Machine Vision and Applications (2021) 32:58
https://doi.org/10.1007/s00138-021-01184-8

ORIGINAL PAPER

Design of a hybrid deep learning system for discriminating between low- and high-grade colorectal cancer lesions, using microscopy images of IHC stained for AIB1 expression biopsy material

Angeliki Theodosi[1] · Sotiris Ouzounis[2] · Spiros Kostopoulos[2] · Dimitris Glotsos[2] · Ioannis Kalatzis[2] · Vassiliki Tzelepi[3] · Panagiota Ravazoula[3] · Pantelis Asvestas[2] · Dionisis Cavouras[2] · George Sakellaropoulos[1]

Received: 19 July 2019 / Revised: 26 January 2021 / Accepted: 12 February 2021 / Published online: 11 March 2021

Abstract

To design a hybrid deep learning system (hDL-system) for discriminating low-grade from high-grade colorectal cancer (CRC) lesions, using immunohistochemically stained biopsy specimens for AIB1 expression. AIB1 has oncogenic function in tumour genesis, and it is an important prognostic factor regarding various types of cancers, including CRC. Clinical material consisted of biopsy specimens of sixty-seven patients with verified CRC (26 low-grade, 41 high-grade cases). From each patient, we digitized images, at × 50 and × 200 lens magnifications. We designed the hDL-system, employing the VGG16 pre-trained convolution neural network for generating DL-features, the SVM classifier, and the bootstrap evaluation method for assessing the discrimination accuracy between low-grade and high-grade CRC lesions. Furthermore, we compared the hDL-system's discrimination accuracy with that of a supervised machine learning system (sML-system). We designed the sML-system by (i) generating sixty-nine (69) textural and colour features from each image, (ii) employing the probabilistic neural network (PNN) classifier, and (iii) using the bootstrapping method for evaluating sML-system performance. The system design was enabled by employing the CUDA platform for programming in parallel the multiprocessors of the Nvidia graphics processing unit card. The hDL-system provided the highest discrimination accuracy of 99.1% using the × 200 lens magnification images as compared to the 92.5.% best accuracy achieved by the sML-system, employing both the × 50 and × 200 lens magnification images. Our results showed that the hDL-system was superior to the sML-system (i) in discriminating low-grade from high-grade CRC-lesions and (ii) by requiring fewer images for its best design, only those at the × 200 lens magnification. The sML-system by employing textural and colour features in its design revealed that high-grade CRC lesions are characterized by (i) loss in the definition of structures, (ii) coarser texture in larger structures, (iii) hazy formless texture, (iv) lower AIB1 uptake, (v) lower local correlation and (vi) slower varying image contrast.

Keywords Machine learning · Deep learning · Colorectal carcinoma · Immunohistochemistry

List of medical and technical abbreviations

AIB1	Amplified in breast cancer 1
CART	Classification and Regression Tree
CNN	Convolutional Neural Network
CRC	Colorectal cancer
DL	Deep Learning
DSS	Decision support system
H&E	Haematoxylin & Eosin
HG_CRC	High-grade Colorectal Lesion
IHC	Immunohistochemistry
KNN	K-Nearest Neighbour
LG_CRC	Low-grade Colorectal Cancer
LDA	Linear Discriminant Analysis
PNN	Probabilistic Neural Network
RF	Random Forest
SVM	Support Vector Machines

✉ Spiros Kostopoulos
skostopoulos@uniwa.gr

[1] Department of Medical Physics, School of Health Sciences, Faculty of Medicine, University of Patras, Rio, Patras, Greece

[2] Medical Image and Signal Processing Laboratory, Department of Biomedical Engineering, University of West Attica, Athens, Greece

[3] Department Pathology, University Hospital of Patras, Rio, Patras, Greece

hDL-system	Hybrid Deep Learning System
sML-system	Supervised Machine Learning System

List of feature abbreviations

CLR MedAD_Ch_a_x50	Median Absolute Deviation of colour channel a
CLR MedV_Ch_a _x200	Min Value of colour channel a
CLR MinV_Ch_b_x500	Min Value of colour channel b
CLR RV_Ch_a_x500	Range Values of colour channel a
CON _x200	Contrast
DENT_x200	Difference entropy
DVAR_x200	Difference variance
ICM1_x200	Information measure of correlation 1
IDM _x200	Inverse different moment
Kurt_x500	Kurtosis
LRE _x200	Long run emphasis
RLNU_x50	Run length non-uniformity
RP_x200	Run percentage
SAV_x200	Sum average
Skew_x200	Skewness
dwt2D_MedV_x50	Median Value of W in diagonal direction (*where W is the 2nd level discrete wavelet transform*):
dwt2H_MedV_x200	Median Value of W in horizontal direction (*where W is the 2nd level discrete wavelet transform*)

1 Introduction

Colorectal cancer (CRC) starts in the colon or rectum of the large intestine, it is lethal, and it affects both men and women. CRC is the third most frequent cancer, and its early detection has been shown to reduce considerably patient mortality [1]. CRC diagnosis may be performed by (a) various diagnostic methods such as colonoscopy, computed tomography (CT), magnetic resonance imaging (MRI), and (b) by spectroscopy, molecular, and histopathology examinations [2].

Final diagnosis is mainly based on the outcome of the histopathology examination. The latter comprises examining under the microscope tissue specimens, resected in colonoscopy and suitably treated and stained with Haematoxylin and Eosin (H&E). Pathologists characterize tissue specimens as normal, benign, or malignant. Diagnosis is based on the textural, structural, and morphology characteristics of the specimen's glandular tissue. In malignant tissues, cancerous glands appear irregular in shape and size, and with loss of complexity. The latter is characterized by luminal bridging and cribriform change. Also, the epithelial malignant cells are characterized by a severe cytologic abnormality with nuclear irregularity and loss of polarity, size variation, increased hyperchromasia and atypical mitotic figures. Malignant tissues are categorized by pathologists into stage and grade. Regarding CRC lesions, there are five stages and four grades. Staging is related to the invasion and metastatic spread of the tumour to other parts such as to regional lymph nodes, liver and lungs. Grading is related to glandular appearance, and it may provide valuable prognostic indication regarding tumour aggressiveness and ultimately patient survival [3–5]. However, histopathology examination is influenced by the examining physician's experience and clinical workload, leading to inter- and intra-observer variability in diagnosis [6–9]. Due to this diagnostic inconsistency, a two-tier CRC classification system has been proposed [5–10], low grade–high grade, for reducing observer diagnostic variability while retaining prognostic value.

Furthermore, to deal with inter- and intra-observer variability, a number of research studies [11–29] have been published proposing computer-based decision support systems (DSS) for characterizing different types of resected colorectal lesions. Those studies have proposed DSS systems for discriminating between normal, benign, and malignant CRC tissues, employing digital images from H&E stained biopsy specimens. The authors employed various classification algorithms such as SVM [18, 20, 21, 23, 24, 26, 28], KNN [16, 23, 26], LDA [12, 15, 18, 25], Naïve Bayes [25], decision trees [25], random forest [8] in the design of the DSS systems. They have also employed in the design textural, morphological and structural features, which were computed from the digital images of the colorectal lesion. DSS classification accuracies varied depending on the colorectal tissue types that those studies were designed to discriminate. A recent comprehensive review on deep learning in colon cancer is given in [30]. Specifically for histopathology, Shaban et al. [31] and Zhou et al. [32] have employed deep learning methods for discriminating between normal-low-grade and high-grade CRC lesions from public available H&E stained histopathology images.

AIB1 (amplified in breast cancer 1) is an immunohistochemical (IHC) dye which is known for its significant prognostic value in various types of cancers, such as in breast, bladder, lung, naso-pharynx, CNS, colon, esophagus, liver, pancreas, stomach, ovaries [33–42]. Concerning the colon, AIB1 staining has been found overexpressed in the nuclei of CRC tissues [34–38, 40], as compared to the nuclei of non-malignant tissues. In the present study, we have employed AIB1 stained biopsy material from verified colorectal cancer lesions, selected by two expert pathologists (P.R & V.T) from the archives of the Department of Pathology, University Hospital of Patras, Greece. The diagnosis of biopsy

specimens had been performed by the two physicians on H&E stained material.

The contribution of the present study is to design a decision support system, based on deep learning (DL) methods, employing digital images of AIB1 stained biopsy material of verified CRC lesions, with the purpose to discriminate with high accuracy low-grade CRC (LG_CRC) from high-grade CRC (HG_CRC) lesions. Because of the AIB1′s oncogenic prognostic value, we considered that the design of the DL-system would facilitate the discrimination between LG_CRC and HG_CRC lesions. To our knowledge, the use of AIB1 stained material in the design of decision support systems for CRC discrimination has not been published before. Our work was influenced by how the expert pathologists examine biopsy specimens, i.e. first inspecting under the microscope images at × 50 magnification and then studying in more detail microscope images at × 200 or higher magnification. Accordingly, we designed our deep learning system by experimentation with image-ROIs at different magnifications from each case, unlike previous studies. We designed a hybrid deep learning (hDL) system by employing a pre-trained convolution neural network for extracting DL features, and we used those features to feed a classifier for classifying CRC lesions. Furthermore, we compared the precision of our hDL-system against the best precision we could achieve by a supervised machine learning (sML) system, designed by computing a large number of textural and colour features and experimenting with various classifiers. The design of the sML-system revealed important textural and colour properties that change with the severity of the disease and differentiate between low-grade and high-grade CRC lesions. To deal with the required enormous processing time workload, we employed GPU and CUDA technologies for parallel processing on the Nvidia's graphics card processors.

