CrossMark

Microscopy image analysis of p63 immunohistochemically stained laryngeal cancer lesions for predicting patient 5-year survival

Konstantinos Ninos[1] · Spiros Kostopoulos[2] · Ioannis Kalatzis[2] ·
Konstantinos Sidiropoulos[2,3] · Panagiota Ravazoula[4] · George Sakellaropoulos[5] ·
George Panayiotakis[5] · George Economou[1] · Dionisis Cavouras[2]

Abstract The aim of the present study was to design a microscopy image analysis (MIA) system for predicting the 5-year survival of patients with laryngeal squamous cell carcinoma, employing histopathology images of lesions, which had been immunohistochemically (IHC) stained for p63 expression. Biopsy materials from 42 patients, with verified laryngeal cancer and follow-up, were selected from the archives of the University Hospital of Patras, Greece. Twenty six patients had survived more than 5 years and 16 less than 5 years after the first diagnosis. Histopathology images were IHC stained for p63 expression. Images were first processed by a segmentation method for isolating the p63-expressed nuclei. Seventy-seven features were evaluated regarding texture, shape, and physical topology of nuclei, p63 staining, and patient-specific data. Those features, the probabilistic neural network classifier, the leave-one-out (LOO), and the bootstrap cross-validation methods, were used to design the MIA-system for assessing the 5-year survival of patients with laryngeal cancer. MIA-system accuracy was about 90 % and 85 %, employing the LOO and the Bootstrap methods, respectively. The image texture of p63-expressed nuclei appeared coarser and contained more edges in the 5-year non-survivor group. These differences were at a statistically significant level ($p < 0.05$). In conclusion, this study has proposed an MIA-system that may be of assistance to physicians, as a second opinion tool in assessing the 5-year survival of patients with laryngeal cancer, and it has revealed useful information regarding differences in nuclei texture between 5-year survivors and non-survivors.

Introduction

Laryngeal cancer is the most frequent cancer among head and neck neoplasms. It amounts to about 3 % of newly diagnosed cancers and has poor prognosis. Risk factors for laryngeal cancer [1] development include smoking, alcohol, heredity, occupational substances, and polluted environment. Laryngeal cancer prognosis depends on factors such as the cancer's stage, involved site on the part of the larynx, its grade, and the patient's well-being and lifestyle after the first diagnosis. In Europe, on average, 62.8 % of patients with laryngeal cancer will survive for a 5-year period after the first diagnosis [2]. The 5-year period is statistically accepted as the time threshold where there is little chance that the cancer of the larynx will reappear if it has not recurred, provided that the patient has adopted a healthy lifestyle. Early and accurate diagnosis are, thus, important for adopting the right treatment and improving patient survival. However, during the last two decades, the

✉ Dionisis Cavouras
cavouras@teiath.gr

1   Department of Physics, School of Natural Sciences, University of Patras, Rio, Patras, Greece

2   Medical Image and Signal Processing Laboratory, Department of Biomedical Engineering, Technological Educational Institute of Athens, Ag.Spyridonos, Egaleo, 12210 Athens, Greece

3   European Molecular Biology Laboratory, Wellcome Trust Genome Campus, European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge CB10 1SD, UK

4   Department of Pathology, University Hospital of Patras, Rio, Greece

5   Department of Medical Physics, Faculty of Medicine, School of Health Sciences, University of Patras, Rio, Patras, Greece

survival of patients with laryngeal cancer has not improved significantly. Thus, researchers have focused on identifying new prognostic factors for laryngeal cancer. One such factor is the production of p63 protein by the TP63 gene and p63 over-expression in epithelial neoplasms of the head and neck squamous cell carcinomas [3–7]. p63 is a nuclear protein homolog of the tumor suppressor p53, involved in embryonic development, and is a marker of non-invasive epithelial tumors. Loss of p63 accelerates tumor genesis and metastasis. The predominant localization of p63 protein is in the basal cells of normal epithelia in the ectocervix, esophagus, prostate, skin, tonsil, urothelium, and vagina and in basal cells in glandular structures of breast, bronchi, prostate, and larynx.

A number of previous studies have attempted to assess the survival of patients with laryngeal cancer by computer-assisted image analysis methods, using hematoxylin and eosin (H&E)-stained microscopy images. In these studies, nuclei parameters were evaluated from microscopy images and parameter thresholds were associated with patient survival, such as the fractal dimension [8] that evaluates texture heterogeneity of nuclei or parameters that evaluate nuclei morphometry [9], such as nuclei concentration, area, perimeter, density, and roundness.

In the present study, a microscopy image analysis (MIA) system was designed for assessing the 5-year survival of patients with laryngeal squamous cell carcinomas, employing histopathology images of lesions, which had been immunohistochemically (IHC) stained for p63 expression. Image analysis systems have been previously developed (i) in two studies [10, 11] for categorizing healthy, nodular, and diffuse color laryngeal images of

vocal folds, employing color microlaryngoscopy images, and (ii) in one recent study by our group [12], for discriminating low- from high-grade laryngeal lesions, employing immunohistochemically staining for p63 expression histopathology images. The contribution of the present study, within the sample of patients studied, is twofold. First, it identifies cell nuclei parameters that sustain statistically significant differences between patients with laryngeal squamous cell carcinomas who survived for more or less than 5 years after the first diagnosis. Second, it proposes a decision support system, designed by employing image analysis methods, to assess with good precision the 5-year survival of patients with laryngeal tumor.

