Tech Science Press

Synovial Sarcoma Classification Technique Using Support Vector Machine and Structure Features

P. Arunachalam[1], N. Janakiraman[1,*], Arun Kumar Sivaraman[2], A. Balasundaram[3], Rajiv Vincent[2], Sita Rani[4], Barnali Dey[5], A. Muralidhar[2] and M. Rajesh[2]

[1]Department of Electronics and Communication, K.L.N. College of Engineering, Madurai, 630612, India
[2]School of Computer Science and Engineering, Vellore Institute of Technology (VIT), Chennai, 600127, India
[3]School of Computer Science and Engineering, Centre for Cyber Physical Systems, Vellore Institute of Technology (VIT), Chennai, 600127, India
[4]Department of Computer Science and Engineering, Gulzar Group of Institutions, Khanna, 141401, India
[5]Department of Information Technology, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Sikkim, 737136, India
*Corresponding Author: N. Janakiraman. Email: janakiraman.n@klnce.edu

Abstract: Digital clinical histopathology technique is used for accurately diagnosing cancer cells and achieving optimal results using Internet of Things (IoT) and blockchain technology. The cell pattern of Synovial Sarcoma (SS) cancer images always appeared as spindle shaped cell (SSC) structures. Identifying the SSC and its prognostic indicator are very crucial problems for computer aided diagnosis, especially in healthcare industry applications. A constructive framework has been proposed for the classification of SSC feature components using Support Vector Machine (SVM) with the assistance of relevant Support Vectors (SVs). This framework used the SS images, and it has been transformed into frequency sub-bands using Discrete Wavelet Transform (DWT). The sub-band wavelet coefficients of SSC and other Structure Features (SF) are extracted using Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA) techniques. Here, the maximum and minimum margin between hyperplane values of the kernel parameters are adjusted periodically as a result of storing the SF values of the SVs in the IoT devices. The performance characteristics of internal cross-validation and its statistical properties are evaluated by cross-entropy measures and compared by nonparametric Mann-Whitney U test. The significant differences in classification performance between the techniques are analyzed using the receiver operating characteristics (ROC) curve. The combination of QDA + SVM technique will be required for intelligent cancer diagnosis in the future, and it gives reduced statistic parameter feature set with greater classification accuracy. The IoT network based QDA + SVM classification technique has led to the improvement of SS cancer prognosis in medical industry applications using blockchain technology.

Keywords: Synovial sarcoma; quadratic discriminant analysis; support vector machine; receiver operating characteristics curve; discrete wavelet transform

1 Introduction

Medical industry applications are one of the fastest growing industries in the world for diagnosing diseases with the help of IoT applications through blockchain technology to protect patients from harmful chronic diseases. In which, the modern advances in digital clinical histopathology images have proposed numerous cancer cell classification techniques for brain, breast, cervix, liver, lung, prostate and colon cancer [1–3]. A Synovial Sarcoma (SS) is often occurring commonly in the extremities of adolescents and young adults [4]. It has a variety of morphological patterns, but its chief forms are *monophasic* and *biphasic* spindle cell patterns [5]. Histologically, in SS cancer image, the spindle cell patterns seem to be a small oval-shape structure, and it is called as *spindle shaped cell* (SSC) [6]. The Histopathological Image Analysis (HIA) for cancer classification is impeded by the reasons such as, (1) nonspecific clinical measurements misleading, (2) stained slide preparation, and (3) different medical imaging modalities [7]. The precise classification of normal and abnormal cancer cells is based on the *diameter*, *shape*, *size*, and *cytoplasm* of features [8]. These features provide the basic and significant information to the pathologists for appropriate treatment plans. Furthermore, the pivotal points of cancer cell classification have been focused in the literature review section.

Over the past three decades, many researchers have examined the problem of identifying which features are useful for differentiating either normal or abnormal cancer cells [9]. Most of the normal cells have a nucleus *diameter* of 5 to 11 microns, and it is suggested to be almost unchanged. Whereas abnormal cells with a diameter of 18 micron nuclei has been reported to have increased the size of the nucleus in the stained bio-marker region [10]. These above features are normally designated in the name of subjective, local, global, morphological, statistical texture feature set, and Haralick by the classification techniques [11]. Numerous mathematical approaches have been adopted, including linear discriminators, Bayesian classifiers, nearest-neighbor classifiers and more, for cancer cell classification [12].

The cancer classification was performed by *kappa statistics* in the HIA of biopsy tissue specimens [13]. Due to the limitations in human practice, a computerized system is needed to diagnose cancer, and it can help the pathologist, as well as it is possible to develop an independent system for cancer classification [14]. A cancer nucleus is classified into one of five staining intensity classes (negative, weak, moderate, strong, and very-strong) using Radial Basis Function (RBF) neural networks [15,16]. The drawbacks in this classification have been rectified using Linear Discriminant Analysis (LDA) and *k*-Nearest Neighbor (*k*-NN) classifiers [17].

This classification accuracy is improved by *fractal analysis* including *correlation*, *entropy*, and *fractal* [18]. Grading the histopathological images using *energy* and *entropy* features calculation from the multiwavelet coefficients of an image is proposed in [19]. Both the global features (color and texture characteristics of the entire image) and the morphometric features (spatial relationship between histological objects) are concentrated in [20]. The performances of several existing classifiers such as Gaussian, *k*-NN, and Support Vector Machine (SVM) have been subsequently evaluated. The characteristics of texture feature using the combination of *entropy* based fractal dimension estimation and Differential Box-Counting (DBC) method has been analyzed in [21]. The Higher Order Statistics (HOS) and spectra based approach provides more valuable information on the variability between cancer and non-cancer cell patterns with respect to different stages and grading [22].

The segmentation of objective features with help of *region of interest* (ROI) using Boosted Bayesian Multiresolution (BBMR) technique has been proposed by [23]. The statistical texture features set, and Gabor filter are used in Adaboost ensemble method to increase the decision boundary for better classification [24]. The classification of cluster-area in white, pink, and purple color features of colon images have been focused in [25]. Subsequently, the geometric features have been proposed to better quantify the elliptical nature of epithelial cells for the classification of colon images using Hybrid Feature

Set (HFS) [26]. The types of training set analysis methods and its corresponding classes of feature extractions using SVM are mentioned in [27]. Spindle Shaped cells (SSC) are specialized cells that are longer than they are wide. They are found both in normal, healthy tissue and in tumors.

In practical situations, unexpected rare tumor is not included in the training data due to different staining (bio-markers) processes. However, in discrimination model, SVMs have also forcibly categorized the above rare tumor into pre-defined classes. Handling of rare tumor is always a complex problem, and it leads to reduce the accuracy of classification between cancer and non-cancer cells [28]. Hence, SVMs seek one-class kernel parameter, and it is determined based on either PCA or LDA. Further, it has been applied to the histopathological images by few researchers so far. According to Tab. 1, most of the researchers have proposed SVM based methods, but they have not yet applied structure based feature components classification particularly in SS cancer images [29,30]. This has paved the way to focus on cancer classification using structure based feature components.

Table 1: Spreading factor determination in PCA + SVM classifier

PCA + SVM with kernel functions	% of PC1 variance with respect spreading factor (γ)				
	58%	61%	64%	68%	70%
Linear	1	1	1	1	1
RBF	0.8571	0.7241	0.6381	0.5117	0.6033
ANOVA RB	0.7563	0.6958	0.6431	0.6683	0.6124

SVM integrated SF and blockchain technologies lead to access to a wide range of histopathological imaging data for processing the cancer classification [31]. Storing and sharing of histopathological imaging data through a decentralized and secure IoT network can lead to in-depth learning in health care centers and medical industry [32]. Here, the input images are decomposed into frequency sub-band coefficients using Discrete Wavelet Transform (DWT). Then the extraction of lower dimensional features set is performed for the distribution of sub-band coefficients using PCA, LDA, and QDA. Those feature sets are considered as inputs to the SVM classifiers. Here, three types of kernel functions are used to classify the cancer and non-cancer cell structure, such as linear, RBF, and Analysis of Variance Randomized Block (ANOVA RB). The rest of this paper is organized as follows. The brief descriptions of basic theories, features components, and SVM formulation are given in Section 2. The classification accuracy of experimental results is analyzed and compared in Section 3. The importance of structure based feature components, SVM classifiers and kernels are discussed in Section 4, and the summary of the analysis is given in Section 5.