2 Material and methods

2.1 Clinical material

Two experienced pathologists collected biopsy material retrospectively from the archives of the Department of Pathology of the University Hospital of Patras, Greece. The material comprised sixty-seven (67) patients. Twenty-six (26) patients had been diagnosed with grade I colorectal cancer, twenty-eight (28) patients with grade II, and thirteen (13) with grade III. For the purposes of the current research, cases were split into two classes, the low-grade CRC (26 cases of grade I) and the 41 high-grade CRC (28 grade II and 13 grade III cases). The pathologists assessed tumour grade by employing H&E staining and by examining the prepared biopsy material under the microscope. Additionally, the pathologists used IHC staining for estimating the AIB1 expression on the biopsy material. The IHC-staining was employed in order to relate textural and colour changes in the glandular tissue, inflicted by cancer, to the histological grade of the tumour. Finally, the pathologists marked regions of interest on the substrates of the AIB1 specimens, indicative of the disease's impact, for further processing. The pathologists followed the guidelines of the American Joint Committee on Cancer (AJCC) [43] for tumour grading and staging and the guidelines of the Declaration of Helsinki and of the ethics committee of the University of Patras, Greece.

2.2 Image acquisition

Original images were digitized at 2592 × 1944x24 bits resolution by means of a Leica DM2500 light microscope fitted with a Leica DFC420C colour digital camera and connected to a dedicated desktop computer. Images were digitized in two magnifications of × 50 and × 200 so as to be used in the design of the system. From each original image, a ROI was extracted at 512 × 512 resolution containing the gland following the pathologist guidance (Fig. 1).

2.3 Computer processing

(1) Design of the Hybrid Deep Learning system

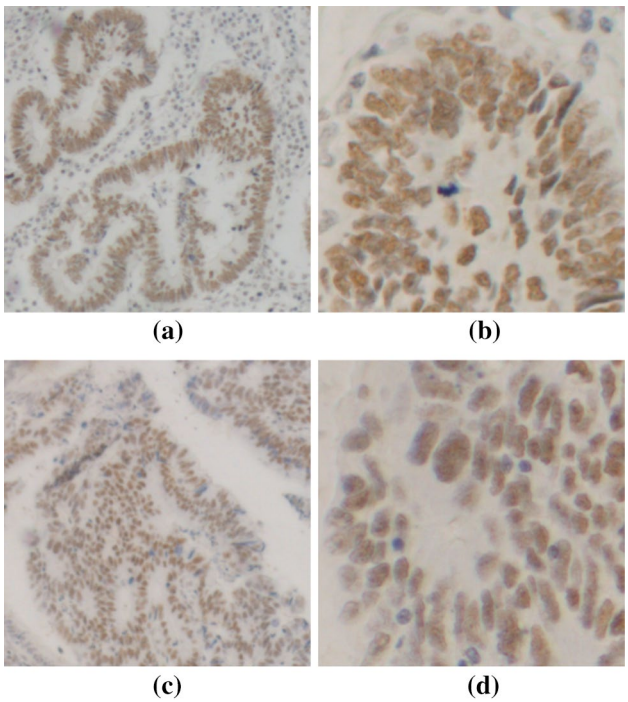


Fig. 1 Digitized Regions of Interest (ROIs) of Colorectal Cancer (CRC), IHC stained for AIB1 expression: Low-grade CRC at **a** × 50 and **b** × 200 lens magnification and high-grade CRC at **c** × 50 and **d** × 200 lens magnification

The hybrid DL-system design was based on a hybrid model design [44], whereby we used the pre-trained VGG16 convolution neural network (CNN) [45] for extracting DL-features from a pooling layer and we used those features to design the DL-system for classifying the CRC lesions. Figure 2 demonstrates the end-to-end deep learning pipeline.

First, we resized our ROIs to 224 × 224 pixels to comply with the input specifications of the VGG16 network. We collected the deep learning features from the fourth pooling layer of the CNN network, since at that layer the hDL-system provided the highest classification accuracy. In more detail, at the fourth CNN pooling layer, each input image was reformatted to 512 images of 14 × 14 pixels each. The DL-features are the mean RGB pixel values of the 512 images, which amounts to a total of 100,352 (512 × 14 × 14) features values per input image. Thus, we formed two class-matrices, one for low-grade and one for high-grade lesions. The class-matrix rows referred to the samples and the columns the DL-features. Next, we reduced the number of DL-features (a) by discarding features of low statistical significance differences between the two classes and (b) by applying PCA reduction. Then, we employed the Recursive Feature Elimination (RFE) in conjunction with the Random Forest classifier [46] for selecting the best features for classification. According to RFE results, two (2) PCA features, (PCA1 and PCA2), were the best choice for best classification at 50 × lens magnification, four (4) features (PC1, PC8, PC9, PC17) at 200 × lens magnification, and two (2) PCA features (PCA1 and PCA2) at mixed (50 × and 200 ×) magnifications. Finally, we proceeded with the design of the hybrid DL system, by (a) using the DL-emanating PCA features, (b) experimenting with different classifiers, and (c) estimating the classification performance by means of the 10-epoch bootstrap performance evaluation method. The classifiers tested are available in the caret (classification and regression training) package and were the following: Linear Discriminant Analysis (LDA), Classification and Regression Tree (CART), k-Nearest Neighbour (kNN), Support Vector Machines (SVM), and Random Forest (RF) Decision Trees. The specifications of the workstation were CPU: Intel i7-3770 K, 3.50 GHz, GPU: EVGA GeForce GTX 980 4 GB, and RAM: 16 GB.

Furthermore, we designed a supervised machine learning system, by extracting a large number of textural and colour features from the CRC-images and employing a classifier. Our purpose was to compare the precision of the hDL-system against that of the sML-system and, additionally, to identify the texture and colour CRC-image properties that change with the severity of the disease and differentiate the low-grade from the high-grade CRC cases.

(2) Design of the supervised machine learning system

To design the sML-system, we considered the following computer processing stages: i) extraction from each selected ROI of a large number of features that evaluate image properties, ii) experimentation with existing classifier algorithms for choosing a high-performance classifier suitable for our data, iii) selection of information rich features to be used in the design of a high-performance SML-system, iv) evaluation of sML-system performance for assessing the SML-system's precision to new, "unseen" by the sML-system, data, 5) employment of GPU parallel processing technologies for

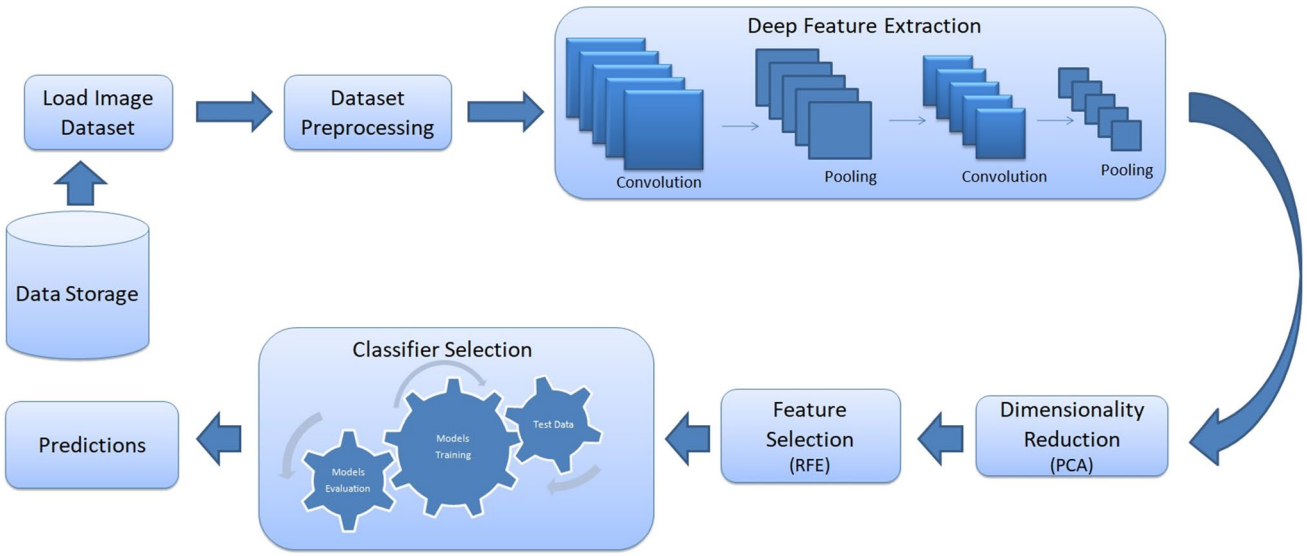


Fig. 2 Schematic pipeline of the hybrid DL-system design

rendering plausible the design of a high performance sML-system. These five steps are analysed below.

(i) Feature extraction. We computed sixty-nine (69) features quantifying textural and colour properties of the AIB1 images [42]. We used the gray-scale version of the digitized colour images for the calculation of fifty three (53) textural features: Four (4) features (*Mean Value,Standard Deviation, Skewness, Kurtosis*) were computed from the histogram of the gray-scale ROI, 14 features (*Angular Second Moment, Contrast, Inverse Different Moment, Entropy, Correlation, Sum Of Squares, Sum Average, Sum Entropy, Sum Variance, Difference Variance, Difference Entropy, Information Measure Of Correlation 1, Information Measure Of Correlation 2, Autocorrelation*) were computed from the co-occurrence matrix [47], 5 features (*Short Run Emphasis, Long Run Emphasis, Gray Level Non-Uniformity, Run Length Non-Uniformity, Run Percentage*) were calculated from the run length matrix [48], 24 features (*Mean Value of wavelet transform (W) in horizontal direction, Median Value of W in horizontal direction, Max Value of W in horizontal direction, Min Value of W in horizontal direction, Range Values of W in horizontal direction, Standard Deviation of W in horizontal direction, Median Absolute Deviation of W in horizontal direction, Mean Absolute Deviation of W in horizontal direction, Mean Value of W in diagonal direction, Median Value of W in diagonal direction, Max Value of W in diagonal direction, Min Value of W in diagonal direction, Range Values of W in diagonal direction, Standard Deviation of W in diagonal direction, Median Absolute Deviation of W in diagonal direction, Mean Absolute Deviation of W in diagonal direction, Mean Value of W in vertical direction, Median Value of W in vertical direction, Max Value of W in vertical direction, Min Value of W in vertical direction, Range Values of W in vertical direction, Standard Deviation of W in vertical direction, Median Absolute Deviation of W in vertical direction, Mean Absolute Deviation of W in vertical direction*) were computed from the 2^{nd} level, two-dimensional discrete wavelet transform using Daubechies wavelets (*W*), and 6 features (*Tamura Coarseness 1, Tamura Coarseness 2, Tamura Coarseness 3, Tamura Coarseness 4, Tamura Contrast, Tamura Roughness*) were from the Tamura method [49]. We also employed the colour images of the digitized ROIs to calculate 16 colour features *(Mean Value of colour channel a, Median Value of colour channel a, Max Value of colour channel a, Min Value of colour channel a, Range Values of colour channel, Standard Deviation of colour channel a, Median Absolute Deviation of colour channel a, Mean Absolute Deviation of colour channel a, Mean Value of colour channel b, Median Value of colour channel b, Max Value of colour channel b, Min Value of colour channel b, Range Values of colour channel b, Standard Deviation of colour channel b, Median Absolute Deviation of colour channel b, Mean Absolute Deviation of colour channel b*) derived from the L*a*b* version of the RGB colour image. Thus, from each ROI we computed 69 features at × 50 magnification and 69 features at × 200 magnification. Finally, from each case, we computed the mean of each feature from all image ROIs of the case and at both × 50 and × 200 magnifications. Thus, we represented each patient with one 69-features vector at × 50 magnification and one 69-features vector at × 200 magnification. We formed two datasets, one at each magnification, and we designed the SML-system with each data set separately and in combination. The later resembles the way that pathologists examine biopsy material under the microscope, first at × 50 then at × 200 and they intergrade information to reach final diagnosis. All features in both datasets were normalized to zero mean and unit standard deviation to avoid bias due to differences in values in different features. Normalization was performed by relation (1):

