Computers in Biology and Medicine 47 (2014) 76–92

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Ensemble classification of colon biopsy images based on information rich hybrid features

Saima Rathore *, Mutawarra Hussain, Muhammad Aksam Iftikhar, Abdul Jalil

Department of Computer & Information Sciences, PIEAS, Pakistan Institute of Engineering and Applied Sciences, P.O. Nilore, Islamabad

ARTICLE INFO

Article history:
Received 22 July 2013
Accepted 23 December 2013

Keywords:
Colon biopsy
Colon cancer
Classification
CCSM, HOG, and Haralick features
SVM

ABSTRACT

In recent years, classification of colon biopsy images has become an active research area. Traditionally, colon cancer is diagnosed using microscopic analysis. However, the process is subjective and leads to considerable inter/intra observer variation. Therefore, reliable computer-aided colon cancer detection techniques are in high demand. In this paper, we propose a colon biopsy image classification system, called CBIC, which benefits from discriminatory capabilities of information rich hybrid feature spaces, and performance enhancement based on ensemble classification methodology. Normal and malignant colon biopsy images differ with each other in terms of the color distribution of different biological constituents. The colors of different constituents are sharp in normal images, whereas the colors diffuse with each other in malignant images. In order to exploit this variation, two feature types, namely color components based statistical moments (CCSM) and Haralick features have been proposed, which are color components based variants of their traditional counterparts. Moreover, in normal colon biopsy images, epithelial cells possess sharp and well-defined edges. Histogram of oriented gradients (HOG) based features have been employed to exploit this information. Different combinations of hybrid features have been constructed from HOG, CCSM, and Haralick features. The minimum Redundancy Maximum Relevance (mRMR) feature selection method has been employed to select meaningful features from individual and hybrid feature sets. Finally, an ensemble classifier based on majority voting has been proposed, which classifies colon biopsy images using the selected features. Linear, RBF, and sigmoid SVM have been employed as base classifiers. The proposed system has been tested on 174 colon biopsy images, and improved performance (=98.85%) has been observed compared to previously reported studies. Additionally, the use of mRMR method has been justified by comparing the performance of CBIC on original and reduced feature sets.

1. Introduction

Colon cancer has become a major cause of deaths in modern and industrialized world. The death toll rate has been raised to 0.5 million deaths per year worldwide [1]. Colon cancer usually arises due to chain smoking, family history, increasing age, and unbalanced consumption of meat and fruits/vegetables [2].

The common and traditional method of colon cancer diagnosis is microscopic analysis of colon biopsy samples. In such an examination, histopathologists analyze the biopsy samples under microscope, and diagnose the tissue as normal/malignant based on the morphology of tissues. Normal and malignant tissues have high contrast in their morphology. Normal colon tissues have well-defined structure. Fig. 1(a) presents microscopic image of a normal colon biopsy sample, wherein all the tissues possess a regular structure. The detailed regular structure of a normal colon tissue is shown in Fig. 1(b), wherein we see that a normal colon tissue has three constituents, namely epithelial cells, non-epithelial cells, and lumen. Epithelial cells usually surround lumen and form glandular structure, whereas non-epithelial cells, called stroma, lie in between these structures. But, cancer heavily disturbs the structure of colon tissues, and makes the structure almost amorphous. The deformation introduced by cancer is clearly visible in the microscopic image of a malignant colon biopsy sample shown in Fig. 1(c). Normal and malignant colon tissues have similar colors, but the distribution of colors heavily varies. Further, normal tissues have well-defined structures such as elliptic shaped epithelial cells and lumen having sharp boundaries. Malignant colon tissues, on the other hand, have no such edges. All the constituents of tissues mix with each other, thereby diminishing the boundaries.

Histopathologists assign two quantitative measures to the malignant samples, namely stages and grades. Stage is the extent to which cancer has reached/spread in the colon or other body

* Corresponding author. Tel.: +92 332 5700789.
E-mail address: saimarathore_2k6@yahoo.com (S. Rathore).

0010-4825/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2013.12.010

parts. There are five stages of colon cancer (0, A–D) according to Duke's scale [3]. Stage 0 is the earliest stage in which cancer just starts to develop. It is still restricted to the innermost lining of colon. In stage A, cancer has reached to the middle layer of colon. In stage B, cancer has reached beyond the middle layer. Cancer has stage C if it reaches lymph nodes, and is found in at least three of them. Stage D is the final stage, wherein cancer has reached other body parts such as lungs and liver. The grade of cancer, on the other hand, is the differentiability level of malignant cells. There are three grades of colon cancer. The lowest grade of colon cancer is 'well differentiated', in which malignant cells are almost similar to the normal ones. It is the grade in which cancer progresses at lowest speed. The second grade of colon cancer is 'moderately differentiated', wherein malignant cells are differentiable from normal cells. In this grade, cancer cells progress at moderate speed. The third grade of colon cancer is 'poorly differentiated', in which malignant cells are totally different from the normal ones, and are easily distinguishable. In this particular grade, cancer cells spread at very high rate. Fig. 2 presents microscopic images of malignant colon biopsy samples having different grades of cancer.

The determination of the grades and stages of colon cancer is a manual process. In order to determine cancer grades, histopathologists analyze the biopsy samples under microscope and assign quantitative cancer grades depending upon the morphology of malignant tissues. On the other hand, cancer stage is determined by microscopic analysis of separate biopsy samples taken from different layers of colon and lymph nodes. The manual process of colon cancer detection has a few limitations. For instance, it consumes precious time of the histopathologists as they have to analyze many images per day. Moreover, the process is subjective, and leads to biased opinion due to workload and experience level of histopathologists. Further, the process leads to inter- and intra-observer variabilities [4,5]. Therefore, an accurate computational system for automatic colon cancer detection is highly desirable.

In the past two decades, a few computer-aided diagnostic systems have been proposed for automatic detection of colon cancer. However, the efforts in case of colon cancer are still deficient compared to other areas of computer-aided diagnosis. Some of the typical approaches for computer-aided diagnosis of colon cancer include analysis of human genes using microarrays [6,7], study of variation in the composition of normal and cancerous blood serum [8,9], and exploitation of textural changes in cancerous and normal colon images. These techniques have been summarized in a recent survey reported by Rathore et al. [10].

Textural variations in colon biopsy images are the emphasis of this research work. Texture analysis of colon biopsy images is characterized by extraction of discriminate features from the observed texture of these images. The extracted features are then

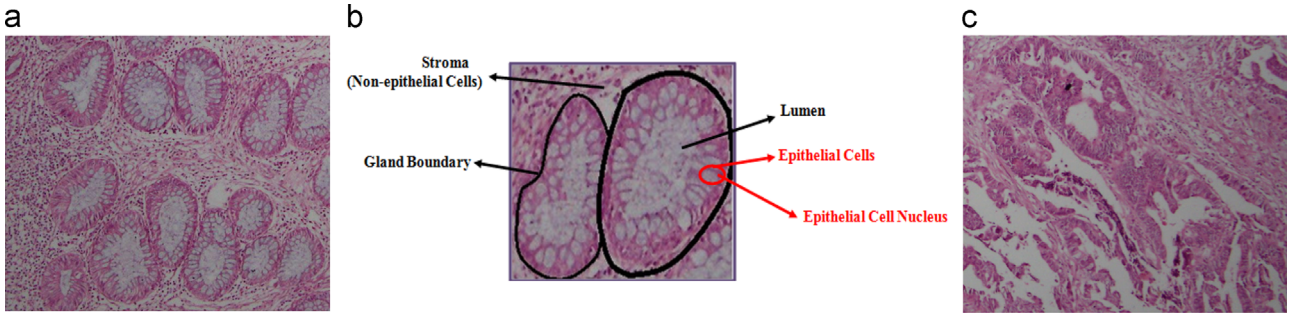

Fig. 1. Microscopic images of (a) normal and (c) malignant colon biopsy samples, and (b) regular structure of normal colon tissue.

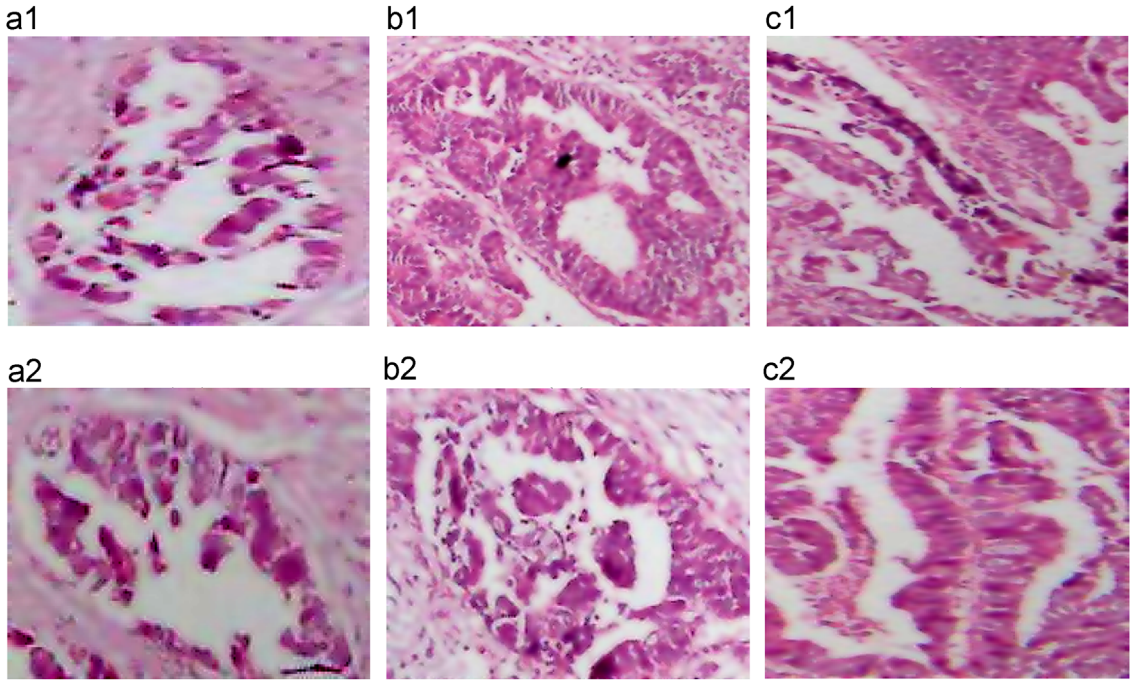

Fig. 2. Different colon cancer grades: (a_1,a_2) well differentiated, (b_1,b_2) moderately differentiated, and (c_1,c_2) poorly differentiated.

used as an input to different classifiers for discerning normal and malignant images. For example, Esgiar et al. calculated six texture features (contrast, entropy, angular second moment, dissimilarity, inverse difference moment and correlation) from gray-level co-occurrence matrix (GLCM) of the input colon biopsy images. They employed linear discriminant analysis (LDA) and *K*-nearest neighbor (KNN) classifiers, and obtained 90.2% classification accuracy [11]. They found correlation and entropy to be the two most distinctive features compared to others. Esgiar et al. [12] further extended their work, and combined features of entropy and correlation with image fractal dimensions. They obtained 94.10% classification accuracy with the same set of classifiers. Masood et al. [13] employed morphological and GLCM based texture features to obtain a classification accuracy of 84% and 90%, respectively. Morphological features comprise features of shape, size and orientation, whereas GLCM based features encompass energy, inertia and local homogeneity. Both types of features are obtained using single spectral band of colon biopsy images. Support vector machines (SVM) with polynomial kernel of degree 3 is employed as a classifier. Masood and Rajpoot [14] further extend their work, and employed circular local binary patterns from single spectral band in order to classify colon biopsy images. They employed Gaussian SVM for classification, and obtained an accuracy of 90%.

In 2010, Altunbay et al. [15] proposed a textural features based technique for classifying colon samples into normal and malignant categories. They constructed a graph on different objects, obtained by using circle fit algorithm [16] on the white, pink and purple clusters of the image. A few structural features such as degree, average clustering coefficient, and diameter are computed from the color graphs, and are used to classify given samples by using SVM classifier with linear kernel. Moreover, Ozdemir et al. [17] presented an interesting method of colon cancer detection. In their work, reference graphs of a few images of normal colon tissues are generated by employing previously used method of graph creation [15,18–20], and are stored for future referencing. Then, query graphs are generated from the test images, and are located in the reference graphs. Query graphs are searched in the reference graphs by placing nucleus node of a query graph on each node of the reference graphs. Three most similar graphs are found in the reference graphs, and then based on the degree of similarity, normal/malignant class is assigned to the test sample.

