methods in literature used to choose the time delay. One of the popular approach is the use of delayed mutual information suggested by Fraser and Swinney.[22] More specifically, the delayed mutual information can be calculated by

$$I(\tau) = -\sum_{ij} p_{ij}(\tau) \ln \frac{p_{ij}(\tau)}{p_i p_j} \tag{2}$$

where p_i is the probability to find a signal value in the i-th interval from some partition of the values, and $p_{ij}(\tau)$ is the joint probability that the value of the current observation and the τ-delayed observation fall into the i-th and j-th interval simultaneously. Since each signal coordinates in the m-th embedding space should be independent in noise free cases, the mutual information calculated from Eq. (2) should have the minimum value at the correct delay.

In order to calculate the embedding dimension, assume that the minimum embedding dimension for a given sequence is m_0. Then, for the embedding dimension estimate m becomes $m \geq m_0$, a sequential path from one m-dimensional delay space never cross itself. However, for the estimate $m < m_0$, various path of the sequential path will cross, and two neighboring points at the crossing point tends to far away in the sequential order of the embedded data series.[23] This behaviour can be captured by the following measure[23]

$$a(i,m) = \frac{||y_i(m+1) - y_{n(i,m)}(m+1)||_\infty}{||y_i(m) - y_{n(i,m)}(m)||_\infty}, \quad i = 1, 2, \cdots, N - m\tau \tag{3}$$

where N is the length of the sample, $||\cdot||_\infty$ is the l_∞ norm and $y_i(m+1)$ is the i-th reconstructed vector in the $m+1$ embedding space given in Eq. (1); $n(i,m)$ is an index such that $y_{n(i,m)}(m)$ is the nearest neighbor of $y_i(m)$ in the m-dimensional reconstructed phase space in the sense of l_∞ norm. In order to avoid any problem dependent thresholds, Cao define the following dimension free parameter[24]

$$E1(m) = E(m+1)/E(m) \tag{4}$$

where $E(m)$ denotes the mean value of all $a(i,m)$'s:

$$E(m) = \frac{1}{N-m\tau} \sum_{i=1}^{N-m\tau} a(i,m) \ , \tag{5}$$

Cao's main idea is that the $E1(m)$ stop changing when $m \geq m_0$. This is the estimate of the embedding dimension.[24]

2.2. Approximate Entropy

Entropy is an information theoretic quantity describing the amount of disorder in a system. Information about how the signal fluctuates with respect to index is obtained by comparing the sequence with itself, but delayed by a specified value. In order to calculate the entropy of the dynamic system in a phase space, Pincus proposed Approximate Entropy (ApEn) and successfully applied it to relatively short and noisy data.[25]

2.3. Curvature and Complexity of Phase Trajectory

While the ApEn measure is an important measure of signal complexity, it is a scalar feature representing global complexity of a signal. In designing a classifier, however, local complexity feature may be more important since it has richer information than the global feature. Hence, our main idea is to exploit the local complexity of signal using the curvature of phase trajectory.

The curvature of a space curve in n-dimensional space can be explicitly calcuated using Frenet frame. Recall that Frenet frame show the kinematic properties of points on a continuous and differentiable curve.[26] The Frenet frame can be described by tangent($\mathbf{T}$), normal($\mathbf{N}$), and binormal($\mathbf{B}$) unit vectors defined in Eq. (6). More specifically, if $\mathbf{r}$ be the position vector of each point, then we have

$$\mathbf{T} = \frac{\mathbf{r}'}{|\mathbf{r}'|}, \quad \mathbf{N} = \frac{\mathbf{T}'}{|\mathbf{T}'|}, \quad \mathbf{B} = \mathbf{T} \times \mathbf{N} \ . \tag{6}$$

Here, curvature κ can be calculated from tangent vector $\mathbf{T}$ as given in Eq. (7).

$$\kappa = |\mathbf{T}'| \tag{7}$$

This curvature κ represents the complexity of a given signal, which is also related to Komogorov complexity in information theory.[27] For example, in image processing or computer vision applications, the minimum curvature flow has been successively used for removing noises.[19] The basic idea of minimum curvature flow is that the true noiseless signal is less complex than the noisy signal. Hence, the use of curvature as a complexity measure can be justified.