$$x' = \frac{(x - \text{mean})}{sd} \tag{1}$$

where x is a feature value, mean and sd are the mean and standard deviation of all values of the specific feature in both classes (LG_CRC and HG_CRC). We normalized the features of each dataset (× 50 and × 200) separately. When we combined both data sets in the design of the SML-system, same features belonging to different magnifications were considered as different features.

Finally, we subjected the features used in the best SML-system design to the Wilcoxon statistical test, for finding differences between the two classes (LG_CRC and HG_CRC). This would reveal important information as to textural and colour alteration caused by advancing grading of the disease.

(ii) Selection of Classification algorithm. The choice of a classifier amongst readily available algorithms is a crucial task in the design process of the SML-system. The latter involves finding the best feature combination that will be used in the classifier to design a high precision SML-system. However, searching amongst a large number of features is a highly demanding task in terms of computer processing time. Optimal system design comprises forming combinations of features and evaluating the performance of the classifier in discriminating between low- and high-grade CRC lesions, with the aim to find the highest classification accuracy with the smallest number of features. This is an enormous task if one considers the number of available features (69 features at magnifications × 50, × 200, × 50 + × 200), the number of cases involved (67 patients), the number of possible feature combinations, the SML-system evaluation method used (see Sect. 4), and the number of classifiers to be examined. This influences the choice of a classification algorithm in that it has to be characterized by properties such as

low complexity, fast execution, and high precision with the available type of data. We put to the test a number of popular classifiers, Bayesian, LDA, KNN, SVM, ANN, PNN [50–52]. Most of these classifiers are readily available in the MATLAB software. We also employed parallel processing methods (see Sect. 5) for speeding up the process of optimum SML-system design. Of the classifiers tested, the PNN proved most efficient in terms of classification accuracy (equally scoring as the SVM), complexity (involving no optimization procedures), and processing time. The PNN algorithm is described below:

The discriminant function of the PNN, $d_j(X)$, of class j is given by the following relation:

$$d_j(x) = k \times \sum_{i=1}^{N_j} e^{\frac{-\left\|x-p_{j,i}\right\|}{2\sigma^2}}$$

where

$$k = \frac{1}{(2\pi)^{\frac{N_f}{2}} \times \sigma^{N_f} \times N_j} \tag{2}$$

where, x is the pattern vector to be classified, N_j is the number of patterns in class j, $p_{j,i}$ is the i-th training pattern vector of class j, σ (set at σ=*0.22*) is a smoothing parameter. N_f is the number of features employed in the feature vector. The input vector x is classified to the class j with the higher discriminant value, $d_j(x)$.

(iii) Methods for selection of features. In each one of the 3 datasets (×50, ×200, and ×50-×200 combined), we ranked features following a score criterion that combined the Wilcoxon statistical test, which indicates each feature's between-class separation ability, and the feature's correlation with the rest of the features of the dataset, which relies on the principle that least correlated features combined are of higher discriminatory power [51, 52]. From the ranked features of each dataset, we picked the top 30% and we subjected them into the exhaustive search procedure, i.e. forming all possible combinations of 2, 3, 4, etc. features to design the sML-system and evaluating each sML-design's accuracy by the leave-one-out method. This process was repeated for each data set and for each classifier, employing parallel processing procedures (see later section). We followed an "engineering rule of thumb" regarding the largest number of features allowed in one combination. That rule states that to avoid overfitting and thus overestimation of the SML-system's performance, the number of features involved in a single combination should not exceed 1/3 of the number of cases in the smallest class [53]. In our case, the smallest class comprised 26 cases of low-grade CRC, so the largest number of features involved in any combination should not exceed 8 features.

(iv) sML-system performance evaluation. In assessing the design of the SML-system, we used two evaluation methods. The leave-one-out (LOO) method and the Bootstrap method [54]. According to the LOO method, the sML-system is designed by all but one case, and the left out case is used as input to be classified by the sML-system. The left-out case is re-introduced into the design data set and another case is excluded and used to be classified by the SML-system. This cycle is repeated until all cases are classified to one of two classes (LG_CRC, HG_CRC). Finally, the results are presented in a truth table to indicate the number of correctly and wrongly classified by the SML-system cases.

According to the bootstrap evaluation method, a subset of say 60% of the total number of cases is randomly formed with re-substitution, meaning that one case may be included in the so formed sample more than once. We used this subset to design the sML-system by means of the LOO evaluation method and the exhaustive search feature-selection method. We next used the best sML-system design, the one providing the best discrimination accuracy with the least number of features, to classify the cases that were not included in the subset. The classification accuracy of the SML-system on the left-out cases was noted. This procedure (sample-subset formation, sML-system design, classification of left-out cases) was repeated for an adequate number of times and the average sML-system precision (overall accuracy, sensitivity, specificity) of the multiple trials was calculated. This average precision is indicative of how the sML-system would perform when presented with new data at its input, as would be the case of testing the performance of the sML-system in a clinical environment.

(v) Parallel processing implementation. Taking into consideration the enormous number of calculations and tasks involved in the design of the sML-system, (a) we employed the inherent parallel processing capabilities of the MATLAB software on the 4-core CPU (central processing unit) of the desktop computer for experimenting with the choice of the optimum classifier algorithm and (b) having chosen the best classification algorithm for building the sML-system, we transferred the design of the sML-system to the multiprocessors (13 multiprocessors of 192 cores each) of the Nvidia Tesla K20c Graphics Processing Unit (GPU) card, housed in the desktop computer, using the programming environment of the CUDA (Compute Unified Device Architecture) toolkit v4.0 and the C/C++programming language. For parallel processing, we divided the sML-design into small tasks, which were then executed in parallel on the different processor cores. One such task was the use of features combination to design the sML-system and then testing its accuracy by means of the LOO evaluation method. Regarding the CPU-based parallel processing, we used the *parfor* property of MATLAB for parallel processing on the cores of the desktop's CPU. For the GPU-based parallel processing, we

loaded each task on a different GPU-thread to be executed concurrently with a large number of similar tasks on the different cores of the GPU-multiprocessors. Similar sML-system design has been previously employed by our group and is described in detail in [55].

3 Results

The classification accuracies obtained by the hybrid DL-system, using different classifiers and the 10-epoch bootstrap evaluation method, are shown in Table 1. Classifications were obtained with images recorded at lens magnifications, 50x, 200x, and using mixed 50 × and 200 × images. Best classification accuracies were achieved by the SVM classifier.

Table 2 shows the classification accuracies achieved by employing the patients' ROIs at × 50 magnification (1st row of Table 2). The third and fourth columns indicate the accuracies of the sML-system in classifying correctly the LG_CRC (specificity) and HG_CRC (sensitivity) cases, the fifth column shows the overall accuracy achieved by the sML-system in correctly discriminating LG_CRC from HG_CRC cases and the last column contains the best features combination used for the sML-system to accomplish the discrimination. At × 50 magnification, the sML-system classified correctly 65.4% (17/26) of the LG_CRC cases, 90.2% (37/41) of the HG_CRC cases, and discriminated correctly 80.6% (54/67) between LG_CRC and HG_CRC cases. The best features combination comprised 7 features (*RLNU_x50, dwt2D_MedV_x50, Tamura contrast_x50, Tamura roughness_x50, CLR RV_Ch_a_x50, CLR MedAD_Ch_a_x50, CLR MinV_Ch_b_x50*). Figure 3 depicts overall discrimination accuracies at × 50, × 200, and × 50 & × 200 lens magnifications by means of three ROC curves. Regarding the accuracy achieved at × 50 lens magnification, the area under the curve (AUC) of the red ROC curve was 0.82, which measures in-between class separation, and the Cohen–Kappa statistic was 0.58, which indicates the validity of the result. The cut-off value for the AUC curve was set at 0.5. Figure 4 shows the variation of the overall accuracy at × 50 lens magnification (red/dashed line) with the number of features involved in achieving the highest classification accuracy at × 50 magnification.

Table 2 also shows the classification accuracies of the sML-system in discriminating between LG_CRC and HG_CRC cases at × 200 magnification (2nd row of Table 1). The sML-system specificity was 80.1% (21/26 cases), its sensitivity was 87.8% (36/41) and the overall accuracy was 85.1% (57/67). The best features combination in achieving highest overall sML-system accuracy comprised 8 features (*Skew_x200, CON _x200, IDM _x200, LRE _x200, RP_x200, Tamura coarseness 1_x200, Tamura coarseness 3_x200, Tamura coarseness 4_x200*). Figure 3 presents the overall discrimination accuracy at the × 200 magnification by means of the green ROC curve (AUC = 0.88, Cohen-Kappa = 0.65). Figure 4 shows (green/dotted line) the variation of the overall sML-system accuracy with the number of features, using × 200 lens magnification images.