The schemes mentioned herein suffer a few drawbacks. For instance, graph based colon image classification schemes [15,17] are computationally expensive, and consume considerable CPU time in feature extraction and classification stages. Further, previous techniques have exploited only one certain aspect for colon biopsy image classification i.e. they have utilized features of only one type. These techniques include texture features, morphological features or object texture based features. But multiple feature types have not been investigated simultaneously to get a more robust and discerning feature set. Therefore, an automatic colon biopsy image classification scheme is highly desirable that is computationally tractable and simultaneously highly rich in terms of discerning features.

In this paper, we propose a colon biopsy image classification (CBIC) system, which performs ensemble classification of samples based on discriminatory capabilities of hybrid feature spaces. In order to exploit the color information present in colon biopsy images, variants of traditional statistical moments and Haralick features have been proposed. Further, traditional histogram of oriented gradients (HOG) based features have been used. These features have been combined to form various hybrid feature sets. The minimum Redundancy Maximum Relevance (mRMR) method has been employed to select discerning feature sets from individual as well as hybrid feature sets. The selected discerning feature sets have been used for classification of samples into normal and malignant classes by employing ensemble classification through majority voting.

The experimental results in this work have been obtained from various aspects. First, the performance of individual as well as hybrid feature types has been investigated. Second, the performance of original feature sets and the feature sets selected by mRMR method has been examined. Third, the performance of individual as well as ensemble classifier has been studied. The experimental results verify that the proposed system is quite suitable for the classification of colon biopsy images. Further, an analysis on computational efficiency of feature extraction and classification stages has been presented in order to validate the suitability of the proposed CBIC system to serve in real-time scenarios where histopathologists receive many images per day.

The remainder of this paper is organized as follows. Section 2 describes proposed system in detail. Section 3 describes performance measures. Section 4 demonstrates experimental results, and Section 5 concludes the paper.

2. Proposed system

The proposed CBIC colon classification system utilizes hybrid features, selected by mRMR, for decision making through ensemble classification. In this paper, we have experimentally validated the proposed CBIC system by evaluating the discerning capability of reduced individual and hybrid feature sets using base and ensemble classifiers. The proposed system comprises four main stages, namely (1) feature extraction, (2) feature selection, (3) training and testing data formulation, and finally, (4) classification of images into normal and malignant categories by using an ensemble classifier. Fig. 3 presents top-level architecture of the proposed system.

2.1. Feature extraction

Features provide a mean to decode an image pattern into a set of discriminatory measurable values. The eventual target of this phase is to articulate a feature vector for every image. Three types of features, namely CCSM, Haralick and HOG, have been extracted from the input image. We describe these feature extraction strategies in the following text.

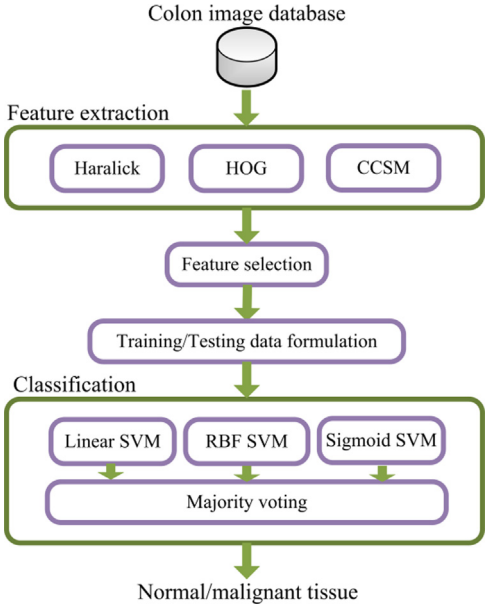

Fig. 3. Top-level architecture of the proposed CBIC system.

2.1.1. Histogram of oriented gradients (HOG) features

HOG features were initially proposed by Dalal et al. in 2005 [21]. In subsequent years, HOG features have been extensively used in various applications of computer vision, like human detection [22,23], face recognition [24], and disease diagnosis [25,26].

HOG features provide discriminative and robust edge-based information. Therefore, HOG features are expected to better exploit the well-defined and sharp edges present in colon biopsy images. In order to compute HOG features, the original image is divided into a pre-defined number of square blocks, and gradient of the blocks is obtained by applying the following two filters:

$$Horizontal: (-1 \quad 0 \quad 1)$$

$$Vertical: (-1 \quad 0 \quad 1)^T$$

Gradient orientations are calculated for each block by using gradients of the blocks obtained by applying the above-mentioned horizontal and vertical filters. Local histograms are then computed for each small block. The histograms count frequency of gradient orientations in local portions of the image. In order to overcome the local variability in the blocks, histograms are usually normalized according to the values of histograms of local neighborhood cells. Several methods exist for histogram normalization, like L_1-norm, L_2-norm, L_1-sqrt, and L_2-hys. It has been proved by Dalal et al. [21] that L_2-norm, L_1-sqrt, and L_2-hys yield almost similar performance, whereas L_1-norm yields slightly lower performance. We have also obtained the same conclusion on our dataset of colon biopsy images, therefore, we have used L_2-norm in this research work owing to the similar performance of L_2-norm, L_2-hys, and L_1-sqrt. The Matlab implementation provided by Ludwig et al. [27] has been used to calculate HOG features in this work. Different factors influence the performance of HOG features, such as number of bins, number of horizontal and vertical partitions of the image. The performance of HOG features has been analyzed over a potential range of values for the said parameters, and is presented in Section 4.3.1.

2.1.2. Proposed CCSM features

Recently, the use of color features has enormously increased in classification related applications owing to their discerning nature for different types of objects. The combination of color features with texture features also improves classification rate [28,29].

Normal and malignant colon tissues though comprise pink, white and purple clusters corresponding to connecting components, epithelial cells, and glands, respectively. But, the overall texture, and color distribution of different biological structures, heavily varies between normal and malignant colon biopsy images. For instance, in normal colon tissue images, epithelial cells (represented by white color) are usually aligned in a circular region as shown in Fig. 1(a) (see Introduction). However, in malignant colon tissue images, epithelial cells are merged with the connecting tissue and lumen, thereby resulting in an amorphous shape in which different colors have random appearance as shown in Fig. 1(c) (see Introduction). Therefore, we need a technique that could capture variation in the texture and color of the normal and malignant colon biopsy images simultaneously. We speculated and experimentally observed that statistical moments of individual color components of the colon images are able to better model such a texture and color variation.

Two color models, namely RGB and HSV, have been used, and statistical moments have been independently extracted for R, G, B, H, S, and V components of these models. The resultant feature set has been named color components based statistical moments (CCSM). The rationale for using six color components is to capture maximum possible information about the texture of images from the perspective of each color channel. Individual color components of the two models for a colon biopsy image have been shown in Fig. 4, wherein we see that each color component has a well-defined texture that better represents the colon biopsy image compared to a gray-scale image.

We have captured this texture based information by calculating statistical moments. The general expression for nth statistical moment [30] about mean is given by the following equation:

$$\mu_n = \sum_{i=0}^{I-1} (z_i - m)^n \boldsymbol{p}(z_i) \tag{1}$$

where z_i is the random variable realized by pixels' intensity values, $\boldsymbol{p}(z_i)$ is the discrete probability of the intensity level z_i, m is the mean intensity level, and I is the number of possible intensity levels in the image. We have used first four statistical moments in this work, namely mean, standard deviation, skewness and kurtosis, corresponding to $n=1, 2, 3, 4$.

The generalized feature vector for individual color components (both in RGB and HSV model) is represented by the following

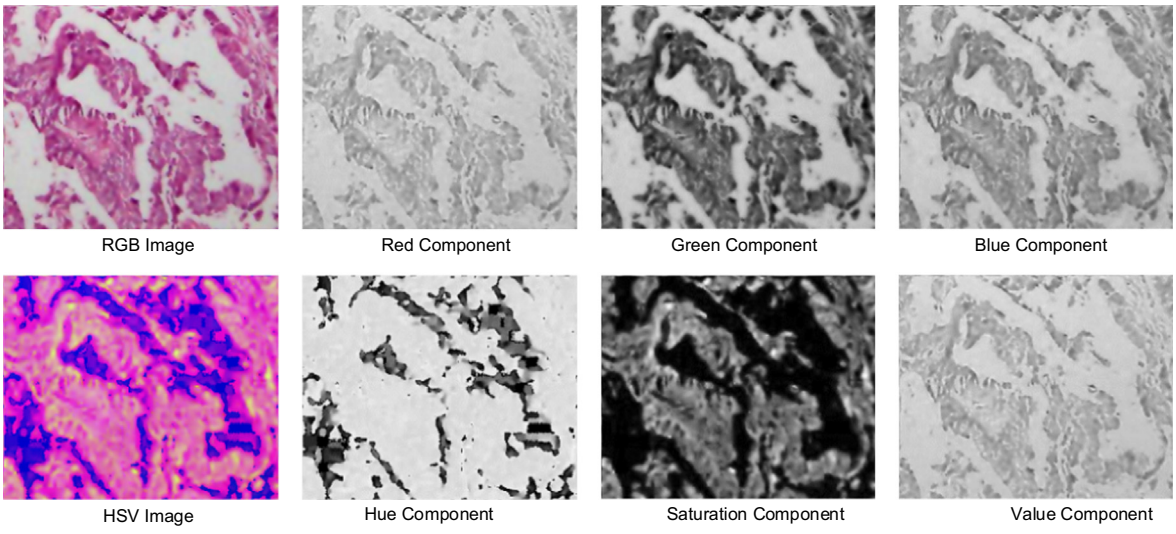

Fig. 4. Individual color components of RGB and HSV model for a colon biopsy image.

expression:

$$\boldsymbol{c}_x = [\mu_1^x \mu_2^x \mu_3^x \mu_4^x d^x]^T \quad \forall x \in \{R, G, B, H, S, V\} \tag{2}$$

where μ_n^x represents nth statistical moment computed from component x of the input image. In addition to the statistical moments, median is also computed for each color component and is represented by d^x in Eq. (2). Individual feature vectors of color components are then combined to form a composite CCSM feature vector $\boldsymbol{c}$, which is represented in the following equation:

$$\boldsymbol{c} = [\boldsymbol{c}_R^T \boldsymbol{c}_G^T \boldsymbol{c}_B^T \boldsymbol{c}_H^T \boldsymbol{c}_S^T \boldsymbol{c}_V^T]^T \tag{3}$$

2.1.3. Proposed Haralick texture features

Haralick texture features, originally proposed by Haralick [31], find applications in various domains of medical image analysis such as diagnosing diseases related to skin [32], carotid artery [33], liver [34], brain [35–37], abdomen [38], and breast [39]. Haralick texture features produce a spatial gray-level dependence matrix (SGLD), which encapsulates the spatial relationship between pixels of an image. Each (i, j)th element in SGLD describes the number of times pixels with intensity values i and j occurred in a certain relationship in the input image. The relationship may be specified in two ways: (1) horizontal and vertical distance of neighbors with the pixel of interest, and (2) the spatial relationship between pixel of interest and neighbors lying at various orientations e.g. $\theta = 0^\circ$, 45°, 90°, 135°. In this research work, we have specified the relationship by method 1 with both horizontal and vertical offset set to 1. Given the SGLD matrix of an input image, the following four statistical features have been computed.

Contrast: Measures the intensity contrast between a pixel and its neighbors over the entire image

$$t = \sum_{i=1}^{K} \sum_{j=1}^{K} (i-j)^2 p_{ij}$$

Correlation: Measures the degree of correlation between a pixel and its neighbors over the entire image

$$\rho = \sum_{i=1}^{K} \sum_{j=1}^{K} \frac{(i-m_r)(j-m_c)p_{ij}}{\sigma_r \sigma_c}$$

Energy: Measures uniformity in the image

$$e = \sum_{i=1}^{K} \sum_{j=1}^{K} {p_{ij}}^2$$

Homogeneity: Measures the spatial closeness of the distribution of elements in **G** to the diagonal

$$o = \sum_{i=1}^{K} \sum_{j=1}^{K} \frac{p_{ij}}{1+|i-j|}$$

where i and j represent row and column indices of the SGLD matrix, respectively. The term K is the number of quantized intensity levels. p_{ij} is the (i, j)th element of SGLD divided by the sum of its elements. The terms m_r, and m_c are the mean, and σ_r and σ_c are the standard deviation of rows and columns of SGLD, respectively.

