

Application of machine learning to predict the recurrence-proneness for cervical cancer

Chih-Jen Tseng · Chi-Jie Lu · Chi-Chang Chang ·
Gin-Den Chen

Received: 18 October 2012 / Accepted: 29 January 2013 / Published online: 14 February 2013
© Springer-Verlag London 2013

Abstract This study applied advanced machine learning techniques, widely considered as the most successful method to produce objective to an inferential problem of recurrent cervical cancer. Traditionally, clinical diagnosis of recurrent cervical cancer was based on physician's clinical experience with various risk factors. Since the risk factors are broad categories, years of clinical study and experience have tried to identify key risk factors for recurrence. In this study, three machine learning approaches including support vector machine, C5.0 and extreme learning machine were considered to find important risk factors to predict the recurrence-proneness for cervical cancer. The medical records and pathology were accessible by the Chung Shan Medical University Hospital Tumor Registry. Experimental results illustrate that C5.0 model is the most useful approach to the discovery of recurrence-proneness factors. Our findings suggest that four most important recurrence-proneness factors were Pathologic

Stage, Pathologic T, Cell Type and RT Target Summary. In particular, Pathologic Stage and Pathologic T were important and independent prognostic factor. To study the benefit of adjuvant therapy, clinical trials should randomize patients stratified by these prognostic factors, and to improve surveillance after treatment might lead to earlier detection of relapse, and precise assessment of recurrent status could improve outcome.

Keywords Recurrent cervical cancer · Support vector machine · Extreme learning machine · C5.0

1 Research problem

Cervical cancer remains one of the leading causes of cancer-related death among women globally [1, 2]. Even though the morbidity and the mortality have been decreasing in recent years, the morbidity rate of cervical cancer is the second leading type in women. In Taiwan, cervical carcinoma is the second most common malignancy for women and contributing to a quarter of all female cancer cases. It remains one of the most pressing medical problems for women. The natural history of cervical carcinoma begins with a normal epithelium which progress through various stages of dysplasia—cervical intraepithelial neoplasia grade CIN 1, CIN 2, CIN 3—and finally, to invasive cervical carcinoma (ICC). There is a long time interval for the progression to ICC, and consensus on the fact that regression occurs in CIN. The most important part of therapy is to detect and eradicate local CIN 3 lesions before the progression to ICC and metastasis can occur [3].

The cure rate of cervical cancer is quite high if detected early, but approximately 30 % of International Federation of Gynecology and Obstetrics (FIGO) stage IB2 to stage IV

C.-J. Tseng · C.-C. Chang (✉)
School of Medical Informatics, Chung Shan Medical University,
No. 110, Sec. 1, Chien-Kuo N. Rd., Taichung 40201, Taiwan, ROC
e-mail: changintw@gmail.com

C.-J. Tseng · G.-D. Chen
Department of Obstetrics and Gynecology, Chung Shan Medical
University Hospital, Taichung, Taiwan, ROC

C.-J. Lu
Department of Industrial Management,
Chien Hsin University of Science and Technology,
Zhongli City, Taiwan, ROC

C.-C. Chang
IT Office, Chung Shan Medical University Hospital,
No. 110, Sec. 1, Chien-Kuo N. Rd., Taichung 40201,
Taiwan, ROC

disease will ultimately recur with modern multimodality treatment [4, 5]. Once the primary treatment has failed, the opportunity of secondary cure is slim. Patients with recurrent disease or pelvic metastases have a poor prognosis with a 1-year survival rate between 15 and 20 % [6]. Since, the treatment for recurrent cervical carcinoma is still a clinical challenge. The existing literature on recurrent cervical cancer reveals that factors include (1) Age, (2) Cell Type, (3) Tumor Grade, (4) Tumor Size, (5) Pathologic T, (6) Pathologic Stage, (7) Surgical Margin Involvement, (8) Lymph Node Metastases (LNM), (9) Number of Fractions of Other RT, (10) RT target Summary, (11) Sequence of Locoregional Therapy and Systemic Therapy and (12) Lympho-Vascular Space Involvement (LVSI). Within these broad categories, many studies have attempted to identify the plethora of factors that could enhance clinical management in the intervention of cervical cancer [7, 8]. Machine learning, which involves the retrieval and analysis of large amounts of data from a data warehouse, has been successfully used to uncover hidden patterns (or rules) among data in a variety of fields. In these circumstances, an alternative solution is to use modern machine learning technology to identify key associations between relevant factors on cervical cancer [9, 10]. In this regard, Louie et al. [11] and Kim et al. [12] used epidemiological analysis to explore the critical factors that might be used by management in the achievement of patients' survivability. However, given that most techniques of prediction involve various combinations of many potential causative factors, it is unlikely that a single factor can be identified that provides a full explanation of the prognosis of serious recurrence in the cervical cancer. Against this background, the present study attempts to improve surveillance after treatment might lead to earlier detection of relapse, and precise assessment of recurrent status could improve outcome.

