J Fluoresc (2012) 22:1281–1289
DOI 10.1007/s10895-012-1070-0

ORIGINAL PAPER

# Support Vector Machine on Fluorescence Landscapes for Breast Cancer Diagnostics

**Tatjana Dramićanin · Lea Lenhardt · Ivana Zeković · Miroslav D. Dramićanin**

Received: 29 November 2011 / Accepted: 29 May 2012 / Published online: 8 June 2012


**Abstract** Excitation-emission matrices (EEM) and total synchronous fluorescence spectra (SFS) of normal and malignant breast tissue specimens are measured in UV–VIS spectral region to serve as data inputs in development of Support Vector Machine (SVM) based breast cancer diagnostics tool. Various input data combinations are tested for classification accuracy using SVM prediction against histopathology findings to discover the best combination regarding diagnostics sensitivity and specificity. It is shown that with EEM data SVM provided 67 % sensitivity and 62 % specificity diagnostics. With SFS data SVM provided 100 % sensitivity and specificity for a several input data combinations. Among these combinations those that require minimal data inputs are identified.

**Keywords** Breast cancer · Fluorescence · Support vector machine · Synchronous fluorescence

## Introduction

Breast cancer is one of the most frequently diagnosed cancers among women in the world of nearly every racial and ethnic group [1]. Breast cancer is also one of the leading causes of deaths from cancer for the female population. However, if detected in early stage this cancer is one of the most treatable forms of cancer. Because of that, the advancement of existing and developing new method for diagnosis cancer is of crucial importance. In last few decades, diagnostic techniques based on optical spectroscopy has been considered as an alternative technique for the conventional diagnostic. Most clinical applications have concentrated on absorption, fluorescence and elastic scattering spectroscopies, because these measurements can be obtained with a good signal-to-noise ratio in reasonably short integration times. These techniques are fast, noninvasive, reproducible and quantitative. Also, these techniques have the potential to link the biochemical and morphological properties of tissues to individual patient care.

Fluorescence has proven to be a versatile tool for studying molecular interactions in analytical chemistry, biochemistry, cell biology, photochemistry, and environmental science. Fluorescence spectroscopy has been widely explored as diagnostic tools in the field of cancer [2]. Its advantages over other light-based methods are high sensitivity, high speed and safety. For all these techniques, fluorescence has been a specially prosperity because all tissues fluoresce in the range 360–600 nm. In any case, it is able to distinguish the normal from abnormal tissues in various organ, such as breast [3, 4], colon [5], head and neck [6], skin [7], etc.

All tissues are rich in complex mixture of substances, involving a large number of molecules that can absorb and emit light of different wavelengths, so-called fluorophores. The endogenous fluorophores, such as nicotinamide adenine dinucleotide (NADH), flavin adenine dinucleotide (FAD), collagen, elastin, amino acids, vitamins, lipids and porphyrins, have a significant variation in the concentration and their distribution in different tissue types. These differences, together with alternations in the local environment within the tissue, are the basis for the discrimination between tumor and normal tissue by fluorescence spectroscopy.

To detect majority of fluorescence changes between normal and decease tissues more complex fluorescence methods are used: excitation–emission matrix (EEM) spectroscopy and synchronous fluorescence spectroscopy (SFS). Fluorescence

T. Dramićanin · L. Lenhardt · I. Zeković · M. D. Dramićanin (✉)
Vinča Institute of Nuclear Sciences, University of Belgrade,
P.O. Box 522, 11001 Belgrade, Serbia
e-mail: dramican@gmail.com

landscape spectroscopy, also known as EEM, utilizing multiple-color illumination, with the full fluorescence spectrum recorded for each excitation wavelength. The different excitation wavelengths might be expected to variously excite different fluorophores, resulting in more complex emission patterns with more information relevant to biochemical changes than for single-color excitation, and with presumed greater likelihood of distinguishing malignancy from normal conditions [8]. Recently, synchronous fluorescence spectroscopy (SFS) has been used in cancer detection [4, 9–16]. The SFS method involves simultaneous scanning of both emission and excitation wavelengths while keeping the interval of wavelengths (constant-wavelength mode) or frequencies (constant-energy mode) between them constant. The synchronous spectrum often has more features and thus provides more information than ordinary emission spectra.

However, observed data are subtly related in ways that are often difficult to express in the form of diagnostic rules and must be processed for tissue classification purposes. Mathematical algorithms then can be developed and optimized to classify tissues into their respective histological category based on their spectral features.

Support vector machine (SVM) is a statistical technology developed by the machine learning community [17, 18] that can be used for both classification and regression. Compared with other machine learning methods, SVM has such advantages as it does not need a large number of training samples for developing model and is not affected by the presence of outliers [19]. Having high generalization procedure and feasibility to extract higher order statistics the SVM [17, 18, 20] has become extremely popular in terms of classification and prediction.

In the context of classification SVM is transforming original data space into much higher dimension to make classification groups linearly separable. Given a set of points that all belong to one of two classes, SVM can find the hyperplane in the transformed space that leaves the largest possible fraction of points of the same class on the same side, while maximizing the distance of either class from the hyperplane. This optimal separating hyperplane can minimize the risk of misclassifying examples of the test set. Neural networks [21] also have been successful in many applications, especially for clustering [22] and pattern recognition [23]. Recent research, however, has suggested that the SVM is superior to the neural network [24–26].