Table 2, third row, shows the classification accuracies reached by combining features extracted from images captured at two lens magnifications, × 50 and × 200. The sML-system achieved the highest accuracies employing an 8-features combination that comprised one feature from the × 50 magnification and seven features from the × 200 magnification. The best 8-features combination comprised the following features: *Kurt_x50, SAV_x200, DVAR_x200, DENT_x200, ICM1_x200, dwt2H_MedV_x200, Tamura coarseness 4_x200, CLR MedV_Ch_a _x200*. The sML-system's specificity was 92.3% (24/26), its sensitivity was 92.7% (38/41) and the overall accuracy was 92.5% (62/67). The sML-system achieved a similar overall accuracy of

Table 1 Results of the hybrid DL-system employing different classifiers, two PCA-features and the 10-epoch bootstrap method

10 Epochs	LDA		CART		KNN		SVM		RF	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Lens magnification 50x *(DL-features used PC1, PC2)*										
Accuracy (%)	92.9	5.7	88.7	3.3	94.7	4.9	95.6	4.4	91.9	3.6
Sensitivity (%)	89.6	7.9	90.6	6.5	92.3	6.9	93.6	6.2	91.5	6.4
Specificity (%)	100	0	90.9	5.2	99.3	2.1	99.3	2.1	95.7	4.9
Lens magnification 200X *(DL-features used PC1, PC8, PC9, PC17)*										
Accuracy (%)	95.8	5.9	95.5	5.8	97.6	3.9	99.1	2.9	95.5	5.8
Sensitivity (%)	94	7.4	96.7	7	100	0	98.8	3.7	96.7	7
Specificity (%)	98.7	2.8	96.3	7.9	96.7	5.6	100	0	96.3	7.9
Lens magnification (50x + 200x) mixed images *(DL-features used PC1, PC2)*										
Accuracy (%)	91.6	5	86.9	6.3	92.2	2.8	94	3	91.7	3.9
Sensitivity (%)	89.7	7.3	90.2	7.7	91.1	5.5	94.3	4.4	93.3	5.4
Specificity (%)	98.4	2.2	89.8	8.5	97.4	2.4	96.5	3.6	93.8	4.3

Table 2 Partial and overall classification accuracies achieved by the sML-system at different camera lens magnifications, ×50 and ×200

	Camera Lens Magnification	LG_CRC (%) (specificities)	HG_CRC (%) (sensitivities)	Overall accuracies (%)	Best feature combinations
1	×50	65.4% (17/26)	90.2% (37/41)	80.6% (54/67)	(1) RLNU _x50 (2) dwt2D_MedV_x50 (3) Tamura contrast_x50 (4) Tamura roughness_x50 (5) CLR RV_Ch_a_x50, (6) CLR MedAD_Ch_a_x50 (7) CLR MinV_Ch_b_x50
2	×200	80.1% (21/26)	87.8% (36/41)	85.1% (57/67)	(1) Skew_x200 (2) CON _x200 (3) IDM _x200 (4) LRE _x200 (5) RP_x200 (6) Tamura coarseness 1_x200 (7) Tamura coarseness 3_x200 (8) Tamura coarseness 4_x200
3	×50 & ×200	92.3% (24/26)	92.7% (38/41)	92.5% (62/67)	(1) Kurt_x50 (2)SAV _x200 (3) DVAR_x200 (4) DENT_x200 (5)ICM1_x200 (6) dwt2H_MedV_x200 (7) Tamura coarseness 4_x200 (8) CLR MedV_Ch_a _x200

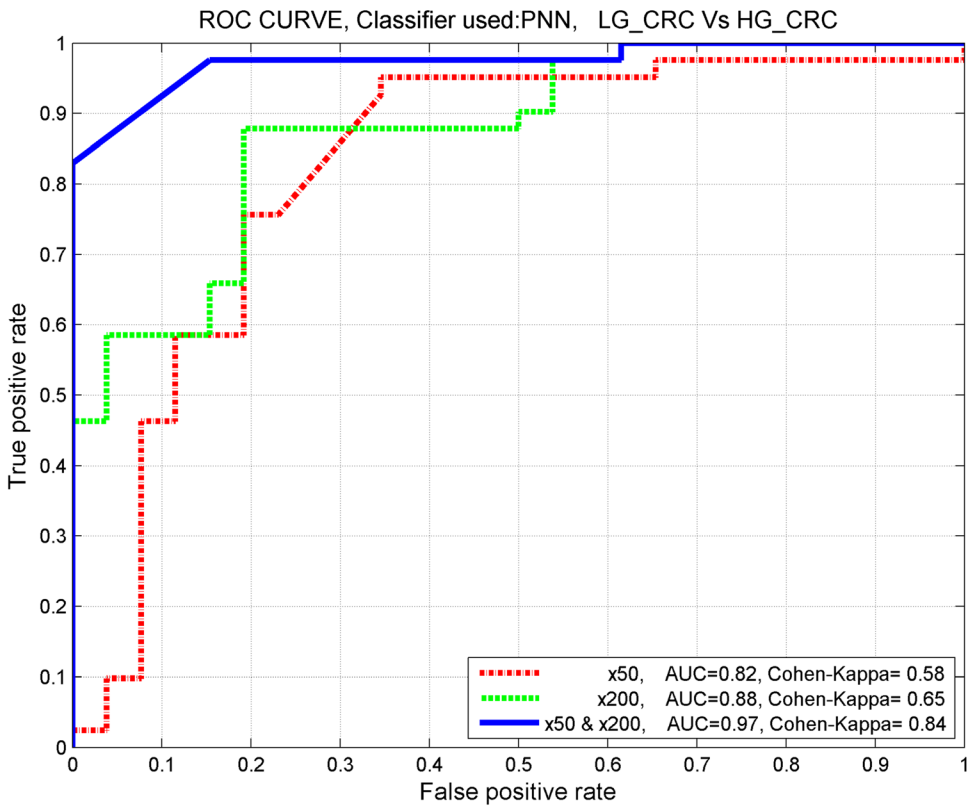


Fig. 3 ROC curves depicting sML-system discrimination accuracies, at different camera lens magnifications ×50, ×200, and ×50 and ×200 combined

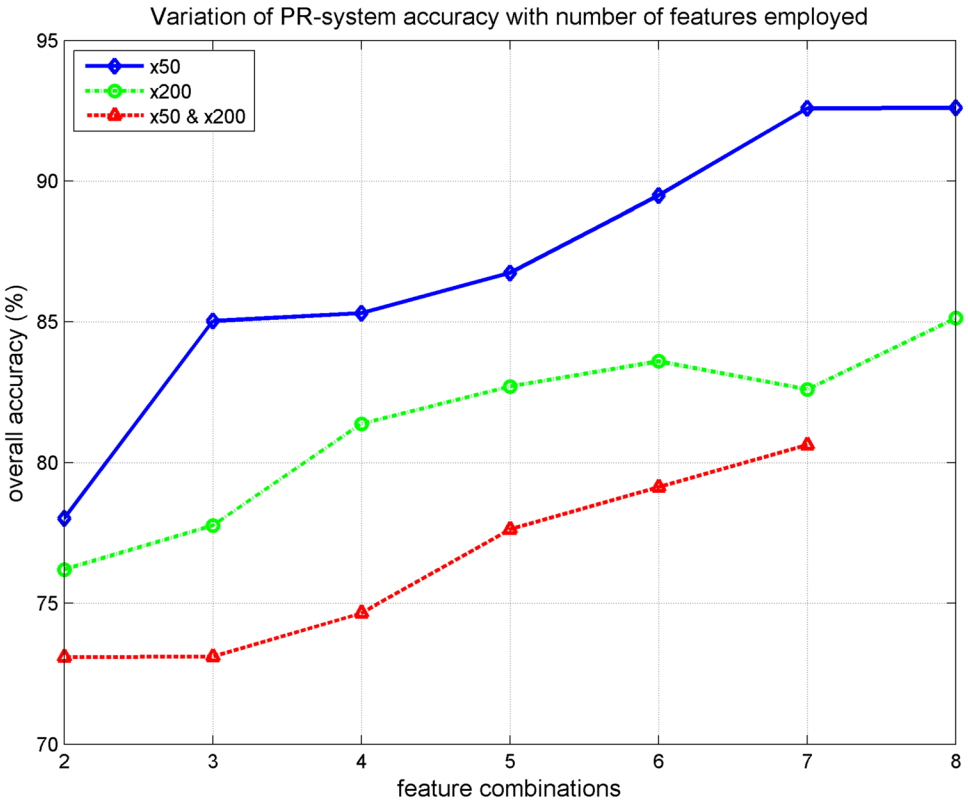


Fig. 4 Variation of sML-system precision with number of features combinations involved in its design, using images captured at different camera lens magnifications × 50, × 200, and × 50 and × 200 combined

92.5% with seven features. However, using 8 features, more balanced accuracies amongst specificity, sensitivity, and overall accuracy were achieved at around 92% (see Fig. 4, blue/solid line). In Fig. 3 the blue ROC curve (AUC = 0.97, Cohen-Kappa = 0.84) depicts the accuracy reached by using features from both magnifications (× 50 and × 200).

All the features used in the best sML-system design sustained statistically significant differences between the LG_CRC and HG_CRC classes. Figure 5 depicts by means of box and whisker plots the distribution of values of each feature in the two classes.