In this work, we have not only computed Haralick features in traditional way (i.e. from gray-level image), but also proposed two variants of Haralick features based on the two color models i.e. RGB and HSV. These variants of Haralick features are described as follows:

1. *Haralick-GL*: These are the traditional Haralick features in which SGLD matrix is calculated from gray-scale colon biopsy image. Four texture measures (contrast, correlation, energy and homogeneity) are computed from the matrix. The feature vector of Haralick-GL features is given in the following equation:

$$\boldsymbol{h}_{gl} = [t \rho e o]^T \tag{4}$$

2. *Haralick-RGB*: This is a color models based variant of Haralick features in which three SGLD matrices are calculated separately from R, G and B components of RGB colon biopsy image. The aforementioned texture measures, separately computed from each SGLD matrix, are combined to form final feature vector of size 12.
3. *Haralick-HSV*: Haralick-HSV features are computed from H, S and V components of HSV color model in a similar fashion as Haralick-RGB features are computed. The texture features, separately calculated from SGLD matrices of H, S and V components, are combined to form a final feature vector of size 12.

The generalized feature vector for individual color components (both in RGB and HSV color models) is given in the following equation:

$$\boldsymbol{h}_x = [t_x \rho_x e_x o_x]^T \quad \forall x \in \{R, G, B, H, S, V\} \tag{5}$$

Individual feature vectors of color components are then combined to form Haralick-RGB ($\boldsymbol{h}_{rgb}$) and Haralick-HSV ($\boldsymbol{h}_{hsv}$) feature vectors, which are given in Eq. (6) and (7), respectively:

$$\boldsymbol{h}_{rgb} = [\boldsymbol{h}_R^T \boldsymbol{h}_G^T \boldsymbol{h}_B^T]^T \tag{6}$$

$$\boldsymbol{h}_{hsv} = [\boldsymbol{h}_H^T \boldsymbol{h}_S^T \boldsymbol{h}_V^T]^T \tag{7}$$

2.2. Feature reduction

In classification problems, high dimensionality of training dataset and the presence of irrelevant/redundant features may cause the classification algorithms to suffer in accurate prediction of the samples and tractable computational performance. Therefore, relevant features must be selected prior to training the classifier. Fundamentally, there exist two types of traditional feature selection methods: filters and wrappers. Filtering methods are essentially data pre-processing methods, wherein features are selected based on their relevance or discerning power with reference to the target classes. Since there may be features, which though have high relevance but are correlated with each other. Therefore, there is a possibility that features selected by these methods may be highly correlated i.e. redundant. In wrapper type methods, feature selection is wrapped around a learning method i.e. the usefulness of a feature is directly judged by the estimated accuracy of the learning method. Therefore, the feature set selected by these methods just matches well with the characteristics of the employed learning method, and thereby limiting the generalization capability of the feature set.

In this research work, mRMR method [40], which is an extension of minimum redundancy based filter method, has been used as a feature selection strategy to select discerning features from individual as well as hybrid feature sets. The mRMR method caters the criteria of both redundancy and relevance for feature selection, thereby resulting in a relevant and non-redundant feature set. Moreover, mRMR is not associated with a particular learning

method, therefore it results in a generalized feature set. Further, mRMR models the problem of feature selection as a series of smaller scale problems, each of which only involves two variables. Hence, the estimation of joint probabilities is more robust, and the results of mRMR are better compared to ranking and filter based methods. Yun and Yang [41] have also experimentally proved the superior performance of mRMR compared to various other feature selection methods.

2.2.1. Minimum Redundancy Maximum Relevance (mRMR)

The mRMR method has the objective to maximize the inter-class and minimize the intra-class variances. It accomplishes this objective by selecting features, which show maximum relevance to the target labels and have minimum redundancy amongst them [40]. Generally, mutual information amongst the features as well as amongst the features and the target labels can be utilized to calculate relevance and redundancy scores of the features. As it is generally not possible to achieve both the objectives simultaneously, therefore, a tradeoff is established between the two objectives.

For a given feature set $\mathbf{F}$ comprising M training samples of N features each, the redundancy of the feature set $R(\mathbf{F})$ is the average value of all mutual information values between all the feature pairs.

$$R(\mathbf{F}) = \frac{1}{N^2}\sum_{i,j=1}^{N} I(\boldsymbol{f}_i,\boldsymbol{f}_j) \quad \text{where } \boldsymbol{f}_i,\boldsymbol{f}_j \in \mathbf{F} \tag{8}$$

where $\boldsymbol{f}_i$ and $\boldsymbol{f}_j$ represent the ith and jth features in $\mathbf{F}$, respectively, and $I(\boldsymbol{f}_i, \boldsymbol{f}_j)$ represents the mutual information between the features f_i and $\boldsymbol{f}_j$, which can be calculated using the following expression:

$$I(\boldsymbol{f}_i\boldsymbol{f}_j) = \sum_{x,y} p(f_{i,x},f_{j,y})\log\left(\frac{p(f_{i,x},f_{j,y})}{p(f_{i,x})p(f_{j,y})}\right) \quad \text{where } x,y = 1,2,3,\ldots,M \tag{9}$$

where $\boldsymbol{f}_{i,x}$ and $\boldsymbol{f}_{j,y}$ are xth and yth elements of feature vectors $\boldsymbol{f}_i$ and $\boldsymbol{f}_j$, respectively. $p(\boldsymbol{f}_{i,x}, \boldsymbol{f}_{j,y})$ shows the joint probability density function of $\boldsymbol{f}_{i,x}$, and $\boldsymbol{f}_{j,y}$. The terms $p(\boldsymbol{f}_{i,x})$ and $p(\boldsymbol{f}_{j,y})$ represent marginal probability density functions of $\boldsymbol{f}_i x$, and $\boldsymbol{f}_{j,y}$, respectively.

Similarly, for the target labels l, where $l=-1$ represents normal and $l=+1$ represents malignant labels, the relevance of the feature set $\mathbf{F}$ for the class l, denoted by $V(\mathbf{F},l)$, is defined by the average value of all mutual information values between the individual feature $\boldsymbol{f}_i$ and the class l as follows:

$$V(\mathbf{F},l) = \frac{1}{N}\sum_{i=1}^{N} I(\boldsymbol{f}_i,l) \quad \text{where } \boldsymbol{f}_i \in \mathbf{F} \text{ and } l \in \{-1,+1\} \tag{10}$$

$I(\boldsymbol{f}i_{,l})$ denotes the mutual information between the feature vector $\boldsymbol{f}_i$ and the class label l. It can be calculated using the following equation:

$$I(\boldsymbol{f}_i,l) = \sum_{x} p(f_{i,x},l)\log\left(\frac{p(f_{i,x},l)}{p(f_{i,x})p(l)}\right) \quad \text{where } x = 1,2,3,\ldots,M \tag{11}$$

Here $\mathbf{p}(f_{i,x},l)$ is the joint probability density function of $f_{i,x}$ and target label l. The terms $p(f_{i,x})$ and $p(l)$ show the marginal probability density functions of $f_{i,x}$ and target label l, respectively.

The objective is to select the set of features which yields maximum relevancy V and minimum redundancy R. As both the objectives are usually not achievable simultaneously, therefore, Eq. (12) establishes a tradeoff between the two objectives by combining Eqs. (8) and (10) as follows:

$$mRMR = \max_{\mathbf{F}}[V(\mathbf{F},l) - R(\mathbf{F})] = \max_{\mathbf{F}}\left[\frac{1}{N}\sum_{i=1}^{N} I(\boldsymbol{f}_i,l) - \frac{1}{N^2}\sum_{i,j=1}^{N} I(\boldsymbol{f}_i,\boldsymbol{f}_j)\right] \tag{12}$$

Let m_i be the set membership indicator function for the feature vector $\boldsymbol{f}_i$, so that $m_i=1$ indicates the presence and $m_i=0$ indicates the absence of the feature $\boldsymbol{f}_i$ in the globally optimal feature set, then Eq. (12) may be written as an optimization problem as follows:

$$mRMR = \max_{m \in \{0,1\}^N}\left[\frac{\sum_{i=1}^{N} I(\boldsymbol{f}_i,l)m_i}{\sum_{i=1}^{N} m_i} - \frac{\sum_{i,j=1}^{N} I(\boldsymbol{f}_i,\boldsymbol{f}_j)m_i m_j}{(\sum_{i=1}^{N} m_i)^2}\right] \tag{13}$$

Thus, the feature set determined using mRMR is expected to contain values which not only bear maximum relevancy to the target labels, but are non-redundant as well [40].

2.3. Training/testing data formulation

An important phase of the proposed CBIC system is training/testing data formulation. In statistical prediction, three cross-validation methods, namely sub-sampling, independent dataset test, and jackknife are used to examine the effectiveness of a classifier in practical applications. A recent review by Chou [42] demonstrates that among the three methods, the jackknife cross-validation test is believed to be least arbitrary, rigorous, and most objective because of its ability to yield a unique result for a given dataset. In medical diagnosis systems, it is highly desirable to yield unique output against a sample no matter how many times we test the sample. Therefore, the jackknife test has been increasingly used by the researchers for investigating the performance of various classifiers in medical diagnostic systems [33,43,44]. Consequently, the jackknife 10-fold cross-validation has been employed in this study to examine the anticipated success rates of the classifiers. In a certain iteration of 10-fold cross-validation, nine folds have been used for training, and the 10th fold has been employed for testing based on the training performed on nine folds. The process has been repeated 10 times until classes of all the samples in all the folds have been determined.

2.4. Classification model

SVM classifier, originally proposed by Vapnik [45], has been quite successfully used in several medical diagnosis applications [46–52]. We have used the SVM classifier with different kernels for classification of colon biopsy images using the features described in Section 2.1. In the following text, we describe the concept of SVM classification in detail.

Consider a training dataset $\mathbf{Q} \in \mathbf{F}$ comprising Z training samples $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_Z$, and target labels $\boldsymbol{t} = [t_1, t_2, \ldots, t_Z]^T$ where $t_Z \in \{-1,+1\}$, $z = 1,2,\ldots,Z$. The data that are being classified using the SVM kernel may be linearly or non-linearly separable. For linearly separable data, the ultimate aim of classification is to design a linear decision surface, which correctly classifies the training samples. Such a decision surface is given in the following equation:

$$f(\boldsymbol{q}) = \boldsymbol{w}^T.\,\boldsymbol{q} + bias = 0 \tag{14}$$

However, such a decision surface, defined by its direction (weight) vector $\boldsymbol{w}$ and position (bias) in the space, may not be unique. Therefore, the objective is to select a direction $\boldsymbol{w}$ of the decision surface such that the distance of the surface to the nearest points of the two classes is maximum. The nearest points are called support vectors, and the distance of the nearest points from the decision surface is called margin. For binary classification of samples into normal and malignant classes, candidate decision surfaces are normalized in such a way that value of $f(\boldsymbol{q})$ for the support vectors is equal to $+1$ for malignant class and -1 for normal class. Therefore, the points of the normal and malignant classes, which are correctly classified, have values of $f(\boldsymbol{q})$ less than -1 and greater than $+1$, respectively. The problem may be formulated as given below, and can be solved using optimization techniques for non-linear objective function subjected to

linear inequalities [53]:

$$\text{minimize: } ||\boldsymbol{w}||^2$$
$$\text{subject to: } t_Z(\boldsymbol{w}^T\boldsymbol{q}_Z+\text{bias}) \geq 1,\ z=1,2,...,Z$$

where weight vector $\boldsymbol{w}$ of the optimal decision surface is a linear combination of the support vectors.

For linearly non-separable data, there are three cases of the samples of the training dataset. First, training points may fall on the correct side of the decision surface and behind margin. Second, training points may fall on the correct side of the decision surface, but inside margin. Third, training points may fall on the wrong side of the decision surface. The objective is to select a decision surface such that points corresponding to second and third cases are minimum. A penalty term is added to the objective function to minimize the points falling inside margin or on the wrong side of the margin. Let $\boldsymbol{\xi}=[\xi_1,\xi_2,...,\xi_z]$ be a vector comprising error terms corresponding to Z training samples in the dataset. Therefore, the problem for linearly non-separable data may be formulated as the following:

$$\text{minimizes: } ||\boldsymbol{w}||^2+c\sum_{z=1}^{Z}\xi_z$$

$$\text{subject to: } t_z(\boldsymbol{w}^T\boldsymbol{q}_z+bias)\geq 1-\xi_z,\ z=1,2,...,Z$$

where $\xi_{z=0}$, $0<\xi_z<1$, and $\xi_z>1$ for points corresponding to first, second and third cases, respectively. The term c is the penalty parameter associated with the penalty term

$$\sum_{z=1}^{Z}\xi_z.$$

When data is not linearly separable, SVM maps the data from lower dimension J to a higher dimension J^* through a non-linear mapping $\Phi(\boldsymbol{q})$so that $\Phi: R^J\rightarrow R^{J^*}$, $J^*\gg J$. The data becomes easily separable in the higher dimension J^*. A non-linear decision surface $f(\boldsymbol{q})$ between the classes can be constructed in terms of kernel functions [53]

$$f(\boldsymbol{q})=\sum_{z=1}^{S^P}\alpha_z t_z K(\boldsymbol{q},\boldsymbol{r})+bias=\sum_{z=1}^{S^P}\alpha_z t_z\Phi(\boldsymbol{q})\cdot\Phi(\boldsymbol{r})+bias \quad (15)$$

where S^P is the number of support vectors. The terms α_z and t_z, respectively, are the Lagrange multipliers and target labels associated with the support vectors.

