Immunohistochemical Analysis and Prognostic Value of Cathepsin D Determination in Laryngeal Squamous Cell Carcinoma

Sven Seiwerth,[†] Nikola Štambuk,*[,‡] Paško Konjevoda,[§] Nikola Mašić,[‡] Ankica Vasilj,[†] Miljenko Bura,[||] Ivo Klapan,[||] Spomenka Manojlović,[†] and Davor Đanić[#]

Institute of Pathology, Department of Pharmacology, and ENT Clinic, Medical Faculty University of Zagreb, Šalata 10, Zagreb, Croatia, Rudjer Bošković Institute, Bijenička cesta 54, HR-10001 Zagreb, Croatia, and ENT Department County Hospital, Andrije Štampara 42, Slavonski Brod, Croatia

Received June 25, 1999

Cathepsin D, a protease with the capability of degrading matrix proteins, is implicated in the process of breast and colorectal cancer invasion and metastasis. Biochemical studies in laryngeal cancer have shown a potential prognostic significance of cathepsin D content determination. We studied immunohistochemical positivity of cathepsin D in tumor epithelium and stroma of 61 surgical specimens of squamous cell laryngeal cancer. Immunohistochemical reaction was quantitatively assessed using a PC-based image analysis system SFORM-VAMS. The results were correlated to clinical and morphological parameters and survival. Immunohistochemical positivity was noted in neoplastic cells and tumor stroma. Significant prognostic value for cathepsin D was established separately for epithelial tumor component and tumor stroma using log-rank test, the Cox proportional hazards regression model, and C4.5 machine learning system. In all groups, patients above the median cathepsin D staining showed significantly shorter survival time. C4.5 machine learning system extracted cutoff values for the decision tree that defines the probabilities of patients survival and death with high sensitivity (92.8% alive, 73.6% dead), 100% specificity, and 86.9% accuracy. This makes immunohistochemical cathepsin D estimation an independent prognostic parameter in laryngeal carcinomas within a 5-year period from the time of tumor surgery.

INTRODUCTION

Cathepsin D is a lysosomal acidic protease[1] thought to be closely associated with tumor invasion or metastasis due to its capability of degrading extracellular matrix.[2] In histopathological and clinical studies, overexpression of cathepsin D was connected with aggressive tumor behavior in different neoplastic diseases.[3−5] Immunochemically the presence of cathepsin D was demonstrated in normal laryngeal mucosa and in primary laryngeal squamous cell carcinomas (SCC).[6] Recently, a study using radioimmunoasay correlated high cathepsin D content with a poor prognosis, independent of lymph node status.[7] Immunohistochemically cathepsin D was demonstrated in neoplastic and normal laryngeal mucosal cells as well as in stromal macrophages.[8] The aim of our study was to correlate immunohistochemical expression of cathepsin D to clinical and morphological parameters as well as survival in a laryngeal SCC. We compared the analysis of the results by means of a standard log-rank test and the Cox proportional hazards regression model to the evaluation of the results by means of the C4.5 machine learning system that extracts the decision tree for the classification of the patients survival.

* Corresponding author phone: +385 1 4680 193; fax: +385 1 4680 094; e-mail: stambuk@rudjer.irb.hr.
[†] Institute of Pathology, Medical Faculty University of Zagreb.
[‡] Rudjer Bošković Institute.
[§] Department of Pharmacology, Medical Faculty University of Zagreb.
[||] ENT Clinic, Medical Faculty University of Zagreb.
[#] ENT Department County Hospital.

MATERIAL AND METHODS

Patients. We investigated 61 consecutive cases of previously untreated laryngeal SCC patients (males and smokers), without detectable distant metastases, randomly selected from our archives. Figures 1 and 2 illustrate data concerning TNM stages[9] and histopathologic type.[9]

Immunohistochemistry. The tissue samples were collected from laryngectomy specimens, fixed in 10% buffered formalin, routinely processed, and embedded in paraffin. Three micrometer sections were mounted on silanised slides and stained immunohistochemically with an anti-cathepsin D antibody (DAKO, Glostrup) using the avidin−biotin method (ABC, Vector, Burlingham) according to the manufacturer's specification (Figure 3, cathepsin D = dark brown staining). All the slides were stained in one batch, by one technical assistant.