Section 2 presents a review of literature relevant to the concept of methodology. Sect. 3 presents empirical application of the cervical cancer. Finally, Sect. 4 concludes the discussion of this study.

2 Method

In the health field, machine learning applications have been growing considerably as it can be used to directly derive patterns, which are relevant to forecast different risk groups among the patients. [13, 14] To the best of our knowledge, machine learning technique such as classification has not been used to analyze the recurrence cervical cancer. Hence, in this paper, we made an attempt to identify patterns from the database of the cervical cancer patients using several advances techniques as follows.

2.1 SVM

The basic idea of applying SVM to classification can be stated briefly as follows. We can initially map the input vectors into one feature space (possible with a higher dimension), either linearly or nonlinearly, which is relevant with the selection of the kernel function [15]. Then, within the feature space from the first, we seek an optimized linear division, that is, construct a hyperplane which separates two classes (this can be extended to multi-class).

A description of SVM algorithm is as follows: Let $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, $\mathbf{x}_i \in R^d$, $y_i \in \{-1, 1\}$ be the training set with input vectors and labels. Here, N is the number of sample observations, d is the dimension of each observation, and y_i is known target. The algorithm is to seek the hyperplane $\mathbf{w} \cdot \mathbf{x}_i + b = 0$, where \mathbf{w} is the vector of hyperplane and b is a bias term, and to separate the data from two classes with maximal margin width $2/\|\mathbf{w}\|^2$, and the all points under the boundary are named support vector. In order to optimal the hyperplane that SVM was to solve, the optimization problem was the following [16].

$$\begin{aligned} \text{Min } \Phi(\mathbf{x}) &= \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{s.t. } y_i(\mathbf{w}^T \mathbf{x}_i + b) &\geq 1, \quad i = 1, 2, \dots, N \end{aligned} \quad (1)$$

It is difficult to solve (1) and to transform the optimization problem to dual problem by Lagrange method. In the Lagrange method, the value of alpha must be nonnegative real coefficients. The (1) is transformed into the following constrained form

$$\begin{aligned} \text{Max } \Phi(\mathbf{w}, b, \xi, \alpha, \beta) &= \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \\ \text{s.t. } \sum_{j=1}^N \alpha_j y_j &= 0, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, N \end{aligned} \quad (2)$$

In (2), C is the penalty factor and determines the degree of penalty assigned to an error. It can be viewed as a tuning parameter which can be used to control the trade-off between maximizing the margin and the classification error. In general, it could not find the linear separate hyperplane in all application data. In the nonlinear data, transforming the original data to higher dimension of linear separate is the best solution. The higher dimension is called feature space, and it improves the data separated by classification. The common kernel functions are linear, polynomial, radial basis function (RBF) and sigmoid. Although several choices for the kernel function are available, the most widely used kernel uncton is the RBF kernel defined as $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$, $\gamma \geq 0$ [16], where γ denotes the width of the RBF. Thus, the RBF is applied in this study. The original SVM was

designed for binary classifications. Constructing multi-class SVM is still an ongoing research issue. In this study, we used multi-class SVM method proposed by Hsu and Lin [17]. For more details, please refer to [17].