In SVM, two types of kernel functions, i.e. local (RBF) kernels and global (linear, polynomial, sigmoid) kernels are commonly used. The measurement of local kernels is based on a distance function. On the other hand, the performance of global kernels depends on the dot product of data samples. To introduce diversity in the ensemble classifier, we have used one local kernel (Gaussian RBF kernel) and two global kernels (linear and sigmoid kernels) for classification. Further, different kernels of SVM adopt different mechanisms for data classification. For instance, sigmoid and RBF kernels draw non-linear decision surfaces for classification of data, whereas linear SVM demarcates a linear boundary between the classes. Linear, RBF, and sigmoid kernels of SVM are mathematically defined by Eqs. (16), (17) and (18), respectively:

$$K(\boldsymbol{q},\boldsymbol{r})=\boldsymbol{q}^T.\boldsymbol{r} \quad (16)$$

$$K(\boldsymbol{q},\boldsymbol{r})=\exp(-\gamma||\boldsymbol{q}-\boldsymbol{r}||^2) \quad (17)$$

$$K(\boldsymbol{q},\boldsymbol{r})=[\gamma<\boldsymbol{q},\boldsymbol{r}>+r] \quad (18)$$

The optimal parameter values of kernels are computed using grid search. All these SVM kernel functions share one common cost parameter c, which is basically the constraint violation cost associated with the data point occurring on the wrong side of the decision surface. The parameter γ in the RBF and sigmoid kernel functions represents the width of Gaussian functions. The parameter r is the offset of sigmoid kernel. The selection of optimal values of SVM parameters will be discussed in Section 4.3.2.

2.5. Ensemble classification

In recent times, ensemble classification has gained much popularity over standalone classification methodologies [54,55]. This popularity is attributed to better classification rate produced by ensemble framework for most applications. In this research study, we have proposed an ensemble classifier for the detection of cancer in colon biopsy images. The proposed ensemble has been developed by stacking the predictions of individual classifiers. In this manner, a discriminating decision space is constructed that helps in better identification of samples compared to original decision space [56]. In this study, linear, RBF and sigmoid kernels of SVM have been used as base classifiers. The final predictions of the proposed ensemble classifier have been acquired by using majority voting on predictions of base classifiers.

3. Performance measures

The proposed CBIC system has been quantitatively evaluated using well-known performance measures such as accuracy, sensitivity, specificity, Mathew's correlation coefficient (MCC), *F*-score, Kappa statistics, and receiver operating characteristics curve (ROC). Generally, a particular measure of accuracy takes into account a certain factor underlying the yielded classification results. However, we use multiple classification measures in order to obtain more reliable comparison. Normal and malignant images, respectively, correspond to negative and positive samples. In this context, true positive (TP) and true negative (TN), respectively, represent the number of malignant and normal samples, which are correctly classified. Likewise, false positive (FP) and false negative (FN), respectively, represent the number of normal and malignant samples, which are incorrectly classified.

3.1. Accuracy

The classification accuracy of a technique depends upon the number of correctly classified samples (i.e. true negative and true positive) [57], and is calculated using the following equation:

$$Accuracy=\frac{TP+TN}{N}\times 100$$

where N is the total number of colon biopsy images.

3.2. Sensitivity

Sensitivity is a measure of the proportion of positive samples which are correctly classified [57]. It can be calculated using the following equation:

$$Sensitivity=\frac{TP}{TP+FN}$$

Its value ranges between 0 and 1, where 0 and 1, respectively, mean worst and best classification.

3.3. Specificity

Specificity is a measure of the proportion of negative samples which are correctly classified [57]. It can be calculated using the

following equation:

$$Specificity = \frac{TN}{TN+FP}$$

Its value ranges between 0 and 1, where 0 and 1, respectively, mean worst and best classification.

3.4. Matthews correlation coefficient (MCC)

MCC is a measure of the eminence of binary class classifications [58]. It can be calculated using the following formula:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{((TP+FN)(TP+FP)(TN+FN)(TN+FP))}}$$

Its value ranges between -1 and $+1$, where -1, $+1$ and 0, respectively, correspond to worst, best, and random prediction.

3.5. F-score

F-score is a measure of the accuracy of classification [57]. *F*-score is a weighted average of precision and recall, and can be calculated using the following equations:

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F-score = 2 \times \frac{Precision \times Recall}{Precision+Recall}$$

The value of *F*-score ranges between 0 and 1, where 0 means the worst classification, and 1 means the best classification.

3.6. Kappa statistic

Kappa statistic (κ) measures the agreement between ground truth and the results of a classification algorithm [59]. The equation for κ is

$$\kappa = \frac{\Pr(a)-\Pr(e)}{1-\Pr(e)}$$

where $\Pr(a)$ is the relative observed agreement among ground truth and the classification algorithm, and $\Pr(e)$ is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly saying each category. The value of κ varies between 0 and 1, where $\kappa=1$ shown near to perfect agreement, and $\kappa=0$ shows no agreement between ground truth and the classification algorithm.

3.7. ROC

An ROC curve is a standard way for graphical representation of the classification performance of a system [57]. It characterizes the system over its entire operating range, and is created by plotting true positive rate (TPR) against false positive rate (FPR). TPR represents the number of correct positive cases divided by the total number of positive cases. FPR, on the other hand, is the number of negative cases predicted as positive cases divided by the total number of negative cases. The terms TPR and FPR are actually sensitivity and $(1-\text{specificity})$ of the classification system. In order to plot the ROC curve, the predicted values (decision values) of a classifier are scaled in the range [0–1], and the curve is plotted by applying a classification threshold T on the decision values. If the decision value is greater than T, then the input sample is allocated to the first class (normal), otherwise to the second class (malignant). The threshold is varied in the range [0–1], and TPR/FPR pair is computed at each threshold value. ROC curve is then plotted between computed TPR and FPR values. In practical applications, the ROC curve provides a degree of freedom to select the operating point which best accomplishes the requirements of the application.

4. Results and discussions

In this section, we present the results of using the proposed system for identifying normal and malignant colon biopsy images from the dataset presented in Section 4.1. Individual features as described in Section 2.1 have been extracted from colon biopsy images, and multiple hybrid feature sets have been constructed from the individual feature sets. Individual as well as hybrid features have been reduced using mRMR method (see Section 4.2). Majority voting based ensemble classifier has been constructed, wherein linear, RBF and sigmoid kernels of SVM have been used as base classifiers. Optimal values of parameters have been selected for these classifiers as well as for feature selection modules (HOG and Haralick texture features), and are discussed in Section 4.3. Then, based on ensemble classification, the performance of reduced individual and hybrid feature sets has been investigated in Sections 4.4 and 4.5, respectively. Computational time requirements of all the feature sets have been summarized in Section 4.6. Finally, the performance enhancement achieved by employing mRMR has been discussed in Section 4.7. All the computations have been performed on Intel Core i7 with 3.4 GHz processor and 16GB RAM. Matlab computational software for 64 bit windows has been used in all the experiments. For classification, data has been scaled in the range 0–1, and 10-fold cross-validation has been used. Experiments have been performed 20 times, and average values of performance measures have been reported.

4.1. Dataset

The biopsy samples used in this work have been collected in the years 2010, 2011, and 2012 from the Pathology Department of Rawalpindi Medical College, Rawalpindi, Pakistan. The available 68 colon biopsy samples for the said time period have been selected without any discrimination of race, gender, and age. The samples have been stained with hematoxylin and eosin, and the thickness of the tissue section in the biopsy slide is 5–6 μm. The dataset and ground truth have been prepared under the guidance of classified histopathologist (Imtiaz Ahmad Qureshi, Assistant Professor) of the department. The confidentiality of the data has been maintained throughout this research work. The hospital has only provided the information about age and gender of the patients along with biopsies. Further, patients have agreed before colonoscopy that the captured biopsy samples may be used for research purposes. The imaging equipment has been provided by PAEC General Hospital, Islamabad. The magnification factor of objective lens of the microscope has been set to $10\times$, and several RGB images of colon have been captured at 600×800 resolutions. A dataset of 174 variable size microscopic RGB images has been extracted from these RGB images. The number of normal and malignant images in the dataset is 82 and 92, respectively. There are 23 well differentiated, 44 moderately differentiated, and 25 poorly differentiated malignant images in the dataset. Further, the distribution of malignant images into stage A, stage B, stage C, and stage D is 15, 24, 21 and 32, respectively. There is no image having stage 0 cancer, because people usually do not carry out routine tests, thereby resulting in late detection of cancer. The age of the patients in the dataset varies from 42 to 68. Mean and standard deviation of the age of patients are 57.11 and 6.35, respectively. The age varies from 43 to 63 and from 42 to 68 for female and male patients, respectively. Fig. 5 shows the distribution of 92

malignant images, wherein we see that most of the cancer patients are above 50 years of age.

4.2. Selection of optimal features through mRMR

The mRMR method has been employed to select discerning features from different feature sets. In this context, mRMR has been applied to each of the individual and hybrid feature sets, and the number of features selected by mRMR for each feature set has been shown in Table 1. Results reveal that mRMR has substantially reduced the size of feature set in all cases except for the Haralick-GL features where there is no room for further feature reduction. The reduced feature set (comprising three or two features) in case of Haralick-GL features leads to deteriorated performance. The percentage decrease of more than 96% for larger datasets (fourth column of Table 1) shows that mRMR has considerably reduced the size of various feature sets.

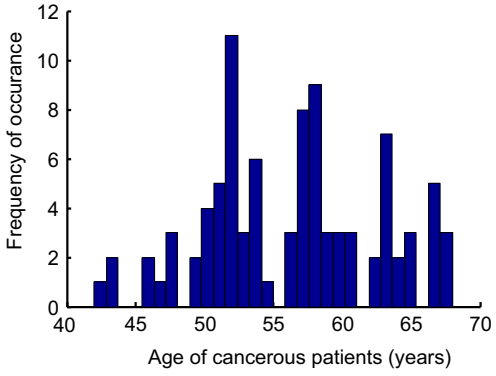

Fig. 5. Distribution of age of cancerous patients.

Table 1
Number of features selected by mRMR.

Features	Number of original features	Number of features selected by mRMR	Percentage decrease
CCSM	30	18	40.00
Haralick-GL	04	04	00.00
Haralick-RGB	12	07	41.67
Haralick-HSV	12	08	33.33
HOG	6840	186	97.28
CCSM + Haralick-HSV	42	26	38.10
CCSM + HOG	6870	211	96.93
HOG + Haralick-HSV	6852	197	97.12
CCSM + HOG + Haralick-HSV	6882	223	96.76

4.3. Selection of optimal values for system parameters

The performance of the proposed CBIC system depends on several parameters, which need to be tuned for optimal performance. In the subsequent text, analysis of optimal values of SVM models and feature selection methods has been presented in detail. These optimal values have been used in all the subsequent experiments. It is worthwhile to note that optimal values of the parameters have been calculated on the reduced feature sets (discussed in Section 4.2).

4.3.1. Feature selection methods

Performance of HOG features depends upon three factors: (a) number of horizontal partitions of the image (H), (b) number of vertical partitions of the image (V), and (c) bin size (B). Therefore, prior to classifying images using HOG, we have experimentally found the optimal values of these parameters. In this context, H, V and B have been varied in the potential ranges of H = 3, 4, 5,…,25, V = 3, 4, 5,…,25 and B = 6, 7, 8,…,25. The step size of 1 has been used for all the variables in order to fully exploit the range of possible values. HOG features have been computed for each particular combination of these parameter values, and the combination that yields maximum classification performance for individual as well as ensemble classifier has been selected. HOG features computed for H = 19, V = 20 and B = 18 yield maximum classification accuracy, therefore, HOG feature vector calculated at these values has been used in further experiments.

In order to further demonstrate the individual effect of each variable on the performance of HOG features, the classification accuracy has been measured by varying one parameter while keeping other two fixed at their maximum values. By maximum values we mean the maximum of the potential set of values i.e. 25 each for H, V and B. Corresponding results are shown in Fig. 6, wherein we see that the classification accuracy using HOG features increases as the values of H, V and B increase. The discriminating capability of HOG features at higher values of H, V and B is enhanced. The results have been shown in Fig. 6 with the interval of 3 for the presentation purposes.