In this work we aimed to investigate possibility for development of breast cancer diagnostic method that is based on SVM classification of breast tissue autofluorescence data and to give first assessment on method specificity and sensitivity. For this purpose data from EEM and SFS, obtained on both normal and malignant breast tissue specimens, are divided into two groups. One data group was used as training set for SVM, and other provided test results. Several combinations of spectral parameters were tested with SVM to find arrangement that gives the best sensitivity and specificity.

## Theory

The principal idea of a SVM is to determine an optimal separating hyperplane that maximizes the margin between two classes in a multidimensional data space. Let first assume linearly separable set $S$ of $m$ objects where each $n$-dimensional object $\mathbf{x}$ has $n$ coordinates, $\mathbf{x}=(x_1, x_2,\ldots, x_n)$, and where each $x_i$ is a real number, $x_i \in R$ Any hyperplane in the $S$ space can be written as

$$\{\mathbf{x} \in S|\mathbf{w}\cdot\mathbf{x}+b=0\},\quad \mathbf{w}\in S,\ b\in R, \tag{1}$$

where dot product is defined by

$$\mathbf{w}\cdot\mathbf{x}=\sum_{i=1}^{n} w_i x_i. \tag{2}$$

Considering that each object $\mathbf{x}_j$, $j=(1, 2, \ldots, m)$, belongs to a class $y_j \in \{-1,+1\}$, the task is to find a hyperplane that divides space $S$ in the way that all objects belonging to different classes lie in space on different sides of the hyperplane. In other words, a pair of ($\mathbf{w}$, $b$) should be found that satisfy

$$y_i(\mathbf{w}\cdot\mathbf{x}_i+b)\geq +1, \tag{3}$$

for each $i=1$ to $n$, and the separating hyperplane equation, Fig. 1a, becomes

$$\mathbf{w}\cdot\mathbf{x}+b=0. \tag{4}$$

In the case of the linearly separable classes, which is under current consideration, there exist at least one or more hyperplanes ($\mathbf{w}$, $b$) that fulfill previous condition (Eq. 3). Each such hyperplane represents a classifier that correctly separates all objects. Among them SVM seeks to find one which has maximal distance to the closest point, that is one that maximize margin between classes for better generalization. This hyperplane is called the optimal separating hyperplane (OSH).

Taking into account that the margin between objects in different classis is $2/\|\mathbf{w}\|$ the second task is to find among all possible separating hyperplanes the one with maximal $2/\|\mathbf{w}\|$, which is equivalent to minimal $\|\mathbf{w}\|^2/2$. So, the optimization problem becomes

$$\min\frac{\|\mathbf{w}\|^2}{2}, \tag{5}$$

with the constraints

$$y_i(\mathbf{w}\cdot\mathbf{x}_i+b)-1\geq 0,\quad i=1,\ldots,m. \tag{6}$$

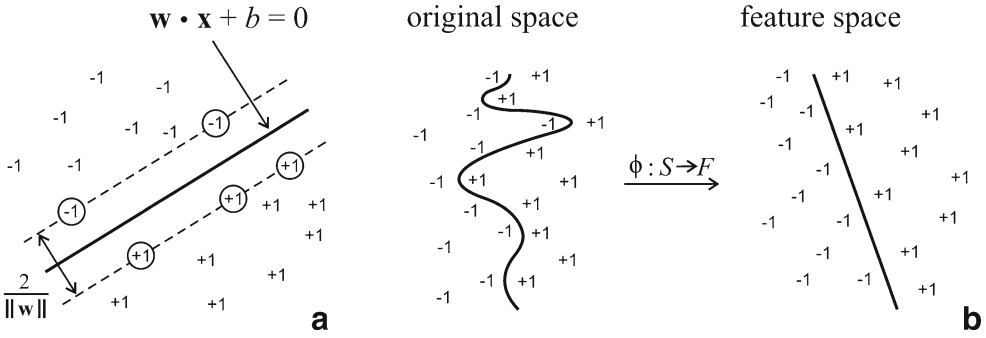


**Fig. 1** **a** optimal separating hyperplane; **b** mapping of linearly nonseparable data into the space of higher dimensionality (feature space) provides linearly separable case

this is actually the minimization of quadratic function under linear constraints (quadratic programming).

This problem can be solved using the classic method of Lagrange multipliers in dual problem formulation:

$$\max L(\mathbf{w}, b, \Lambda) = \sum_{i=1}^{m} \lambda_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j, \tag{7}$$

under constraints

$$\sum_{i=1}^{m} \lambda_i y_i = 0, \tag{8}$$

where $\Lambda=(\lambda_1, \lambda_2, \ldots, \lambda_m)$ are Lagrange multipliers, $\lambda_i \geq 0$, and represent solution to the optimization problem, Eq. (5). The objects from the SVM solution that have $\lambda_i>0$ are support vectors, while the objects that have $\lambda_i=0$ are not important and do not contribute to the SVM model.

Now, parameters ($\mathbf{w}_o$, $b_o$) of optimal separating hyperplane can be computed as:

$$\mathbf{w}_o = \sum_{i=1}^{m} \lambda_i y_i \mathbf{x}_i = \sum_{i=1}^{Ns} \lambda_i y_i \mathbf{s}_i. \tag{9}$$

and $b_o$ is an average of $b$ values obtained for all support vector objects

$$b = y_j - \sum_{i=1}^{m} \lambda_i y_i \mathbf{x}_i \cdot \mathbf{x}_j = y_j - \sum_{i=1}^{Ns} \lambda_i y_i \cdot \mathbf{s}_i \cdot \mathbf{x}_j. \tag{10}$$

Here, **s** represents support vectors and *Ns* is their number.