Image Analysis. Immunoreactivity was analyzed by two of the authors (SS and AV) on a Leitz Diaplan microscope, using a PC based image analysis system SFORM-VAMS (Zagreb, Croatia; http://www.vams.com)[10] and a CCD camera (JVC TK 1270). Background lightening was kept constant and uniform, and a standard blue filter was used. The immunoreactive area was accessed as the percentage of the total area analyzed (four fields, objective ×25). Immunoreactivity was analyzed separately for epithelial and stromal cells.

Data Analysis. Analyzed variables were as follows: cathepsin D immunoreactivity (separately for tumor stroma and tumor epithelium), histopathologic grade (grade I and

10.1021/ci990075q CCC: $19.00
Published on Web 04/27/2000

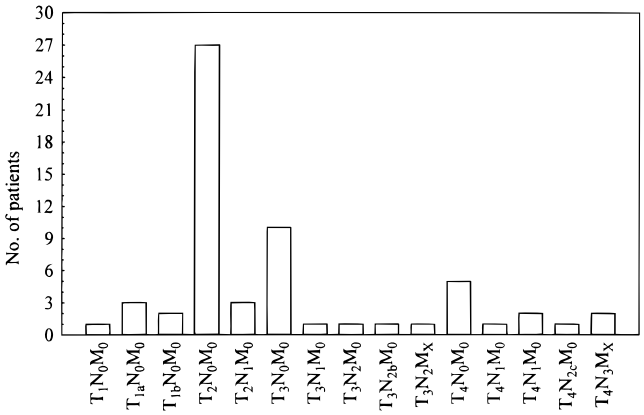

Figure 1. Distribution of patients by TNM classification.

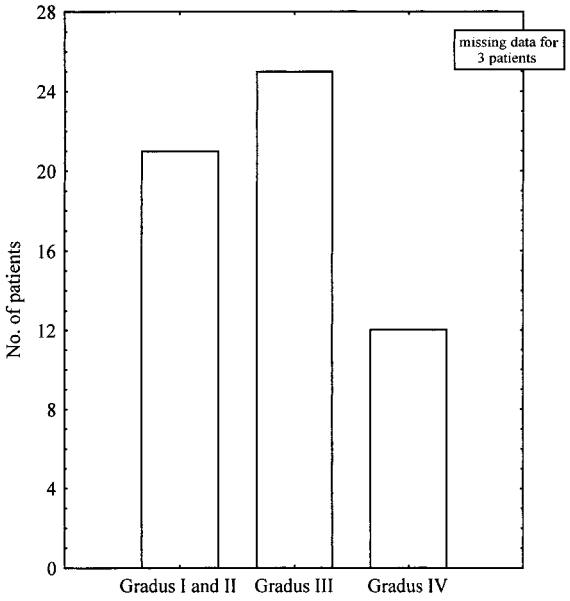

Figure 2. Distribution of patients by Gradus (I−IV).

II together), clinical TNM stage, and survival. Analysis was made using the Kaplan-Meier curves and the log-rank test,[11] the Cox proportional hazards regression model,[12] and the C4.5 decision tree learning algorithm.[13−18]

C4.5 Machine Learning Program. The program C4.5 is a successor of the basic ID3 decision tree learning algorithm.[13,15,18]

ID3 and C4.5 define the possible decision tree by means of a hill-climbing search based on the statistical property measure called information gain.[13,15,16,18] Information gain measure defines how well a given attribute separates the training examples according to their target classification and selects candidate attributes at each step of the tree.[13,15,18] Consequently, this measure is the expected reduction in Shannon's entropy caused by partitioning the examples according to the attribute in classifying the training data.[13,15,18] ID3 algorithm is specified to learning boolean-valued functions, and it grows the tree top-down at each node selecting the attribute that best classifies local training examples.[13,15,18] The process continues until the tree accurately classifies the training data or until all attributes are used.[13,15,18]

The elements of the tree generated by ID3 and C4.5 are either leafs or decision nodes.[13−18] The leaf shows a class, and the decision node specifies the test to be implemented on an attribute value, with one branch and subtree for each