2.2 C5.0

C5.0 classifier is a process for the classification and analysis of information hidden in large data sets/databases, which retrieves useful information in the form of a decision tree, that is, a flowchart-like tree structure [18]. The algorithm adopts a greedy approach in which the decision trees are constructed in a top-down recursive divide-and-conquer manner on the basis of a training set employing an attribute selection measure. C5.0 is more efficient than C4.5 such as: faster, more memory efficient, similar results by smaller decision trees, supports for more accuracy, weight different attributes and reduce more noise [18, 19]. C4.5 Quinlan [24] builds decision trees from a set of training data in the same way as ID3 (Iterative Dichotomiser 3), using the concept of information entropy. The training data are a set of already classified samples. Each sample is a vector including attributes or features. The training data are augmented with a vector representing the class that each sample belongs to. Each attribute of the data can be used to make a decision. [20, 21] C4.5 examines the normalized information gain that results from choosing an attribute for splitting the data. The attribute with the highest normalized information gain is the one used to make the decision. The algorithm then recurs on the smaller sub-lists.

Take calculating evaluation properties of A as an example, calculate information gain ratio $\text{GainRatio}(A)$, S represents a set of samples, p_i is the probability that an arbitrary sample belongs to B_i . Suppose that categorical attributes have n different values, which define n different classes B_i , ($i = 1, \dots, n$). Suppose S_j is the number of samples in the class B . $\text{Info}(S)$ indicates the information entropy in the current sample. The calculation process is as follows:

$$\text{Info}(S) = \sum_{i=1}^n p_i \log(p_i) \tag{3}$$

Suppose attribute A has n different values $\{A_1, A_2, \dots, A_n\}$, uses A to divide S into n subsets $\{S_1, S_2, \dots, S_n\}$, and S_j is the sample that has A_j in A , S_{ij} is the sample number of class B_i in subset S_j . $\text{Info}(S, A)$ is the needed information entropy.

The calculation progress is as follows:

$$\text{Info}(S, A) = \sum_{j=1}^n \frac{S_{1j} + S_{2j} + \dots + S_{nj}}{S} \text{Info}(A) \tag{4}$$

The split information $\text{SplitInfo}(A)$ is the entropy of each value of attribute A about S , and it is used to eliminate deviation of attribute that has a large number of value attributes. The calculation progress is as follows:

$$\text{SplitInfo}(A) = - \sum_{i=1}^n \frac{|S_j|}{|S|} \log\left(\frac{|S_j|}{|S|}\right) \tag{5}$$

$$\text{Gain}(A) = \text{Info}(S) - \text{Info}(S, A) \tag{6}$$

$$\text{GainRatio}(A) = \text{Gain}(A) / \text{SplitInfo}(A) \tag{7}$$

2.3 ELM

Extreme learning machine (ELM) proposed by Huang et al. [17] is a single hidden layer feed-forward neural networks (SLFNs) which randomly selected the input weights and analytically determines the output weights of SLFNs. ELM has been successfully applied to many real-world applications [21, 22, 23, 24]. It not only can be thousands of times faster than traditional feed-forward network learning algorithms like back-propagation (BP) algorithm while obtaining better generalization performance, but also avoids many difficulties presented to gradient-based methods such as stopping criteria, learning rate, learning epochs, local minimal and overtuning issues

Consider N arbitrary distinct samples (x_i, t_i) where $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$, and $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$. SLFNs with \tilde{N} hidden neurons and activation function $g(x)$ can approximate N samples with zero error. This means that

$$H\beta = T \tag{8}$$

where

$$H(w_1, \dots, w_{\tilde{N}}, b_1, \dots, b_{\tilde{N}}, x_1, \dots, x_N) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_1) & \dots & g(w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}}; \beta_{N \times m} = (\beta_1^T, \dots, \beta_{\tilde{N}}^T)^t; T_{N \times m} = (T_1^T, \dots, T_N^T)^t, \text{ where}$$

$w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$, $i = 1, 2, \dots, \tilde{N}$, is the weight vector connecting the i th hidden node and the input nodes, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ is the weight vector connecting the i th hidden node and the output nodes, and b_i is the threshold of the i th hidden node. $w_i \cdot x_j$ denotes the inner product of w_i and x_j . H is called the hidden layer output matrix of the neural network; the i th column of H is the i th hidden node output with respect to inputs x_1, x_2, \dots, x_N .

Thus, the determination of the output weights (linking the hidden layer to the output layer) is as simple as finding the least-square solution to the given linear system. The minimum norm least-square (LS) solution to the linear system [i.e., Eq. (8)] is

$$\hat{\beta} = H^{\Psi}T \quad (9)$$

where H^{Ψ} is the Moore–Penrose generalized inverse of matrix H . The minimum norm LS solution is unique and has the smallest norm among all the LS solutions. Steps of ELM algorithm can be summarized as follows:

Step 1: Randomly assign input weight w_i , and bias b_i ;

Step 2: Calculate the hidden layer output matrix H ;

Step 3: Calculate the output weight β , $\hat{\beta} = H^{\Psi}T$, where $T = (t_1, \dots, t_N)^t$.