Any new object can be classified only on the sign of the expression $\mathbf{w}_o \cdot \mathbf{x} + b$. However, it is also possible to make classification without computing $\mathbf{w}_o$. In this case one can use supporting vectors from the training set and corresponding Lagrange multipliers to find decision function $f(\mathbf{x}_{\text{new}})$ of classifying a new object $\mathbf{x}_{\text{new}}$:

$$f(\mathbf{x}_{new}) = \operatorname{sgn}\left(\sum_{i=1}^{m} \lambda_i y_i \mathbf{x}_i \cdot \mathbf{x}_{new} + b\right) = \operatorname{sgn}\left(\sum_{i=1}^{Ns} \lambda_i y_i \mathbf{s}_i \cdot \mathbf{x}_{new} + b\right). \tag{11}$$

Most of the object sets are not linearly separable. Then, the data need to be mapped into some other dot product space of higher dimensionality (called *feature space, F*) using feature function $\Phi(\mathbf{x})$, $\Phi$:$S \rightarrow F$, so that a reliable linear separation may be achieved, Fig. 1b. Now, the training algorithm would only depend on the data through dot products in *F*, i.e. on functions of the form $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$:

$$f(\mathbf{x}_{new}) = \operatorname{sgn}\left(\sum_{i=1}^{m} \lambda_i y_i \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_{new}) + b\right). \tag{12}$$

However, since the $F$ is high-dimensional this product will be very expensive to compute. Fortunately, one can overcome computation in high-dimensional space by applying the simple trick, so called *kernel trick* [10], which enables calculation of dot product in initial data space:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \bullet \Phi(\mathbf{x}_j). \tag{13}$$

In this way a support vector machine that lives in a high-dimensional space is produced, and will work in roughly the same amount of time it would take for un-mapped data:

$$f(\mathbf{x}_{new}) = \operatorname{sgn}\left(\sum_{i=1}^{Ns} \lambda_i y_i K(s_i, \mathbf{x}_{new}) + b\right), \tag{14}$$

The choice of kernel function is restricted to some conditions, like Mercer's condition [27], which will not be discussed here. Some of the most popular kernels used by SVM practitioners are [11]:

- Linear (dot) kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$;
- Polynomial kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = (1 + \mathbf{x}_i \cdot \mathbf{x}_j)^d$;
- Neural (sigmoid) kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(a\mathbf{x}_i \cdot \mathbf{x}_j + b)$;
- Gaussian radial basis function (RBF) kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2\right)$; etc.

## Experimental (Materials and Methods)

The study involved tissue samples from female patients with breast cancer. Both tumor and normal human breast tissue

specimens were obtained from patients at the Institute of Oncology and Radiology of Serbia (Belgrade, Serbia) after Informed Consents had been signed by the patients. Samples were collected soon after surgical resection, identified, assessed by a pathologist for diagnosis and stored at -80 °C until fluorescence measurements. Specimen sizes varied from 0.2×0.5×0.5 cm to 0.3×1.0×1.5 cm.

Fluorescence emissions from breast tissue samples were measured at room temperature using Perkin Elmer Fluorescence Spectrophotometer LS45. Samples are placed in Perkin Elmer LS Series Front Surface Accessory which enables detection of fluorescence from sample surface at 90˚angle in respect to excitation beam. EEM spectra are measured in two spectral regions to avoid overlap between excitation and emission: (I) excitation from 335 to 400 nm and emission from 430 to 625 nm, and (II) excitation from 400 to 470 nm and emission from 500 to 625 nm. Synchronous fluorescence spectra are measured in a constant wavelength mode in the excitation range from 330 to 650 nm. Total synchronous fluorescence spectra are obtained by repeatedly measuring synchronous spectra while varying constant wavelength interval (wavelength difference between positions of excitation and emission monochromators) from 30 to 120 nm. Data were collected at 200 nm/min scan rate, automatically corrected with respect to excitation power by the instrument and normalized in a manner that the point of maximal fluorescence emission has a value of 100 in order to account for differences in specimen morphologies. The system provided ±1.0 nm wavelength accuracy and ±0.5 nm wavelength reproducibility.

Spectral domain volumes (volumes below luminescence intensity surface) are calculated numerically. Spectral domain of interest ($x$-$y$ surface→$\lambda_{exc}$-$\Delta\lambda$ domain) is represented with the grid of discrete points ($x_i$, $y_j$), each of which has a unique