3. METHOD

3.1. Data Collection

The samples were excited by 488-nm wavelength plane polarized light from an Ar-ion laser (Spectra Physics 165, 5W). The polarized fluorescence spectra were collected in right angle geometry using triple-mate monochromator. For polarized fluorescence, a depolarizer was used after the analyzer, in order to ensure that there was no preference of the selected directions of polarized fluorescence by the detection system as shown in Figure 1.

The components of fluorescence light that are parallel and perpendicular to the incident polarized light were measured in the 500- to 700-nm wavelength region using triplemate monochromator (SPEX-1877E) and

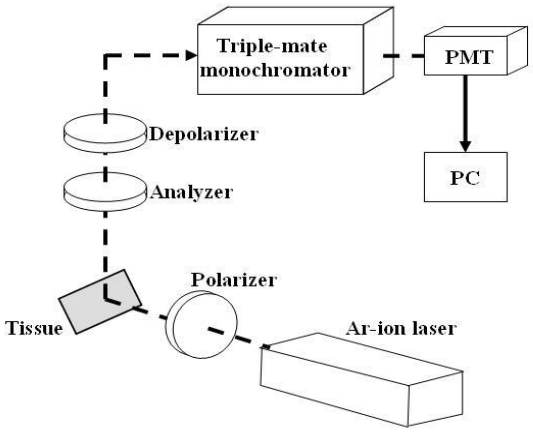

Figure 1. Experimental diagram used for data collection.

PMT (RCA C-31034). In this wavelength region, FAD is the dominant fluorophore in the fluorescence spectra from tissues, which has a fluorescence peak at 530 nm. The main absorbers in this region are hemoglobin and porphyrin, which absorb at 540-nm and 580-nm wavelengths.[1,2] Porphyrin also acts as a weak fluorophore, with an emission peak at 630 nm.[2] In total, 36 tumor tissue samples with their normal counterparts from each subjects were supplied by Ganesh Shankar Vidyarthi Memorial Medical College, Kanpur, India, after surgery. These were analyzed in close collaboration with the pathologist of the hospital. The age of patients spanned a broad range, from 16 to 85 years, coming from varied economic backgrounds. The collected samples were analyzed on the same day, without any chemical treatment.

3.2. SVM Classifier Design

For optical diagnostics of cancerous cells, support vector machines (SVM)[12,13,28] have been recently proposed, which outperformed the existing classifiers in the literatures.[15] Our SVM classifier design is, however, different from the conventional SVM approaches since our classifier also incorporates the curvature of phase space trajectory as a complexity measure in addition to the conventional intensity features. As will be explained later, the additional complexity measure significantly improves the specificity of the classifier.

A support vector machine is a learning algorithm based on statistical learning theory. Training SVM for pattern recognition problem can be represented as a quadratic optimization problem. For a given set of data (x_i,y_i) and a capacity constant C, support vectors can be calculated by solving the following optimization problem: minimize the following with respect to $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_l]^T$

$$W(\alpha) = -\sum_{i=1}^{l} \alpha_i + \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j k(x_i, x_j) \tag{8}$$

subject to

$$\sum_{i=1}^{l} y_i \alpha_i = 0 \quad , \quad \forall i : 0 \leq \alpha_i \leq C \tag{9}$$

where $k(x_i, x_j)$ denotes the kernel. For a fair evaluation, a public domain SVM tool - mySVM[32] has been used, which is designed based on the algorithm suggested by Joachims.[29]

Since not all the features from whole intensity or curvature are supportive for the SVM classifier, we select a few dominant features using *random forest* algorithm.[30] Random forest is a voting procedure for most popular class among large number of trees that are randomly generated. According to the strength of individual trees and the correlation between them, the generalization error of a forest of tree classifiers is effected. There are internal estimates that monitor error, strength and correlation. The importance of each variable can be measured from these estimates.

Various numbers of features were chosen from random forests result. To construct random forest, randomForest R package[33] based on Breiman's algorithm[30] has been used.