2 Materials and Methods

In this section, mathematical definitions of DWT, PCA, LDA, QDA and SVM with three kernel trick functions are intuitively explained.

2.1 Discrete Wavelet Transforms (DWT)

The DWT is a discrete set of wavelet *scales* and *translations* with some rules, and it is an implementation of discrete time series known as DWT. The discrete sample function $f(n)$ and the resulting coefficients [33,34] are known and the DWT transform pair of $f(n)$ is defined by

$$W_S(j_0, \ k) = \frac{1}{\sqrt{M}} \sum_n f(n) s_{j_0,k}(n) \tag{1}$$

$$W_\psi(j, \ k) = \frac{1}{\sqrt{M}} \sum_n f(n) \psi_{j,k}(n) \tag{2}$$

where j_0 is an arbitrary starting scaling, for $j \geq j_0$ and the DWT of function $f \in L^2(R)$ related to wavelet function $\psi(n)$ and scaling function $s(n)$ is defined as

$$Wf(x) = \frac{1}{\sqrt{M}} \sum_k W_s(j_0, \ k) s_{j_0,k}(n) + \frac{1}{\sqrt{M}} \sum_{j=j_0}^{\infty} \sum_k W_\psi(j_0, \ k) \psi_{j,k}(n) \tag{3}$$

where $s_{j_0,k}(n)$ and $\psi_{j,k}(n)$ are the functions of discrete variables and normally $j_0 = 0$, (i.e., $M = 2^j$), $n = 0, 1, 2, \ldots, M-1, j = 0, 1, 2, \ldots, J-1$ and $k = 0, 1, 2, \ldots, 2^j - 1$. These two functions are mathematically expressed by *scaling* and *translation* operation using

$$s_{j,k}(n) = 2^{j/2} S(2^j n - k) \tag{4}$$

$$\psi_{j,k}(n) = 2^{j/2} \psi(2^j n - k) \tag{5}$$

The standard Haar wavelet and the discretized scaling and wavelet functions are employed corresponding to the Haar transformation matrix of the row $M \times M$ which *is* obtained from Haar basis function, and it is defined as

$$\psi_0(n) = \psi_{00}(n) = \frac{1}{\sqrt{N}}, \qquad n \in [0, \ 1] \tag{6}$$

and

$$\psi_k(n) = \psi_{pq}(n) = \frac{1}{\sqrt{N}} \begin{cases} +2^{p/2} & (q-1)/2^p \leq n < (q-0.5)/2^p \\ -2^{p/2} & (q-0.5)/2^p \leq n < q/2^p \\ 0 & Otherwise, \ n \in [0, \ 1] \end{cases} \tag{7}$$

where, $k = 2^p + q - 1$, $0 \leq p \leq n-1$, $q = 0 \ or \ 1 \ for \ p = 0$, and $1 \leq q \leq 2^p \ for \ p \neq 0$.

2.2 Principal Component Analysis (PCA)

In PCA, mathematical representation of linear transformation of original image vector into projection feature vector [35] is given by

$$Y = W^T X \tag{8}$$

where, Y is the $m \times N$ feature vector matrix, m is the dimension of feature vector, and transformation matrix W is an $n \times m$, whose columns are the *eigenvectors* (I) corresponding to the *eigenvalues* (λ) computed using $\lambda I = SI$. Here, the total scatter matrix S and the average image of all samples are defined as

$$S = \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T \tag{9}$$

where, $\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$. After applying the linear transformation W^T, the scatter of the transformed feature vectors $\{y_1, y_2 \ldots y_n\}$ is $W^T SW$. In PCA, the projection W is chosen to maximize the determinant of total

scatter matrix of the projected samples, that is

$$W = arg^{MAX_w} |W^T SW| = [w_1, \ w_2, \ldots w_m] \tag{10}$$

where, $w_1, w_2, \ldots w_m$ is the set of n-dimensional *eigenvectors* of S, and m is corresponding largest *eigenvalues* of S. In other words, the input vector in n-dimensional space is reduced to a feature vector in m-dimensional subspace. As a result, the dimension of the reduced feature vector m is much less than the dimension of the input image vector n.

2.3 Linear Discriminant Analysis (LDA)

The main goal of the LDA is to discriminate the classes by projecting class samples from p-dimensional space into finely orientated line. For a N-class problem, $m = min \ (N - 1; \ d)$ different orientated lines will be involved [36]. For example, we have N-classes, $X_1, X_2, \ldots, X_N$. Let the ith observation vector from the X_j be x_{ji}, where $j = 1, 2, \ldots, J$ and $i = 1, 2 \ldots K_j$. J is the number of classes and K_j is the number of observations from class j. The co-variance matrix is determined in two ways (1) within-class co-variance matrix C_W and (2) between-class co-variance matrix C_B. They are given as

$$C_w = \sum_{j=1}^{N} C_j = \sum_{j=1}^{N} \frac{1}{K_j} \sum_{i=1}^{K_j} (x_{ji} - \mu_j)(x_{ji} - \mu_j)^T \tag{11}$$

$$C_B = \sum_{j=1}^{N} N_j(\mu_j - \mu)(\mu_j - \mu)^T \tag{12}$$

where, $\mu_j = \frac{1}{K_j}\sum_{i=1}^{K_j} x_{ji}$ and $\mu = \frac{1}{K}\sum_{i=0}^{K} x_i$ are the mean of class j and global mean, respectively.

The projection of observation space into feature is accomplished through a linear transformation matrix T:

$$y = T^T x \tag{13}$$

The corresponding *within-class* covariance matrix $\bar{C}_W$ and *between-class* covariance matrix $\bar{C}_B$ in the feature space are given as

$$\bar{C}_W = \sum_{j=1}^{N} \sum_{i=0}^{K_j} (y_{ji} - \bar{\mu}_j)(y_{ji} - \bar{\mu}_j)^T \tag{14}$$

$$\bar{C}_B = \sum_{j=1}^{N} K_j(\bar{\mu}_j - \bar{\mu})(\bar{\mu}_j - \bar{\mu})^T \tag{15}$$

where, $\bar{\mu}_j = \frac{1}{K_j}\sum_{i=1}^{K_j} y_{ji}$ and $\bar{\mu} = \frac{1}{K}\sum_{i=0}^{K} \bar{y}_i$. It is honest to show that:

$$\bar{C}_W = T^T S_W T \tag{16}$$

$$\bar{C}_B = T^T S_B T \tag{17}$$

The linear discriminant is then defined as a linear function for which the objectives function $J(T)$ is given as

$$J(T) = \frac{|\bar{C}_B|}{|\bar{C}_W|} = \frac{|T^T C_B T|}{|T^T C_W T|} \tag{18}$$

Here, the value of $J(T)$ is maximum, and it can be shown that the solution of Eq. (17) is that i^{th} column of an optimal transformation matrix T is the generalized eigenvector corresponding to the ith largest eigenvalue of matrix $C_W^{-1} W S_B$. To obtain the discriminant profile, the LDA classification score (L_{ik}) for a given class k is calculated by the Eq. (19) and it is considered equivalent to the class covariance matrices:

$$L_{ik} = (X_i - \bar{X}_k)^T \sum_{pooled}^{-1} (X_i - \bar{X}_k) - 2log_e \pi_k \tag{19}$$

where, X_i is a measurement vector unknown to a sample I; X_k is the mean measurement vector of class k; Σ_{pooled} is a pooled covariance matrix; and π_k is the prior probability of class k [36].