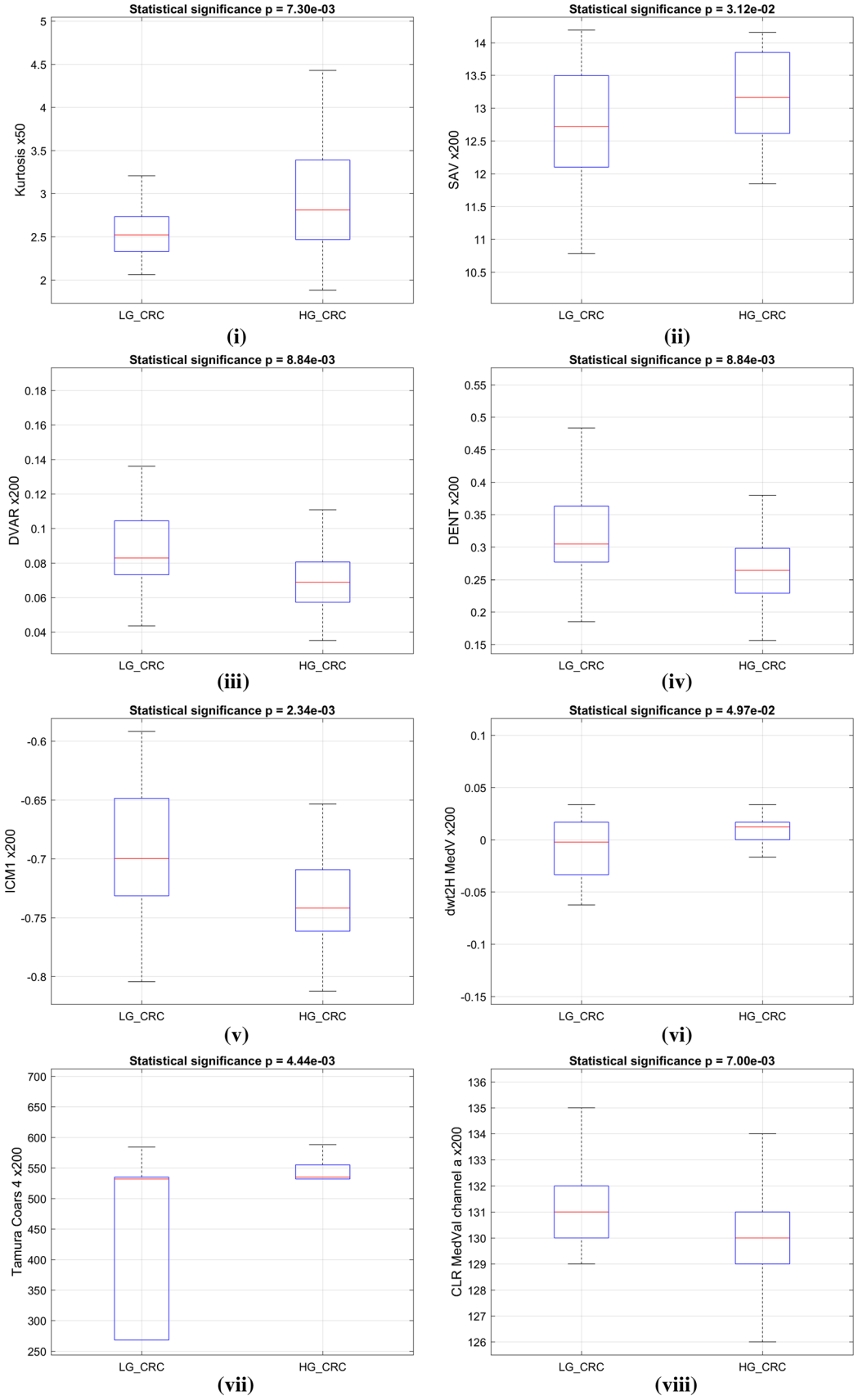
Table 3 demonstrates the sML-system's performance employing the ten-fold bootstrap evaluation method, for assessing sML-system's precision when presented to new data. Mean overall accuracy achieved was 86.2% and the mean specificity and sensitivity were of the same level (around 86%) and well balanced.

4 Discussion

We undertook the task of collecting biopsy material of patients with verified colorectal cancer, from the Department of Pathology of the University of Patras, with the purpose to design a decision support system that would discriminate with high accuracy low-grade from high-grade CRC biopsy specimens. We used microscopy digitized images captured from regions of interest of AIB1 stained biopsy material from each patient, as indicated by the pathologists at × 50 and × 200 camera lens magnifications. We experimented with deep learning and with supervised machine learning methods. First, we designed a hybrid deep learning system by employing a pre-trained VGG16 convolution neural network, which was only used to extract DL-features from each CRC-image, and then we used those features to feed a classifier to assign each image-lesion to either the low-grade or high-grade classes. Next, we designed a supervised machine learning system, by generating textural and colour features from each digitized image-ROI and by employing a classifier to classify CRC image-lesions into low-grade or high-grade classes.

4.1 The hybrid deep learning system

The design of the hybrid DL-system resulted in significantly increased classification accuracies, as compared to those achieved by the sML-system, at different lens magnifications, employing deep-learning features, the SVM classifier, and the bootstrap validation method. We achieved similar high accuracies by designing the hybrid DL-system with other classifiers, which is indicative of the stability and supremacy of hybrid deep learning designs. In a previous study by Shaban et al. [31] on CRC-grading using deep learning methods, the authors employed H&E histology


Statistical significance p = 7.30e-03
Kurtosis x50
LG_CRC
HG_CRC
(i)
Statistical significance p = 3.12e-02
SAV x200
LG_CRC
HG_CRC
(ii)
Statistical significance p = 8.84e-03
DVAR x200
LG_CRC
HG_CRC
(iii)
Statistical significance p = 8.84e-03
DENT x200
LG_CRC
HG_CRC
(iv)
Statistical significance p = 2.34e-03
ICM1 x200
LG_CRC
HG_CRC
(v)
Statistical significance p = 4.97e-02
dwt2H MedV x200
LG_CRC
HG_CRC
(vi)
Statistical significance p = 4.44e-03
Tamura Coars 4 x200
LG_CRC
HG_CRC
(vii)
Statistical significance p = 7.00e-03
CLR MedVal channel a x200
LG_CRC
HG_CRC
(viii)

◂**Fig. 5** Features used in the best sML-system design sustaining statistically significant differences between the low-grade and high-grade CRC classes (i) Kurtosis at × 50 lens magnification, features at × 200 lens magnification (ii) sum average, (iii) difference variance, (iv) difference entropy, (v) information correlation measure 1, (vi) Median Value of the 2nd level discrete wavelet transform in the horizontal direction, (vii) Tamura coarseness 4, (viii) Median Value of colour channel a

images, labeled as normal, low-grade and high-grade CRC lesions. They have compared the classification accuracies of four convolutional networks using the threefold cross validation method. Classification accuracies obtained ranged between 91 and 96%. Zhou et al. [32] have proposed the Cell Graph Convolutional network for CRC-grading. The authors have used H&E stained CRC-histology images divided into three classes of normal, low-grade, and high-grade CRC lesions. They have compared their method with other state-of-art methods deep learning methods using the threefold evaluation method. They have reported 97% accuracy against those of other methods that ranged between 88 and 96% accuracies. Those accuracies are comparable to our hybrid DL-system results, which ranged between 94 and 99%, depending on the lens-magnification used. However, deep learning methods do not employ meaningful features in system-designs, as is the case with the sML-systems.

4.2 The supervised machine learning system

Concerning the design of the sML-system, we tested the design using (a) × 50 magnification images, (b) × 200 magnification images, and (c) a combination of × 50 and × 200 magnification images. At × 50 magnification we found that sML-system's highest accuracy was 80.6%, with low specificity 65.4% and high sensitivity 90.2%. These accuracies were achieved employing the features exhaustive search method on 21 (30%) of the highest-ranked features and the LOO evaluation method. Low specificity indicates that the system is not performing accurately enough with regards to low-grade CRC cases. This is undesirable since wrong estimation of CRC grading may seriously affect patient management. It seems that low resolution images at × 50 lens magnification do not provide enough discriminatory information regarding low-grade cases. However, we adopted sML-system design with images at × 50 magnification since it is a procedure that is followed by physicians in diagnosis in conjunction with viewing ROIs at higher lens magnifications.

Next, we designed the sML-system employing only images at × 200 magnification. We found that the best design × 200 precision of the sML-system did not differ significantly from the accuracy achieved by the × 50 design. The overall accuracy of 85.1% with a low specificity of 80.1% using the exhaustive search and LOO methods is still low and, thus, we could not accept it as a reliable design.

Then, we combined the × 50 and × 200 lens magnifications data, in the same way that the pathologist would first get an overview of the images at × 50 magnification and then would examine regions of interest at higher magnifications such as the × 200 magnification. We employed the first 41 (30%) highest ranked features, of the features combined (× 50 and × 200). A mixture of × 50 and × 200 features was found that provided the highest sML-system classification accuracy of 92.5%, with balanced specificity (92.3%) and sensitivity (92.7%), a ROC AUC of 0.97 and Cohen Kappa statistic of 0.84, both indicating good class separation and that findings are not accidental. These findings indicate that the sML-system's precision in classifying LG_CRC and HG_CRC cases could be relied upon. However, this performance relies on the choice of the ROIs by the physician.

The best features combination that provided the highest classification accuracy consisted of one × 50 feature and seven × 200 features. Those comprised one 1st order statistics feature (*Kurtosis_x50*), four co-occurrence matrix features (*SAV_x200, DVAR_x200, DENT_x200, ICM1_x200*), one wavelet transform feature (*dwt2H_MedV_x200*), one Tamura feature (*Tamura coarseness 4_x200*), and one colour feature (*CLR MedV_Ch_a _x200*). Each one of those features sustained statistically significant differences between the two classes. The mathematical formulation of those features is presented in the Appendix.