Materials and methods

Clinical material

Biopsy materials from 42 lesions from an equal number of patients, diagnosed with laryngeal squamous carcinomas with known follow-up, were selected from the archives of the Department of Pathology, University Hospital of Patras, Rio, Greece. Twenty six of the patients had survived a 5-year period from the first diagnosis, while 16 died before the end of the 5-year period. Patients had a mean age of 63 years, ranging between 44 and 89 years; 42 were male and there were no female patients, and most were smokers (41/42). Laryngeal lesion sites (see Table 1) were 29 glottic, 10 supraglottic, and 3 were spread to more than one laryngeal subsite. The 5-year non-survivors died due to the disease. Clinical and pathological staging were

Table 1 Distribution of squamous cell laryngeal tumors with respect to lesion site, stage, and histological grade and among 5-year survivors and 5-year non-survivors

		5-year survivors 26	5-year non-survivors 16	Total 42
Lesion site	Glottis	20	9	29
	Supraglottic	4	6	10
	Spread to subsites	2	1	3
Stage	T2	6	2	8
	T3	15	7	22
	T4	4	7	11
	N0	21	13	34
	N1	2		2
	N2	2	3	5
	N/A	1		1
	II	5	2	7
	III	14	6	20
	IV	7	8	15
Histological tumor grade	Grade I	12	4	16
	Grade II	9	6	15
	Grade III	5	6	11

determined following the American Joint Committee on Cancer (AJCC) guidelines [13]. There were 8 T2, 22 T3, and 11 T4 cases. There were 34 N0, 2 N1, and 5 N2 cases. Seven cases were stage II, 20 stage III, and 15 stage IV. In one case, the TNM (T: tumor size, N: lymph node involvement, M: distant metastasis) evaluation was not filed.

Biopsy sections were formalin fixed, paraffin embedded, and specimens were hematoxylin and eosin (HE) stained for histological tumor grade and stage assessment and immunohistochemically (IHC) stained for p63 expression. IHC staining was performed not automatically, but by a standard streptavidin peroxidase method, using a Dako Production, ready-to-use, monoclonal mouse antibody clone DAK-p63/IsotypeIgG2a, kappa.

Sixteen cases were diagnosed as grade I, 15 as grade II, and 11 as grade III. Tumor grade and p63 expression were visually assessed by the same experienced histopathologist (P.R.). The percentage of positive p63 expression was above 50 % (most were within 60–90 %). This referred to the percentage of positively expressed nuclei present in the whole region (most heterogeneous region from where images were captured) of the slide under examination. Areas within that region with poor p63 expression were discarded from the material of the present study. The study was conducted in accordance with the guidelines of the Declaration of Helsinki and of the ethics committee of the University of Patras, Greece.

Capturing of p63-stained microscopy images

All areas of tumor were examined under the microscope and the histopathologist selected and marked the most heterogeneous region on each patient's p63-stained slide. From this region four images on average, with p63 nuclei expression were digitized, using a Leica DM2500 light microscope, equipped with a Leica DFC420C digital camera. Images were captured at magnification of 400× and resolution of 1728 × 1296 × 24bit (see Fig. 1a) and stored on the hard disk of a dedicated host computer, which was connected to the digital camera of the microscope. Image capturing and storage were executed by means of the Leica Application Suite (LAS) program, which also was responsible for automatically regulating image-capturing parameters such as focus, exposure time, gamma value, and white balance. All areas of tumor on the slide were examined and the most heterogeneous area selected.

Processing of p63-stained digital images

The digitized RGB images were next processed with the purpose of identifying and outlining the p63-stained cell nuclei within the image, which were obscured since they
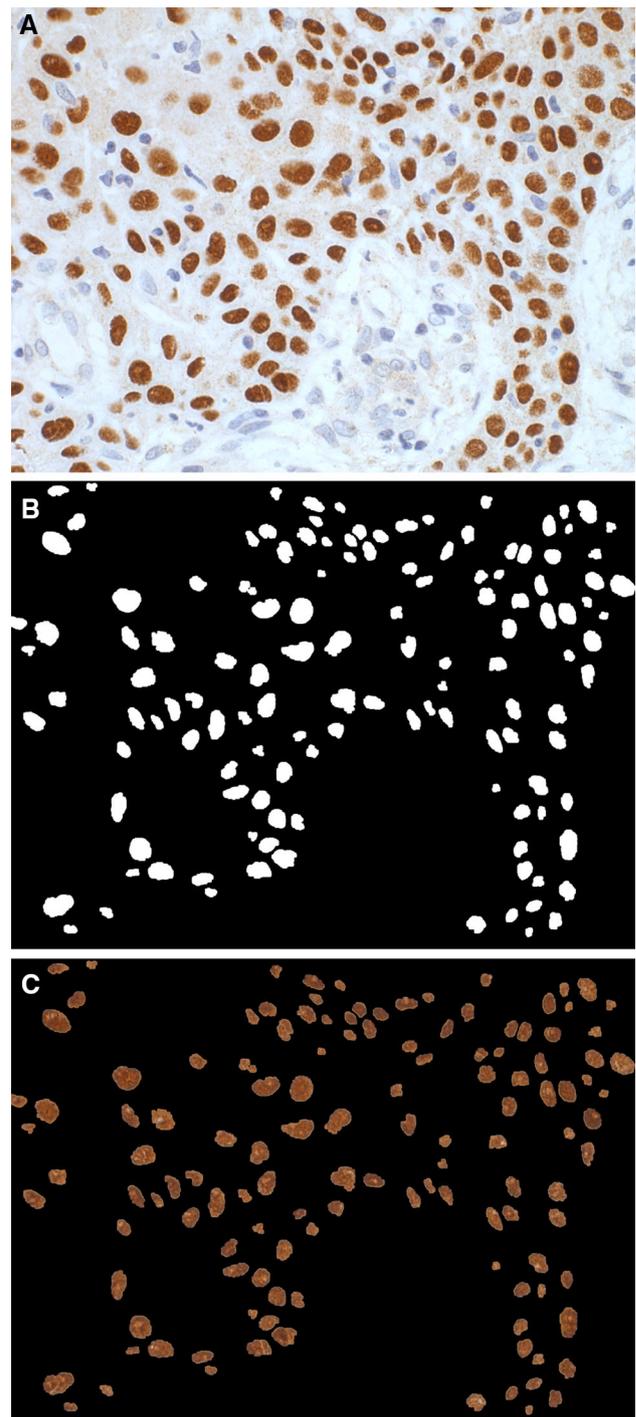
Fig. 1 a Digitized frame from p63-stained specimen. b Binary image of p63-expressed nuclei; c the final segmented image, where only the nuclei are visible