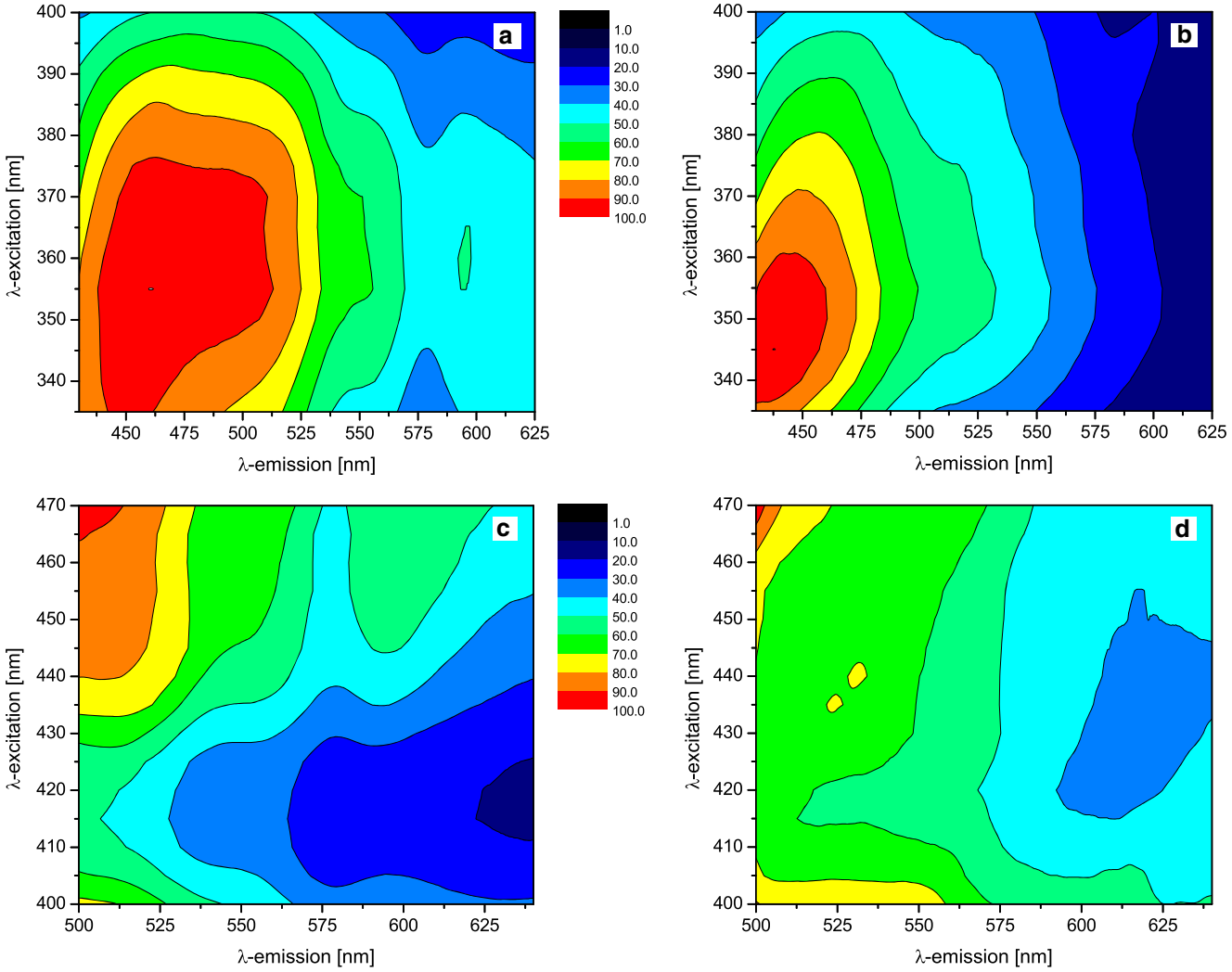


**Fig. 2** EEM contour graphs of normal and malignant tissue specimen: excitation from 335 to 400 nm and emission from 430 to 625 nm, (**a**) and (**b**) for normal and malignant, respectively; excitation from 400 to 470 nm and emission from 500 to 625 nm, (**c**) and (**d**) for normal and malignant, respectively

value taken from the spectral domain: $x_i=x_{min}+i\Delta x$; $y_i=y_{min}+j\Delta y$, $\Delta x=(x_{max}-x_{min})/N$, $\Delta y=(y_{max}-y_{min})/M$, where 'max' and 'min' denote upper and lower spectral bounds and $N$ and $M$ are discretization factors. Then, spectral domain volumes are calculated using the following equation:

$$\int_{y_{\min}}^{y_{\max}}\int_{x_{\min}}^{x_{\max}} I(x,y)dxdy \cong (\Delta x\Delta y)\sum_{j=1}^{M}\sum_{i=1}^{N} I(x_i,y_j). \qquad (15)$$

First derivative total synchronous fluorescence patterns are obtained by differentiation of measured total synchronous spectra with respect to excitation wavelength. Spectral slopes are calculated from first derivate TSFS patterns as a mean value of intensities over regions of interest.

R™ software is used for calculations and for SVM training and exploitation.

## Results and Discussion

### Fluorescence Landscapes and Fluorescence Synchronous Spectra of Normal and Malignant Tissue

Breast tissue fluorescence is measured on two sets of specimens, malignant and normal, and the EEM and SFS spectra are recorded for each specimen. Representative EEM and SFS spectra are shown in Figs. 2 and 3, respectively, in the