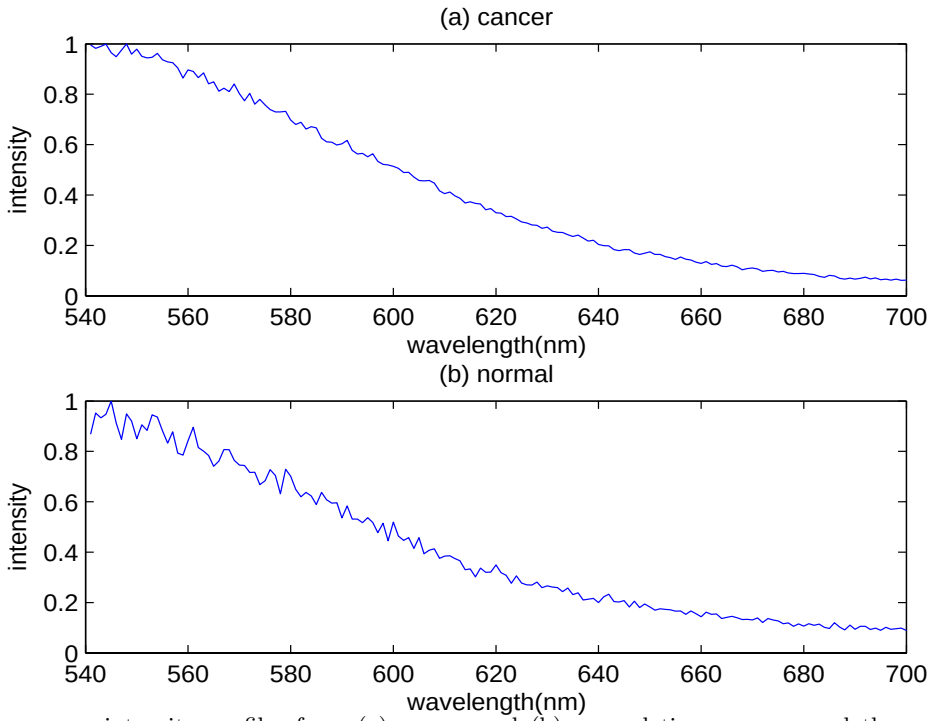

Figure 2. Fluorescence intensity profiles from (a) cancer and (b) normal tissues, measured through perpendicularly polarized channel.

4. RESULTS

The fluorescence data from normal tissue has remarkable peaks around 520nm wavelength. However, since the peaks are due to the fat, it is more dependent to whether the extracted tissue contains fat or not, rather than whether the tissue is malignant or not. Therefore, we focused on the wavelength after 540nm to eliminate the peaks.

4.1. Nonlinear Complexity Analysis

Figures 2(a)(b) illustrate the typical examples of fluorescence profiles from cancerous and normal cells, respectively. By inspection, two signal intensities are noticeably different at the wavelength between 541nm-620nm. However, only few variations are observed between 621nm-700nm. We want to explore these minute differences furthermore using phase space trajectory.

After calculating the optimal delays using Eq. (2), the embedding dimensions were estimated using TSTOOL[31] implementation of Cao's algorithm, cao. For illustration, we plot typical phase space trajectories of fluorescence profiles in the range of 621nm-700nm from cancer and normal tissues in Figures 3(a)(b), respectively. In our experiments, the delay had been fixed as 3 to keep the data length constantly since the average delay was about 3 regardless of different types of signals. Also, the embedding dimension was fixed to 3. From the phase space