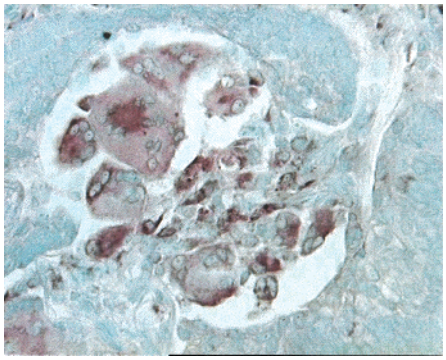

Figure 3. Cathepsin D positivity in tumor cells (×40).

possible result of the test.[13,15−18] The starting node is the root node, and a tree is used to predict a case by starting at the root and moving through the tree until the leaf is encountered.[13,15,17,18] For any tree, all paths lead to a leaf corresponding to a decision rule that is a logical conjunction of various tests.[15,18] If there are multiple paths for a given class, then the paths represent logical disjunctions.[13,15,17,18] All paths are mutually exclusive.[13,15,18] For any new case, one and only one path in the tree will always have to be satisfied.[13,15,18]

C4.5 is an outgrowth of the original ID3 algorithm based on the technique called "rule post pruning".[15,18] This procedure enables the finding of high accuracy hypothesis and prevents overfitting (i.e. misclassification of the data by 10−25%) when there is noise in the data or when the size of the training sample is too small.[18] Each attribute test along the path from the root to the leaf of ID3 is pruned (generalized) by removing those leaves that do not improve the accuracy of the tree prediction.[15,18] Then, the pruned results are sorted according to their estimated accuracy to obtain the final decision tree algorithm.[15,18]

C4.5 additionally improves the accuracy of the prediction by using a pessimistic estimate of the rule performance (observed accuracy over the training set minus 1.96 times the estimated standard deviation).[15,18] This heuristic method has been proved to be very useful in practice, especially when we have larger data sets.[18]

Sensitivity, specificity, and accuracy of the procedures were obtained in a standard way.[19] Predictivity, i.e., reliability of the classifier predictions, was calculated as a ratio of the number of true predictions to the size of appropriate prediction class (*alive* or *dead*).[13,15,17]

RESULTS

Histopathological Analysis. Follow up of the patients was from 4 to 108 months with a median of 60 months. Forty-two patients were censored (group *alive*), and 19 were completely observed (group *dead*).[12] Cathepsin D immunoreactivity was histologically observed in normal and neoplastic tissue. In normal laryngeal mucosa next to a tumor, diffuse, weak cytoplasmatic positivity was noted. Scarce reactivity was also present in stromal macrophages. Neoplastic epithelial cells showed mostly diffuse positivity ranging from an occasional cell to the majority of cells, with a slightly stronger expression in more dedifferentiated cells. In the tumor stroma abundant immunoreactive cells (macrophages) were noted. Scarce apical reaction in the stroma of the salivary gland cells was also noted.

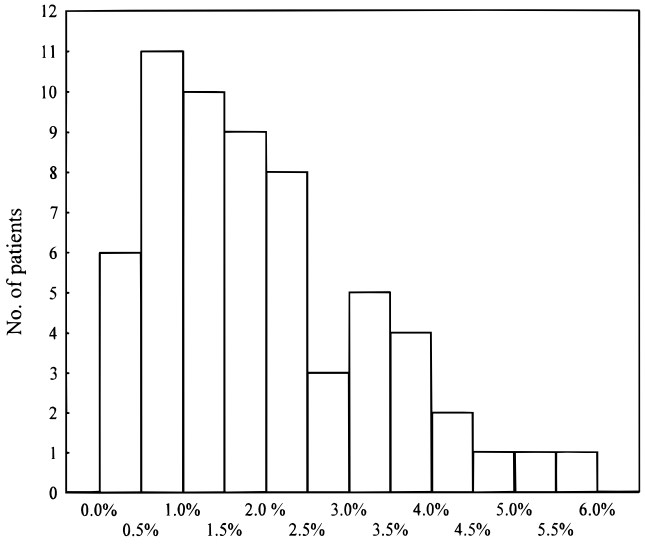

Figure 4. Distribution of patients by cathepsin D positive area (%) in the epithelial component of tumor.