2.4 Quadratic Discriminant Analysis (QDA)

The QDA classification score (Q_{ik}) is estimated for each class k using the variance-covariance matrix and the additional natural logarithmic term, as defined as follows;

$$Q_{ik} = (X_i - \bar{X}_k)^T \sum_{k}^{-1} (X_i - \bar{X}_k) + log_k |\Sigma_k| - 2log_e \pi_k \tag{20}$$

where, $log_k |\Sigma_k|$ is the nature logarithm of the determinant of the variance-covariance matrix Σ_k; and Σ_k is the variance-covariance matrix of class k. The variance-covariance matrix (Σ_k), the pooled covariance matrix (Σ_{pooled}), and the prior probability (π_k) are calculated as follows;

$$\Sigma_k = \frac{1}{N_k} \sum_{i=1}^{N_k} (X_i - \bar{X}_k)(X_i - \bar{X}_k)^T \tag{21}$$

$$\Sigma_{pooled} = \frac{1}{N} \sum_{k=1}^{K} N_k \Sigma_k \tag{22}$$

$$\pi_k = \frac{N_k}{N} \tag{23}$$

Here, K is the total number of classes, N is the total number of objects in the training set, and N_k is the number of objects in class k [37].

2.5 The Constructing of Support Vector Machine (SVM)

Let us consider two classes case α_1 and α_2 and we have a set of training data $X = \{x_1, x_2, \ldots, x_N\} \subset P^R$. The following rules are labelled by the training data.

$$y_i = \begin{cases} +1, & x_i \in \alpha_1, \\ -1, & x_i \in \alpha_2. \end{cases} \tag{24}$$

In appearance, SVM is a binary classifier, and it can evaluate image data points and assign them to one of two classes. The input observation vectors are projected into higher dimensional feature *space F*. The function $f(x)$ takes the form

$$f(x) = (v \cdot \theta(x)) + b \tag{25}$$

With $\theta: P^R \to F$ and $v \in F$, where $(\cdot)$ is denoted by dot product. Mostly, all the preferable data in these two classes satisfy the following constraints.

$$y_j(v \cdot \theta(x_i) + b - 1 \geq 0, \qquad \forall_i. \tag{26}$$

Consider the points $\theta(x_i)$ in F for which the equality in Eq. (26) holds that these points lie on two hyperplanes, the first one is $H_1:(v \cdot \theta(x_i)) + b = +1$ and the other one is $H_2:(v \cdot \theta(x_i)) + b = -1$. These two hyperplanes are parallel and no training points fall between them. The margin between them is $2/\|v\|$. Hence, a pair of hyperplanes with a maximum margin can be found by minimizing $\|v\|^2$ subject to Eq. (26). This problem can be written as a convex optimization technique

$$\textit{Minimize } \frac{1}{2} \|v\|^2 \tag{27}$$

$$\textit{Subject to } y_i(v \cdot \theta(x_i) + b - 1 \geq 0 \qquad \forall_i , \tag{28}$$

where Eq. (27) is a primal objective function and Eq. (28) is a corresponding constraint. Both can be solved by constructing a *Lagrange* function. Hence, the positive *Lagrange* multipliers are taken into consideration ω_i; $i = 1, 2, \ldots, N$, one for each constraint in both equations. The *Lagrange* function [38] is defined by

$$L_p = \frac{1}{2} \|v\|^2 - \sum_{i=0}^{N} \omega_i y_i(v \cdot \theta(x_i) + b) + \sum_{i=1}^{N} \omega_i, \tag{29}$$

The gradient of L_p must be minimized with respect to both v and b values, and they must vanish. The gradients are given by

$$\frac{\partial L_P}{\partial v_s} = v_s - \sum_{i=1}^{N} \omega_i y_i \theta(x_{is}) = 0, \qquad s = 1, 2, \ldots, p \tag{30}$$

$$\frac{\partial L_P}{\partial b} = -\sum_{i=1}^{N} \omega_i y_i = 0 \tag{31}$$

where p is the dimension of space F. By combining these conditions and other constraints on primal functions and *Lagrange multipliers*, we obtain the *Karush-Kuhn-Tucker* (KKT) condition is obtained and it's given by

$$\frac{\partial L_P}{\partial v_s} = v_s - \sum_{i=1}^{N} \omega_i y_i \theta(x_{is}) = 0, \qquad s = 1, 2, \ldots, p \tag{32}$$

$$\frac{\partial L_P}{\partial b} = -\sum_{i=0}^{n} \omega_i y_i = 0 \tag{33}$$

$$y_i(v \cdot \theta(x_i) + b - 1 \geq 0 \textit{ and } \omega_i \geq 0 \qquad \forall_i \tag{34}$$

$$\omega_i(y_i(v \cdot \theta(x_i) + b - 1) = 0 \qquad \forall_i \tag{35}$$

Here, v, b and ω are the variables to be solved. From the KKT Eq. (31), the following is obtained

$$v = \sum_{i=1}^{N} \omega_i y_i \theta(x_i) \quad \sum_{i=1}^{N} \omega_i y_i = 0 \tag{36}$$

Therefore,

$$f(x) = \sum_{i=0}^{N} \omega_i y_i (\theta(x_i) \cdot \theta(x)) + b = \sum_{i=0}^{N} \omega_i y_i l(x_i, \ x) + b \tag{37}$$

where $l(x_i, x) = (\theta(x_i) \cdot \theta(x_j))$ is a kernel function that uses the dot product in the feature space.

In this study, three different kernel functions have been used, and they are calculated as follows:

   i) Linear kernel :

$$l(x_i, \ x) = (x_i^T \cdot x_{j)} \tag{38}$$

   ii) Gaussian Radial Basis Function (RBF) :

$$l(x_i, \ x) = \exp(-\gamma \| x_i - x_j \|^2) \tag{39}$$

   iii) Analysis of variance randomized block (ANOVA RB) kernel : z

$$l(x_i, \ x) = \sum_{k=1}^{n} \exp\left(-\frac{1}{\sqrt{2\gamma}}(x_i^k - x_j^k)^2\right)^d \tag{40}$$

where $l(x_i, x)$ is positive definite for values in RBF kernel case, $\gamma = \frac{1}{2\sigma^2}$ is a spreading factor that controls the width of the kernel function, d is the dimensionality or degree of input space, and σ is a standard division. Substitute Eq. (36) into Eq. (29). This leads to maximization of dual function L_D as given by

$$L_D = -\frac{1}{2} \sum_{i=0}^{N} \sum_{j=1}^{N} \omega_i \omega_j y_i y_j l_{ij} + \sum_{i=0}^{N} \omega_i. \tag{41}$$

Here, the dual function incorporates the constraints, and the dual optimization technique is obtained as defined by

$$Maximize \ -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_i \omega_j y_i y_j l_{ij} + \sum_{i=1}^{N} \omega_i \tag{42}$$

$$Subject \ to \sum_{i=1}^{N} \omega_i y_i = 0, \tag{43}$$

$$\omega_i \geq 0 \quad \forall_i.$$

Both primal L_p and dual L_D constraints are represented with same objective function, but with different constructions. On this, optimization of the primal and dual constraints are a type of convex optimization problem and it can be rectified by using the gird-point search algorithm. Finally, the SVM classifier classifies x by the following classification criteria and it is defined by

$$x \in Class \ i \ if f^{(i)}(x) = \max_{for \ all \ j \in M} f^{(j)}(x). \tag{44}$$

where x an input data vector which is entering the classifier and M-dimensional value vector $f^{(i)}(x)$, $i = 1, 2,$ …, M (one dimension for each class) is generated.