were surrounded by other nuclear tissue. Following an image-processing procedure, published in a previous study by our group [14] and given here in brief, digitized images were first transformed from the RGB scale into the $L*a*b*$ color space in accordance with the guidelines issued by CIELAB (Commission International de l'Eclairage) [15],

where $L*$ is the difference between light and dark intensities, $a*$ is the red-green scale, and $b*$ is the blue-yellow scale. This facilitated the application of image-processing algorithms, for making nuclei localization and outlining easier. Image segmentation for nuclei detection was accomplished by means of the fuzzy c-means algorithm [16]. The latter formed three pixel clusters: image pixels of p63-expressed nuclei (brownish), of non-expressed nuclei (bluish), and of non-nuclei tissue (whitish). Using the pixel coordinates of the brownish pixels and in conjunction with morphological filters, a binary image (see Fig. 1b) was formed, in which pixels that were set (white) corresponded to pixels of the p63-expressed nuclei. By combining the binary image with the original RGB image, via a logical AND operation, an RGB image resulted, containing only segmented nuclei (see Fig. 1c). The correctness of the process was evaluated by inspection by an experienced pathologist (P.R.) using a specially designed custom software. Accordingly, the original RGB image and its corresponding segmented image were concurrently displayed on the PC monitor. The segmented image contained only objects that had been identified by the segmentation software as nuclei. The physician had the task of inspecting both images and identifying on the segmented image true nuclei, by clicking on them and excluding objects that were mistakenly considered by the segmentation software as nuclei.

Feature generation

Seventy-one features were generated from the nuclei, and each case was represented by a 71-feature vector, the mean feature values of all nuclei contained in the patient images. Additionally, 6 more features were included, which were patient specific, thus forming a 77-feature vector. Two files were formed, a 16×77 file, consisting of the feature vectors of the 16 patients of the "5-year non-survivors" class and a 26×77 file comprising the feature vectors of the 26 patients of the "5-year survivors" class. The 71 nuclei-related features were generated using the grayscale version of the RGB-processed images. Four features were calculated from the nucleus image matrix; the mean value, standard deviation, skewness, and kurtosis; 26 features from the four co-occurrence matrices (mean and range) of the nucleus image [17]; 10 features from the four run-length matrices (mean and range) [18] of the nucleus image. Ten features were generated that evaluated the morphology of the nuclei (area, perimeter, eccentricity, length major axis, length of minor axis, convex area, solidity, equivalent diameter, rectangularity, compactness). Seven features were calculated from the minimum spanning tree (MST) that reflects the topology of the nuclei within the image (mean, range, standard deviation,

maximum, minimum, sum, and total length) [19]. One feature was evaluated from the fractal dimension, which has been shown to be a prognostic factor in laryngeal carcinoma [8]. Six features were calculated from the outline of the nucleus (mean, standard deviation, range and entropy of the radial distance, circularity ratio, and roughness index). Six features were computed from the two-level two-dimensional discrete wavelet transform (energy of level 1 and level 2 horizontal, vertical, and diagonal detail coefficients). One feature evaluated the percentage of p63-expressed nuclei in the patients' four digital images, since it has been claimed in a previous study [6, 7] that p63 under-expression has been correlated with poor prognosis. Most of the above 71 features are functions readily available in Matlab. Nuclear features were also normalized to zero mean and unit standard deviation, employing $\bar{f}_i = (f_i - \mu)/\sigma$ where $\bar{f}_i$ is the normalized version of feature f_i, and μ and σ are the mean and standard deviations of the feature, both calculated over all patterns of both classes (low or high laryngeal lesions).

Six patient-related features were also quantified, four concerning patient's age, smoking, alcohol consumption, and work-related risk factors; two were disease-specific features, related to tumor stage and grade. Smoking and alcohol consumption were ranked on a scale of 0–4, with 4 corresponding to longtime heavy smokers or drinkers. Work-related risks were measured on a scale 0–2, with 2 referring to working for a very long period in environments posing risks for laryngeal cancer, such as diesel exhaust, asbestos, organic solvents, metal dust, asphalt, wood dust, stone dust, mineral wool, cement dust, and pesticide. Age was evaluated as a continuous feature, histological tumor grade was ranked on a three-grade scale (I, II, III), and tumor stage on a four-stage scale (I–IV). All patient-related features, with the exemption of age, were normalized to a standardized value between zero and one [0, 1] by means of $f = \frac{r-1}{R-1}$ [20], where $r = 1$ to R refers to the feature's ranking scale.

Statistical analysis

Statistical analysis consisted of determining those features that sustained statistically significant differences between the classes of 5-year non-survivors and survivors, employing the Wilcoxon statistical test [21]. This was an important measure undertaken to reveal the distinct properties in nuclear texture, shape, nuclei distribution, or patient-specific features that differentiate the two classes. Additionally, the correlation between each feature and patients' survival, in terms of a two-stage decision, "5-year non-survivors" and "5-year survivors", was evaluated

employing the Point Biserial Correlation, designed for such dichotomous variables [21]. The importance of such measure was to verify the validity of trends in feature values with increasing patient survival.