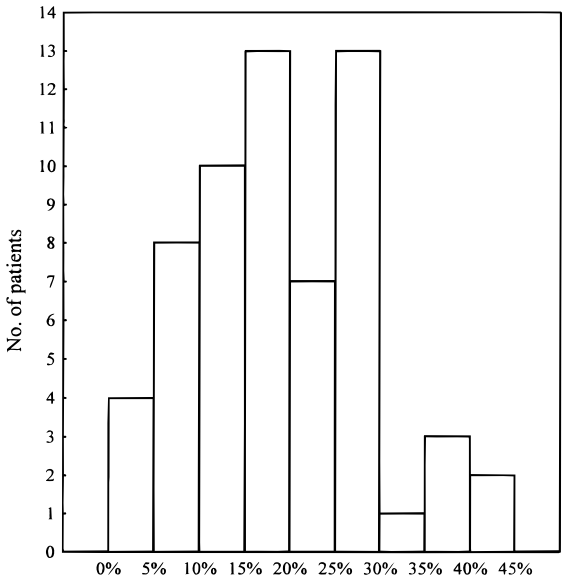

Figure 5. Distribution of patients by cathepsin D positive area (%) in tumor stroma.

Kaplan-Meier Curves and Log-Rank Test. In the tumor epithelium cathepsin reactivity ranged from 0.06% to 5.63% with a median of 1.35%, while in the stroma its range was from 0.62% to 42.02% with a median of 9.86% (Figures 4 and 5.). We compared cathepsin D expression with conventional prognostic factors. There was no significant correlation of cathepsin D immunoreactivity with clinical TNM stages (Spearman rank order correlation = 0.24 and $p > 0.05$) as well as with histopathological grading (Spearman rank order correlation = 0.08 and $p > 0.05$). Immunoreactivity for cathepsin D, in both tumor epithelial cells and stroma respectively, showed strong influence on patient survival (Figures 6 and 7). Clinical TNM status (Chi-square = 3.896, df = 3, $p = 0.273$) and a histological grade (Chi-square = 3.739, df = 2, $p = 0.154$) showed no significant influence on patient survival.

The Cox Proportional Hazards Regression Model. The Cox proportional hazards regression model suggested that only cathepsin D immunoreactivity in epithelial cells has a statistically significant effect ($p < 0.05$, Table 1).

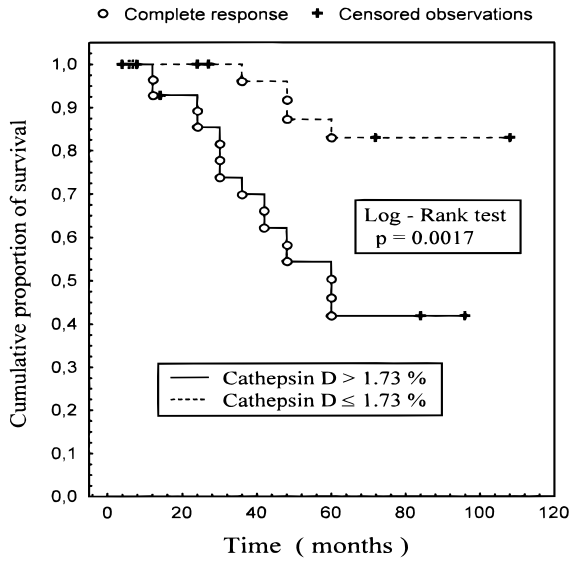

Figure 6. Survival analysis for cathepsin D positive area (%) in the epithelial component of tumor (Kaplan-Meier).