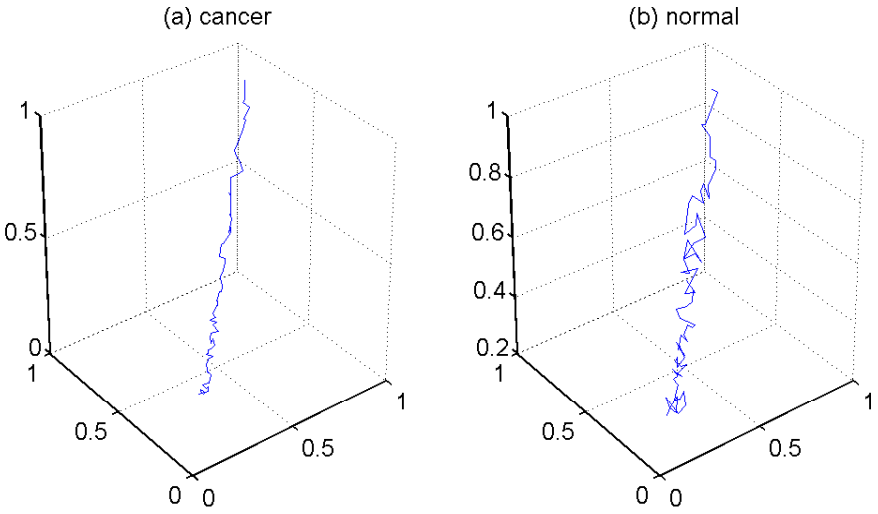

Figure 3. Phase space trajectories of fluorescence profile at 621nm-700nm with embedding dimension m=3, and delay τ=3: (a) cancer tissue and (b) normal tissue.

trajectory, the curvature was calcuated using Eq. (7). An example of curvatures is shown in Figure 4. By inspection, we can easily observe that the phase space trajectory of cancer tissue is more linear compared to the complex trajectory of the normal tissue. Hence, the curvature is used for classification as complexity features.

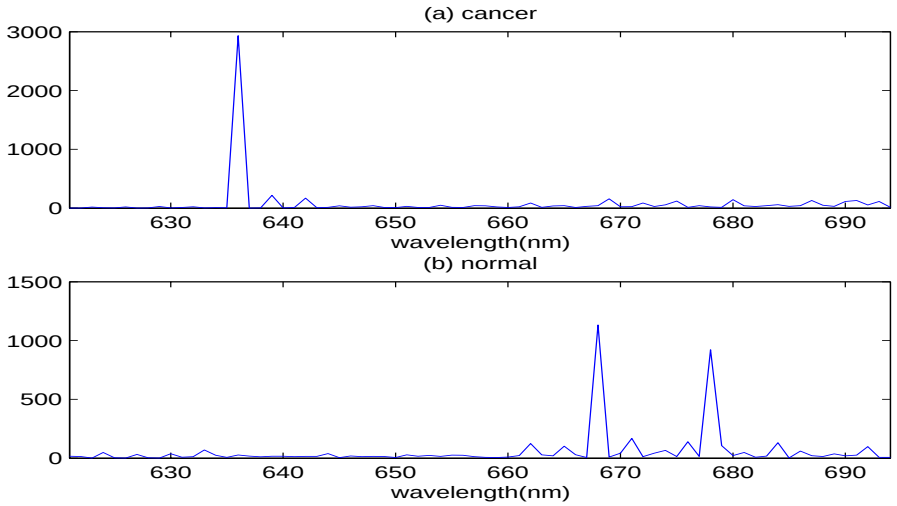

Figure 4. Curvature of phase space trajectory of fluorescence profile at 621nm-700nm: (a) cancer tissue and (b) normal tissue.

Table 1 illustrates the percentage of sample sets whose ApEn of normal tissue is bigger than that of cancerous one. Note that at 541nm-620nm regions, no consistent behaviors of ApEns are observed that can distinguish cancer and normal cells. This may be caused due to strong intensity features such as hemoglobin and pophyrin - that absorb 540nm and 580nm. However, for the wavelength between 621nm-700nm, as shown at the second row of Table 1, we now start to see that ApEns for the normal tissues are consistently higher than their cancerous counterparts. These results confirm the physiological findings that there exists morphological changes such as enlargement and hypercromasia of nuclei, overcrowding and irregular cellur arrangement, which alters light propagation and scattering properties of tissues and hence affect the complexity of fluorescence spectra.[18]

Another observation is that these behaviors are more dominant in perpendicular channels. Since the PE

Approximate Entropy	parallel	perpendicular
541-620 nm	47.2%	59.3%
621-700 nm	75.0%	80.6%

Table 1. Percentage of tissue samples whose ApEns for normal tissue is bigger than those of cancer tissue.