Performance of Haralick features significantly depends upon the size of SGLD matrix. Finding the optimal size of SGLD matrix for a particular application is a critical task. In this study, we have extracted different variants of Haralick features at n = 2, 3,…,7 where the size of SGLD matrix is 2^n, and measured the classification performance by employing base as well as proposed ensemble classifier. Corresponding results are given in Fig. 7, which reveal that classification performance keeps on increasing until n = 5. An increase in the size of SGLD matrix beyond this point leads to little or no change in the performance. Therefore, the size of SGLD matrix is set to 32×32 corresponding to n = 5.

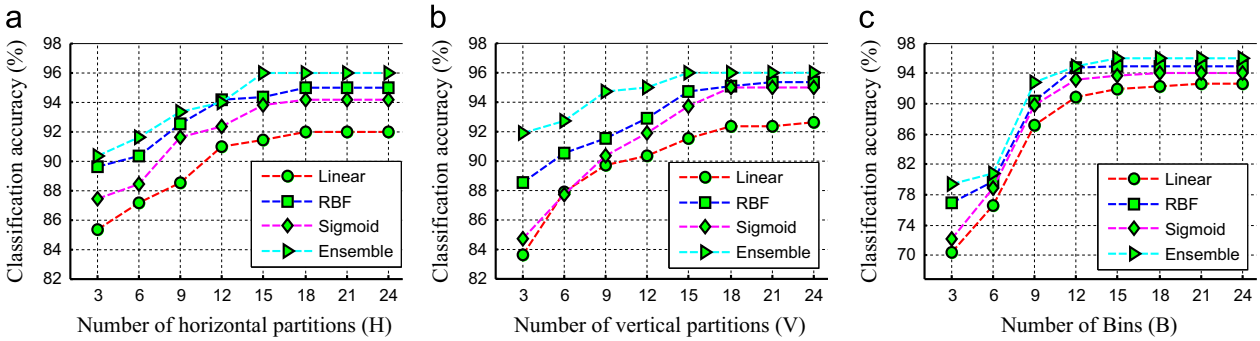

Fig. 6. Classification performance of HOG features as a function of number of (a) horizontal partitions, (b) vertical partitions, and (c) bins.

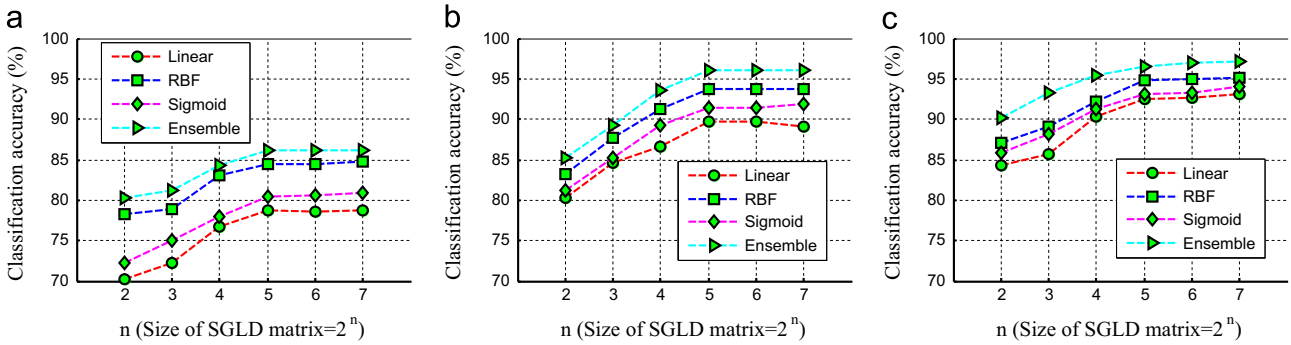

Fig. 7. Classification performance of (a) Haralick-GL, (b) Haralick-RGB and (c) Haralick-HSV features as a function of size of SGLD matrix.

4.3.2. Classification models

Performance of SVM classifiers depends on several parameters. In this research study, the grid search method [60] has been employed for selection of optimal parameter values by carefully setting grid range and step size. The parameter c is common to all the classifiers. Its optimal value has been obtained by adjusting the grid range of $c=[0,...,10^2]$ with $\Delta c=1$ for all the kernels. The parameter γ is involved in RBF and sigmoid kernels. Its optimal value has been obtained by adjusting the grid range of $\gamma=[0.001,...,0.1]$ with $\Delta\gamma=0.002$ for both the kernels. A parameter r is specific to sigmoid kernel only, and its default value is used.

4.4. Performance analysis of individual feature sets

In this section, we discuss our findings about the performance of the individual feature sets, which have been acquired after feature reduction through the mRMR method.

4.4.1. Performance analysis of HOG features

The reduced HOG features have been employed for classification of colon biopsy images. Linear, RBF, sigmoid and the proposed ensembles have been used for this purpose. Corresponding performance measures have been reported in Table 2, which indicate that HOG features are able to classify the given dataset with good accuracy. In terms of individual classifiers, the best accuracy, achieved using HOG features, is 94.86%. However, due to the exploitation of strengths of multiple classification models in the proposed ensemble classification, accuracy further increases up to 95.98%, thereby showing a percentage increase of 1.13%. Such a noteworthy performance validates the suitability of HOG features for classifying colon cancer dataset. HOG features have been able to produce excellent results because they realistically model the sharp edges present in colon biopsy images. Moreover, in normal colon biopsy tissues, epithelial cells, glands and lumen have an elliptical structure, which is destroyed in malignant tissues. HOG features exploit local image structure by dividing an image into multiple horizontal and vertical partitions, therefore, are better able to capture the structural variations in normal and malignant colon tissues. Better sensitivity (=0.95) and specificity (=0.98) prove that HOG features are able to equally distinguish normal and malignant classes. Further, superior MCC (=0.92) and *F*-score (=0.96) values show the goodness of HOG features for classification of colon biopsy images.

4.4.2. Performance analysis of proposed Haralick texture features

This section presents the capabilities of reduced Haralick-GL, Haralick-HSV, and Haralick-RGB features for the classification of colon biopsy images. Linear, RBF, sigmoid and the proposed ensembles have been employed for the classification of colon biopsy images. Table 3 demonstrates the performance of different variants of Haralick texture features.

Table 2
Performance of HOG features.

Classifier\ Performance measures	Accuracy	Sensitivity	Specificity	MCC	*F*-score
Linear	91.95	0.98	0.85	0.84	0.93
RBF	94.86	0.97	0.93	0.89	0.95
Sigmoid	94.25	0.98	0.90	0.89	0.95
Ensemble	95.98	0.95	0.98	0.92	0.96

Table 3
Performance of Haralick-GL, Haralick-RGB and Haralick-HSV features.

Classifier\ Performance measures	Accuracy	Sensitivity	Specificity	MCC	*F*-score
Haralick-GL					
Linear	78.74	0.77	0.80	0.58	0.79
RBF	84.48	0.84	0.85	0.69	0.85
Sigmoid	80.46	0.77	0.84	0.61	0.81
Ensemble	86.21	0.87	0.85	0.72	0.87
Haralick-RGB					
Linear	89.66	0.88	0.91	0.79	0.90
RBF	93.68	0.90	0.98	0.88	0.94
Sigmoid	91.38	0.96	0.87	0.83	0.92
Ensemble	95.98	0.93	0.99	0.92	0.96
Haralick-HSV					
Linear	92.53	0.89	0.96	0.85	0.93
RBF	94.83	0.91	0.99	0. 90	0.95
Sigmoid	93.10	0.87	1.00	0.87	0.93
Ensemble	96.55	0.97	0.96	0.93	0.97

Table 3 demonstrates that the proposed variants of Haralick features classify the colon biopsy images with good accuracy. Haralick-RGB and Haralick-HSV yield 95.98% and 96.55% classification accuracy, respectively. This accuracy value is much higher compared to that of Haralick-GL features (86.21%). These classification results also reveal the significance of information contained by different color components in colon biopsy images. Fig. 4, shown in Section 2.1.2, demonstrates the changes between different color channels of a sample colon biopsy image. In the Haralick-GL features, this valuable color information is not exploited. Whereas Haralick-HSV and Haralick-RGB quite handily exploit color information along with texture information in the colon biopsy images, which results in much improved classification performance compared to Haralick-GL features. Among the features employed for capturing texture variation, energy and homogeneity have good variation in normal and malignant colon tissues. Therefore, these two features play a vital role among the four features for classification of normal and malignant tissues.

Only Haralick-HSV features among the two variants of Haralick features will be used in subsequent sections owing to their better

Table 4
Performance of CCSM features.

Classifier\ Performance measures	Accuracy	Sensitivity	Specificity	MCC	*F*-score
Linear	95.40	0.93	0.98	0.91	0.96
RBF	96.55	0.97	0.96	0.93	0.97
Sigmoid	95.98	0.93	0.99	0.92	0.96
Ensemble	96.68	0.98	0.96	0.95	0.96

performance. Almost equal values of sensitivity (=0.97) and (=0.96) prove that Haralick-HSV features are capable of identifying normal and malignant tissues equally well. Similar phenomenon is observed among other performance measures, which verifies the reliability of Haralick-HSV features. Particularly, MCC value for Haralick-HSV is very good which shows the effectiveness of the classification results.

4.4.3. *Performance analysis of proposed CCSM features*

In this section, a detailed analysis on the performance of reduced CCSM features for classification of colon biopsy images is presented. Linear, RBF, sigmoid kernels of SVM classifier, and the proposed ensemble strategies have been employed for the classification of colon biopsy images using CCSM features. Table 4 shows the performance measures for CCSM features.

The results in Table 4 demonstrate that the proposed CCSM features are quite discerning for the classification of colon biopsy images, and have been able to yield as high as 96.68% classification accuracy. The values of other performance measures are also in accordance with the accuracy values, thereby emphasizing the consistency of classification results. CCSM features have been able to yield better sensitivity (=0.98) and specificity (=0.96) values which prove the fact that CCSM features can equally identify normal and malignant samples. Note that among different types of individual features, CCSM features yield overall maximum performance both for base and ensemble classifiers. This is due to the combination of statistical measures from different color channels, which capture the underlying model of our classification problem more vividly. Among the six different color channels, hue and saturation are the most discerning for the classification of normal and malignant colon biopsy images compared to other statistical features.

4.5. *Performance analysis of hybrid feature sets*

In this section, we present the performance analysis on hybrid feature sets constructed by the combination of features discussed in Section 2.1. The hybrid features have been developed by simple concatenation of individual feature sets. Hybrid feature sets have been reduced using mRMR as discussed in Section 4.2. Colon biopsy images have been classified into normal and malignant classes based on reduced hybrid feature sets. The classification results have been reported for ensemble of the base classifiers (linear, RBF and sigmoid SVM). Among different variants of Haralick features, only Haralick-HSV, owing to its better discerning characteristics (see Table 3), has been used for hybridization. Table 5 presents classification results for different hybrid feature sets.

The first three hybrid feature sets, each comprising two feature types, have been developed by hybridizing the three individual feature types. The classification accuracy of CCSM+HOG, CCSM+Haralick-HSV and Haralick-HSV+HOG features is 97.13%, 97.13% and 97.70%, respectively. This classification accuracy is marginally higher compared to that of HOG (=95.98%), CCSM (=96.68%), and Haralick-HSV (=96.55%) features. Among the hybrid feature sets, the feature set comprising Haralick-HSV and HOG produces better classification performance for all the evaluation measures. Finally, for the 3-tuple combination (i.e. CCSM+Haralick-HSV+HOG), overall highest classification accuracy (=98.85%) has been obtained.

Table 5
Performance of hybrid feature sets.

Hybrid feature sets	Accuracy	Sensitivity	Specificity	MCC	*F*-score
CCSM+HOG	97.13	0.99	0.95	0.94	0.97
CCSM+Haralick-HSV	97.13	0.96	0.99	0.94	0.96
Haralick-HSV+HOG	97.70	0.97	0.98	0.96	0.97
CCSM+HOG+Haralick-HSV	98.85	0.98	1.00	0.98	0.99

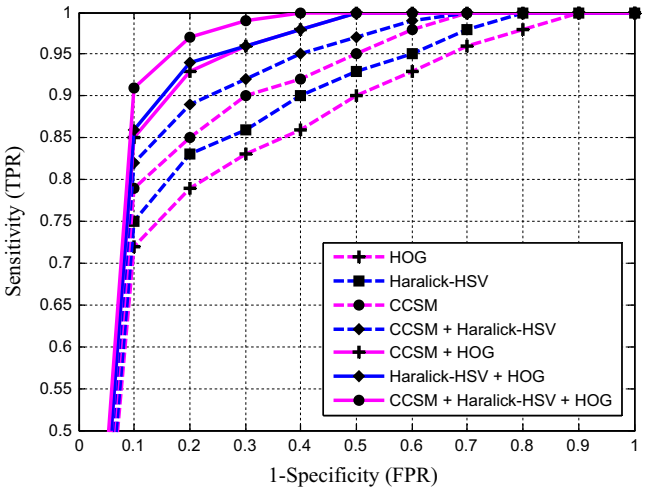

Fig. 8. ROC curves for various hybrid feature vectors.