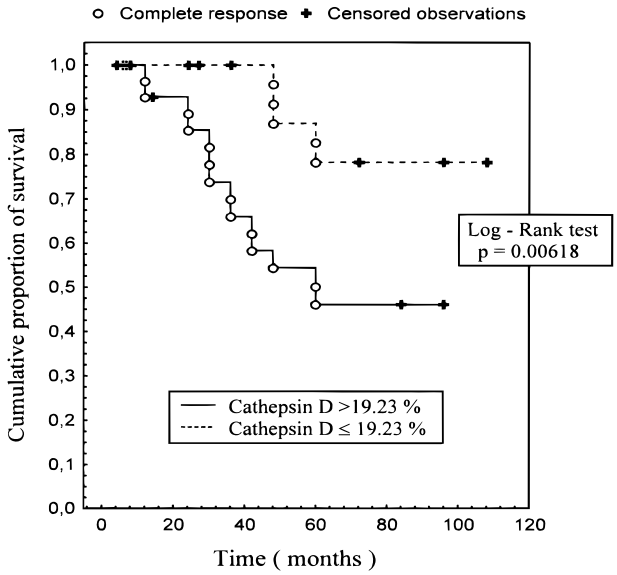

Figure 7. Survival analysis for cathepsin D positive area (%) in tumor stroma (Kaplan-Meier).

C4.5 Machine Learning System. For the analysis with the C4.5 classifier the patients were divided into two groups (*dead* or *alive* after 60 months, i.e., 5 years following the surgical procedure). A decision tree extracted cathepsin D epithelial and stromal tumor cell staining as the most significant for the classification of the *alive* or *dead* groups of patients, i.e., the prognosis. The classification rules depicted in Figure 8 can be read as follows:

1. The first rule says that if the value of RED (epithelial cathepsin D) is less than or equal to 2.33%, then a patient belongs to a group *alive*.

2. If that rule is not satisfied (i.e. RED (epithelial cathepsin D) is more than 2.33%), then the group is alive when BLUE (stromal cathepsin D) is above 38%.

3. The group is *dead* when RED (epithelial cathepsin D) is more than 2.33% and BLUE (stromal cathepsin D) is equal or less than 38%.

The test showed high sensitivity by accurately predicting 5-year survival following the surgical procedure in 92.8%

Table 1. Cox Proportional Hazard Risk Model for Clinical Stage—Cathepsin D Content in Epithelial and Stromal Component of Tumor and Histopathological Gradus

Cox model	Chi-square = 18.3673, df = 4, $p = 0.0015$			
variables	estimate	SE	*t*-value	two-sided *p* value
clinical stage	0.43	0.28	1.51	0.13
cathepsin D (epithel)	0.83	0.32	2.30	0.009
cathepsin D (stroma)	−0.04	0.05	0.96	0.46
histopathological grade	0.16	0.29	0.54	0.59

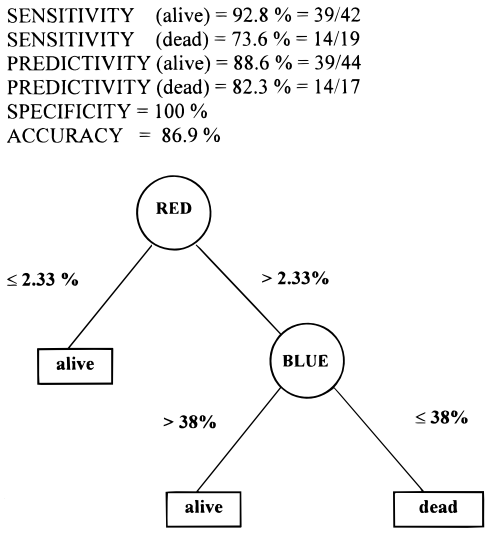

Figure 8. Decision tree obtained by C4.5 machine learning system.

of the patients with laryngeal squamous cell carcinoma (Figure 8). The classifier's sensitivity in predicting death due to the tumor progression, within a 5-year period following surgery, was satisfactory at 73.6% (Figure 8). The specificity of the test was 100% since the decision tree evaluation is made on the tumor tissue, which is absent in the normal laryngeal immunohistochemical sample. Therefore all patients without the tumor have a priori negative test results with respect to the specificity evaluation.[19] The accuracy of the test was also high (86.9%) and the reliability of the classifier's prediction (predictivity) was 88.6% for the group *alive* and 82.3% for the group *dead* (Figure 8).