	Intensity		Curvature	
	parallel	perpendicular	parallel	perpendicular
541-620 nm	35.9%	27.2%	18.4%	18.4%
621-700 nm	30.5%	26.2%	19.9%	23.4%

Table 2. Percentage of specific features selected by random forest algorithm as the top 30 most important features.

channel measures the multiple scattered light intensities, the increased discrimination based on ApEns in perpendicular channel coincides our conjecture that the multiple scattering manifests more complicated dynamics.

4.2. Classifier Performance

4.2.1. Random Forest

In order to extract the dominant features for SVM classifier design, random forest algorithm was first applied to the four data sets, i.e. the fluorescence intensity and the curvature features for parallel and perpendicular components, respectively. Specifically, the four data sets (parallel intensity and curvature, perpendicular intensity and curvature) were divided according to the corresponding wavelength: i.e. 541-620nm, and 621-700nm. Then, random forest has been generated ten times for each mixed data set, and the percentage of selecting each features are calculated. The result was obtained by counting how many times the specific feature are ranked among the top most important thirty features.

The result of random forest for whole signal is drawn in Table 2. Note that the more curvature features are obtained at 621nm-700nm regions especially at the perpendicular polarized channels. This result coincides with those of ApEn such that in the wavelength region between 541nm and 620nm, the intensity data is more discriminating features, whereas in the wavelength between 621nm and 700nm, the complexity feature becomes important features, especially in perpendicular polarized channel.

4.2.2. Support Vector Machine

SVM classifiers were designed using the significant features from the random forest algorithm. Various kernels have been tested and the dot product kernel is found most robust. The number of features for SVM were determined by number of features among the top thirty features using the random forest; hence the specific number of features could be slightly different. Cross validation technique was used to obtain an unbiased estimate of the classifier performance. Specifically, among 36 sets of cancer and normal tissues, we randomly picked 3 sets from each group as test sets. The remaining sets were used for training SVM. This random shuffling of training and test set was repeated for 120 times, and the average values were obtained as an estimate of the classifier performance.

Interestingly, SVM using intensity data only has higher sensitivity of cancer detection whereas SVM using curvature data only has higher specificity for normal tissue (See Table 3). However, if the SVM is designed using

541-700 (nm)	Intensity (57)		Curvature (60)		All (57)	
	sensitivity	specificity	sensitivity	specificity	sensitivity	specificity
training set	95.1 %	83.9%	79.4 %	96.1%	94.7 %	90.2%
test set	94.4 %	84.8%	78.0 %	94.2%	92.6 %	91.2%

Table 3. Classification results using SVM using various such as intensity, curvature, and all of them at the 541nm-700nm. Both parallel and perpendicular channels are used for feature selection. The numbers in blanket in first row denotes the number of actual features used for classification of the signal.

541-620 (nm)	Intensity (54)		Curvature (64)		All (57)	
	sensitivity	specificity	sensitivity	specificity	sensitivity	specificity
training set	90.9 %	83.3%	78.5 %	97.1%	90.5 %	89.6%
test set	91.2%	82.6%	77.3 %	96.5%	90.3 %	86.8%

Table 4. Classification results using SVM using various such as intensity, curvature, and all of them at the 541nm-620nm. Both parallel and perpendicular channels are used for feature selection. The numbers in blanket in first row denotes the number of actual features used for classification of the signal.

621-700 (nm)	Intensity (62)		Curvature (62)		All (63)	
	sensitivity	specificity	sensitivity	specificity	sensitivity	specificity
training set	78.1 %	80.7%	70.5 %	94.9%	90.0%	90.7%
test set	75.2 %	79.4%	68.1 %	91.9%	90.5%	90.7%

Table 5. Classification results using SVM using various such as intensity, curvature, and all of them at the 621nm-700nm. Both parallel and perpendicular channels are used for feature selection. The numbers in blanket in first row denotes the number of actual features used for classification of the signal.

both intensity and curvature data, the overall performance of the classifier are improved in both sensitivity and specificity.