As shown in Fig. 5, the features-values of Kurtosis_x50, *dwt2H_MedV_x200*, *SAV _x200*, and *Tamura coarseness 4_x200* were higher in the high-grade CRC cases. Kurtosis (*Kurt_x50*) expresses the shape of the distribution of the gray-tones in an image. High kurtosis values indicate thinner distributions around the mean gray-tone value and thick tails (leptokurtic). Thick tails imply that there are large numbers of pixels spreading across the gray-tones spectrum. This is probably due to the breakdown of the glandular structure in HG_CRC lesions, leading to the loss of definition of structures in the image. Feature *dwt2H_MedV_x200* evaluates the discrete wavelet transform along the second level horizontal direction. It gives an estimate of image edges along the horizontal direction of the image. Our results showed (see Fig. 5vi) that HG_CRC images displayed marginally statistically higher edges ($p = 0.049$). The sum average (*SAV _x200*) feature evaluates the presence of structures with variation in their grey-tones. The feature attains higher values in images containing larger elements of varying gray-tones and of coarser texture. Our findings showed that the texture of the HG_CRC lesions appeared coarser (see Fig. 5ii). The Tamura coarseness (*Tamura coarseness 4_x200*) feature evaluates image coarseness [43] and, according to our findings, HG_CRC lesions had higher values than LG_CRC lesions and, thus, appeared coarser (see Fig. 5vii). This is in

Table 3 Ten-fold bootstrap evaluation method for assessing the generalization accuracy of the sML-system using ROIs at both × 50 and × 200 lens magnifications

Runs	Overall accuracy (%)	Sensitivity (%)	Specificity (%)	#features
1	92.3	100.0	80.0	7
2	92.3	100.0	80.0	7
3	84.6	75.0	100.0	7
4	76.9	75.0	80.0	5
5	84.6	87.5	80.0	8
6	76.9	75.0	80.0	7
7	76.9	87.5	60.0	6
8	92.3	87.5	100.0	6
9	92.3	87.5	100.0	5
10	92.3	87.5	100.0	7
Mean ± std	86.2 ± 7.0	86.3 ± 9.2	86.0 ± 13.5	

line with the findings of the *SAV _x200* feature. The existence of larger structures in HG_CRC lesions is probably due to cell destruction and cell merging in the glandular texture of HG_CRC lesions.

We found that the rest four of the best eight features, *DVAR_x200, DENT_x200, ICM1_x200,* and *CLR MedV_Ch_a _x200*, displayed lower values at high-grade CRC images. The Difference Variance (*DVAR_x200*) feature assesses the variation in image contrast (or local gray-tone differences) across the image. The feature attains low values for equally distributed contrast transitions or image gray-level differences. Our findings showed (Fig. 5iii) that HG_CRC lesions displayed lower *DVAR_x200* values, which indicates that gray-tone differences were more evenly distributed leading to a slowly varying local image contrast. The Difference Entropy (*DENT_x200*) feature, shown in Fig. 5iv, assesses the variation in the lack of structural order across the image. HG_CRC lesions displayed lower values, which is probably due to the merging of cells and to the loss of fine structural definition in the glandular texture, giving the impression of a hazy and formless texture. The Information Measure I (*ICM1_x200*) feature is related to the correlation content of an image and it attains low values for highly uncorrelated texture. Our findings (see Fig. 5v) showed that HG_CRC lesions attained smaller values. This is due to the gland attaining formless texture at high-grades and to losing its pixel-to-pixel association (loss of correlation). The Median Value of colour channel a feature (*CLR MedV_Ch_a _x200*), which was calculated from the L*a*b* colour transformation of images, evaluates the median value of channel a (the red to green colour scale) and is related to AIB1 uptake. Our findings showed (see Fig. 5viii) that the feature attained lower feature values in the HG_CRC lesions, which is the result of lower AIB1 uptake (or staining) due to the destruction of AIB1 receptors in high-grade lesions.

Summarizing, we found that LG_CRC lesions may be well differentiated from HG_CRC lesions by a combining into the sML-system eight (8) features that quantify the image properties of intensity, texture, and colour, at two (2) different microscope lens magnifications, × 50 and × 200. We found that HG_CRC lesions displayed loss of definition of structures, coarser texture of larger structures, hazy formless texture, lower AIB1 uptake, lower local correlation, and slower varying image contrast.

To assess how the system would perform when new, "unseen" by the system, data were presented at its input, such as those obtained in a clinical environment, we put to the test the sML-system by employing the bootstrapping cross-validation method. We found that the overall classification accuracy as well as the sensitivity and the specificity were 86% (see Table 3), which is about 6% lower to the accuracies obtained by the sML-system designed by the LOO evaluation method. The drop in classification accuracies was expected, and it was due to the bias introduced by the LOO method in assessing the sML-system's precision [47].

Previous studies that have designed decision support systems, have mainly used H&E stained biopsy material for discriminating between normal, benign and malignant CRC tissues [11–29]. Regarding the discrimination between grades in CRC lesions, in one study by Awan et al. [27], the authors have examined the discrimination between low- and high-grade CRC lesions. The authors have used images from H&E tissue slides and have employed morphology features, the SVM classifier and a threefold cross validation method to distinguish between normal, low-grade and high-grade lesions reaching discrimination precision of 91%. We have used a more elaborate and computer processing demanding method of validation, the bootstrapping cross-validation method with multiple experimentation, so that to get a good representative estimate (86% overall accuracy) on to how our sML-system would perform on new, unseen by the system, data.

Regarding the use of IHC-staining in the discrimination of colorectal lesions, Esgiar et al. [56] have used images of IHC stained biopsy specimens for cytokeratins

to differentiate normal from cancerous colon mucosa. The authors have employed the linear discriminant analysis and, alternatively, the KNN classifier, six (6) textural features from the co-occurrence matrix, and half the dataset were used to design the classifiers and the other data half to test the classifiers' performance. They have achieved an overall accuracy of about 90%. In a later study by the same group [15], the authors have designed a decision support system for discriminating normal mucosa from adenocarcinomas. The authors have employed the KNN classifier, the fractal dimension, the correlation, and the entropy features, and they have achieved about 94% overall accuracy by using the LOO evaluation method. In the present study, we have used images from AIB1, IHC-stained, biopsy material, captured at different lens magnifications similar to those used by physicians, to discriminate between LG_CRC and HG_CRC lesions, we have tested a large number of intensity, textural, and colour features, we have experimented with a large number of readily available classifiers, and we have achieved discrimination accuracies of 92.5%, using the LOO evaluation method and about 86% using the bootstrapping evaluation method. The features that produced the best classifier design express important properties that facilitate the differentiation between the two classes, LG_CRC and HG_CRC.

To deal with the sML-system's design high dimensionality and excessive computer processing time demands, we have employed GPU and CUDA technologies for programming the multiprocessors of the Nvidia graphics card to execute concurrently similar tasks. This provision rendered plausible for us to experiment with different classifiers, to fine-tune the classifier of choice (PNN), and to achieve optimal sML-system design by testing a multitude of feature combinations. However, it has to be stressed that once the optimum sML-system has been designed it can operate on an ordinary desktop computer. This allows for the sML-system to be easily transferred into different work environments, such as histopathology laboratories, using inexpensive computer technology.

In conclusion, we employed digital images of AIB1 stained colorectal cancer lesions to construct a high-performance hybrid deep learning system, based on the pre-trained VGG16 convolution neural network, for discriminating between low-grade and high-grade CRC lesions with high accuracy (99.1%). In addition, we designed a supervised machine learning system, using a combination of eight features that quantify intensity, texture, and colour image properties, to discriminate between LG_CRC and HG_CRC with and accuracy of 92.5%. Those image properties indicate that the HG_CRC lesions are characterized by (i) loss in the definition of structures, (ii) coarser texture in larger structures, (iii) hazy formless texture, (iv) lower AIB1 uptake, (v) lower local correlation, and (vi) slower varying image contrast.

Appendix

The mathematical formulations of the eight (8) best-features combination of the sML-system design are presented below.

A.1 Histogram feature

(1) Kurtosis (Kurt)

$$Kurt = \frac{1}{N}\frac{\sum_i \sum_j (g(i,j) - m)^4}{std^4}$$

where $g(i,j)$ is the pixel intensity in position (i,j), N the total number of pixels, m is the mean value of the g, and std is the standard deviation of g.

A.2 Co-occurrence matrix based features

(2) Sum average (SAV)

$$SAV = \sum_{i=2}^{2N_g} i p_{x+y}(i)$$

where N_g is the number of gray levels in the image, $i,j = 1,\ldots,N_g$, and $p(i,j)$ is the co-occurrence matrix, and p_{x+y} is $p_{x+y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j), i+j=k, k=2,3,\ldots,2N_g$.

(3) Difference variance(DVAR)

$$DVAR = \sum_{i=2}^{2N_g} (i - SAV)^2 p_{x-y}(i)$$

where p_{x-y} is $p_{x-y}(k) = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j), |i-j| = k, k = 2,3,\ldots,N_g - 1.$

(4) Difference entropy (DENT)

$$DENT = -\sum_{i=0}^{N_g-1} p_{x-y}(i) \log\left(p_{x-y}(i)\right)$$

(5) Information measure of correlation 1 (ICM1)

$$ICM1 = \frac{HXY - HXY1}{\max\{HX, HY\}}$$

where $HXY = -\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} p(i,j) \log\left(p(i,j)\right)$

$$HXY1 = -\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} p(i,j)\log\left(p_x(i)p_y(j)\right)$$

A.4 Wavelet based feature

(6) Median Value of the 2nd level discrete wavelet transform in the horizontal direction (dwt2H_MedV).

The discrete wavelet function of an image $f(x, y)$ of size $M\ x\ N$ is defined as:

$$W_\varphi(j_0,m,n) = \frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\varphi_{j_0,m,n}(x,y)$$

$$W_\psi^i(j,m,n) = \frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\psi_{j,m,n}^i(x,y)$$

$$i = \{H, V, D\}$$

where $\varphi(x, y)$ is a scaling function, with j = 0,1,2…

W_φ are the coefficients define an approximation of image $f(x, y)$ at level (scale) j_0.

and W_ψ are the coefficients that add horizontal, vertical and diagonal details for levels $j \geq j_0$.

A.5 Tamura-based features

(7) Tamura coarseness 4

$$F_{crs} = \frac{1}{m \times n}\sum_{i}^{m}\sum_{j}^{n} S_{\text{best}}(i,j)$$

where m, n are region dimensions and

$$S_{best}(i,j) = 2^k$$

where k is the best scaling for highest neighborhood average.

The particular feature is the value of the 3rd bin histogram of *Sbest*.