Class discrimination

In designing an MIA-system to estimate 5-year survival of patients with laryngeal cancer, there were the following stages.

First, a classifier had to be chosen that was fast in execution and of high discriminatory ability. Following repeated experimentation, the probabilistic neural network (PNN) [22] classifier with Gaussian kernel was employed, as it is simple to design and fast in execution, since it does not incorporate sophisticated convergence algorithms and, for the particular problem at hand, performed equally well in precision as other more sophisticated classifiers, such as the SVM [23]. The PNN discriminant function is given by $G_j(X) = A \sum_{i=1}^{N_j} \exp(-||X - F_{ji}||/2s^2)$, where $A = 1/\left((2\pi)^{(n/2)} s^n N_j\right)$, X is the test pattern vector to be classified, n the number of features employed in the feature vector, s a smoothing parameter, here set to 0.2 after experimentation, N_j the number of patterns in class j, and $F_{j,i}$ the i-th training feature vector of class j. The input feature vector X belongs to the class j with the higher discriminant value, $G_j(X)$. As it may be deducted from the equation, for the classifier to be optimally designed for highest precision, a combination of features $F_{j,i}$ of high discriminatory power would have to be selected among the 77 available features. That best feature vector would have to be as small as possible and should additionally provide the highest MIA-system discrimination accuracy. One way to build high-precision MIA-systems is to examine the classification accuracy, by forming all possible combinations of the available features, and at each combination to test the precision of the designed system by means of a cross-validation method, such as the leave-one-out (LOO) method. The latter requires that the system is built by all but one pattern, which is then classified, and this process is repeated by sequentially testing all available patterns in both classes, eventually ending up with a number of correctly classified patterns that determines the MIA-system's precision at a particular feature combination. The optimal MIA-system design is then chosen as the design that employs a combination with the least number of features and provides the highest classification accuracy. However, testing all possible feature combinations would have required a large number of feature combinations to be formed, as dictated by the formula $\sum_{k=1}^{n} \frac{n!}{(n-k)!k!}$, where $n = 77$ and k the number of features in the combination [24]. The number of

feature combinations was reduced by following a rule of thumb [24], which states that to avoid overfitting, the maximum number of features employed in any combination should be less than one-third of the number of patterns in the smallest of the two classes. Considering that in the present study the smallest class contained 16 patterns, k in this equation should not exceed five-feature combinations. Nevertheless, the number of feature combinations to be tested was still large enough and would have become even larger when, at each feature combination, the LOO validation method was applied, which requires that the MIA-system is redesigned as many times as the total number of patterns in both classes (42 patterns). Practically, the task could not have been accomplished on a desktop computer, using sequential programming techniques, due to substantial time-processing requirements. Consequently, the design of the MIA-system was transferred to the multi-processors of an Nvidia Graphics Processing Unit (GPU) card, using CUDA (Compute Unified Device Architecture) toolkit v4.0, the C/C++ programming environment [25], and employing parallel programming techniques. The whole process has been presented in previous studies published by our group [12, 26]. In short, the graphics card used was the Nvidia Tesla K20c, equipped with 13 multiprocessors of 192 cores each, which were programmed in parallel. The GPU card was housed in a desktop computer equipped with the Intel QUAD Core CPU at 2.83 GHz processor and 4 GB of RAM. Parallel processing involved (a) breaking down the problem into small tasks, such as the design of the PNN by one feature combination and evaluation of its accuracy by the LOO validation method, (b) loading each task on a single GPU thread that runs concurrently along with numerous such tasks on other threads of the different cores of the GPU multiprocessors, and (c) transferring the results to the desktop computer, where the best design is isolated.

The MIA-system's performance was also verified by the AUC (area under the curve) of the ROC curve [24], using the PNN classifier, the LOO cross-validation method, and the same feature combination as the combination used in the best MIA-system design. AUC may vary between 0.5 and 1 and for AUC values greater than 0.9 the accuracy of discrimination is rated as "excellent". AUC is indicative of how well the PNN classifier, designed by a particular feature combination, can discriminate between the two classes of "5-year non-survivors" and "5-year survivors".

The MIA-system's generalization performance was tested by means of the Bootstrap [24] method, where the original dataset was split ten times in training and test sets, and at each split a contingency classification table (truth table) was formed. The mean overall accuracy would be an indicator of the proposed system's performance to new data.

Results

The expert physician examined the segmented images from all patients and it was found that the accuracy of the segmentation algorithm, as compared to the expert physician in correctly identifying nuclei was approximately close to 90 % (88.7 %), which was in line with similar findings of previous studies [14, 27–29]. The resulting segmented images were stored to be employed for further processing.

Features were subjected to the Wilcoxon non-parametric statistical test to identify those features that sustained statistically significant differences between the two classes of patients. Six such features were found, all emanating from the same family of features, calculated by means of the discrete wavelet transform (energy of the dwt-features) [24]. As shown in Table 2, statistically significant differences ($p < 0.05$) between the two classes were found for features calculated from the first-level decomposition (H_L1, V_L1, D_L1) and second-level decomposition (H_L2, V_L2, D_L2) of the nuclei images; H, V, and D stand for horizontal, vertical, and diagonal two-dimensional detail coefficient matrices, respectively, and L1 and L2 represent first-level and second-level decompositions, respectively. The values of all those features decreased with increasing laryngeal cancer survival, and this negative correlation was at a statistical confidence level of $p < 0.05$, as shown in the third column of Table 2. The last two columns of Table 2 show class means, standard deviations, and medians of the 5-year non-survivors and 5-year survivors.