In order to investigate wavelength dependency of the behaviours, another sets of SVMs are designed using intensity and curvature data at the wavelength of 541-620nm and 621nm-700nm, respectively (See Table 4 and Table 5). Note that curvature based SVMs have clearly higher specificity compared to their counterparts. Curvature based SVMs have specificity over 90% in all wavelength regions. Intensity based SVMs have sensitivity over 90% at 541-620nm; however, in wavelength range between 621nm and 700nm the sensitivity are degraded. Again by combining intensity and curvature features together in designing SVMs, the overall classifier performance were improved in both sensitivity and specificity. Especially, this synergic effect is more noticeable in Table 5, where each SVM using a single intensity or curvature feature alone has sensitivity less than 80%; however, by combining both features in SVM, the overall sensitivity and specificity were significantly improved. This result confirms that complexity information is a useful feature for cancer diagnosis.

5. DISCUSSIONS AND CONCLUSION

A novel SVM classifiler incorporating the complexity of a signal was designed for optical diagnosis of breast cancer. The analysis was carried out with parallel and perpendicularly polarized fluorescence data obtained from 36 normal and 36 cancerous tissues samples. Random forest results showed that the conventional intensity based classifier captures the intensity peaks at 541nm-620nm as the major discriminating features, whereas the complexity feature from the curvature of phase space trajectory was important at the wavelength of 621nm-700nm. Extensive SVM classifier results showed that the intensity based classification has high sensitivity wheras the curvature based classification has high specificity. Classification using both signal has better specificity in all cases, and also has better sensitivity in 621-700nm wavelength region. These experimental results confirmed our conjecture that complexity information of a fluorescence data contributes to improve the classification performance of cancer tissues from normal tissues.

ACKNOWLEDGMENTS

This work is supported by grant No. 10017755 from Korea Ministry of Commerce, Industry and Energy (MOCIE). IP and DL were supported by the Korea Science and Engineering Foundation(KOSEF) grant funded by the Korea government(MOST)(No. 2005- 01450).

REFERENCES

1. R. R. Alfano, D. Tata, J. Cordero, P. Tomashefshy, F. Longo, and M. A. Alfano, "Laser induced fluorescence spectroscopy from native cancerous and normal tissues," *IEEE J. Quantum Electron.*, vol. 20, pp. 1507–1511, 1984.