3 Experiment Results

Histopathological Synovial Sarcoma (SS) cancer images have been downloaded from the online link http://www.pathologyoutlines.com/topic/softtissuesynovialsarc.html for training the classifier model. The stained slide SS images are collected from the Kilpauk Medical College and Hospital, Government of Tamilnadu, Chennai, India for external validation purpose. In MATLAB, the multispectral color images are stored individually in red, green, and blue components and the size of each image is 128×128 pixels, respectively. The total number of images is 5400, the number of samples per second in each image is 49152, and eight iterations have been considered in each image for computation. Most of the medical image classification methods used the ratio of 70% for training the classifier, 15% for *internal validation*, and 15% for *external validation* from the overall dataset. The overall procedure and the implementation flow of SSC classification are represented as a block diagram, Fig. 1. A multiclass kernel machine has been implemented to operate on all groups of features simultaneously and combine them without increasing the number of required classifiers.
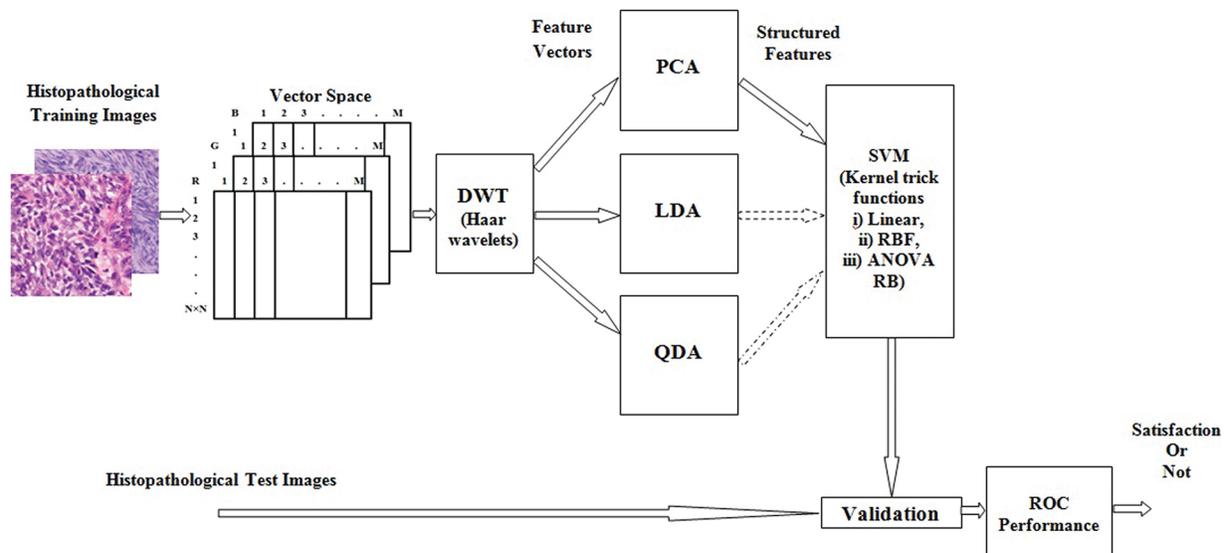
Figure 1: Block diagram of overall procedure and its internal operation of SSC structure classification

More interestingly, an SVM has two adjustable parameters that should be optimized for margin maximization and error minimization. One is the determination of spreading factor (γ) in each kernel function and *grid-search* based optimization technique. The grid search is performed by 256 samples with uniform resolution in log_2-space, for example, γ is $log_2\gamma \in \{-126, -124, -122, \ldots, 128\}$. Second is the linear representation and its corresponding weighting coefficients. It is determined by the *Lagrangian multiplier* (ω_i). The higher weighting coefficient with relevant input vectors is called as SVs. This is done by *linear algebra*, in which the features of basis vectors are considered as a matrix form with a linear transformation as positive and negative labeled data points.

The statistical testing for the internal cross validation work, the non-parametric test such as Mann Whitney U test has been applied with dependent cross-entropy measures to analyze the matching characteristics of decision-making with respect to null-hypothesis (H_o), and alternate-hypothesis (H_1). The

cross-entropy measures are used to find the cross-information between the observed and the predicted sample variables. The null-hypothesis represents the sample means of minimum cross-entropy, and it provides valuable information about the well separation of two classes based on the hyperplane. The above-mentioned systematic procedure is applied to PCA + SVM, LDA + SVM, and QDA + SVM for the purpose of spreading factor (γ), cross-entropy and *p-value* determination.

3.1 PCA + SVM

In PCA, the transformed space, in descending order of accounting variance is orthogonal and the sub-band coefficients are jointly-normally distributed. With the help of *principal component regression* (PCR), it is capable to find out the best fit of regressor coefficients based on the *direction of variations* among the other pairs of eigenvalues. The eigenvalues and the eigenvectors are determined from the covariance or correlation matrix. When the PC variance is high, it is easier to extract the coefficients which are linearly correlated with each other. It is very helpful to find different multiple groups of pixels based on its intensity relevant to cancer (SSC) or non-cancer (non -SSC) pixels.

The characteristics of data classification based on its highest 68% PC variance and the corresponding kernel functions are shown in Fig. 2a which has most of the data points are inclined to the first component regressor line. Based on the consistent cell structures in PCR with iterative procedure, it is capable to find the regressor lines and these lines are helpful to extract the hidden feature components for cancer classification. The SVM classification with three different kernel functions is represented in the form of inverse mapping Γ^{-1} for convenience Figs. 2b, 3c, 3d.
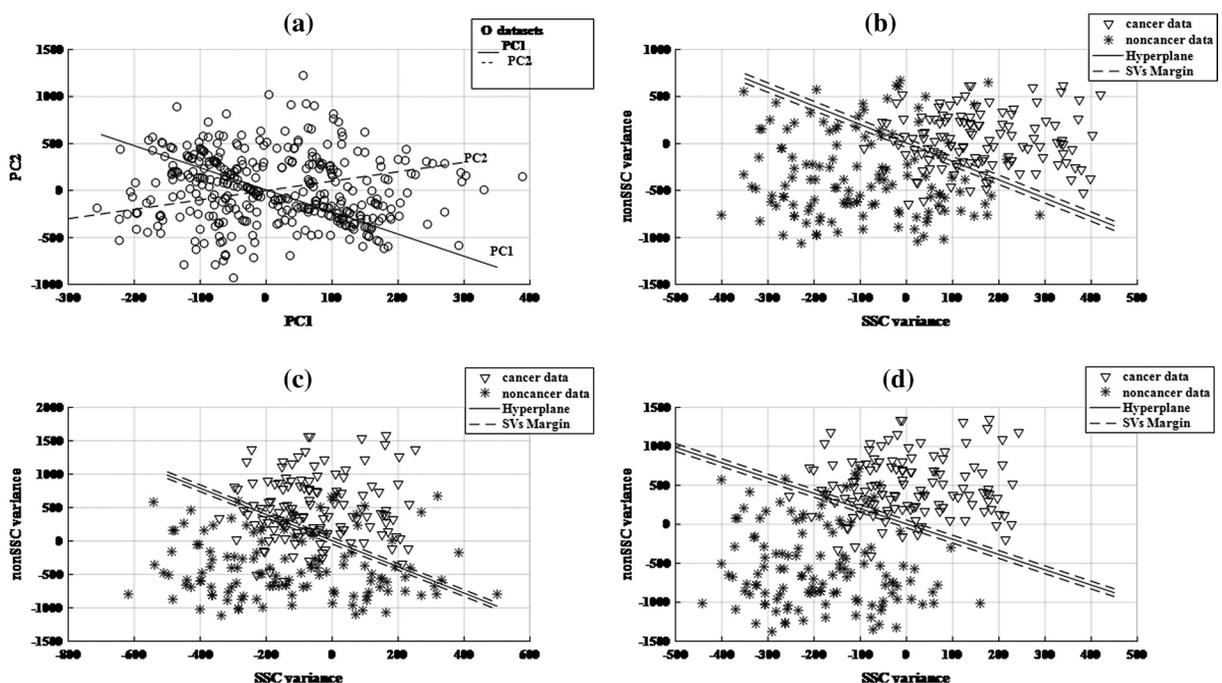
Figure 2: PCA + SVM classification model. (a) PCI 68% variance, (b) Linear, (c) RBF, (d) ANOVA RB

The performance of discrimination between the structured groups of sub-band features depends on the effective prediction of *regressor* coefficients using the least mean square approach. The *null hypothesis* of F-statistics is represented in the corresponding feature variables. The spreading factor determination is defined by the proportion of variability in Tab. 1 and that has been explained by the corresponding

highest PC. The determination of very narrow spreading factor γ among the data samples is the most important for designing the classifier models perfectly.