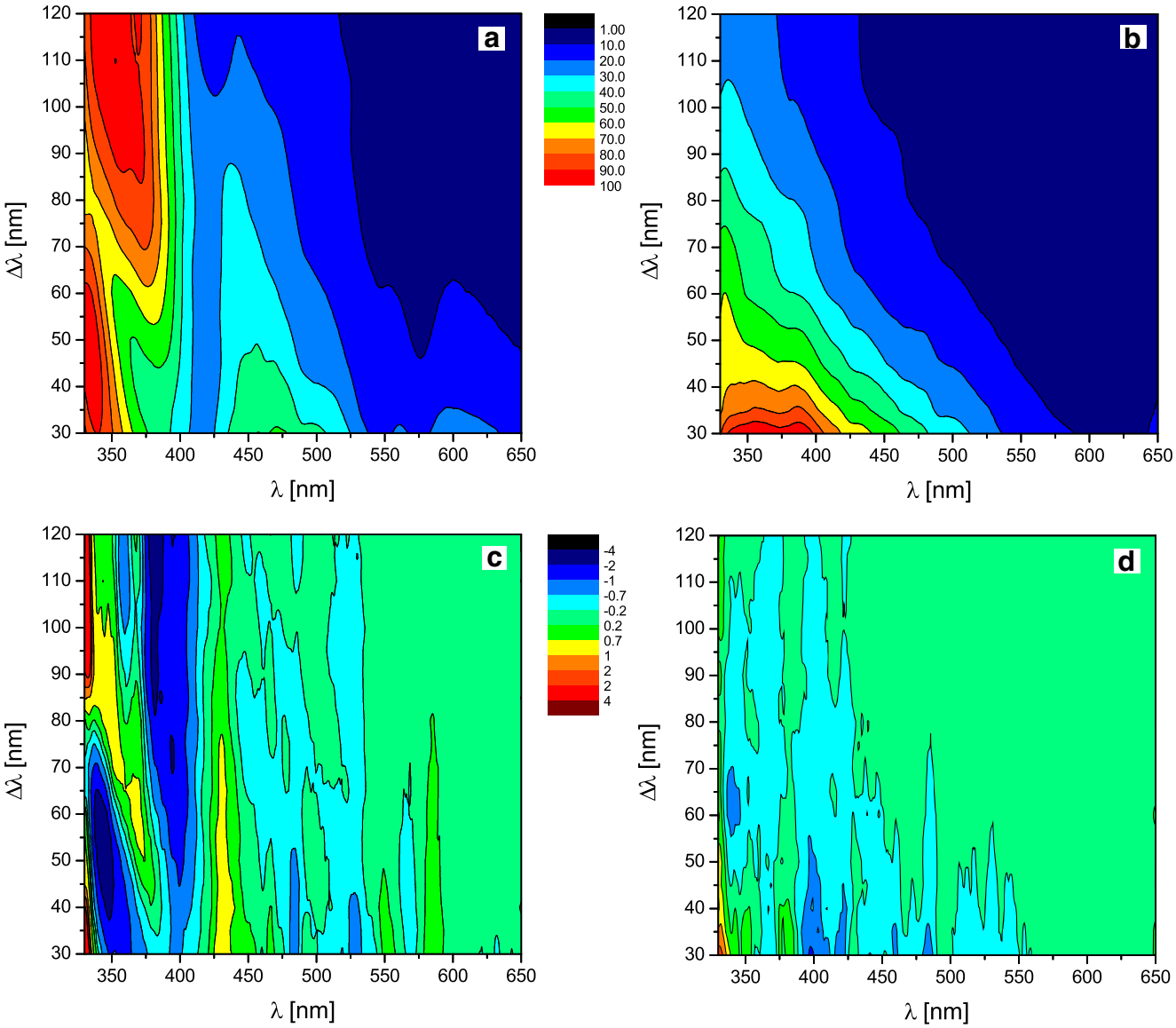


**Fig. 3** SFS and first derivate SFS contour graphs of normal and malignant tissue specimen measure in excitation range from 335 to 400 nm and synchronous interval range from 30 to 120 nm (**a**) and (**b**) SFS for normal and malignant tissue, respectively; (**c**) and (**d**) first derivate SFS for normal and malignant tissue, respectively

form of contour graphs. In these graphs different colors stand for different fluorescence intensities.

Fluorescence has good capability for use in the diagnosis of tissue conditions because fluorescence depends on tissue biochemical make-up, and because of its strong sensitivity to changes in endogenous fluorophore concentrations and/or local environment. Therefore, tissue fluorescence spectra may be regarded as tissue characteristic "fingerprint" specific for each tissue specimen. Breast tissue is mainly composed of epithelial cells, an extracellular matrix, and fat. In addition, it contains a number of optically important fluorophores and chromophores.

The endogenous fluorophores include tryptophan, type I collagen (the primary fluorophore in the extracellular matrix), NAD(P)H, flavin adenine dinucleotide (FAD) and elastin [2, 28, 29]. Tryptophan fluorescence is an indicator of protein or free tryptophan content, while NAD(P)H and flavoproteins are indicators of metabolic activity and are maximally excited at 351 and 450 nm, respectively. Finally, collagen is the primary structural protein in the extracellular matrix and the fluorescence from its cross-links is maximally excited at 325 nm [2]. Other major factors that influence breast tissue fluorescence spectra in the ultraviolet–visible (UV–vis) spectrum are endogenous absorbers and scatterers in the tissue [30, 31]. The more detailed description of tissue fluorescence is given in the literature, for example in Ref. [32].