A.6 Lab colour transform-based features

(8) Median Value of colour channel a* (CLR MedV_Ch_a)

According to the CIE, the coordinates of the Lab colour space are derived by a nonlinear transformation of the three primary colours X, Y and Z. The linear transformation of RGB space to X, Y and Z is defined as [57]

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.114 \\ 0.000 & 0.066 & 1.116 \end{pmatrix}\begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

$$L* = 116\left(\sqrt[3]{\frac{Y}{Y_0}}\right) - 16$$

$$a* = 500\left[\sqrt[3]{\frac{X}{X_0}} - \sqrt[3]{\frac{Y}{Y_0}}\right]$$

$$b* = 200\left[\sqrt[3]{\frac{Y}{Y_0}} - \sqrt[3]{\frac{Z}{Z_0}}\right]$$

where *Ch_x* is the **a* channel of *L*a*b* colour transform.

Acknowledgement The authors would like to thank the Department of Pathology of the University Hospital of Patras, Greece, for supplying the IHC stained for AIB1 expression material.

Declarations

Conflict of interest The authors declare that there is no conflict of interests regarding the publication of this study.

References

1. Siegel, R.L., Miller, K.D., Fedewa, S.A., Ahnen, D.J., Meester, R.G.S., Barzi, A., Jemal, A.: Colorectal cancer statistics, 2017. CA: a cancer journal for clinicians, 67(3):177–193 (2017)
2. Kunhoth, S., Al Maadeed, S., Bouridane, A.: Medical and computing insights into colorectal tumors. Int. J. Life Sci. Biotech. Pharm. Res. **4**(2), 122–126 (2015)
3. Phillips, R.K.S., Hittinger, R., Blesovsky, L., Fry, J.S., Fielding, L.P.: Large bowel cancer: Surgical pathology and its relationship to survival. Br. J. Surg. **71**, 604–610 (1984)
4. Freedman, L.S., Macaskill, P., Smith, A.N.: Multivariate analysis of prognostic factors for operable rectal cancer. Lancet **2**(8405), 733–736 (1984)
5. Jass, J.R., Atkin, W.S., Cuzick, J., Bussey, H.J., Morson, B.C., Northover, J.M., Todd, I.P.: The grading of rectal cancer: historical perspectives and a multivariate analysis of 447 cases. Histopathology **41**(3A), 59–81 (2002)
6. Thomas, G.D., Dixon, M.F., Smeeton, N.C., Williams, N.S.: Observer variation in the histological grading of rectal carcinoma. J. Clin. Pathol. **36**(4), 385–391 (1983)
7. Andrion, A., Magnani, C., Betta, P.G., Donna, A., Mollo, F., Scelsi, M., Bernardi, P., Botta, M., Terracini, B.: Malignant mesothelioma of the pleura: interobserver variability. J. Clin. Pathol. **48**(9), 856–860 (1995)
8. Rathore, S., Hussain, M., Ali, A., Khan, A.: A recent survey on colon cancer detection techniques. IEEE/ACM Trans. Comput. Biol. Bioinf. **10**(3), 545–563 (2013)
9. Turner, J.K., Williams, G.T., Morgan, M., Wright, M., Dolwani, S.: Interobserver agreement in the reporting of colorectal polyp

pathology among bowel cancer screening pathologists in Wales. Histopathology **62**(6), 916–924 (2013)
10. Blenkinsopp, W.K., Stewart-Brown, S., Blesovsky, L., Kearney, G., Fielding, L.P.: Histopathology reporting in large bowel cancer. J. Clin. Pathol. **34**(5), 509–513 (1981)
11. Allen, D.C., Hamilton, P.W., Watt, P.C., Biggart, J.D.: Morphometrical analysis in ulcerative colitis with dysplasia and carcinoma. Histopathology **11**(9), 913–926 (1987)
12. Hamilton, P.W., Allen, D.C., Watt, P.C., Patterson, C.C., Biggart, J.D.: Classification of normal colorectal mucosa and adenocarcinoma by morphometry. Histopathology **11**(9), 901–911 (1987)
13. Allen, D.C., Hamilton, P.W., Watt, P.C., Biggart, J.D.: Architectural morphometry in ulcerative colitis with dysplasia. Histopathology **12**(6), 611–621 (1988)
14. Hamilton, P.W., Allen, D.C., Watt, P.C.: A combination of cytological and architectural morphometry in assessing regenerative hyperplasia and dysplasia in ulcerative colitis. Histopathology **17**(1), 59–68 (1990)
15. Hamilton, P.W., Bartels, P.H., Thompson, D., Anderson, N.H., Montironi, R., Sloan, J.M.: Automated location of dysplastic fields in colorectal histology using image texture analysis. J. Pathol. **182**(1), 68–75 (1997)
16. Esgiar, A.N., Naguib, R.N., Sharif, B.S., Bennett, M.K., Murray, A.: Fractal analysis in the detection of colonic cancer images. IEEE Trans. Inf. Technol. Biomed. Publ. IEEE Eng. Med. Biol. Soc. **6**(1), 54–58 (2002)
17. Ficsor, L., Varga, V.S., Tagscherer, A., Tulassay, Z., Molnar, B.: Automated classification of inflammation in colon histological sections based on digital microscopy and advanced image analysis. Cytometry Part A J. Int. Soc. Anal. Cytol **73**(3), 230–237 (2008)
18. Masood, K., Rajpoot, N.: Texture based classification of hyperspectral colon biopsy samples using CLBP. In: Proceedings of International Symposiumon Biomedical Imaging: From Nano to Macro, Boston 2009. Pp. 1011–1014 (2009)
19. Kalkan, H., Nap, M., Duin, R. P., Loog, M.: Automated classification of local patches in colon histopathology. In: 21st Int'l Conf. Pattern Recognition, pp. 61–64 (2012).
20. Jiao, L., Chen, Q., Li, S.: Xu Y Colon cancer detection using whole slide histopathological images. In: World Congress on Medical Physics and Biomedical Engineering, pp. 1283–1286. Springer, Berlin Heidelberg (2013)
21. Rathore, S., Hussain, M., Aksam Iftikhar, M., Jalil, A.: Ensemble classification of colon biopsy images based on information rich hybrid features. Comput. Biol. Med. **47**, 76–92 (2014)
22. Xu, Y., Mo, T., Feng, Q., Zhong, P., Lai, M., EI-C C Deep learning of feature representation with multiple instance learning for medical image analysis. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1626–1630 (2014).
23. Kumar, R., Srivastava, R., Srivastava, S.: Detection and classification of cancer from microscopic biopsy images using clinically significant and biologically interpretable features. J. Med. Eng. **2015**, 457906 (2015)
24. Rathore, S., Hussain, M., Khan, A.: Automated colon cancer detection using hybrid of novel geometric features and some traditional features. Comput. Biol. Med. **65**, 279–296 (2015)
25. Chaddad, A., Desrosiers, C., Bouridane, A., Toews, M., Hassan, L., Tanougast, C.: Multi texture analysis of colorectal cancer continuum using multispectral imagery. PLoS ONE **11**(2), 1–17 (2016)
26. Kather, J.N., Weis, C.A., Bianconi, F., Melchers, S.M., Schad, L.R., Gaiser, T., Marx, A., Zollner, F.G.: Multi-class texture analysis in colorectal cancer histology. Sci. Rep. **6**, 27988 (2016)
27. Chaddad, A., Tanougast, C.: Texture analysis of abnormal cell images for predicting the continuum of colorectal cancer. Anal. Cell. Pathol. **2017**, 8428102 (2017)
28. Awan, R., Sirinukunwattana, K., Epstein, D., Jefferyes, S., Qidwai, U., Aftab, Z., Mujeeb, I., Snead, D., Rajpoot, N.: Glandular morphometrics for objective grading of colorectal adenocarcinoma histology images. Scientific reports **7**(1), 16852 (2017)
29. Peyret, R., Bouridane, A., Khelifi, F., M. Atif Tahir b, Al-Maadeed S, : Automatic classification of colorectal and prostatic histologic tumor images using multiscale multispectral local binary pattern texture features and stacked generalization. Neurocomputing **275**, 83–93 (2018)
30. Pacal, I., Karaboga, D., Basturk, A., Akay, B., Nalbantoglu, U.: A comprehensive review of deep learning in colon cancer. Comput. Biol. Med. **126**, 104003 (2020)
31. Shaban, M., Awan, R., Fraz, M.M., Azam, A., Tsang, Y.W., Snead, D., Rajpoot, N.M.: Context-Aware Convolutional Neural Network for Grading of Colorectal Cancer Histology Images. IEEE Trans. Med. Imaging **39**(7), 2395–2405 (2020)
32. Zhou, Y., Graham, S., Alemi Koohbanani, N., Shaban, M., Heng, P.A., Rajpoot, N.: CGC-net: Cell graph convolutional network for grading of colorectal cancer histology images. In: *Proceedings - 2019 International Conference on Computer Vision Workshop,* ICCVW 2019, 2019. pp. 388–398.
33. Anzick, S.L., Kononen, J., Walker, R.L., Azorsa, D.O., Tanner, M.M., Guan, X.Y., Sauter, G., Kallioniemi, O.P., Trent, J.M., Meltzer, P.S.: AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. Science **277**(5328), 965–968 (1997)
34. Ghadimi, B.M., Schrock, E., Walker, R.L., Wangsa, D., Jauho, A., Meltzer, P.S., Ried, T.: Specific chromosomal aberrations and amplification of the AIB1 nuclear receptor coactivator gene in pancreatic carcinomas. Am. J. Pathol. **154**(2), 525–536 (1999)
35. Sakakura, C., Hagiwara, A., Yasuoka, R., Fujita, Y., Nakanishi, M., Masuda, K., Kimura, A., Nakamura, Y., Inazawa, J., Abe, T., Yamagishi, H.: Amplification and over-expression of the AIB1 nuclear receptor co-activator gene in primary gastric cancers. Int. J. Cancer **89**(3), 217–223 (2000)
36. Wang, Y., Wu, M.C., Sham, J.S., Zhang, W., Wu, W.Q., Guan, X.Y.: Prognostic significance of c-myc and AIB1 amplification in hepatocellular carcinoma. A broad survey using high-throughput tissue microarray. Cancer **95**(11), 2346–2352 (2002)
37. Xie, D., Sham, J.S., Zeng, W.F., Lin, H.L., Bi, J., Che, L.H., Hu, L., Zeng, Y.X., Guan, X.Y.: Correlation of AIB1 overexpression with advanced clinical stage of human colorectal carcinoma. Hum. Pathol. **36**(7), 777–783 (2005)
38. Xu, F.P., Xie, D., Wen, J.M., Wu, H.X., Liu, Y.D., Bi, J., Lv, Z.L., Zeng, Y.X., Guan, X.Y.: SRC-3/AIB1 protein and gene amplification levels in human esophageal squamous cell carcinomas. Cancer Lett. **245**(1–2), 69–74 (2007)
39. Liu, M.Z., Xie, D., Mai, S.J., Tong, Z.T., Shao, J.Y., Fu, Y.S., Xia, W.J., Kung, H.F., Guan, X.Y., Zeng, Y.X.: Overexpression of AIB1 in nasopharyngeal carcinomas correlates closely with advanced tumor stage. Am. J. Clin. Pathol. **129**(5), 728–734 (2008)
40. Tzelepi, V., Grivas, P., Kefalopoulou, Z., Kalofonos, H., Varakis, J.N., Melachrinou, M., Sotiropoulou-Bonikou, G.: Estrogen signaling in colorectal carcinoma microenvironment: expression of ERbeta1, AIB-1, and TIF-2 is upregulated in cancer-associated myofibroblasts and correlates with disease progression. Virchows Archiv Int. J. Pathol. **454**(4), 389–399 (2009)
41. Chen, L., Wang, C., Zhang, X., Gao, K., Liu, R., Shi, B., Hou, P.: AIB1 Genomic Amplification Predicts Poor Clinical Outcomes in Female Glioma Patients. J. Cancer **7**(14), 2052–2060 (2016)
42. Theodosi, A., Glotsos, D., Kostopoulos, S., Kalatzis, I., Tzelepi, V., Ravazoula, P., Asvestas, P., Cavouras, D., Sakellaropoulos, G.: Correlating Changes in the Epithelial Gland Tissue With