The *p-value* for accepting the null-hypothesis based on the cross-entropy measures is tabulated in Tab. 2. If the *p-value* is greater than the significant level ($\alpha = 0.05$), then the null-hypothesis is accepted. The *p-value* of 0.0676 is very different to accept the null-hypothesis, even RBF kernel function is used at the same time consistently. It is possible to take the decision to reject the null-hypothesis, when the linear kernel is used. The cross-entropy measure is always less during the transformation using ANOVA RB kernel function, and it provides better performance than other kernels. The ANOVA RB gives better results compared to other two kernel functions and it is a very time-consuming process, when more PCs take influence in PCA.

Table 2: *p*-value for cross-entropy measures using Mann Whitney U test for PCA + SVM classifier

PCA+SVM classification parameters	Kernel trick functions		
	Linear	RBF	ANOVA RB
Correctly classified	118	182	209
Incorrectly classified	76	45	37
Cross entropy measure	0.8652	0.7957	0.6539
p-value	0.0082	0.0676	0.2867

3.2 LDA + SVM

LDA finds the directions with maximize the variance in between-classes, and minimize the variance in within-classes. The determination of separating the hyperplane between multiple classes is as follows: (1) to find first and second order moments of all classes, (2) to find the individual co-variance matrix for each class, (3) to construct linear equation using mean, standard deviation and co-variance matrix using Eq. (19), and (4) to find the linear line for the separation between two classes using equivalent discriminant function. The major assumption is that identically independently distributed (i.i.d) random variables are represented with either unimodal or multimodal characteristics and its prior probabilities are known in advance. It becomes easier to calculate the between-class C_b and within-class C_w scatter matrices even the cumulative normalized-distance between two classes is very less. It reduces the computation complexity based on relevant discriminant feature extraction.

The dataset and the corresponding discriminator linear line are determined by the identity covariance matrix with a Gaussian distribution of the LDA, as shown in Fig. 3a. The mass of data points is normally distributed towards the middle portion of the scatter plot. The mass of data is separated into two classes using SVM hyperplane, and each of the classes occupies half of the space. The best boundary discrimination is done by example of linear, RBF, and ANOVA RB kernels as shown in Figs. 3b–3d respectively. The boundary is determined by the Bayes rule using the *true (normal)* distribution.

The determination of spreading factor in LDA + SVM with different kernel functions and cross-entropy measures are given in the Tabs. 3 and 4. The average spreading factor is 0.5684 for RBF kernel and 0.52014 for ANOVA RB kernel, when the variance of inverse covariance matrix varies from 58% to 70%. In linear kernel, an alternate-hypothesis is accepted, due to the insignificance *p-value* during Mann Whitney U test. The spreading factor γ is always unity in linear kernel, and much significance is not possible to attain. The good indication is that the cross-entropy is less than 0.5 in ANOVA RB kernel function. Hence, the performance of LDA depends only on the exact determination of mean and variance

of the individual group pixels. The resulting boundaries are well separated between the SSC (∇ plus) over the non-SSC (*) and the best boundary line is improved in ANOVA RB. The discrimination of class L_{ik} gives successive separation between two class feature vectors. Here, it is imperative to increase the performance of linear discrimination.
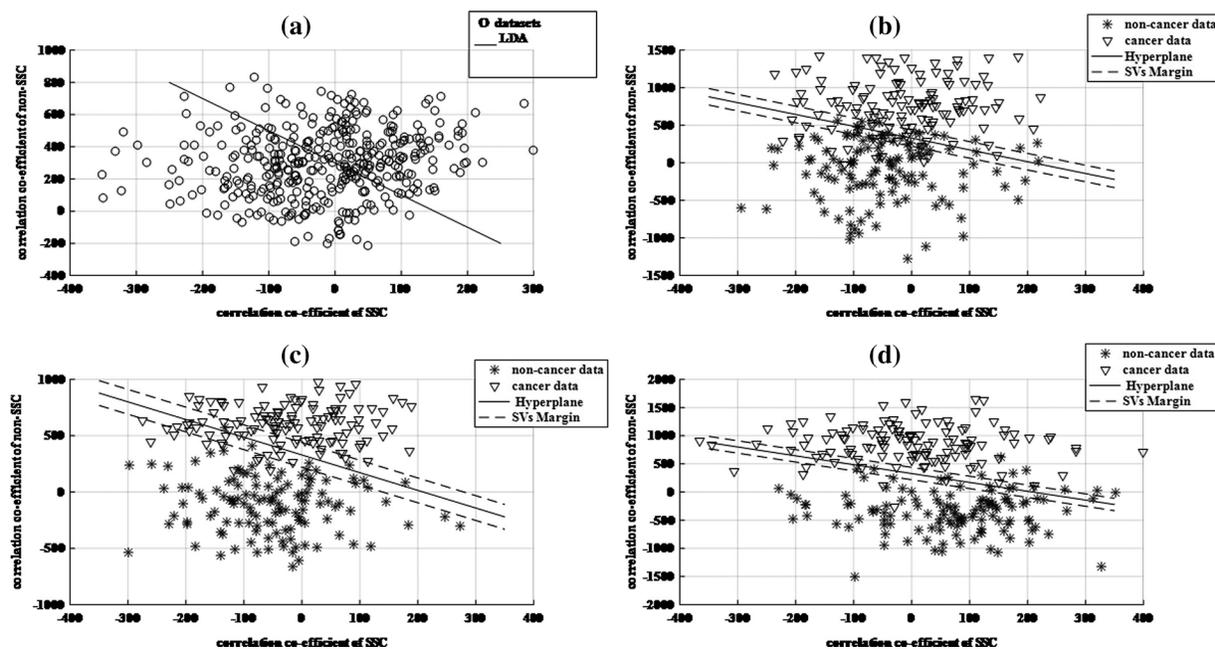
Figure 3: LDA + SVM classification model. (a) LDA feature components, (b) Linear, (c) RBF, (d) ANOVA RB

Table 3: Spreading factor determination in LDA + SVM classifier

LDA+SVM with kernel functions	% of inverse co-variance with respect to spreading factor (γ) for PC1				
	58%	61%	64%	68%	70%
Linear	1	1	1	1	1
RBF	0.6527	0.6289	0.5380	0.5192	0.5033
ANOVA RB	0.6831	0.6015	0.5140	0.4521	0.3511

Table 4: *p*-value for cross-entropy measures using Mann Whitney U test for LDA + SVM classifier

LDA+SVM classification parameters	Kernel trick functions		
	Linear	RBF	ANOVA RB
Correctly classified	114	186	124
Incorrectly classified	78	56	110
Cross entropy measure	0.9252	0.8459	0.4971
p-value	0.0417	0.0947	0.6282