Performance of the proposed features has also been investigated using ROC curves. Fig. 8 demonstrates the performance of various individual and hybrid feature vectors. The ROC curves of all the features (individual and hybrid) are well above the diagonal line, which proves the fact that all the features yield good classification results. Further, Fig. 8 demonstrates that hybrid features marginally improve performance compared to their individual counterparts because ROC curves of all the hybrid feature sets are slightly above the curves of individual feature sets.

Finally, we provide visual classification results for a few colon biopsy images in Fig. 9. These results have been obtained by using hybrid feature vector comprising CCSM, HOG and Haralick-HSV features. Fig. 9 presents 12 sample normal and malignant images, which are correctly classified by the proposed CBIC system. Labels below the figures are assigned by the proposed classification system. Normal images in the first row of Fig. 9 have almost similar structure. Epithelial cells are lined in circular regions, and the lumen is surrounded by the epithelial cells. As the proposed system was trained on this standard regular structure of normal tissues, therefore, it has successfully identified the unseen normal images as well. Second and third rows in Fig. 9 demonstrate the malignant images, which are correctly classified by the proposed system. These images have amorphous shape, and the regular structure of normal tissues is not visible in these images. Boundaries between different constituents of the tissues have disappeared. Therefore, the system has rightly identified these images as malignant. Moreover, malignant images with diverse grades (poorly differentiated, moderately differentiated, and well differentiated) have been shown in Fig. 9 to demonstrate the capability of the system to successfully identify a malignant image regardless of its grade.

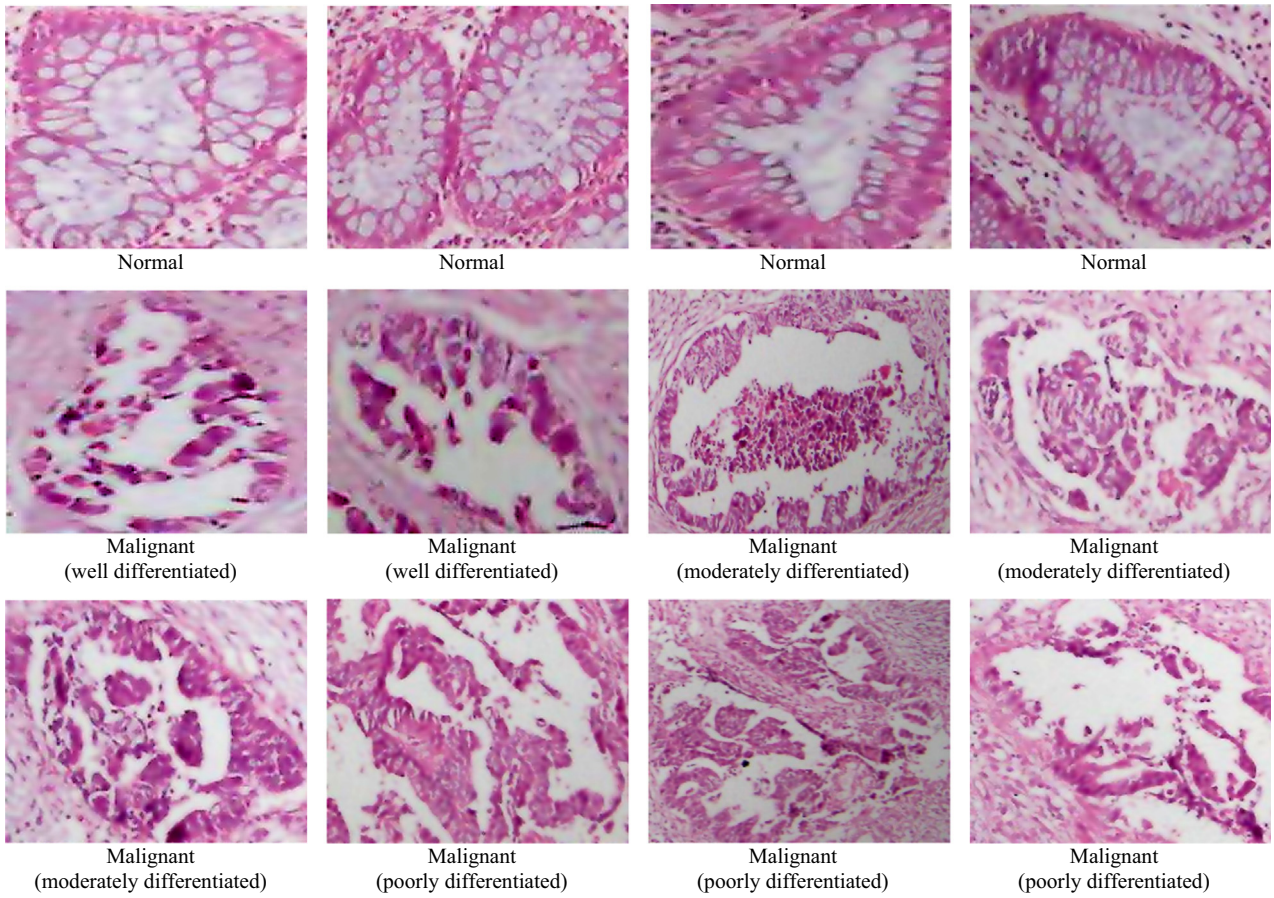

Fig. 9. Examples of the normal, poorly differentiated, moderately differentiated, and well differentiated malignant colon tissues that are correctly classified by the proposed CBIC system.

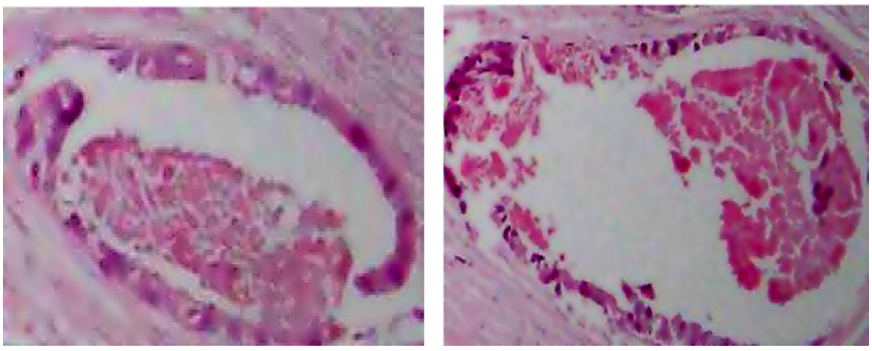

Fig. 10. The two well differentiated malignant colon tissues, which are incorrectly classified as normal.

Out of 174 images, there are only two images which are misclassified by the proposed CBIC system. These images are shown in Fig. 10. These are actually well differentiated malignant colon tissues, which are incorrectly classified as normal by the proposed system. We can observe from the structure of these tissues that they lie at the boundary between normal and well differentiated malignant tissues. In these tissues, gland boundary is preserved, and the overall appearance of these tissues resembles with the normal ones, therefore, the proposed system finds difficulty in identifying such images. It is worthwhile to note that these two images have been acquired from the same biopsy slide. Therefore, we can conclude that only one sample (patient) is not correctly diagnosed among the 68 by the proposed CBIC system.

4.6. Computational time requirements of various feature sets

In this section, we analyze the computational time requirements of various individual and hybrid feature sets. In this context, the CPU time involved in the extraction of features and the classification time taken by different classifiers for each reduced feature set have been measured in seconds. Corresponding results are provided in Table 6.

Feature extraction time has been separately calculated for each image, and the average time involved in the extraction of a single image is shown here. The extraction time for hybrid feature sets has been calculated by adding the extraction time of their individual constituent feature sets. The results in Table 6 reveal

that the average computational time required for the extraction of Haralick-GL, Haralick-RGB, and Haralick-HSV features from a colon biopsy image is 0.037, 0.065 and 0.078 s, respectively, which is very small. Likewise, the average time to compute CCSM and HOG features of a colon biopsy image is 0.05 and 1.283 s, respectively. The average time to compute a hybrid feature set comprising all the three feature types is 1.411 s only which proves that the proposed technique is very fast.

The third column of Table 6 shows the size of reduced feature vectors. The classification time has been separately calculated on reduced feature sets for each classifier, and is shown in column 4, 5 and 6 of Table 6. It is actually the time of applying cross-validation on total dataset to calculate labels of all the samples by a particular classifier. The results reveal that sigmoid classifier consistently takes more time for classification compared to other classifiers. Further, we can conclude from the results that classification time increases with an increase in the size of a feature set.

4.7. Analysis of performance boost up using mRMR

In this section, we present an analysis on the performance enhancement achieved by employing mRMR. The performance has been measured on original feature sets, and then compared with the performance on reduced feature sets (discussed in Sections 4.4 and 4.5). The performance has been analyzed in two respects: classification accuracy and computational time required for classification. Fig. 11 shows the classification accuracy for original and reduced feature sets. It is observed from Fig. 11 that features selected by mRMR enhance the performance both for individual and for hybrid feature sets. This is because mRMR selects features, which bear maximum relevancy to the target labels and share minimum redundancy/overlapping among themselves.

The mRMR method not only boosts the classification performance of different feature sets, but also cuts down the time needed in the classification. Fig. 12 shows the time taken by different classifiers in the classification of original as well as reduced feature sets. Fig. 12 further demonstrates that mRMR has substantially reduced the classification time of all the classifiers.

Table 6
Computational time required for the classification of individual and hybrid feature sets.

Feature sets	Feature extraction time (s)	Size of reduced feature vector	Classification time (s)		
			Linear	RBF	Sigmoid
CCSM	0.050	18	06	18	21
Haralick-GL	0.037	04	02	03	05
Haralick-RGB	0.065	07	03	04	07
Haralick-HSV	0.078	08	04	05	10
HOG	1.283	186	25	42	55
CCSM + Haralick-HSV	0.128	26	09	16	22
CCSM + HOG	1.333	211	51	65	75
HOG + Haralick-HSV	1.361	197	45	53	67
CCSM + HOG + Haralick-HSV	1.411	223	58	69	83

4.8. Performance comparison of CBIC with existing schemes

The performance of the proposed CBIC system has been compared with previously proposed approaches of colon biopsy image classification. In this context, five techniques [11–15] have been selected from the contemporary literature for comparison. We have implemented these techniques in Matlab, and evaluated classification performance measures on the dataset described in Section 4.1. In order to obtain a fair comparison with CBIC, we have used optimal values of the parameters used in these techniques. The selection of optimal values of these parameters is summarized in Table 7. The techniques proposed by Esgiar et al. [11] and Masood et al. [14] employ multiple classifiers. The second column contains the classifiers employed by these techniques, and bold face entry within the second column represents the classifier which yields maximum performance among the employed classifiers by a particular technique. The bold face entries within the third column correspond to the parameters' values where maximum performance is achieved for each classifier by a particular technique.

Table 8 demonstrates the performance comparison of CBIC with these techniques in terms of various quantitative performance evaluation measures. The data samples have been permutated 20 times to get 20 different versions of the dataset. For each combination of the parameter values given in Table 7, classification results have been collected over 20 different permutations of the dataset. Mean rates and standard deviations have been calculated for 20 different results against each combination of the parameter values. Table 8 shows the mean rates and standard deviations for that combination of parameter values where

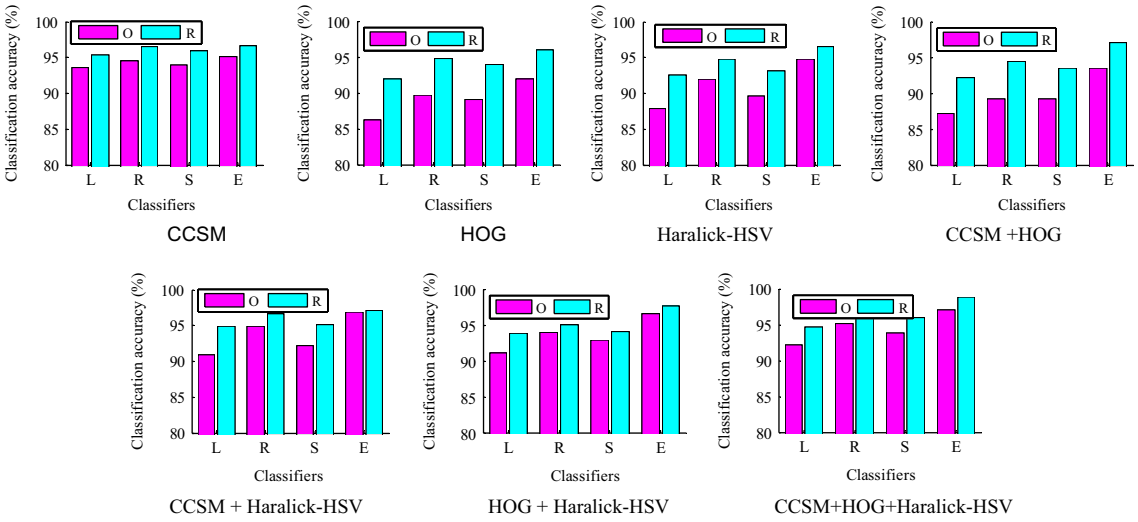

Fig. 11. Performance enhancement in terms of classification accuracy for individual and hybrid feature sets reduced through mRMR (on *x*-axis, *L*, linear; *R*, RBF; *S*, sigmoid; *E*, ensemble, and in legend, *O*, original feature set; *R*, reduced feature set).