2. R. R. Alfano, G. C. Tang, A. Pradhan, W. Lam, D. S. J. Choy, and E. Opher, "Fluorescence spectra from malignant and normal human breast and lung tissues," *IEEE J. Quantum Electron.*, vol. 23, no. 10, pp. 1806– 1811, 1987.
3. G. A. Wagnieres, W. M. Star, and B. C. Wilson, "In vivo fluorescence spectroscopy and imaging for oncological applications," *Photochem. Photobiol.*, vol. 68, pp. 603–632, 1998.
4. E. Servick-Muraca and R. Richards Kortum, "Quantitative optical spectroscopy for tissue diagnosis," *Annu. Rev. Phys. Chem.*, vol. 47, pp. 556–606, 1996.
5. N. Ramanujam, "Fluorescence spectroscopy of neoplastic and nonneoplastic tissues," *Neoplasia*, vol. 2, no. 1, pp. 1–29, 2000.
6. A. Mahadevan-Jansen and R. Richards-Kortum, "Raman spectroscopy for the detection of cancers and precancers," *J. Biomed. Opt.*, vol. 1, no. 1, pp. 31–70, 1996.
7. D. B. Tata, M. Foresti, J. Cordero, P. Tomashefsky, M. Alfano, and R. R. Alfano, "Fluorescence polarization spectroscopy and time resolved fluorescence kinetics of native cancerous and normal rat kidney tissues," *Biophys. J.*, vol. 50, pp. 463–469, 1986.
8. N. Ramanujam, M. Follen-Mitchell, A. Mahadevan-Jansen, S. Thomson, G. Staerkel, A. Malpica, T. Wright, N. Atkinson, and R. Richards-Kortum, "Cervical precancer detection using multivariate statistical algorithm based on laser-induced fluorescence spectra at multiple excitation wavelengths," *Photochem. Photobiol.*, vol. 64, pp. 720–735, 1996.
9. A. Zuluaga, U. Utzinger, H. Durkin, A. Fuchs, A. Gillenwater, R. Jacob, B. Kemp, J. Fan, and R. Richards-Kortum, "Fluorescence excitation-emission matrices of human tissue: a system for in-vivo measurement and method of data analysis," *Appl. Spectrosc.*, vol. 53, pp. 302–311, 1999.
10. H. J. van Staveren, R. L. van Veen, O. C. Speelman, M. J. Witjes, W. M. Star, and J. L. Roodenburg, "Classification of clinical autofluorescence spectra of oral leukoplakia using an artificial neural network: a pilot study," *Oral Oncol.*, vol. 36, no. 3, pp. 286–293, 2000.
11. K. Tumer, N. Ramanujam, J. Ghosh, and R. Richards-Kortum, "Ensembles of radial basis function networks for spectroscopic detection of cervical pre-cancer," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 8, pp. 953–961, 2001.
12. V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
13. C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998.
14. I. Guyon, J. Weston, S. Barnhill, and V. N. Vapnik, "Gene selection for cancer classification using support vector machines," *J. Mach. Learn.*, vol. 46, pp. 389–422, 2002.
15. S. K. Majumder, N. Ghosh, , and P. K. Gupta, "Support vector machine for optical diagnosis of cancer," *Journal of Biomedical Optics*, vol. 10, no. 2, pp. 024034, March/April 2005.
16. Hoyer D, Schmidt K, Bauer R, Zwiener U, Kohler M, Luthke B, and Eiselt M, "Nonlinear analysis of heart rate and respiratory dynamics," *IEEE Engineering in Medicine and Biology Magazine*, vol. 16, no. 1, pp. 31–39, 1997.
17. M. Le Van Quyen, M. Chavez, D. Rudrauf, , and J. Martinerie, "Exploring the nonlinear dynamics of the brain," *Journal of Physics*, vol. 97, no. 4-6, pp. 629–39, 2003.
18. V. Backman, M. B. Wallace, L. T. Perelman, J. T. Arendt, R. Gurjar, M. G. Muller, Q. Zhang, G. Zonios, E. Kline, T. McGillican, S. Shapshay, T. Valdez, K. Badizadegan, J. M. Crawford, M. Fitzmaurice, S. Kabani, H. S. Levin, M. Seiler, R. R. Dasari, I. Itzkan, J. Van Dam, and M. S. Feld, "Detection of preinvasive cancer cells," *Nature*, vol. 406, pp. 35–36, 2000.
19. J. A. Sethian, *Level Set Methods and Fast Marching Methods*, Cambridge University Press, United Kingdom, 1996.
20. T. Schreiber, "Interdisciplinary application of nonlinear time series methods," *Physics*, 1998.
21. Boccaletti S, Grebogi C, Lai YC, Mancini H, and Mazaet D, "The control of chaos: Theory and applications," *Physics Reports*, vol. 329, pp. 108–109, 2000.
22. Fraser AM and Swinney HL, "Independent coordinates for strange attractors from mutual information," *Phys Rev A*, vol. 33, pp. 1134–1140, 1986.

23. M. Kennel, R. Brown, and H. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A*, vol. 45, pp. 3403–3411, 1992.
24. L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D.*, vol. 110, pp. 43–50, 1997.
25. Pincus SM, "Approximate entropy as a measure of system complexity," *Proc. Natl. Acad. Sci. USA*, vol. 88, pp. 2297–2301, 1991.
26. B. O'Neill, *Elementary Differential Geometry*, Academic Press, Inc, New York, 1966.
27. T. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., New York, 2 edition, 1991.
28. C. Cortes and V. N. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
29. T. Joachims, *Advances in Kernel Methods-Support Vector Learning*, chapter Making large-scale SVM learning practical, MIT Press, Cambridge, USA, 1999.
30. L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, October 2001.
31. TSTOOL can be found at http://www.physik3.gwdg.de/tstool/.
32. mySVM can be found at http://www-ai.cs.uni-dortmund.de/SOFTWARE/MYSVM/.
33. randomForest R package can be found at http://www.r-project.org/index.html.