Figure 2a–f shows the box plots of the six dwt-features and the class median values, represented by the horizontal line within each box, from where the class differences in dwt-feature values become evident. However, a closer look at the box plots would reveal that, although each dwt-feature sustained statistically significant difference between the two classes, there was a significant overlap in the distribution of the dwt-feature values in the two classes. This also becomes evident from the mean and standard

deviation values in Table 2. This limits the option of employing one such dwt-feature to propose an index for assessing the probability of a new patient as belonging to either class. For this to be feasible, one would have to combine information from different features, not necessarily only from those features sustaining statistically significant difference between the two classes, to separate the classes as further apart as possible in a multidimensional feature space. Such a feature combination could then be used to derive a multi-dimensional index or discriminant function that would predict with high precision if a new case would belong to either the 5-year non-survivor or 5-year survivor classes.

Such a design was achieved using the PNN classifier and the exhaustive search method for forming feature combinations, and the LOO evaluation method for assessing MIA-system precision, on the multiprocessors of the GPU. This best MIA-system design was achieved by a five-feature combination; sum average (mean), difference entropy (range), roughness index, level one diagonal detail coefficients matrix of discrete wavelet transform decomposition, and the tumor grade; the mean stands for the average of four feature values and range signify the maximum spread between the four feature values, which were calculated from the four co-occurrence matrices along the four main directions (E, NE, N, NW) on the nucleus surface [17]. The first two of the best textural features (sum average and difference entropy) were calculated from the co-occurrence matrices of the nucleus image, the third was evaluated from the shape of the nucleus outline (roughness index), the fourth (level 1 diagonal dwt-feature) was computed from the texture of the nucleus image, and the last (tumor grade) was quantified from the histological grade of the tumor.

Table 3 is a classification table that gives the accuracy with which the MIA-system assigned the laryngeal tumor cases into the two classes. As it may be observed, two 5-year survivor cases were mistakenly classified as 5-year non-survivors, resulting in a partial system accuracy (specificity) in survivors of 92.3 % and two 5-year non-

Table 2 Mean, median, and standard deviation of features that sustained statistically significant differences between 5-year non-survivors and 5-year survivors of laryngeal squamous cell carcinoma (see Table 3 for feature label explanation)

Feature label	Statistically significant difference ($p < 0.05$)	Correlation r/at p confidence level ($p < 0.05$)	(5-year non-survivors) mean value ± std/median	(5-year survivors) mean value ± std/median
H_L1	0.030564	−0.340/0.028	234,174 ± 40,494/227,166	198,714 ± 64,838/186,551
V_L1	0.020444	−0.364/0.018	244,623 ± 56,699/233,429	201,920 ± 71,032/183,427
D_L1	0.034781	−0.332/0.032	97,010 ± 18,404/93,262	83,387 ± 25,315/77,453
H_L2	0.025065	−0.352/0.022	314,493 ± 55,359/313,915	265,306 ± 89,339/248,935
V_L2	0.028626	−0.344/0.026	335,065 ± 83,859/320,681	275,758 ± 101,236/252,668
D_L2	0.034781	−0.332/0.032	85,037 ± 16,426/81,059	72,561 ± 22,858/66,331

H, V, and D stand for horizontal, vertical, and diagonal detail coefficients of the two-level (L1 and L2), two-dimensional discrete wavelet transform features
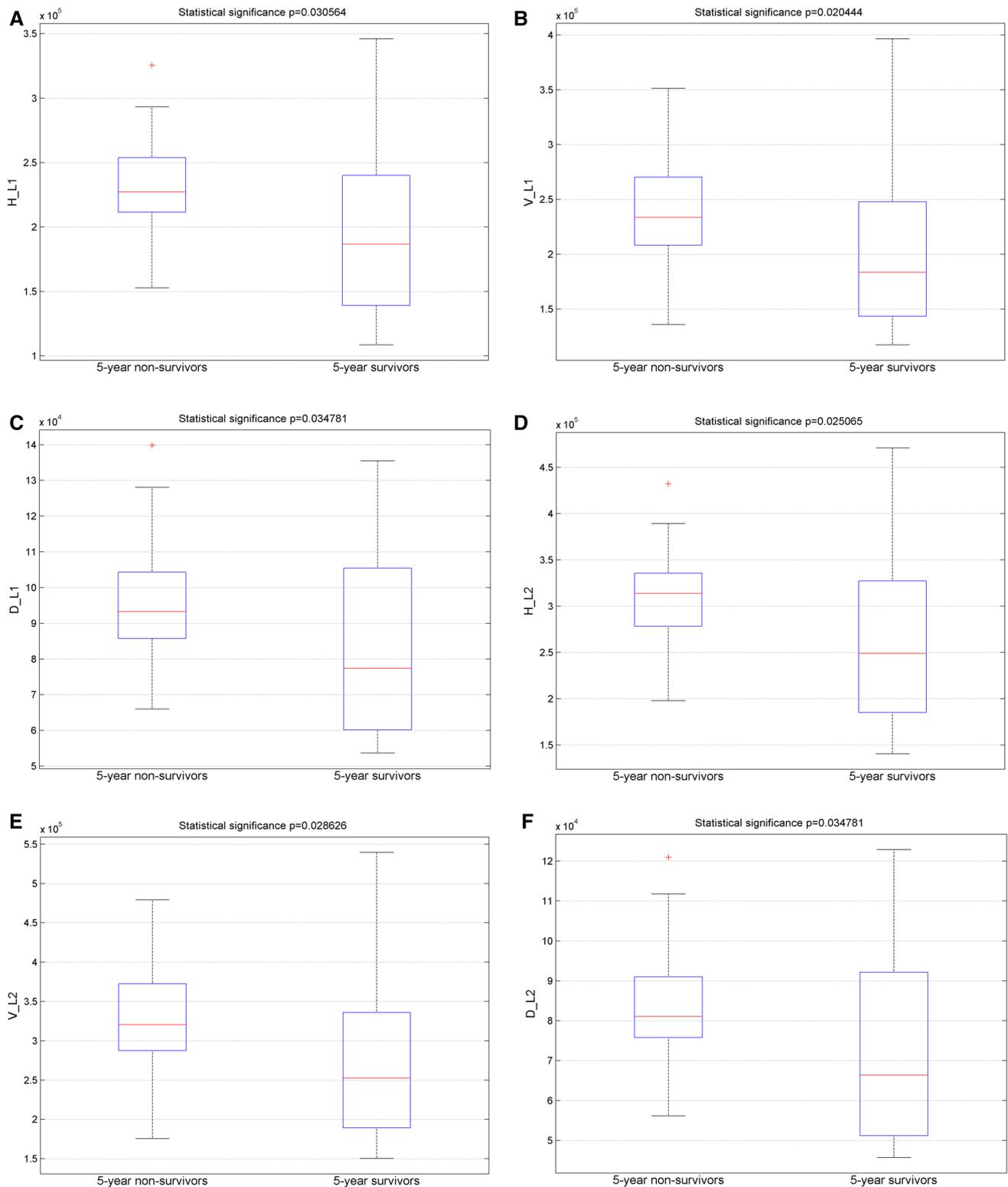
Fig. 2 Box plots of the discrete wavelet transform features, sustaining statistically significant differences ($p < 0.05$) between the '5-year non-survivors' and the '5-year survivors' with laryngeal squamous cell carcinomas. The H, V, D, L1, and L2 stand for the energy features calculated from the horizontal, vertical, diagonal, level 1, and level 2 detail coefficients of the discrete wavelet transformation, respectively