## Classification Results

Support vector machine classification is performed with data extracted from selected spectral regions of EEM and SFS for each considered tissue specimen. Regarding EEM, vector components are taken as volume below intensity surface using Eq. (15) in five regions: 1) excitation from 335 to 400 nm and emission from 460 to 530 nm – denoted as EEM-1; 2) excitation from 350 to 385 nm and emission from 580 to 625 nm – denoted as EEM-2; 3) excitation from 440 to 470 nm and emission from 500 to 525 nm – denoted as EEM-3; 4) excitation from 450 to 470 nm and emission from 590 to 615 nm – denoted as EEM-4; and 5) excitation from 400 to 440 nm and emission from 590 to 615 nm – denoted as EEM-5. Regarding SFS, vector components are taken from spectral regions as volume below intensity surface for ordinary SFS and as an average value for the first derivate SFS. For ordinary SFS three regions are considered: 1) excitation from 330 to 400 nm and $\Delta\lambda$ from 65 to 120 nm – denoted as SFS-1; 2) excitation from 330 to 425 nm and $\Delta\lambda$ from 30 to 55 nm – denoted as SFS-2; and 3) excitation from 430 to 530 nm and $\Delta\lambda$ from 30 to 90 nm – denoted as SFS-3. For first derivate SFS: 1) excitation from 340 to 375 nm and $\Delta\lambda$ from 30 to 65 nm – denoted as FDS-1; 2) excitation from 380 to 420 nm and $\Delta\lambda$ from 65 to 120 nm – denoted as FDS-2; and 3) excitation from 425 to 440 nm and $\Delta\lambda$ from 30 to 85 nm – denoted as FDS-3. In addition, arithmetic combinations of spectral components are calculated to be also tested as an input vectors for support vector machine, Table 1.

To perform classification and to test classification accuracy samples are divided into two, equal size groups, training group and test group. Then, SVM classification and testing were done for different sets of vector inputs chosen from 42 possible spectral data (11 spectral components and 31 spectral combinations listed in Table 1.). It is important to note that combinations are made only from spectral data from either EEM or SFS measurements. In each case SVM utilized linear (dot) kernel. Classification accuracy for different sets of input vector components is listed in Table 2.

In general, we found that SVM machine can classify malignant breast tissues samples using fluorescence data. In the case of data extracted from EEM spectra the accuracy is relatively poor, and the sensitivity of 66.67 % and specificity of 62.50 % is found for all tested EEM data combinations. On the contrast, SVM classifications using SFS and FD SFS data were quite successful, yielding in a number of input data combinations 100 % sensitivity and specificity. This result supports previous findings of SFS superiority over EEM for analysis of multicomponent samples [33] due to SFS increased selectivity and decreased emission bandwidths. It is important to note that there exist two minimal data combinations that provide 100 % sensitivity and specificity. First combination comprises RT12, RT13 and RT23, which is in

**Table 1** Combinations of spectral components used as input vector components in support vector classification of normal and malignant breast tissue specimens

| Symbol | Combination | Symbol | Combination |
|---|---|---|---|
| Etot | EEM-1 + EEM-2 + EEM-3 + EEM-4 + EEM-5 | RT2 | SFS-2 / Stot |
| E12 | EEM-1 / EEM-2 | RT3 | SFS-3 / Stot |
| E13 | EEM-1 / EEM-3 | RT12 | (SFS-1+SFS-2) / Stot |
| E14 | EEM-1 / EEM-4 | RT13 | (SFS-1+SFS-3) / Stot |
| E15 | EEM-1 / EEM-5 | RT23 | (SFS-2+SFS-3) / Stot |
| E23 | EEM-2 / EEM-3 | PR112 | SFS-1 / (SFS-1+SFS-2) |
| E24 | EEM-2 / EEM-4 | PR113 | SFS-1 / (SFS-1+SFS-3) |
| E25 | EEM-2 / EEM-5 | PR223 | SFS-2 / (SFS-2+SFS-3) |
| E34 | EEM-3 / EEM-4 | Ftot | FDS-1+FDS-2+FDS-3 |
| E35 | EEM-3 / EEM-5 | D1 | (FDS-1 – Ftot) / FDS-1 |
| E45 | EEM-4 / EEM-5 | D2 | (FDS-2 – Ftot) / FDS-2 |
| Stot | SFS-1+SFS-2+SFS-3 | D3 | (FDS-3 – Ftot) / FDS-3 |
| S12 | SFS-1 / SFS-2 | D12 | (FDS-1 – FDS-2) / FDS-1 |
| S13 | SFS-1 / SFS-3 | D13 | (FDS-1 – FDS-3) / FDS-1 |
| S23 | SFS-2 / SFS-3 | D23 | (FDS-2 – FDS-3) / FDS-2 |
| RT1 | SFS-1 / Stot | | |