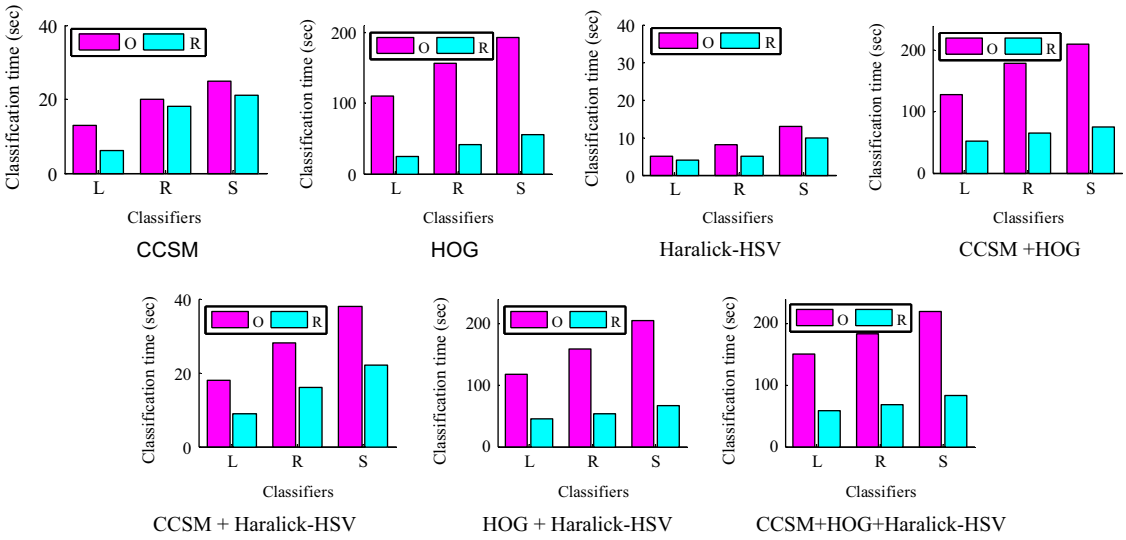

Fig. 12. Performance enhancement in terms of classification time for individual and hybrid feature sets reduced through mRMR (on *x*-axis, *L*, linear; *R*, RBF; *S*, sigmoid, and in legend, *O*, original feature set; *R*, reduced feature set).

Table 7
Optimal parameters used for the evaluation of existing colon cancer detection technique.

Classification techniques	Classifier	Parameters
Esgiar et al. [11]	KNN	Horizontal window size=5 × 5, 7 × 7, 9 × 9, 11 × 11, **13 × 13**,…,19 × 19, 25 × 25 Vertical window size=5 × 5, 7 × 7, 9 × 9, **11 × 11**, 13 × 13,…,19 × 19, 25 × 25 Number of neighbors for KNN classifier=1, **3**, 5, 7, 9, 11
	LDA	Horizontal window size=5 × 5, 7 × 7, **9 × 9**, 11 × 11, 13 × 13,…,19 × 19, 25 × 25 Vertical window size=5 × 5, 7 × 7, 9 × 9, **11 × 11**, 13 × 13,…,19 × 19, 25 × 25
Esgiar et al. [12]	KNN	Number of neighbors for KNN classifier=1, 3, **5**, 7, 9, 11
Masood et al. [13]	SVM	c (cost of polynomial SVM)== 1, 2, 3, 4, 5,…,**86**,…,1000 ε=0.001, 0.002,…,**0.008**,…,2.000 c (cost of Gaussian SVM)== 1, 2, 3, 4, 5,…,**92**,…,1000 γ=0.001, 0.002,…,**0.079**,…,2.000
Masood and Rajpoot [14]	PCA	a (used for compactness)=1, **2**, 3, 4,…,20 r (radius)=2, 3, **4**, 5,…,15 b (number of neighbors)=8, 12, **16**, 20, 24 Principal components of PCA=32
	SVM	a (used for compactness)=1, **2,** 3, 4,…,20 r (radius)=2, 3, 4, **5**,…,15 b (number of neighbors)=8, **12**, 16, 20, 24 c (cost of Gaussian SVM)=1, 2, 3, 4, 5,…46, **47**,…,1000 γ (Gaussian SVM)=0.001, 0.002,…,**0.056**, 0.057,…,2.000
	LDA	a (used for compactness)=1, **2**, 3, 4,…,20 r (radius)=2, 3, **4**, 5,…,15 b (number of neighbors)=8, 12, 16, **20**, 24
Altunbay et al. [15]	SVM	Size of the square structuring element=2, **3**, 4, 5 Area threshold to remove smaller components=15, 20, 25, **30**, 35,…,95, 100 c (cost of linear SVM)=1, 2, 3, 4, 5,…,**99**, 100, 101,…,1000

maximum average classification accuracy is achieved. It is noteworthy that results reported for [11,14] are corresponding to LDA and SVM classifiers since they yield maximum performance as discussed in the previous paragraph.

Our proposed approach has produced highest classification accuracy (98.85%) for colon biopsy dataset, which is 6.90% higher than the highest accuracy yielded by Altunbay et al. [15]. Similarly, there is a significant increase in sensitivity (=0.98), specificity (=1.00), MCC (=0.98) and *F*-score (=0.99) values. The performance of the proposed approach is boosted due to the proposed hybrid feature space and ensemble classification. It can be concluded reasonably that the better performance of CBIC is attributed to the fact that previous techniques consider features which take into account only a certain aspect of input images e.g. image texture. On the other hand, our proposed scheme uses rich hybrid feature vector wherein each individual feature category captures potentially exclusive information about the image. For instance, HOG features capture the variation in the shape of epithelial cells, which are elliptic in normal tissues, but have almost amorphous shape in malignant tissues. Moreover, normal colon tissues have sharp and well-defined boundaries of different constituents, whereas these boundaries are merged in malignant tissues. HOG features quite handily capture the variation in boundary (see Section 4.4.1), therefore, are better able to discriminate normal and malignant colon biopsy images. The Haralick features are commonly used for measuring texture features in an image. The four different Haralick features (correlation, contrast, energy and homogeneity) employed for capturing texture information discriminate the texture of normal and malignant images quite effectively. For example, normal colon tissues have regular structure, whereas this structure is strongly disturbed in malignant images. The textural feature of 'energy' captures this information

Table 8
Performance comparison of the proposed CBIC with existing techniques.

Classification techniques	Accuracy	Sensitivity	Specificity	MCC	*F*-score
Esgiar et al. [11]	86.00 ± 1.32	0.89 ± 0.09	0.95 ± 0.04	0.54 ± 0.07	0.84 ± 0.12
Esgiar et al. [12]	73.18 ± 1.78	0.82 ± 0.02	0.67 ± 0.02	0.48 ± 0.03	0.71 ± 0.02
Masood et al. [13]	87.93 ± 1.52	0.91 ± 0.02	0.83 ± 0.01	0.78 ± 0.02	0.88 ± 0.02
Masood and Rajpoot [14]	89.88 ± 1.32	0.92 ± 0.03	0.88 ± 0.02	0.79 ± 0.03	0.91 ± 0.03
Altunbay et al. [15]	91.95 ± 0.95	0.91 ± 0.02	0.93 ± 0.02	0.84 ± 0.03	0.92 ± 0.01
Proposed CBIC (hybrid features + ensemble classification)	98.85 ± 0.51	0.98 ± 0.01	1.00 ± 0.00	0.98 ± 0.01	0.99 ± 0.00

Table 9
Performance of individual and hybrid feature sets in terms of Kappa statistics.

Feature sets	Linear	RBF	Sigmoid	Ensemble
HOG	0.8394	0.8960	0.8847	0.9310
CCSM	0.9080	0.9308	0.9195	0.9423
Haralick-HSV	0.8508	0.8853	0.8627	0.9308
CCSM + HOG + Haralick-HSV	0.9484	0.9510	0.9498	0.9771

Table 10
P-values for McNemar's exact test and Fisher's exact test.

Feature sets	McNemar's exact test			Fisher's exact test		
	$2-t$	$1-t$	Confidence level (%)	$2-t$	$1-t$	Confidence level (%)
HOG	0.180	0.090	> 90	0.174	0.087	> 90
CCSM	0.289	0.145	> 80	0.283	0.142	> 80
Haralick-HSV	0.289	0.145	> 80	0.283	0.142	> 80

effectively, and performs a major role to discriminate the normal/malignant images. Moreover, combining the texture information from multiple color channels is also a major factor of improving the classification performance. Finally, CCSM features exploit the color information from different channels of the colored colon biopsy images, therefore, yield good classification results. When individual feature vectors are combined in an ensemble paradigm, they all reinforce each other and produce superior classification results.

4.9. Statistical significance analysis of feature sets and classifiers

This section analyzes the significance of the classification results presented in this paper using Kappa statistics. The performance significance has been evaluated from two perspectives. First, we have analyzed the statistical significance of the performance achieved using individual and hybrid features sets. Second, the performance improvement has been discussed from various viewpoints.

The performance of various individual and hybrid feature sets has been shown in Table 9. From the kappa statistics values, we can see that the individual and hybrid feature sets show good performance. The values of all these feature sets exceed the threshold of 0.8, which is supposed to be excellent agreement between the ground truth and the classification results of the corresponding feature set [59]. Further, the hybrid feature set, having kappa statistic of 0.977, shows almost a perfect agreement with the ground truth.

The significance of the performance improvement has been discussed from various perspectives such as the improvement of the proposed scheme compared to previously reported studies, improvement of the proposed ensemble compared to individual classifiers, and improvement of the proposed hybrid feature set compared to individual feature sets.

The values of kappa statistics for Esgiar et al. [11,12], Masood et al. [13], Masood and Rajpoot [14], Altunbay et al. [15], and the proposed CBIC system are 0.839, 0.489, 0.745, 0.804, 0.839, and 0.977, respectively. The kappa statistic value of the proposed ensemble (=0.977) is better compared to the kappa statistics value of Altunbay et al. [15] (=0.839), which performs best amongst previously reported studies. The results in Table 9 reveal better performance of ensemble compared to individual classifiers despite small margin of performance improvement in some cases.

Further, the performance improvement of hybrid feature set (CCSM + HOG + Haralick-HSV) has been investigated by comparing its performance with each of the individual features. Two well-known statistical significance tests, namely McNemar's exact test and Fisher's exact test, have been used. The 1-tailed and 2-tailed *p*-values of these results are given in Table 10. These results indicate that the performance improvement of the hybrid feature set is marginal only. Based on these *p*-values, the level of confidence on the hypothesis "Performance of hybrid feature set is better compared to individual features" is > 90%, > 80%, > 80% for HOG, CCSM and Haralick-HSV features, respectively. The marginal increase in classification performance may be attributed to the small margin of performance improvement that is left after more than 96% classification rate of individual features, and the smaller data size.

5. Conclusion

In this research study, a classification system (CBIC) has been proposed for predicting cancer in colon tissues. In the proposed system, hybrid feature set comprising CCSM, Haralick-HSV, and HOG is constructed. The mRMR method is employed to select discerning features from the hybrid feature set. The discerning features are then used in different SVM kernels based ensemble classification. Working with colon biopsy images, highest classification accuracy of 98.85% and 96.68% has been observed with hybrid and individual feature set (CCSM), respectively. Results prove that the proposed variants of traditional Haralick features and statistical moments are promising feature types for classification of colon biopsy images. Further, results demonstrate that hybrid and rich feature space marginally improves the classification performance compared to the performance achieved by using individual features. Likewise, results verify that the ensemble classifier improves performance compared to individual classifiers. The efficacy of mRMR method has also been validated in terms of reduction in classification time and boost up in classification accuracy. Proposed scheme has also been compared with previously known colon cancer detection techniques, and an increase in classification accuracy is observed. The future directions along this research work are to extend the proposed technique on colon biopsy images captured at multiple magnification factors, and to further classify the malignant images into different cancer grades.