Table 3 Classification table of PR-system's best design

Truth table			
	Classified cases >5-year Survival	Classified cases >5-year Survival	Accuracy (%)
>5-year survival	24	2	92.3
≤5-year survival	2	14	87.5
			90.5

Five features involved: sum average (average), difference entropy (range), level 1 diagonal detail coefficients matrix of the two-dimensional discrete wavelet transform, roughness index, and tumor grade

survivors were incorrectly assigned to the wrong class, resulting in a partial system accuracy (sensitivity) for non-survivors of 87.5 %, thus, giving an overall MIA-system classification accuracy of 90.5 %. The Cohen–Kappa test statistic was also applied and gave an index value of 0.8, which was rated as a 'very good' indicator that the result was not achieved by chance.
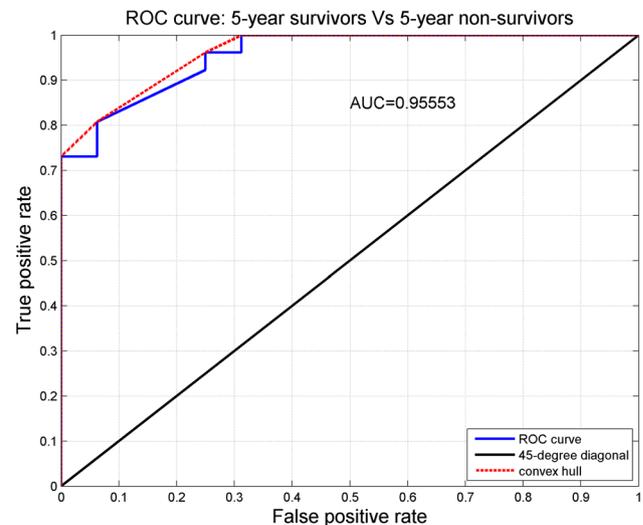
Class separability was also obtained by means of the ROC curve, when using those five best features, the PNN classifier, and the LOO cross-validation method. Figure 3 shows the ROC curve and the area under the curve (AUC = 0.96), which indicate an excellent degree of class separation.

The generalization performance of the MIA-system was evaluated by means of the bootstrap method over ten repetitions and the mean overall accuracy was found to be 85.2 ± 7.6 %.

Fig. 3 The receiver operating characteristic (ROC) curve between the '5-year non-survivors' and the '5-year survivors' with laryngeal squamous cell carcinomas using features determined in the best MIA-system design. The area under the curve is also presented

Discussion

The task of the present study was to examine the biopsy specimens of patients diagnosed with laryngeal squamous cell carcinoma, analyze microscopy image parameters, examine patient-related data for identifying differences between short-time (<5 years) and long-time (>5 years) survivors of laryngeal cancer, and finally design a microscopy image analysis system for assessing the patients' 5-year survival. H&E and p63 staining were used on specimens of the biopsy material. Patients were split into two groups: those who had survived the 5-year threshold and those who had not.

First, using p63-stained digitized images from each patient, associations were sought of the nuclei properties or features, such as texture, shape, morphology, and distribution of p63-stained nuclei, as well as patient-related data with the patient's 5-year survival; second, a decision support system was designed, based on image analysis methods and features, for estimating the patient's 5-year survival. Image-processing and statistical analysis tasks were performed on a conventional desktop computer and image analysis tasks were accomplished on the

microprocessors of a graphics-processing unit card using parallel-processing techniques.

Previous studies that have investigated the survival of patients with laryngeal cancer, employing computer analysis of histopathologic images [8, 9], have evaluated cell nuclei features regarding texture, morphology, or shape, using H&E-stained images of biopsy specimens. They have found that fractal dimension [8], cell concentration, nuclear area, perimeter, density, and roundness [9] were important cell nuclei features, and associated threshold values in those features with patient survival. In the present study, we used images from laryngeal cancer biopsy material that had been IHC stained for p63 expression, since p63 mainly targets the cell nuclei and probably renders the nuclei's texture and outline to be more discernible. Statistical analysis showed that out of 77 features, only those calculated from the two-dimensional discrete wavelet transform sustained statistically significant differences between the two classes of patients. There were no significant differences found between classes regarding fractal dimension, as in [8], or morphological features as in [9], although

different IHC-staining methods were employed. However, the importance of textural, shape, and patient-related features was revealed in the present study, by means of the five features used in the best design of the MIA-system.