3.3 QDA+SVM

More Gaussian mixtures of feature components are added to provide the information on each class density by single multivariate Gaussian density function. The covariance matrix of each class is equal to the sum of the data points of all classes. Furthermore, the probability of each class π_k is determined at the given data point of k classes and the total number of comparisons is $k-1$. The variance-covariance matrix of the cell structure must have a multivariate normal distribution. Fig. 4a clearly shows two different quadratic discriminant lines which are needed to differentiate the cell structures. The exact classification of kernel functions is shown in Figs. 4b–4d. The individual class probability (π_k) takes major contribution to find the best discriminator boundary line. It requires more observations per class $(n_i > p) \ \forall \ i$. The variance-covariance matrix is estimated separately for each class, whereas LDA uses a pooled estimation of a common variance-covariance matrix.
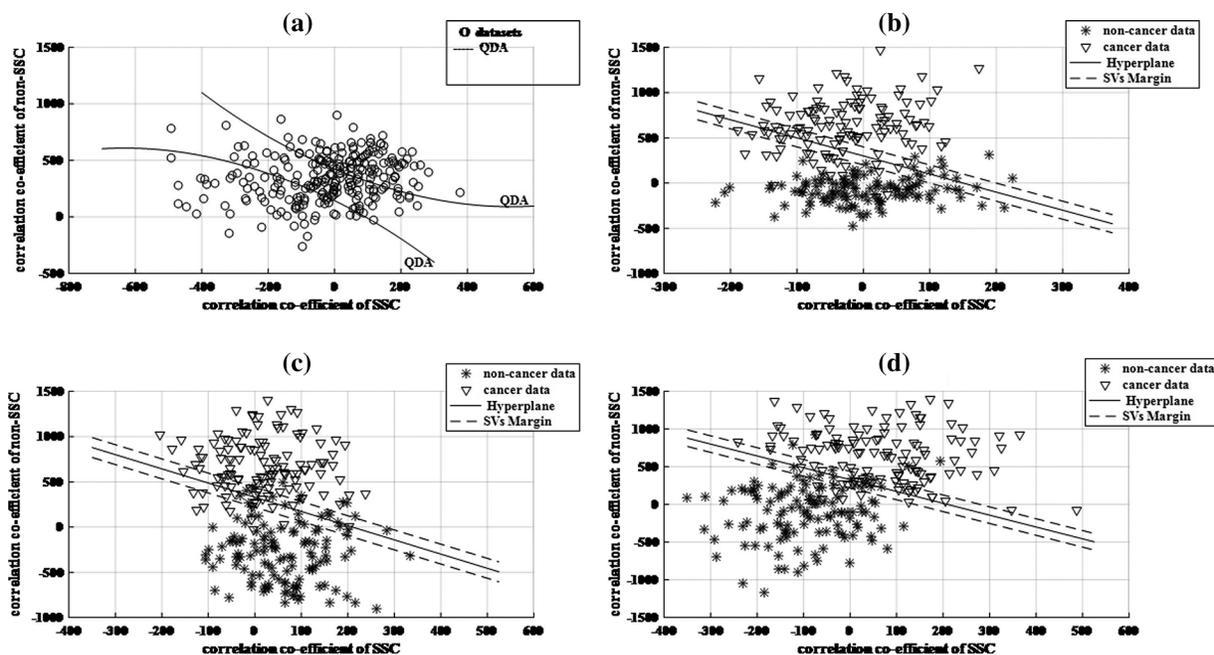
Figure 4: QDA + SVM classification model. (a) QDA feature components, (b) Linear, (c) RBF, (d) ANOVA RB

The determination of spreading factor in terms of inverse pooled co-variance matrix is represented in Tab. 5. Due to the less percentage of pooled covariance matrices and the narrowed γ value, all the kernel trick functions are performed well and it has been given in Tab. 6. The mean value of spreading factor is minimum, even the variance of pooled covariance matrix has been increased from 58% to 70%. The transformed data sample using ANOVA RB function does not affect the performance of cross-entropy measure and it always leads to minimum values. Hence, the technique is very good option compared to other two discriminator models for classification.

Table 5: Spreading factor determination in QDA + SVM classifier

QDA+SVM with kernel functions	% of inverse pooled co-variance with respect to spreading factor (γ)				
	58%	61%	64%	68%	70%
Linear	1	1	1	1	1
RBF	0.5742	0.4805	0.4831	0.3175	0.2133
ANOVA RB	0.5013	0.4413	0.3251	0.2531	0.1563

Table 6: *p*-value for cross-entropy measures using Mann Whitney U test for QDA + SVM classifier

QDA + SVM classification parameters	Kernel trick functions		
	Linear	RBF	ANOVA RB
Correctly classified	189	212	118
Incorrectly classified	46	32	126
Cross entropy measure	0.8265	0.7759	0.6935
p-value	0.0829	0.1769	0.5867

3.4 Classifiers Evaluation Using ROC

The purpose of external validation is mainly to evaluate the classification performance between the concern feature extraction methods and the kernel functions by ROC curve [38,39]. It is a curve which depicts the interaction between the *x* coordinate as a *1- specificity* and the *y* coordinate as a *sensitivity*. These two parameters are mathematically expressed as *sensitivity = TP/(FN + TP)* and *specificity = TN/(FP + TN)* using the confusion matrix. The threshold for the best cut-off range is in-between *0 to 1* with 0.05 uniform interval and 20 samples per parameter. The most common quantitative index for describing the accuracy is expressed by the *area under the ROC curve* (AUC). It provides a useful parameter for estimating and comparing the classifier performance. The AUC can be determined by Eq. (45). The AUC performance results with different kernel functions are summarized in Tab. 7.

$$AUC = \sum_{i}^{N-1}((1 - specificity)_{i+1} - (1 - specificity_i) * (sensitivity_i + sensitivity_{i+1}))/2 \qquad (45)$$

Table 7: Classifiers estimation of the AUC, and Approximate Standard Error

SF Components + SVM + Kernel trick functions	AUC	Approximate Standard Error	Grade
PCA + SVM + Linear	0.5753	0.1105	not good
PCA + SVM + RBF	0.6569	0.1062	not good
LDA + SVM + Linear	0.6898	0.1034	not good
PCA + SVM + ANOVA RB	0.7014	0.1023	worthless
LDA + SVM + RBF	0.7487	0.0970	worthless
QDA + SVM + Linear	0.7525	0.0965	worthless
LDA + SVM + ANOVA RB	0.8091	0.0879	good
QDA + SVM + RBF	0.8196	0.0860	good
QDA + SVM + ANOVA RB	0.9042	0.0658	excellent

The grade system of AUC is categorized into four groups, which are excellent, good, worthless and not good, and their common range values are $0.9 < AUC < 1.0$, $0.8 < AUC < 0.9$, $0.7 < AUC < 0.8$ and $0.6 < AUC < 0.7$, respectively. Hence, Fig. 5a shows the result of AUC values for PCA + SVM with Linear, RBF, and ANOVA RB kernel functions, and they have an AUC of 0.5753, 0.6569, and 0.7014, respectively. Similarly, LDA + SVM and QDA + SVM with their three different functions, and the ROC curves are shown in Figs. 5b and 5c. The results of AUC values are 0.6898, 0.7487, 0.8091 and 0.7525,

0.8196, 0.9042, respectively. Depending on the AUC value, the QDA + SVM with the ANOVA RB has achieved as the excellent classification model, whereas the LDA + SVM with ANOVA RB and the QDA + SVM with the RBF are good models. Based on these results, it is demonstrated that QDA + SVM with ANOVA RB has been chosen to be applied in the proposed SSC structure classification, because this classifier QDA + SVM with ANOVA RB model has the greatest AUC than the others.
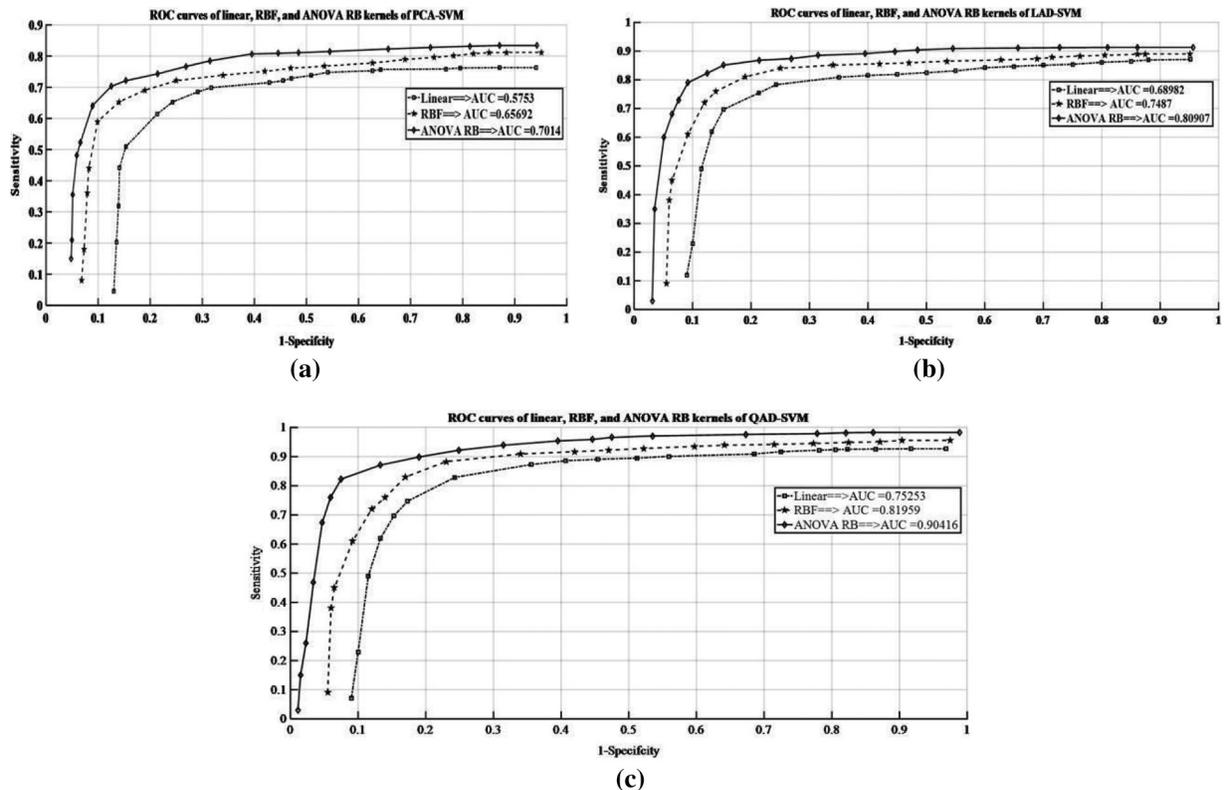
Figure 5: ROC curves of PCA + SVM, LDA + SVM and QDA + SVM classifier models (a) PCA + SVM classifier model (b) LDA + SVM classifier model (c) QDA + SVM classifier model