3 Empirical study

In this study, the cervical cancer data set provided by the Chung Shan Medical University Hospital Tumor Registry is used in order to verify the feasibility and effectiveness of C5.0, SVM and ELM. Each patient in the data set contains 12 predictor variables, namely Age, Cell Type, Tumor Grade, Tumor Size, Pathologic T, Pathologic Stage, Surgical Margin Involvement, Lymph Node Metastases (LNM), Number of Fractions of Other RT, RT target Summary, Sequence of Locoregional Therapy and Systemic Therapy and Lympho-Vascular Space Involvement (LVSI). And the response variable is recurrent or no. There are totally 168 patients in the data set. Among them, 118 datasets with respect to the ratio of recurrent to non-recurrent patients (the prior probabilities or simply priors) were randomly selected as the training sample (estimating the parameters of the corresponding built classification models), while the remaining 50 will be retained as the testing sample (evaluating the classification capability of the built models). Note that the original data sets are first scaled into the range of [0, 1] before building C5.0, ELM and SVM classification models. The purpose of doing so is to ensure that large value input variables do not overwhelm smaller value inputs, thus helping to reduce classification errors.

In the modeling of C5.0 classification model, the predictor (or independent) variables should first be selected. Two significant independent variables were included in the final C5.0 model, namely Pathologic T and RT target Summary. The classification results (the confusion matrix)

Table 1 Classification results using C5.0 model

Actual class	Classified class	
	1 (recurrent)	2 (non-recurrent)
1 (recurrent)	34 (95.00 %)	0 (0.00 %)
2 (non-recurrent)	2 (12.50 %)	14 (87.50 %)
Average correct classification rate: 96.00 %		

of the testing sample using the obtained C5.0 model are summarized in Table 1. From the results revealed in Table 1, we can observe that the average correct classification rate is 96.0 % with 0 (2) class 1 (2) patients misclassified as class 2 (1) patients (here a class 1 patient is defined as a patient with recurrent, while a class 2 patient is a patient with non-recurrent).

For modeling the SVM classification model, all of the twelve predictor variables are used as inputs. The performance of the SVM model is mainly affected by the setting of two parameters (C and γ) since the RBF kernel function is adopted in this study. There are no general rules for the choice of the parameters. In this study, the grid search proposed by Hsu and Lin [14] is used for parameters setting. After using the Grid search method, the parameter set ($C = 2^{13}$, $\gamma = 2^{-5}$) is the best parameter set for the SVM model. Note that the model selection details of the three models are omitted for saving space.

The classification results of the testing sample using the obtained SVM model are summarized in Table 2. From the results in Table 2, it is observed that the average correct classification rate is 68.00 % with 2 (14) class 1 (2) patients misclassified as class 2 (1) patients.

In the modeling of the ELM model, the input layer has twelve nodes as twelve independent variables are used. It is known that the most important and critical parameter of ELM is the number of hidden nodes and ELM tends to be unstable in a single run classification. Therefore, the ELM model with different numbers of hidden nodes varying from 1 to 30 was constructed. For each number of nodes, the ELM model is repeated 30 times and the number of hidden nodes that gives the smallest testing RMSE value is selected. In this study, the ELM model with 23 hidden nodes has smaller RMSE value and therefore is the best models of the ELM model.

Table 2 Classification results using SVM model

Actual class	Classified class	
	1 (recurrent)	2 (non-recurrent)
1 (recurrent)	32 (94.12 %)	2 (5.88 %)
2 (non-recurrent)	14 (87.50 %)	2 (12.50 %)
Average correct classification rate: 68.00 %		

Table 3 Classification results using ELM model

Actual class	Classified class	
	1 (recurrent)	2 (non-recurrent)
1 (recurrent)	31 (91.18 %)	3 (8.82 %)
2 (non-recurrent)	0 (0.00 %)	16 (100.00 %)
Average correct classification rate: 94.00 %		