From the box plots of Fig. 2 and the contents of Table 2, it can be deduced that the values of all six dwt-features decreased with good prognosis; 5-year survivors demonstrated lower values of dwt-features. This decrease was also verified by examining the correlation of those features with improving prognosis. Negative correlation with higher survival was established at a statistically confident level ($p < 0.05$). Regarding the textural information that these features hold, dwt-features depict the energy (or the sum of squares of pixel values) of images produced at two levels of decomposition as follows. At the first level (L1), the 'horizontal detail' (H) image was produced by processing the original nucleus image by vertical high-pass and horizontal low-pass filtering and it contains mainly the horizontal edges of the original image. The 'vertical detail' image (V) was obtained by vertical low-pass and horizontal high-pass filtering and mainly contains the vertical edges of the original image. The 'diagonal detail' image (D) was obtained by vertical and horizontal high-pass filtering and contains the edges of the original image in the diagonal direction. At the second level (L2), features were calculated from detailed images (H, V, and D), produced in a similar way that images were formed at the first level, but at a lower resolution. In accordance with our findings, and since edges reflect high-frequency image content as well as image coarseness, it may be assumed that patients who survived 5 years after laryngeal cancer diagnosis appeared to have smoother nuclei images as compared to those who did not survive and whose nuclei structure appeared coarser.

Regarding the design of an MIA-system to predict patient's 5-year survival, there were few factors to consider that influenced such an attempt. Such factors included the dimensionality of the problem, the choice of the classifier, and the need to design a system that would have high precision in classifying data not used in the system design [30].

Dimensionality poses serious computer time demands, which were dealt with by transferring system design onto the microprocessors of a graphics processing unit card using parallel processing techniques. It took about 3.3 min for a desktop computer, equipped with the Nvidia GPU, to design and test the MIA-system precision, using the LOO cross-validation method.

Regarding the choice of classification algorithm, those considered were [24, 31] the nearest neighbor, the Bayesian, the probabilistic neural network, the linear discriminant analysis, and the support vector machine. Classifiers were tested with regard to speed of execution, robustness, computational complexity, and accuracy in discriminating between the classes of the present study. The PNN classifier, employing the Gaussian kernel, was the classifier of choice, due to its speed of execution and accuracy, around which the classification system was built.

Results showed that the best feature combination that produced the MIA-system design with the highest classification accuracy included features from the nucleus's texture (sum average mean, difference entropy range, energy at level-one dwt diagonal detail image), the shape (roughness index), and the lesion's severity (grade). Sum average assesses gray-level inhomogeneity, in terms of the existence of structures with variation in gray levels and attains higher values in coarser textures. Difference entropy range is a measure of anisotropy in gray-levels randomness and attains high values for equally distributed gray-tone differences or for images with low variation in image contrast. Energy at level-one dwt diagonal detail image assesses image coarseness and was the only feature in the best feature combination that sustained statistically significant difference between 5-year survivors and non-survivors. Roughness index is a measure of the nucleus outline irregularity, and tumor grade is a feature related to the severity of the disease. The results are presented in Table 3. Two cases of the "5-year survivor" class were wrongly assigned to the "5-year non-survivor" class and two cases of the latter were wrongly classified as 5-year survivors, giving an overall accuracy of 90.5 %. The generalization of such performance was put to test by a less biased method (bootstrap) that revealed a mean accuracy of about 85 %. Another way of looking at the between-classes discrimination is by constructing the ROC curve using the PNN classifier, the same best feature combination, and the LOO cross-validation method. The area under the curve of 0.96 indicates excellent class separation by the particular MIA-system design.

Conclusions

Summarizing, an image analysis system has been designed to discriminate the laryngeal cancer cases of the present study, with reasonably high accuracy, into "5-year non-survivors" and "5-year survivors", employing immunohistochemical staining for p63 expression and histopathology images. MIA-system design was feasible by transferring data processing onto the processors of a graphics processing unit and applying parallel processing methods. Features related to the nucleus image high-frequency content, the anisotropy of the nucleus image gray-level randomness, the nucleus shape, and the tumor grade contributed in the design of the best MIA-system. Discrete wavelet transform features, related to the existence of

edges in the nucleus texture, revealed that nucleus texture in 5-year non-survivors was coarser at a statistically significant level. The proposed system was designed so as to function in a clinical environment, as a second opinion diagnostic tool.