DISCUSSION

Expression of cathepsin D was analyzed in different malignancies such as breast,[3,20,21] melanoma,[22] colorectal,[23] and head and neck cancer[24] as well as in childhood or nervous system neoplasm.[25,26] It was suggested that cathepsin D plays a role in tumor cell proliferation by growth factor activation or promotes tumor invasion and metastasis by activating proteolytic enzymes.[1,2]

In laryngeal carcinomas different studies established by immunometric assays show a higher cathepsin D content in tumor tissue samples, as compared with normal laryngeal mucosa.[6,7] Immunohistochemically strong reactivity was demonstrated for cathepsin D in tumor cells and in tumor stroma macrophages that infiltrate the tissue. In this context a very high stromal cathepsin D value linked to the patients survival in a subgroup of patients (Figure 8) may be due to the enhanced local immune response to tumor antigens.[27]

The C4.5 decision tree learning algorithm (Figure 8) was superior to Cox's model (Table 1), regarding the analysis of data structure, since it extracted cutoff values of both epithelial and stromal cathepsin D content relevant for survival. Cox's model did not extract stromal cathepsin D as an important prognostic parameter due to the small group of patients with an extremely high content of stromal cathepsin D. The log-rank test showed the statistical significance of epithelial and stromal cathepsin D content for patients survival (Figures 6 and 7) but failed to explain good prognosis for the subgroup of the patients with extremely high stromal cathepsin D content. C4.5 based analysis was particularly appropriate for the data set investigated because it enabled accurate prediction of all attributes (epithelial and stromal cathepsin D) and all groups of patients, including the one with an extremely high content of stromal cathepsin D.

Stromal cathepsin D values of ≤ 38%, linked to the tumor progression (Figure 8), probably reflect enhanced activity of protease concerning the metastasis[1,2] and less pronounced immune response. It is worth mentioning that, with respect to C4.5 based classification, the cutoff prognostic value of stromal cathepsin D is the descendant node of the best predicting attribute[18] (i.e. epithelial tumor cathepsin D) which represents the most informative and root node of the tree.[15,18] The best attribute defining of tumor epithelial cathepsin D by means of the C4.5 classifier is in agreement with the fact that laryngeal squamous cell carcinoma is an epithelial neoplasm.

Prognostic significance was recently assumed for radiometrically measured cathepsin D levels;[6,27,28] however, to our knowledge this is the first immunohistochemical study demonstrating the strong prognostic significance of cathepsin D in laryngeal cancer. Besides being highly prognostic, this type of immunochemical test is relatively cheap and easy to perform which makes the combination of quantitative immunohistochemical analysis of cathepsin D and C4.5 based data classification a potent prognostic tool in laryngeal cancer patients.

Although the surgical treatment is *condicio sine qua non* (prerequisite) in the therapy of laryngeal squamous cells carcinoma, it seems that cathepsin D content in histopathology samples of the tumor cells represents an important predictive factor for tumor recidives and aggressive behavior. In this study the data analysis was not influenced by other therapeutic procedures (e.g. chemotherapy or radiotherapy) due to the fact that surgery is the primary therapeutic procedure for the type and stage of laryngeal neoplasm we observed (Figure 1). It is worth mentioning that all of our patients were males and smokers.

In our study expression and localization of cathepsin D immunoreactivity correlated with data obtained by others for laryngeal neoplasms.[27,28] From our results, cathepsin D seems to be an independent prognostic marker in primary laryngeal carcinomas, which confirms the hypothesis of Marsigliante et al.[28] Decision tree extracted by means of C4.5 classifier was shown to be a valuable tool to define highly sensitive, specific, accurate, and predictive cutoff values for immunohistochemical cathepsin D data. It remains an open question if this method of analysis could be applied to other tumors (e.g. breast, colorectal, and gactric cancer)[3−5,21,29−31] with an established link between aggressive neoplastic behavior and cathepsin D content.

REFERENCES AND NOTES

(1) Capony, F.; Morrisset, M.; Barrett, A. J.; Capony, J. P.; Broquet, P.; Vignon, F.; Chambon, M.; Louist, P.; Rochefort, H. Phosphorylation, glycosilation and proteolytic activity of the 52-kD estrogen-induced protein secreted by MCF7 cells. *J. Cell Biol.* **1987**, *104*, 253−262.