Table 4 Robustness evaluation of the C5.0, SVM, ELM models

Model Runs	{1-1}			{2-2}			Overall		
	C5.0	SVM	ELM	C5.0	SVM	ELM	C5.0	SVM	ELM
1	100.00	94.12	91.18	87.50	12.50	100.00	96.00	68.00	94.00
2	91.67	94.44	94.44	100.00	7.14	100.00	94.00	70.00	96.00
3	95.00	92.50	92.50	80.00	10.00	70.00	96.00	76.00	88.00
4	89.47	89.47	94.74	100.00	16.67	83.33	92.00	72.00	92.00
5	91.89	91.89	91.89	92.31	7.69	92.31	92.00	76.00	92.00
6	94.87	89.74	92.31	81.82	18.18	90.10	92.00	94.00	92.00
7	84.38	96.88	96.88	94.44	11.11	83.33	88.00	68.00	92.00
8	97.14	94.29	91.43	80.00	0.00	86.67	92.00	66.00	90.00
9	95.12	92.68	95.15	100.00	0.00	77.78	96.00	76.00	92.00
10	88.89	88.89	88.89	92.86	0.00	92.86	90.00	72.00	90.00
Average	92.05	92.31	93.14	91.27	7.87	86.26	92.44	74.44	91.56

Table 3 shows the classification result of the ELM model. The average correct classification rate is 94.00 % with 3 (0) class 1 (2) patients misclassified as class 2 (1) patients.

From Tables 1, 2 and 3, it can be found that the average correct classification rates of the C5.0, SVM and ELM models are 96.00, 68.00 and 94.00 %, respectively. The C5.0 model has the best classification capability in terms of the average correct classification rate. It outperforms the SVM and ELM models and hence provides an efficient alternative in conducting cervical cancer classification tasks.

In order to assess the robustness of the C5.0 method, the performance of the C5.0, SVM, ELM models was tested using 10 independent runs. Based on the findings in Table 4, it can be observed that the ELM model has the highest average correct classification rate 93.14 % in {1-1} (a patient with recurrent is classified as recurrent), and the C5.0 model generates the highest average correct classification rate 91.27 % in {2-2} (a patient with non-recurrent is classified as non-recurrent). The highest average correct classification rate for the Overall is 92.44 % which is provided by the C5.0 model. Since the C5.0 model outperforms the SVM and ELM models under {2-2} and overall situations, it indicates that the C5.0 model indeed provides better classification accuracy than the other two approaches. Consequently, based on the results from this dataset, we can conclude that the C5.0 model is an effective alternative for cervical cancer classification.

In this study, the C5.0 model not only generates the best classification result, but also can be used to select important independent variables for cervical cancer classification. The selected important independent variables can provide useful information for cervical cancer treatment. In this study, after 10 runs, the selected important independent

variables are Pathologic Stage, Pathologic T, Cell Type and RT target Summary.

4 Conclusion

The cervical cancer can hide for a long time but it would be helpful if the doctor could predict the recurrence on a basis of a few variables shortly after the treatment. That is, in recurrent cervical cancer, the most important question is why the decision was made; when this is understood. The doctor is always expected to make the right decision. Years of clinical study and experience are needed to select the correct answer yet mistakes are always possible. To better predict the prognosis, many investigators have tried to identify risk factors for recurrence. Indeed, these include tumor size, lymphovascular space involvement, depth of tumor invasion and lymph node metastasis. Nevertheless, these factors are interrelated, but analyses using these factors do not reflect the true prognoses. The problem we investigate in this study is how to support a physician's decision of whether variables are warranted. In our study, Pathologic Stage deeply invasive tumors and Pathologic T were independent risk factors, in contrast to other similar analyses [7, 8].

As a result, our findings support that Pathologic Stage and Pathologic T are important and independent prognostic factor. In particular, Cell Type and RT target Summary were significantly related to the recurrence. Due to the small number of patients studied, outcomes of different treatment modalities in patients with LVSI and deeply invasive tumors without LNM could not be further analyzed. The presented results suggest the decision tree is a good decision model. For medical interpretation, however, a further clinical cooperation with doctors is needed to

verify the proposed model. Perhaps by analysis existing or easily measured data about a patient, we can develop some results by which a physician caring a patient can better decide when to take the critical intervention.

Acknowledgments This work is supported by the Chung Shan Medical University Hospital: CSH-2012-C-012.