**Table 2** SVM tissue type prediction versus histopathology findings for various input data combinations

| Combination: | SVM test results | | | | |
|---|---|---|---|---|---|
| All combinations with EEM data | | Normal | Malignant | Sensitivity | 66.67 % |
| | Normal | 15 | 6 | Specificity | 62.50 % |
| | Malignant | 9 | 12 | Error | 35.71 % |
| SFS-1, SFS-2, SFS-3, FDS-1, FDS-2, FDS-3 | | Normal | Malignant | Sensitivity | 100 % |
| | Normal | 20 | 0 | Specificity | 95.23 % |
| | Malignant | 1 | 21 | Error | 2.38 % |
| SFS-1, SFS-2, SFS-3, RT-12, RT-13, RT-23 | | Normal | Malignant | Sensitivity | 100 % |
| | Normal | 20 | 0 | Specificity | 95.23 % |
| | Malignant | 1 | 21 | Error | 2.38 % |
| FDS-1, FDS-2, FDS-3 | | Normal | Malignant | Sensitivity | 100 % |
| | Normal | 20 | 0 | Specificity | 95.23 % |
| | Malignant | 1 | 21 | Error | 2.38 % |
| FDS-1, FDS-2, FDS-3, RT12, RT13, RT23 | | Normal | Malignant | Sensitivity | 100 % |
| | Normal | 20 | 0 | Specificity | 95.23 % |
| | Malignant | 1 | 21 | Error | 2.38 % |
| SFS-1, SFS-2, SFS-3 | | Normal | Malignant | Sensitivity | 76.19 % |
| | Normal | 18 | 5 | Specificity | 85.71 % |
| | Malignant | 3 | 16 | Error | 21.43 % |
| SFS-1, SFS-2, SFS-3, FDS-1, FDS-2, FDS-3, RT12, RT13, RT23 | | Normal | Malignant | Sensitivity | 95.24 % |
| | Normal | 20 | 1 | Specificity | 95.24 % |
| | Malignant | 1 | 20 | Error | 4.76 % |
| D1, D2, D3, D12, D13, D23 | | Normal | Malignant | Sensitivity | 100 % |
| | Normal | 21 | 0 | Specificity | 100 % |
| | Malignant | 0 | 21 | Error | 0.00 % |
| RT12, RT23, SFS-3, FDS-1 | | Normal | Malignant | Sensitivity | 100 % |
| | Normal | 21 | 0 | Specificity | 100 % |
| | Malignant | 0 | 21 | Error | 0.00 % |
| FDS-3, P112, P113, D2, D3 | | Normal | Malignant | Sensitivity | 100 % |
| | Normal | 21 | 0 | Specificity | 100 % |
| | Malignant | 0 | 21 | Error | 0.00 % |
| FDS-3, P112, P113, RT23, RT2 | | Normal | Malignant | Sensitivity | 100 % |
| | Normal | 21 | 0 | Specificity | 100 % |
| | Malignant | 0 | 21 | Error | 0.00 % |
| FDS-3, P112, P113, RT23 | | Normal | Malignant | Sensitivity | 100 % |
| | Normal | 21 | 0 | Specificity | 100 % |
| | Malignant | 0 | 21 | Error | 0.00 % |
| FDS-3, P112, P113 | | Normal | Malignant | Sensitivity | 100 % |
| | Normal | 21 | 0 | Specificity | 100 % |
| | Malignant | 0 | 21 | Error | 0.00 % |
| SFS-1, SFS-2, SFS-3, FDS-1, FDS-2, FDS-3, RT1, RT2, RT3 | | Normal | Malignant | Sensitivity | 100 % |
| | Normal | 21 | 0 | Specificity | 100 % |
| | Malignant | 0 | 21 | Error | 0.00 % |
| FDS-1, FDS-2, FDS-3, D12, D13, D23 | | Normal | Malignant | Sensitivity | 100 % |
| | Normal | 21 | 0 | Specificity | 100 % |
| | Malignant | 0 | 21 | Error | 0.00 % |
| D12, D13, D23 | | Normal | Malignant | Sensitivity | 100 % |
| | Normal | 20 | 0 | Specificity | 95.23 % |
| | Malignant | 1 | 21 | Error | 2.38 % |
| FDS-3, PR-113 | | Normal | Malignant | Sensitivity | 100 % |
| | Normal | 21 | 0 | Specificity | 100 % |
| | Malignant | 0 | 21 | Error | 0.00 % |

fact combination of spectral volumes under three SFS regions (SFS-1, SFS-2 and SFS-3). The second minimal combination is comprised of one FD SFS region (FDS-3) and one ratio from SFS region, RT113. Finding minimal input data combination for classification is important because it lessen spectral regions that need to be measured and in this way reduce necessary work and time needed for fluorescence measurements in everyday applications.

## Conclusions

Fluorescence of normal and malignant breast tissue samples exhibits differences in UV–VIS spectral regions. These differences come from different concentrations of tissue endogenous fluorophores and chromophores, and from changes in fluorophore local environment. Fluorescence emission patterns, EEM and SFS, are complex and characteristic for each individual tissue specimen. Therefore, visual inspection of emission patterns may be difficult and inappropriate for diagnostics. However, we show here that it is possible to develop and optimize Support Vector Machine to classify tissues into their respective histological category from their fluoresecence spectral features. Support Vector Machine prediction from EEM spectral features was relatively poor with about 67 % sensitivity and 62 % specificity. On the other hand, 100 % sensitivity and specificity is achieved with SFS data from several spectral regions. These finding indirectly support thesis of SFS advantage over EEM for analysis of complex, multicomponent samples.

**Acknowledgments** Authors acknowledge the support of the Ministry of Education and Science of the Republic of Serbia (project numbers 45020 and 173049).