(2) Montcurrier, P.; Mangeant, PH.; Salazar, G.; Morisset, M.; Sahuquet, A.; Rochefort, H. Cathepsin D in breast cancer can digest extracellular matrix in large acid vesicles. *Cancer Res.* **1990**, *50*, 6045−6054.

(3) Kute, T. E.; Shao, Z. M.; Sugg, K.; Long, R. T.; Russell, G. B.; Case, D. Cathepsin D as prognostic indicator for node-negative breast cancer patients using both immunoassay and enzymatic assay. *Cancer Res.* **1992**, *52*, 5198−5203.

(4) Rochefort H.; Capony, F.; Garcia, M. Cathepsin D: A protease involved in breast cancer metastasis. *Cancer Metastasis Rev.* **1990**, *9*, 321−331.

(5) Leto, G.; Gebbia, N.; Rausa, L.; Tuminello, FM. Cathepsin D in the malignant progression of neoplastic diseases. *Anticancer Res.* **1992**, *12*, 235−240.

(6) Marsigliante, S.; Resta, L.; Leo, G.; Mazzotta, D.; d'Amore, R.; Biscozzo, L.; Storelli, C. Expression of cathepsin D in malignant an in the corresponding nonmalignant node-negative laryngeal samples: correlation with receptors for androgen, glucocorticoid, oestrogen and progesterone. *Cancer Lett.* **1993**, *68*, 135−142.

(7) Maurizi, M.; Almadori G.; Cadoni, G.; Scambia, G.; Ottaviani, F.; Ferrandina, G.; Paludetti, G.; D'Abramo, G.; Manusco, S. Cathepsin D concentration in primary laryngeal cancer: correlation with clinicopathological parameters, EGFR status and prognosis. *Int. J. Cancer* **1996**, *69*, 105−109.

(8) Ferrandina, G.; Scambia, G.; Benedetti, P.; Almadori, G.; Paludetti, G.; Cadoni, G.; Distefano, M.; Maurizi, M.; Mancuso, S. Cathepsin D in primary squamous laryngeal tumors: correlation with clinicopathological parameters and receptor status. *Cancer Lett.* **1992**, *67*, 133−138.

(9) American Joint Commitee on Cancer. Manual for Staging on Cancer; JB Lippincot: Philadelphia, PA, 1988; p 43.

(10) Konjevoda, P.; Nasić, M.; Ćurković, T.; Sikirić, P.; Seiwerth, S.; Štambuk, N. Effects of BPC157 on the healing of corneal lesions. In *Uveitis Today*; Ohno, S., Aoki, K., Usui, M., Uchio, E., Eds.; Elsevier: Amsterdam, 1998; pp 311−314.

(11) Bartels, P. H.; Thompson, D.; Weber, J. E. In *Image Analysis − A primer for Pathologist*; Marchevsky, A. M., Bartels P. H., Eds.; Raven Press: New York, 1994; Chapter 7, p 207.

(12) Afifi, A. A.; Clark, V. *Computer-aided Multivariate Analysis*; Chapman & Hall: London, 1996; Chapter 13, p 306.

(13) Quinlan, J. R. Induction of Decision Trees. *Machine Learning* **1986**, *1*, 81−106.

(14) Blazadonakis, M.; Moustakis, V.; Charissis, G. Deep assessment of machine learning techniques using patient treatment in acute abdominal pain in children. *Artif. Intell. Med.* **1996**, *8*, 527−542.

(15) Quinlan, J. R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann: San Mateo, 1993; pp 3−41.

(16) Quinlan, J. R. Rule induction with statistical data—a comparison with multiple regression. *J. Operat. Res. Soc.* **1987**, *38*, 347−352.

(17) Štambuk, N.; Mašić, N.; Brinar, V.; Trbojević-Čepe, M.; Štambuk, V.; Rabatić, S.; Marotti, T.; Šverko, V.; Marušić-Della Marina, B.; Brzović, Z.; Zurak, N.; Svoboda-Beusan, I.; Mažuran, R.; Martinić, R.; Štambuk, A.; Karaman, K.; Gagro, A.; Rudolf, M.; Malenica, B.; Pokrić, B. Machine learning approach defines immune response during met-enkephalin immunotherapy in multiple sclerosis and optic neuritis. In *Advances in Systems, Signals, Control and Computers, vol. II*; Bajić, V. B., Ed.; International Association for the Advancement of Methods for System Analysis and Design (IAAMSAD) and Academy of Nonlinear Sciences (ANS): South Africa, 1998; pp 302−305.