References

- Parkin DM, Bray FI, Devesa SS (2001) Cancer burden in the year 2000: the global picture. *Eur J Cancer* 37:S4–S66
- Goldie SJ, Kuhn L, Denny L, Pollack A, Wright T (2001) Policy analysis of cervical cancer screening strategies in low-resource setting: clinical benefits and cost effectiveness. *J Am Med Assoc* 285:3107–3115
- Delgado G, Bundy B, Zaino R, Sevin BU, Creasman WT, Major F (1990) Prospective surgical—pathological study of disease-free interval in patients with stage Ib squamous cell carcinoma of the cervix: a gynecologic oncology group study. *Gynecol Oncol* 38:352–357
- Lai CH, Hong JH, Hsueh S (1999) Preoperative prognostic variables and the impact of postoperative adjuvant therapy on the outcomes of stage IB or II cervical carcinoma patients with or without pelvic lymph node metastases. *Cancer* 85:1537–1546
- Waggoner SE (2003) Cervical cancer. *Lancet* 361:2217–2225
- Berek JS, Hacker NF (2005) *Practical gynaecologic oncology*. Lippincott Williams & Wilkins, New York
- Kamura T, Tsukamoto N, Tsuruchi N, Saito T, Matsuyama T, Akazawa K (1992) Multivariate analysis of the histopathologic prognostic factors of cervical cancer in patients undergoing radical hysterectomy. *Cancer* 69:181–186
- Grisaru DA, Covens A, Fransen E, Chapman W, Shaw P, Colgan T (2003) Histopathologic score predicts recurrence free survival after radical surgery in patients with stage IA2-IB1–2 cervical carcinoma. *Cancer* 97:1904–1908
- Ho SH, Jee SH, Lee JE, Park JS (2004) Analysis on risk factors for cervical cancer using induction technique. *Expert Syst Appl* 27(1):97–105
- Thangavel K, Jaganathan PP, Easmi PO (2006) Data mining approach to cervical cancer patients analysis using clustering technique. *Asian J Inf Technol* 5(4):413–417
- Louie KS, de Sanjose S, Mayaud P (2009) Epidemiology and prevention of human papillomavirus and cervical cancer in sub-Saharan Africa: a comprehensive review. *Trop Med Int Health* 14(10):1287–1302
- Kim HS, Park NH, Kang SB (2008) Rare metastases of recurrent cervical cancer to the pericardium and abdominal muscle. *Arch Gynecol Obstet* 278:479–482
- Kruppa J, Ziegler A, König IR (2012) Risk estimation and risk prediction using machine-learning methods. *Hum Genet* 131(10):1639–1654
- Hu YH, Wu F, Lo CL, Tai CT (2012) Predicting warfarin dosage from clinical data: a supervised learning approach. *Artif Intell Med* 56(1):27–34
- Van Belle V, Pelckmans K, Van Huffel S, Suykens JAK (2011) Support vector methods for survival analysis: a comparison between ranking and regression approaches. *Artif Intell Med* 53(2):107–118
- Vapnik VN (2000) *The nature of statistical learning theory*. Springer, Berlin
- Huang GR, Zhu QY, Siew CX (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70:489–501
- Gomathi M, Thangaraj P (2011) A computer aided diagnosis system for lung cancer detection using machine learning technique. *Eur J Sci Res* 51:260–275
- Malar E, Kandaswamy A, Chakravarthy D, Giri Dharan A (2012) A novel approach for detection and classification of mammographic microcalcifications using wavelet analysis and extreme learning machine. *Comput Biol Med* 42:898–905
- Vanneschi L, Farinaccio A, Mauri G, Antoniotti M, Provero P, Giacobini M (2011) A comparison of machine learning techniques for survival prediction in breast cancer. *BioData Mining* 4(1):12
- Bharathi A, Natarajan AM (2012) Efficient classification of cancer using support vector machines and modified extreme learning machine based on analysis of variance features. *Am J Appl Sci* 8(12):1295–1301
- Hsu CW, Lin CJ (2002) A comparison of methods for multiclass support vector machines. *IEEE Trans Neural Netw* 13(2):415–425
- Larose DT (2005) *Discovering knowledge in data: an introduction to data mining*. Wiley, New Jersey
- Quinlan JR (1993) *C4.5: programs for machine learning*. Morgan Kaufmann, San Mateo