## References

1. Komen GS (2010) Facts for life. Racial & ethic differences. Web. http://www.komen.org/bci. accessed 2011
2. Ramanujam N (2000) Fluorescence spectroscopy in vivo. In: Meyers RA (ed) Encyclopedia of analytical chemistry. John Willey & Sons, Ltd., Chichester, NewYork, pp 20–56
3. Alfano RR, Tang GC, Pradhan A, Lam W, Choy DSJ, Opher E (1987) Fluorescence spectra from cancerous and normal human breast and lung tissues. IEEE J Quantum Electron 23:1806–1811
4. Dramicanin T, Dramicanin MD, Jokanovic V, Nikolic-Vukosavljevic D, Dimitrijevic B (2005) Three-dimensional total synchronous luminescence spectroscopy criteria for discrimination between normal and malignant breast tissues. Photochem Photobiol 81:1554–1558
5. Schomacker KT, Frisoli JK, Compton CC et al (1992) Ultraviolet laser-induced fluorescence of colonic tissue: basic biology and diagnostic potential. Lasers Surg Med 12:63–78
6. Kamath SD, Mahato KK (2007) Optical pathology using oral tissue fluorescence spectra: classification by principal component analysis and k-means nearest neighbor analysis. J Biomed Opt 12:014028
7. Sterenborg HJCM, Motamedi M, Wagner RF, Duvic M, Thomsen S, Jacques SL (1994) In vivo fluorescence spectroscopy and imaging of human skin tumours. Lasers Med Sci 9:191–201
8. Koteeswaran D, Venkatesan P, Ganesan S (2003) Native fluorescence spectroscopy of blood plasma in the characterization of oral malignancy. Photochem Photobiol 78:197–204
9. Dramicanin T, Dimitrijevic B, Dramicanin MD (2011) Application of supervised self-organizing maps in breast cancer diagnosis by total synchronous fluorescence spectroscopy. Appl Spectrosc 65 (3):293–297
10. Dramicanin T, Dramicanin MD, Dimitrijevic B, Jokanovic V, Lukic S (2006) Discrimination between normal and malignant breast tissues by synchronous luminescence spectroscopy. Acta Chim Slov 53:444–449
11. Ivancuic O (2007) Applications of support vector machines in chemistry. In: Lipkowitz KB, Cundari TR (eds) Reviews in computational chemistry, Volume 23. Wiley-VCH, Weinheim, pp 291–400
12. Liu Q, Chen K, Martin M, Wintenberg A, Lenarduzzi R, Panjehpour B, Overholt F, Vo-Dinh T (2007) Development of a synchronous fluorescence imaging system and data analysis methods. Opt Exp 15:12583–12594
13. Mujamder SK, Gupta PK (1998) Synchronous luminescence spectroscopy of human breast tissues. Proc SPIE 3252:169–178
14. Mujamder SK, Gupta PK (2000) Synchronous luminescence spectroscopy for oral cancer diagnosis. Lasers Life Sci 9:143–151
15. Vengadesan N, Anbupalam T, Hemamalini S, Ebenezar J, Muthvelu K, Koteeswaran D, Aruna PR, Ganesan SC (2002) Characterization of cervical normal and abnormal tissues by synchronous luminescence spectroscopy. Proc SPIE 4613:13–17
16. Vo-Dinh T (2000) Principle of Synchronous Luminescence (SL) technique for biomedical diagnostics. Proc SPIE 3911:42–49
17. Chapelle O, Haffner P, Vapnik VN (1999) Support vector machines for histogram-based image classification. IEEE Trans Neural Network 10:1055–1064
18. Vapnik VN (1995) The nature of statistical learning theory. Springer, New York
19. Burges CJC (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Disc 2:121–167
20. Pontil M, Verri A (1998) Support vector machines for 3D object recognition. IEEE Trans Pattern Anal Mach Intell 20:637–646
21. Hagan MT, Demuth HB, Beale MH (1996) Neural network design. PWS Publishing Co., Boston
22. Park DC (2000) Centroid neural network for unsupervised competitive learning. IEEE Trans Neural Network 11:520–528
23. Gader PD, Keller JM, Krishnapuram R, Chiang JH, Mohamed MA (1997) Neural and fuzzy methods in handwriting recognition. Computer 30:79–86
24. Burbidge R, Trotter M, Buxton B, Holden S (2001) Drug design by machine learning: support vector machines for pharmaceutical data analysis. Comput Chem 26:5–14
25. Ding CH, Dubchak I (2001) Multi-class protein fold recognition using support vector machines and neural networks. Bioinformatics 17:349–358
26. Liang H, Lin Z (2001) Detection of delayed gastric emptying from electrogastrograms with support vector machine. IEEE Trans Biomed Eng 48:601–604
27. Mercer J (1909) Functions of positive and negative type and their connection with the theory of integral equations. Phil Trans Roy Soc Lond A 209:415–446
28. Palmer GM, Keely PJ, Breslin TM, Ramanujam N (2003) Autofluorescence spectroscopy of normal and malignant human breast cell lines. Photochem Photobiol 78:462–469
29. Palmer GM, Ramanujam N (2003) Diagnosis of breast cancer using optical spetroscopy. Med Laser Appl 18:233–248

30. Chance B, Salkovitz IA, Kovach AG (1972) Kinetics of mitochondrial flavoprotein and pyridine nucleotide in perfused heart. Am J Physiol 1:207–218
31. Tamura M, Hazeki O, Nioka S, Chance B (1989) In vivo study of tissue oxygen metabolism using optical and nuclear magnetic resonance spectroscopies. Annu Rev Physiol 51:813–834
32. Vo-Dinh T, Cullum BM (2003) Fluorescence spectroscopy for biomedical diagnostics. In: Vo-Dinh T (ed) Biomedical photonics handbook, Volume 28. CRC Press, Boca Raton, pp 1–50
33. Vo-Dinh T (1978) Multicomponent analysis by synchronous luminescence spectrometry. Anal Chem 50:396–401