References

1. Talamini R, Bosetti C, La Vecchia C, Dal Maso L, Levi F, Bidoli E, Negri E, Pasche C, Vaccarella S, Barzan L, Franceschi S (2002) Combined effect of tobacco and alcohol on laryngeal cancer risk: a case-control study. Cancer Causes Control 13(10):957–964
2. Sant M, Allemani C, Santaquilani M, Knijn A, Marchesi F, Capocaccia R (2009) EUROCARE-4. Survival of cancer patients diagnosed in 1995–1999. Results and commentary. Eur J Cancer 45(6):931–991
3. Pruneri G, Pignataro L, Manzotti M, Carboni N, Ronchetti D, Neri A, Cesana BM, Viale G (2002) p63 in laryngeal squamous cell carcinoma: evidence for a role of TA-p63 down-regulation in tumorigenesis and lack of prognostic implications of p63 immunoreactivity. Lab Invest 82(10):1327–1334
4. Dong P, Li X, Zhu Z, Yu Z, Lu G, Sun Z, Wang S (2004) Application of tissue microarray: evaluation of the expression of S-100-positive dendritic cells, tumor suppressor gene p63 and tissue inhibitor of metalloproteinase-1 in laryngeal carcinoma. Acta Otolaryngol 124(10):1204–1207
5. Takahashi Y, Noguchi T, Takeno S, Kimura Y, Okubo M, Kawahara K (2006) Reduced expression of p63 has prognostic implications for patients with esophageal squamous cell carcinoma. Oncol Rep 15(2):323–328
6. Borba M, Cernea C, Dias F, Faria P, Bacchi C, Brandao L, Costa A (2010) Expression profile of p63 in 127 patients with laryngeal squamous cell carcinoma. ORL J Otorhinolaryngol Relat Spec 72(6):319–324
7. Re M, Zizzi A, Ferrante L, Stramazzotti D, Goteri G, Gioacchini FM, Olivieri F, Magliulo G, Rubini C (2014) p63 and Ki-67 immunostainings in laryngeal squamous cell carcinoma are related to survival. Eur Arch Otorhinolaryngol 271(6):1641–1651
8. Delides A, Panayiotides I, Alegakis A, Kyroudi A, Banis C, Pavlaki A, Helidonis E, Kittas C (2005) Fractal dimension as a prognostic factor for laryngeal carcinoma. Anticancer Res 25(3B):2141–2144
9. Dobroś W, Gil K, Chłap Z, Olszewski E (1999) The use of nuclear morphometry for the prediction of survival in patients with advanced cancer of the larynx. Eur Arch Otorhinolaryngol 256(5):257–261
10. Bacauskiene M, Verikas A, Gelzinis A, Valincius D (2009) A feature selection technique for generation of classification committees and its application to categorization of laryngeal images. Pattern Recogn 42(5):645–654
11. Verikas A, Gelzinis A, Valincius D, Bacauskiene M, Uloza V (2007) Multiple feature sets based categorization of laryngeal images. Comput Methods Programs Biomed 85(3):257–266
12. Ninos K, Kostopoulos S, Sidiropoulos K, Kalatzis I, Glotsos D, Athanasiadis E, Ravazoula P, Panayiotakis G, Economou G, Cavouras D (2013) Computer-based image analysis system designed to differentiate between low-grade and high-grade laryngeal cancer cases. Anal Quant Cytol Histol 35(5):261–272
13. Egner JR (2010) AJcc cancer staging manual. JAMA 304(15):1726–1727
14. Kostopoulos S, Cavouras D, Daskalakis A, Kagadis GC, Kalatzis I, Georgiadis P, Ravazoula P, Nikiforidis G (2008) Cascade pattern recognition structure for improving quantitative assessment of estrogen receptor status in breast tissue carcinomas. Anal Quant Cytol Histol 30(4):218–225
15. Cheng HD, Sun Y (2000) A hierarchical approach to color image segmentation using homogeneity. IEEE Trans Image Process 9(12):2071–2082
16. Jain AK (1989) Fundamentals of digital image processing. Prentice-Hall Inc, USA
17. Haralick RM, Shanmugam K, Dinstein IH (1973) Textural features for image classification. IEEE Trans Sys Man Cyb 3:610–621
18. Galloway MM (1975) Texture analysis using gray-level run lengths. Comput Graph Image Process 4:172–179
19. Huang PW, Lai YH (2010) Effective segmentation and classification for HCC biopsy images. Pattern Recogn 43(4):1550–1563
20. Bandyopadhyay S, Saha S (2013) Unsupervised classification: similarity measures, classical and metaheuristic approaches, and applications. Springer-Verlag, Berlin
21. Le CT (2003) Introductory biostatistics. Wiley, New York
22. Specht DF (1990) Probabilistic neural networks. Neural Netw 3:109–118
23. Kechman V (2001) Support vector machines. Learning and soft computing. MIT, Cambridge, pp 121–184
24. Theodoridis S, Koutroumbas K (2008) Pattern recognition, 4th edn. Academic Press, USA
25. Kirk DB, W-mW Hwu (2010) Programming massively parallel processors a hands-on approach. Morgan Kaufmann, Amsterdam
26. Sidiropoulos KP, Kostopoulos SA, Glotsos DT, Athanasiadis EI, Dimitropoulos ND, Stonham JT, Cavouras DA (2013) Multi-modality GPU-based computer-assisted diagnosis of breast cancer using ultrasound and digital mammography images. Int J Comput Assist Radiol Surg 8(4):547–560
27. Glotsos D, Spyridonos P, Cavouras D, Ravazoula P, Dadioti PA, Nikiforidis G (2004) Automated segmentation of routinely hematoxylin-eosin-stained microscopic images by combining support vector machine clustering and active contour models. Anal Quant Cytol Histol 26(6):331–340
28. Latson L, Sebek B, Powell KA (2003) Automated Cell Nuclear Segmentation in Color Images of Hematoxylin and Eosin-Stained Breast Biopsy. Anal Quant Cytol Histol 25(6):321–331
29. Spyridonos P, Cavouras D, Ravazoula P, Nikiforidis G (2002) A computer-based diagnostic and prognostic system for assessing urinary bladder tumour grade and predicting cancer recurrence. Med Inform Internet Med 27(2):111–122
30. Shamir L, Delaney JD, Orlov N, Eckley DM, Goldberg IG (2010) Pattern recognition software and techniques for biological image analysis. PLoS Comput Biol 6(11):e1000974
31. Theodoridis S, Pikrakis A, Koutroumbas K, Cavouras D (2010) Introduction to pattern recognition: a matlab approach. Academic Press