Conflict of interest statement

This paper is the authors' original work and has not been published nor has it been submitted simultaneously elsewhere. All authors have checked the paper and have agreed to the submission. We do not have any financial/personal relationships with other people/organizations which could inappropriately influence this work.

Acknowledgment

This work is supported by the Higher Education Commission of Pakistan (HEC) under indigenous PhD scholarship program as per Award no. 117-7931-Eg7-037. We are thankful to Mr. Imtiaz Ahmad Qureshi (Assistant Professor, Histopathology Department, Rawalpindi Medical College) for providing data and relevant expert opinion. We also appreciate the support provided by Histopathology department of PAEC general hospital, Islamabad, Pakistan for providing imaging equipment.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.compbiomed.2013.12.010.

References

[1] Cancer Facts and Figures, American Cancer Society, ⟨http://www.cancer.org/research/cancerfactsstatistics⟩, October 2013.
[2] Colon Cancer Risk Factors, C.C. Alliance, ⟨http://www.ccalliance.org/colorectal_cancer/riskfactors.html⟩, October 2013.
[3] D. Myers, Colon Cancer Stages: Basics of Each Colon Cancer Stage, ⟨http://coloncancer.about.com/od/stagesandsurvivalrate1/a/ColonCancerStag.htm⟩, October 2013.
[4] G.D. Thomas, M.F. Dixon, N.C. Smeeton, N.S. Williams, Observer variation in the histological grading of rectal carcinoma, J. Clin. Pathol. 36 (1983) 385–391.
[5] A. Andrion, C. Magnani, P.G. Betta, A. Donna, F. Mollo, M. Scelsi, P. Bernardi, M Botta, B. Terracini, Malignant mesothelioma of the pleura: inter observer variability, J. Clin. Pathol. 48 (1995) 856–860.
[6] E.T. Venkatesh, P. Thangaraj, S. Chitra, An improved neural approach for malignant and normal colon tissue classification from oligonucleotide arrays, Eur. J. Sci. Res. 54 (2011) 159–164.
[7] H.S. Shon, G. Sohn, K.S. Jung, S.Y. Kim, E.J. Cha, K.H. Ryu, Gene expression data classification using discrete wavelet transform, in: Proceedings of the International Conference on Bioinformatics & Computational Biology, Las Vegas, Nevada, USA, 2009, pp. 204–208.
[8] X. Li, X. Li, M. Lei, D. Wang, J. Lin, Detection of colon cancer by laser induced fluorescence and raman spectroscopy, in: Proceedings of the 27th International Conference of Engineering in Medicine and Biology Society, Shanghai, China, 2005, pp. 6961–6964.
[9] A. Molckovsky, L.M. Song, M.G. Shim, N.E. Marcon, B.C. Wilson, Diagnostic potential of near-infrared raman spectroscopy in the colon: differentiating adenomatous from hyperplastic polyps, Gastrointest. Endosc. 57 (2003) 396–402.
[10] S. Rathore, M. Hussain, A. Ali, A. Khan, A recent survey on colon cancer detection techniques, IEEE/ACM Trans. Comput. Biol. Bioinf. 10 (2013) 545–563.
[11] A.N. Esgiar, R.N.G. Naguib, B.S. Sharif, M.K. Bennett, A. Murray, Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa, IEEE Trans. Inf. Technol. Biomed. 2 (1998) 197–203.
[12] A.N. Esgiar, R.N.G. Naguib, B.S. Sahrif, M.K. Bennett, Fractal analysis in the detection of colonic cancer images, IEEE Trans. Inf. Technol. Biomed. 6 (2002) 54–58.
[13] K. Masood, N. Rajpoot, H. Qureshi, K. Rajpoot, Co-occurrence and morphological analysis for colon tissue biopsy classification, in: Proceedings of the Fourth International Workshop on Frontiers of Information Technology, Islamabad, Pakistan, 2006, pp. 211–216.
[14] K. Masood, N. Rajpoot, Texture based classification of hyperspectral colon biopsy samples using CLBP, in: Proceedings of the International Symposium on Biomedical Imaging: From Nano to Macro, Boston, MA, USA, 2009, pp. 1011–1014.
[15] D. Altunbay, C.D. Altunbay, C. Cigir, C. Sokmensuer, C.G. Demir, Color graphs for automated cancer diagnosis and grading, IEEE Trans. Biomed. Eng. 57 (2010) 665–674.
[16] A.B. Tosun, M. Kandemir, C. Sokmensuer, C.G. Demir, Object-oriented texture analysis for the unsupervised segmentation of biopsy images, J. Pattern Recogn. 42 (2009) 1104–1112.
[17] E. Ozdemir, C.G. Demir, A hybrid classification model for digital pathology using structural and statistical pattern recognition, IEEE Trans. Med. Imaging 32 (2013) 474–483.
[18] D. Altunbay, C. Cigir, C. Sokmensuer, C.G. Demir, Color graphs for automated cancer diagnosis and grading, IEEE Trans. Biomed. Eng. 57 (2010) 665–674.
[19] C.G. Demir, M. Kandemir, A.B. Tosun, C. Sokmensuer, Automatic segmentation of colon glands using object-graphs, Med. Image Anal. 14 (2010) 01–12.
[20] A.B. Tosun, C.G. Demir, Graph run-length matrices for histopathological image segmentation, IEEE Trans. Med. Imag. 30 (2011) 721–732.
[21] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proceedings of the International Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005, pp. 886–893.
[22] L. Meng, L. Li, S. Mei, W. Wu, Directional entropy feature for human detection, in: Proceedings of the 19th International Conference on Pattern Recognition, Tampa, FL, 2008, pp. 1–4.
[23] K. Lee, C.Y. Choo, H.Q. See, Z.J. Tan, Y. Lee, Human detection using histogram of oriented gradients and human body ratio estimation, in: Proceedings of the Third International Conference on Computer Science and Information Technology, Chengdu, China, 2010, pp. 18–22.
[24] S. Paisitkriangkrai, C. Shen, J. Zhang, Face detection with effective feature extraction, in: Proceedings of the 10th Asian conference on Computer vision, Queenstown, Newzealand, 2010, pp. 460–470.
[25] D. Unay, A. Ekin, Dementia diagnosis using similar and dissimilar retrieval items, in: Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Chicago, IL, 2011, pp. 1889–1892.
[26] L. Song, X. Liu, L. Ma, C. Zhou, X. Zhao, Y. Zhao, Using HOG-LBP features and MMP learning to recognize imaging signs of lung lesions, in: Proceedings of the 25th International Symposium on Computer-Based Medical Systems (CBMS), Rome, Italy, 2012, pp. 1–4.
[27] O. Ludwig, D. Delgado, V. Goncalves, U. Nunes, Trainable classifier-fusion schemes: an application to pedestrian detection, in: Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems, St. Louis, 2009, pp. 432–437.
[28] H.Y. Yang, X.Y. Wang, X.Y. Zhang, J. Bu, Color texture segmentation based on image pixel classification, Eng. Appl. Artif. Intell. 25 (2012) 1656–1669.
[29] A. Sengur, Color texture classification using wavelet transform and neural network ensembles, Arabian J. Sci. Eng. 34 (2009) 491–502.
[30] R.C. Gonzalez, R.E. Woods, Digital Image Processing, 2nd ed., Prentice-Hall, New Jersey, USA, 2002.
[31] R.M. Haralick, Statistical and structural approaches to texture, Proc. IEEE (1979) 784–804.
[32] A.K. Mittra, R. Parekh, Automated detection of skin diseases using texture features, Int. J. Eng. Sci. Technol. 3 (2011) 4801–4808.
[33] M. Hassan, A. Chaudhry, A. Khan, J.Y. Kim, Carotid artery image segmentation using modified spatial fuzzy c-means and ensemble clustering, Comput. Methods Programs Biomed. 108 (2012) 1261–1276.
[34] C.C. Lee, S.H. Chen, Y.C. Chiang, Classification of liver disease from CT images using a support vector machine, J. Adv. Comput. Intell. Intell. Inf. 11 (2007) 396–402.
[35] K. Dhanalakshmi, V. Rajamani, An intelligent mining system for diagnosing medical images using combined texture-histogram features, Int. J. Imag. Syst. Technol. 23 (2013) 194–203.
[36] S. Jafarpour, Z. Sedghi, M.C. Amirani, A robust brain MRI classification with GLCM features, Int. J. Comput. Appl. 37 (2012) 1–5.
[37] N. Zulpe, V. Pawar, GLCM textural features for brain tumor classification, Int. J. Comput. Sci. Issues 9 (2012) 354–359.
[38] D. Mitrea, M. Socaciu, R. Badea, A. Golea, Texture based characterization and automatic diagnosis of the abdominal tumors from ultrasound images using third order GLCM features, in: Proceedings of the Fourth International Congress on Image and Signal Processing, 2011, pp. 1558–1562.
[39] C.E. Mclaren, W.P. Chen, K. Nie, M.Y. Su, Prediction of malignant breast leisons from MRI features: a comparison of artificial neural network and logistic regression techniques, Acad. Radiol. 16 (2009) 842–851.
[40] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and minredundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 1226–1238.
[41] C. Yun, J. Yang, Experimental comparison of feature subset selection methods, in: Proceedings of the Seventh International Conference on Data Mining, Omaha, NE, 2007, pp. 367–372.
[42] K.C. Chou, Some remarks on protein attribute prediction and pseudo amino acid composition, J. Theor. Biol. 273 (2011) 236–247.
[43] L. Nanni, A. Lumini, D. Gupta, A. Garg, Identifying bacterial virulent proteins by fusing a set of classifiers based on variants of Chou's pseudo amino acid composition and on evolutionary information, IEEE/ACM Trans. Comput. Biol. Bioinf. 9 (2012) 467–475.
[44] S.S. Sahu, G. Panda, A novel feature representation method based on Chou's pseudo amino acid composition for protein structural class prediction, Comput. Biol. Chem. 34 (2010) 320–327.
[45] V.N. Vapnik, Statistical Learning Theory, 1st ed., Wiley-Interscience, New York, USA, 1998 (September).
[46] E.D. Ubeyli, Comparison of different classification algorithms in clinical decision-making, Expert Syst. 24 (2007) 17–31.

[47] A. Jayachandran, R. Dhanasekaran, Automatic detection of brain tumor in magnetic resonance images using multi-texton histogram and support vector machine, Int. J. Imag. Syst. Technol. 23 (2013) 97–103.
[48] E.L. Torre, B. Caputo, T. Tommasi, Learning methods for melanoma recognition, Int. J. Imag. Syst. Technol. 20 (2010) 316–322.
[49] W. Chmielnicki, I. Roterman-Konieczna, K. Stąpor, An improved protein fold recognition with support vector machines, Expert Syst. 29 (2012) 200–211.
[50] S. Osowski, K. Siwek, R. Siroic, Neural system for heartbeats recognition using genetically integrated ensemble of classifiersComput. Biol. Med. 41 (2011).
[51] W.B. Sampaioa, E.M. Diniza, A.C. Silvaa, A.C. Paivaa, M. Gattassb, Detection of masses in mammogram images using CNN, geostatistic functions and SVM, Comput. Biol. Med. 41 (2011) 653–664.
[52] A. Subasi, Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders, Comput. Biol. Med. 43 (2013) 576–586.
[53] S. Theodoridis, K. Koutroumbas, Pattern Recognition, 4th ed., Academic Press, Massachusetts, USA, 2008.
[54] M. Kurzynski, M. Wozniak, Combining classifiers under probabilistic models: experimental comparative analysis of methods, Expert Syst. 29 (2012) 374–393.
[55] M. Tong, K.H. Liu, C. Xu, W. Ju, An ensemble of SVM classifiers based on gene pairs, Comput. Biol. Med. 43 (2013) 729–737.
[56] J. Kittler, F. Roli, Multiple classifier systems, in: Proceedings of the Second International Workshop on Multi Classifiers System, Cambridge, UK, 2001, pp. 369–377.
[57] I.H. Witten, E. Frank, M.A. Hall, Data mining: Practical Machine Learning Tools and Techniques, 2nd ed., Morgan Kaufmann Publishers, London, 2005.
[58] B.W. Matthews, Comparison of the predicted and observed secondary structure of T4 phase lysozyme, Biochem. Biophys. Acta 405 (1975) 442–451.
[59] J. Carletta, Assessing agreement on classification tasks: the kappa statistic, Comput. Linguist. 22 (1996) 249–254.
[60] C.W. Hsu, C.C. Chang, C.J. Lin, A practical Guide to Support Vector Machines, Department of Computer Science & Information Engineering, National Taiwan University, Taiwan, 2003.