This ANOVA RB kernel is consistent (Tab. 7), and it possesses the lowest classification error to build a hyperplane model. There is an attractive difference between the two analyses, that is, the classification error and the ROC curve against the occurrence of other two kernels. As a result of approximate standard error, the RBF kernel is better than the linear kernel. With respect to ROC curve analysis, it is clear that ANOVA RB is better than RBF kernel. However, it is stated that the ROC curve analysis is particularly useful for the performance evaluation in the current framework, because it has widely reported that it reflects the true state of a classification problem rather than a measurement of classification error.

4 Discussion

A new approach has been proposed and the SSC structure feature components are extracted and statistically classified using the combination of SVM with different kernel trick functions. When differentiating cancer and non-cancer cells in SS images, the third-level decomposed cell SF is discriminated using PCA, LDA, and QDA techniques, and the nonlinear features are transformed into linear using kernel functions. The *direction of variance* and its different characteristics of the structured

cell pixels are key parameters for the classification. It depends on the input data variables subject to *centering*. The PCA technique finds the *direction of variance* among the feature coefficients, but in QDA and LDA, the variances of *within-classes* and *between-classes* are considered respectively. Here, LDA uses a pooled covariance matrix and QDA uses each group's variance-covariance matrix. Both techniques are tried to maximize the between-class variability, and minimize the within-class variability based on a *Mahalanobis* distance calculation.

The kernel trick functions are used to discover the 'right set of features' when the mapping function l is non-trivial. The performance analysis is compared based on the transformation of feature coefficients using, (1) Inner product between $f(x)$ and $f(x^2)$, (2) Exponential operation with single spreading parameter, and (3) Sum of exponential operation with multiple spreading parameters into the higher dimensionality of the input space. Then, SVM automatically discovers the optimal separating hyperplane in the complex decision surface during mapping back into input space via l^{-1}. The overall comparison performance between SF techniques and kernel trick functions are given in Tab. 8.

Table 8: Overall comparison parameters with PCA + SVM, LDA + SVM, and QDA + SVM

Classification	PCA + SVM			LDA + SVM			QDA + SVM		
Algorithm Parameters	Linear [29]	RBF [28]	ANOVA RB	Linear [30]	RBF [28], [30]	ANOVA RB	Linear	RBF	ANOVA RB (Our study)
Correctly classified	205	209	217	211	223	239	236	240	242
Incorrectly classified	51	47	39	45	33	17	20	16	14
SE $\pm$ SERR	0.8692 ± 0.0296	0.8143 ± 0.0329	0.8613 ± 0.0295	0.7442 ± 0.0384	0.8462 ± 0.0302	0.9242 ± 0.0230	0.9237 ± 0.0232	0.9462 ± 0.0198	0.9370 ± 0.0216
SP $\pm$ SERR	0.6774 ± 0.0395	0.8190 ± 0.0358	0.8190 ± 0.0343	0.7442 ± 0.0358	0.9027 ± 0.0279	0.9435 ± 0.0207	0.9200 ± 0.0243	0.9286 ± 0.0229	0.9535 ± 0.0185
Gini coefficient	0.3191	0.2998	0.2583	0.2898	0.2246	0.1240	0.1440	0.1172	0.1034
Total number of data	256	256	256	256	256	256	256	256	256
Accuracy in %	80.08	81.64	84.77	82.42	87.11	93.36	92.19	93.75	94.53
PRESS	206.53	185.64	210.13	181.56	110.56	140.56	114.21	132.16	96.58
SS$_{RES}$	159.31	168.51	155.83	148.11	148.10	147.16	156.53	151.31	144.83

The Gini coefficient is an elementary parameter that measures the superior performance between the classifiers. It is based on the area dominated by the ROC surfaces. The determination of Gini coefficient from the ROC curve is given by Eq. (46).

$$G = 1 - \sum_{k=1}^{n} ((1 - specificity)_k - (1 - specificity)_{k-1}) * (sensitivity_k + sensitivity_{k-1}) \qquad (46)$$

The proposed classification models showed the robustness of classifying SS cancer effectively and accurately in QDA + SVM with ANOVA RB kernel function. The prediction residual error sum of squares (PRESS) statistics shows that QDA + SVM is lower than the sum of square residuals (SS$_{RES}$) and its average variations of PCA + SVM, LDA + SVM, and QDA + SVM. The main purpose of PRESS statistics is to measure how well the proposed classifier models will perform in predicting the new data. The standard error (SERR) is also specified in each parameter to understand the maximum and minimum

variations in sensitivity (SE) and specificity (SP). The QDA + SVM have obtained a higher classification rate than the other classifier models.

ANOVA RB kernel trick function requires less computational time over other kernels functions. This can be used to increase the accuracy of classification, reduce the timing of diagnostic procedures, to speed up treatment plans, and improve prognosis and patient survival rates. This is an opportunity to improve the results with deep learning techniques.

5 Conclusion

The SSC structure classification has been proposed as a new approach and it is totally based on the combination of PCA + SVM, LDA + SVM, and QDA + SVM, for the prediction of relevant structured SVs with the help of kernel trick functions. Based on the experiments carried out in this study, following three conclusions have been drawn such as, (1) QDA based technique does not affect the 'dependence-observer' and 'intra-inter observer' variability, since there is no general assumption about any practical feature vector datasets, (2) QDA + SVM always trying to maximize the between-class variability while minimizing within class variability based on a distance calculation, and (3) The significant differences in SF component between *shape* and *similarity* are studied especially in ANOVA RB kernel trick function. The evidence for structured support vectors and its randomized block variance of a related trick function as few parameters are enough for easier recognition of cancer and non-cancer cells in SS cancer images. In the future, deep learning investigation about SVM can be examined. Therefore, recent support vectors and different structural features will enable future works to achieve integrated blockchain technology based SS image data classification barriers and medical device industrial applications.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] W. Wang, J. A. Ozolek, D. Slepcev, A. B. Lee, C. Chen *et al.,* "An optimal transportation approach for nuclear structure-based pathology," *IEEE Transactions on Medical Imaging*, vol. 30, no. 3, pp. 621–631, 2015.

[2] H. Irshad, A. Veillard, L. Roux and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: A review-current status and future potential," *IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 97–114, 2012.

[3] J. Barker, A. Hoogi, A. Depeursinge and D. L. Rubin, "Automated classification of brain tumour type in whole-slide digital pathology images using local representative tiles," *Medical Image Analysis*, vol. 30, pp. 60–71, 2016.

[4] A. Kerouanton, I. Jimenez, C. Cellier, V. Laurence, S. Helfre *et al.,* "Synovial sarcoma in children and adolescents," *Journal of Pediatric Hematologyl/Oncology*, vol. 36, no. 4, pp. 257–262, 2014.