(18) Mitchell, T. M. *Machine Learning*; McGraw-Hill: Singapore, 1997; pp 55−80.

(19) Campbell, M. J.; Machin, D. *Medical Statistics*; J. Wiley & Sons: Chicester, 1995; Chapter 3, pp 34−35, 74.

(20) Tandon, A. K.; Clark, G. M.; Chamness, G. C.; Chirgwin, J. M.; McGuire, W. L. Cathepsin D and prognosis in breast cancer. *N. Engl. J. Med.* **1990**, *322*, 297−302.

(21) Schultz, D. C.; Bazel, S.; Wright, L. M.; Tucker, S.; Lange, M. K.; Tachovsky, T.; Lomgo, S.; Niedbala, S.; Alhadeff, J. A. Western blotting and enzymatic activity analysis of cathepsin D in breast tissue and sera of patients with breast cancer and benign breast disease and of normal controls. *Cancer Res.* **1994**, *54*, 48−54.

(22) Goldmann, F. J.; Biess, B.; Lippold, A.; Suter, L.; Westhoff, U. Prognostic classification of malignant melanomas by combining clinical, histological, and immunohistochemical parameters. *Oncology* **1999**, *56*, 208−214.

(23) Tuminello, F. M.; Gebbia, N.; Pizzolanti, G.; Russo, A.; Bazan, V.; Leto, G. Cathepsin D content in colorectal cancer. *Oncology* **1995**, *52*, 237−242.

(24) Zeillinger, R.; Swoboda, H.; Machacek, E.; Nekahm, D.; Sliutz, G.; Knogler, W.; Speiser, P.; Swoboda, E.; Kubista, E. Expression of cathepsin D in head and neck cancer. *Eur. J. Cancer* **1992**, *28A*, 1413−1415.

(25) Parham, D.; Whitaker, J. N.; Costan, W.; Berard, W. Cellular distribution of cathepsin D in childhood tumors. *Arch. Pathol. Lab. Med.* **1985**, *109*, 250−255.

(26) Robson, D. K.; Ironside, J. W.; Reid, W. A.; Bogue, P. R. Immunolocalization of cathepsin D in the human central nervous system and central nervous system neoplasms. *Neuropathol. Appl. Neurobiol.* **1990**, *16*, 39−44.

(27) Resta, L.; Fiorella, R.; di Nicola, V.; Marzullo, A.; Boticella, M. A.; Maiorano, E. Cathepsin D in laryngeal carcinoma. Preliminary report. *Bolletino-Societa Italiana Biologia Sperimentale* **1995**, *71*, 257−261.

(28) Marsigliante, S.; Biscozzo, L.; Resta, L.; Leo, G.; Mottaghi, A.; Maiorano, E.; Collucci, G.; Storelli, C. Immunohistochemical and immunoradiometric evaluations of total cathepsin-D in human larynx. *Eur. J. Cancer* **1994**, *30B*, 51−55.

(29) Marsigliante, S.; Mottaghi, A.; Muscella, A.; Cirado, V.; Leo, G.; Storelli, C. Effect of tumor size on the association between PS2 and cathepsin D in primary breast cancer. *Int. J. Oncol.* **1995**, *6*, 69−73.

(30) Theodoropoulos, G. E.; Panoussopoulos, D.; Lazaris, A. C.; Golematis, B. C. Evaluation of cathepsin D immunostaining in colorectal adenocarcinoma. *J. Surg. Oncol.* **1997**, *65*, 242−248.

(31) Allgayer, H.; Babic, R.; Grutzner, K. U.; Beyer, B. C. M.; Tarabichi, A.; Schildberg, F. W.; Heiss, M. M. Tumor-associated proteases and inhibitors in gastric cancer − analysis of prognostic impact and individual risk protease patterns. *Clin. Exp. Metastasis* **1998**, *16*, 62−73.

CI990075Q