Advancing Colorectal Cancer Histologic Grade, Using IHC Stained for AIB1 Expression Biopsy Material. Appl Immunohistochem Mol Morphol AIMM (2018)
43. Greene, F.L.: AJCC Cancer Staging Manual. Springer, New York (2002)
44. Xu, Y., Jia, Z., Wang, L.B., Ai, Y., Zhang, F., Lai, M., Chang, E.I.: Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. BMC Bioinf **18**(1), 281 (2017)
45. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (2015).
46. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for Cancer classification using support vector machines. Mach. Learn. **46**(1), 389–422 (2002)
47. Haralick, R., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. **3**, 610–621 (1973)
48. Galloway, M.M.: Texture analysis using grey level run lengths. Comput. Graph. Image Process. **4**, 172–179 (1975)
49. H. Tamura SM, and T. Yamawaki: Textural features corresponding to visual perception. IEEE Trans. Syst. Man Cybern. **8**(6), 460–473 (1978)
50. Specht, D.: Probabilistic neural networks. Neural Netw. **3**, 109–118 (1990)
51. Theodoridis, S., Pikrakis, A., Koutroumbas, K., Cavouras, D.: An Introduction to Pattern Recognition: A Matlab Approach. Academic Press, Oxford (2010)
52. Theodoridis, S., Koutroumbas, K.: Pattern Recognition, 2nd edn. Elsevier, San Diego (2003)
53. Foley, D.: Considerations of sample and feature size. IEEE Trans. Inf. Theory **18**(5), 618–626 (1972)
54. Ambroise, C., McLachlan, G.J.: Selection bias in gene extraction on the basis of microarray gene-expression data. Proc. Natl. Acad. Sci. U.S.A. **99**(10), 6562–6566 (2002)
55. Sidiropoulos, K., Glotsos, D., Kostopoulos, S., Ravazoula, P., Kalatzis, I., Cavouras, D., Stonham, J.: Real time decision support system for diagnosis of rare cancers, trained in parallel, on a graphics processing unit. Comput. Biol. Med. **42**(4), 376–386 (2012)
56. Esgiar, A.N., Naguib, R.N., Bennett, M.K., Murray, A.: Automated feature extraction and identification of colon carcinoma. Anal. Quant. Cytol. Histol. **20**(4), 297–301 (1998)
57. Cheng, H.D., Sun, Y.: A Hierarchical Approach to Color Image Segmentation Using Homogeneity. IEEE Trans. Image Process. **9**(12), 2071–2082 (2000)

Angeliki Theodosi was born in Zakynthos, Greece, in 1977. She received her B.Sc. degree in Informatics from the Department of Informatics, University of Piraeus, Greece, in 2001, M.Sc. degree in Bioinformatics from the University of Athens, Greece, in 2012. She is Ph.D. candidate in the Department of Medical Physics, Faculty of Medicine, University of Patras, Greece. Since 2002, she works as ICT educator in secondary and primary education. Her main research interests include the fields of medical image analysis, machine learning, bioinformatics and STEM education.

Sotiris Ouzounis was born in Veria, Greece, in 1997. He received an integrated masters degree from the Department of Biomedical Engineering of the University of West Attica, Athens, Greece, in 2020. He is currently a Ph.D. candidate in the same department and his dissertation is focused on computational drug repositioning. His primary research interests comprise data analysis, bioinformatics, machine learning, deep learning and medical image analysis.

Spiros Kostopoulos was born in Athens, Greece, in 1977. He received the B.Sc. degree in Biomedical Engineering from the Department of Medical Instruments Technology, TEI, Athens, in 2000, the M.Sc. degree in Medical Physics from the University of Surrey, U.K., in 2004, and the Ph.D. degree in Medical Physics from the University of Patras, Greece, in 2009. Since 2017, he holds an assistant professor position in the Department of Biomedical Engineering of University of West Attica, Greece. His main research activities lie in the field of medical image analysis, machine learning, and bioinformatics.

Dimitris Glotsos was born in Athens, Greece, in 1977. He received his B.Sc. degree in Biomedical Engineering from the Department of Biomedical Engineering (former Medical Instruments Technology) of the Technological Educational Institute of Athens, Greece, in 2000. He received his M.Sc. degree in Medical Physics from the University of Patras, Greece, in 2002. He was awarded with the Ph.D. degree by the University of Patras, Greece, in 2006. His thesis concerned the development of machine learning methods for processing of microscopy images and signals. His research interests include pattern recognition, medical image processing and analysis, bioinformatics and microscopy. Currently, he holds the position of Associate Professor in Medical Image Analysis at the Department of Biomedical Engineering of the University of West Attica, Greece.

Ioannis Kalatzis was born in Athens, Greece in 1965. He received B.Sc. degree in Physics from the University of Athens, Greece, in 1987 and Ph.D. degree in Medical Physics from the Medical School of the University of Athens, Greece, in 2000. Since 2018, he holds an Associate Professor position in the Department of Biomedical Engineering of the University of West Attica, Greece, being tenured faculty member in the same department since 2009. His main research activities lie in the fields of medical image and signal processing and analysis, medical physics, medical statistics and pattern recognition with applications in biomedicine.

Vasiliki Tzelepi was born in Patras in 1979. She received her M.D. diploma in 2003, her B.Sc. in Basic Medical Sciences diploma in 2006 and her Ph.D. in Medicine in 2008, all from the School of Medicine of the University of Patras. She also completed her subspecialization in Pathology at the University Hospital of Patras, Greece, in 2009. Subsequently, she was appointed as a Postdoctoral Fellow at the Department of Genitourinary Medical Oncology, The University of Texas, MD Anderson Cancer Center, Houston, USA. She currently works as an Assistant Professor of Pathology at the School of Medicine of the University of Patras with a joint appointment at the Department of Pathology of the University Hospital of Patras (Instructor 2014–2019, Assistant Professor 2019–to date) and is also an Adjunct Assistant Professor at the Department of Genitourinary Medical Oncology, The University of Texas, MD Anderson Cancer Center, Houston, USA (since 2015). Her main research interests are prostate cancer pathology, morphology and therapy resistance and the effect of epigenetic mechanisms in disease progression. She is also involved in investigating the epigenetic alterations underlying a variety of neoplasms, i.e., colon, ovarian and endometrial carcinoma.

Panagiota Ravazoula was born in Patras, Greece, in 1954. She received the medical degree from the Medical School, University of Thessaloniki, Greece. She completed her subspecialization in Pathology at the University Hospital of Patras, Greece, and received her PhD at the Medical School University of Patras, Greece. She works at the Department of Pathology, University Hospital of Patras, Greece. Her

main research interests are breast, cervix and colon cancer pathology and image analysis.

Pantelis Asvestas was born in 1973. He received the diploma in engineering from the Department of Electrical and Computer Engineering, National Technical University of Athens, in 1996, and the Doctor of Engineering degree in medical image processing from the Department of Electrical and Computer Engineering in 2001. Since 2003, he has been an electrical engineer participating in R&D projects. Since 2008, he is with the Department of Biomedical Engineering, University of West Attica, Athens, Greece. His research interests are in the fields of classification of biosignals and medical images using artificial intelligence, image processing, biometrics, and real-time video processing.

Dionisis Cavouras was born in Kalamata, Greece. He received the B.Sc. degree in electronic engineering, the M.Sc. and Ph.D. degrees in systems engineering from the City University London, U.K. He is currently an Emeritus Professor of Medical Imaging Processing with the Department of Biomedical Engineering, University of West Attica, Athens, Greece. His research interests include medical image processing, image analysis, pattern recognition, medical statistics, and medical physics. He has published numerous technical and medical papers in refereed journals and in conference proceedings.

George Sakellaropoulos is Professor of Medical Physics—Medical Informatics at the Department of Medical Physics, School of Medicine, University of Patras, Greece. Born in Thessaloniki, Greece, in 1971, he obtained his Diploma in Physics from the University of Athens in 1993, M.Sc. (1995) and Ph.D. (2000) degrees in Medical Physics from the University of Patras and M.Sc. in Health Care Management (2006) from the Hellenic Open University. He teaches medical physics, biostatistics and medical informatics to students of the School of Medicine and the Physics Department as well as in postgraduate courses. His research interests lie in the field of medical physics, medical informatics, statistical learning and decision support systems. He has authored or co-authored more than 90 papers in peer-reviewed scientific journals and has been member of the program committee of many international conferences.