[5] K. Thway and C. Fisher, "Synovial sarcoma: Defining features and diagnostic evolution," *Annals of Diagnostic Pathology*, vol. 18, no. 6, pp. 369–380, 2014.

[6] C. Che Chang and P. Ying Chang, "Primary pulmonary synovial sarcoma," *Journal of Cancer Research and Practice*, vol. 5, no. 1, pp. 24–26, 2018.

[7] K. Sirinukunwattana, S. E. Ahmed Raza, Y. W. Tsang, D. R. J. Snead, I. A. Cree *et al.,* "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 1196–1206, 2016.

[8]   F. Li, X. Zhou, J. Ma and S. T. C. Wong, "Multiple nuclei tracking using integer programming for quantitative cancer cell cycle analysis," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 96–105, 2010.

[9]   B. J. Cho, Y. J. Choi, M. J. Lee, J. H. Kim, G. H. Son *et al.,* "Classification of cervical neoplasms on colposcopic photography using deep learning," *Scientific Reports*, vol. 10, no. 3, pp. 347–361, 2020.

[10] M. T. Rezende, A. G. C. Bianchi and C. M. Carneiro, "Cervical cancer: Automation of Pap test screening," *Diagnostic Cytopathology*, vol. 49, no. 4, pp. 559–574, 2021.

[11] R. S. Majdar and H. Ghassemian, "A probabilistic SVM approach for hyperspectral image classification using spectral and texture features," *International Journal of Remote Sensing*, vol. 38, no. 15, pp. 4265–4284, 2017.

[12] M. Salvia, U. Rajendra Acharya, F. Molinaria and K. M. Meiburger, "The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis," *Computers in Biology and Medicine*, vol. 128, no. 5, pp. 162–179, 2021.

[13] A. R. Feinstein and D. V. Cicchetti, "High agreement but low kappa: I. the problems of two paradoxes," *Journal of Clinical Epidemiology*, vol. 43, no. 6, pp. 543–549, 1990.

[14] D. V. Cicchetti and A. R. Feinstein, "High agreement but low kappa: II. resolving paradoxes," *Journal of Clinical Epidemiology*, vol. 43, no. 6, pp. 551–558, 1990.

[15] F. Schnorrenberg, C. S. Pattichis, K. C. Kyriacou and C. N. Schizas, "Computer-aided detections of breast cancer nuclei," *IEEE Transaction on Information Technology in Biomedicine*, vol. 2, pp. 128–140, 1997.

[16] G. Huang, S. Song and C. Wu, "Orthogonal least squares algorithm for training cascade neural networks," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 11, pp. 2629–2637, 2012.

[17] A. N. Esgiar, R. N. G. Nauib, B. S. Sharif, M. K. Bennett and A. Murray, "Microscopic image analysis for quantitative measurements and feature identification of normal and cancerous colonic mucosa," *IEEE Transactions on Information Technology in Biomedicine*, vol. 2, no. 3, pp. 197–203, 1998.

[18] A. N. Esgiar, R. N. G. Nauib, B. S. Sharif, M. K. Bennett and A. Murray, "Fractal analysis in the detection of colonic cancer images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 6, no. 1, pp. 54–58, 2002.

[19] K. J. Khouzani and H. S. Zadeh, "Multiwavelet grading of pathological images of prostate," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 6, pp. 697–704, 2003.

[20] A. Tabesh, M. Teverovskiy, H. Y. Pang, V. P. Kumar, D. Verbel *et al.,* "Mutlifeature prostate cancer diagnosis and gleason grading of histological images," *IEEE Transactions on Medical Imaging*, vol. 26, no. 10, pp. 1366–1378, 2007.

[21] P. W. Huang and C. H. Lee, "Automated classification for pathological prostate images based on fractal analysis," *IEEE Transactions on Medical Imaging*, vol. 28, no. 7, pp. 1037–1050, 2009.

[22] M. M. Krishnan, V. Venkatraghavan, U. R. Acharya, M. Pal, R. R. Paul *et al.,* "Automated oral cancer identification using histopathological images: A hybrid feature extraction paradigm," *Micron*, vol. 43, no. 2–3, pp. 352–364, 2012.

[23] S. Doyle, M. Feldman, J. Tomaszewski and A. Madabhushi, "A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1205–1218, 2012.

[24] C. Hinrichs, V. Singh, L. Mukherjee, G. Xu, M. K. Chung *et al.,* "Spatially augmented lpboosting for AD classification with evaluations on the ADNI dataset," *Neuroimage*, vol. 48, no. 1, pp. 138–149, 2009.

[25] S. Rathore, M. Hussain, M. A. Iftikhar and A. Jalil, "Ensemble classification of colon biopsy images based on information rich hybrid features," *Computer in Biology and Medicine*, vol. 47, pp. 76–92, 2013.

[26] S. Rathore, M. Hussian and A. Khan, "Automated colon cancer detection using hybrid of novel geometric features and some traditional features," *Computer in Biology and Medicine*, vol. 65, pp. 279–296, 2015.

[27] J. Nalepa and M. Kawulok, "Selecting training sets for support vector machines: A review," *Artificial Intelligence Review*, vol. 52, no. 2, pp. 857–900, 2019.

[28] M. N. Gurcan, L. E. Bouncheron, A. Can, A. Madabhushi, N. M. Rajpoot *et al.,* "Histopathological image analysis: A review," *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147–171, 2009.

[29] D. Bardou, K. Zhang and S. M. Ahmad, "Classification of breast cancer based on histology images using convolutional neural networks," *IEEE Access*, vol. 6, pp. 24680–24693, 2018.

[30] I. Ahmad, M. Basheri, M. J. Iqbal and A. Rahim, "Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection," *IEEE Access*, vol. 6, pp. 33789–33795, 2018.

[31] R. Kumar, W. Y. Wang, J. Kumar, T. Yang, A. Khan *et al.,* "An integration of blockchain and AI for secure data sharing and detection of CT images for the hospitals," *Computerized Medical Imaging and Graphics*, vol. 87, pp. 1–41, 2021.

[32] U. Bodkhe, S. Tanwar, K. Parekh, P. Khanpara, S. Tyagi *et al.,* "Blockchain for industry 4.0: A comprehensive review," *IEEE Access*, vol. 8, pp. 79764–79800, 2020.

[33] A. Khatami, A. Khosravi, T. Nguyen, C. P. Lim and S. Nahavandi, "Medical image analysis using wavelet transform and deep belief networks," *Expert Systems with Applications*, vol. 86, pp. 190–198, 2017.

[34] S. M. Salaken, A. Khosravi, T. Nguyen and S. Nahavandi, "Extreme learning machine based transfer learning algorithms: A survey," *Neurocomputing*, vol. 267, pp. 516–524, 2017.

[35] H. Kong, M. Gurcan and K. B. Boussaid, "Partitioning histopathology images: An integrated framework for supervised color-texture segmentation and cell splitting," *IEEE Transactions on Medical Imaging*, vol. 30, no. 9, pp. 1661–1677, 2011.

[36] W. Wu, Y. Mallet, B. Walczak, W. Penninckx, D. L. Massart *et al.,* "Comparison of regularized discriminant analysis, linear discriminant analysis and quadratic discriminant analysis applied to NIR data," *Analytica Chimica Acta*, vol. 329, no. 3, pp. 257–265, 1996.

[37] S. J. Dixon and R. G. Brereton, "Comparison of performance of five common classifiers represented as boundary methods: Euclidean distance to cencroids, linear discriminant analysis, quadratic discriminant analysis, learning vector quantization and support vector machines, as dependent on data structure," *Chemometrics and Intelligent Laboratory Systems*, vol. 95, no. 1, pp. 1–17, 2009.

[38] N. H. Sweilam, A. A. Tharwat and N. K. A. Moniem, "Support vector machine for diagnosis cancer disease: A comparative study," *Egyptian Informatics Journal*, vol. 11, no. 2, pp. 81–92, 2010.

[39] A. N. Kamarudin, T. Cox and R. Kolamunnage-Dona, "Time-dependent ROC curve analysis in medical research: Current methods and applications," *BMC Medical Research Methodology*, vol. 17, no. 7, pp. 623–641, 2017.