Appl Intell (2009) 30: 24–36
DOI 10.1007/s10489-007-0073-z

An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers

Ilias Maglogiannis · Elias Zafiropoulos · Ioannis Anagnostopoulos

Published online: 12 July 2007
© Springer Science+Business Media, LLC 2007

Abstract In recent years, computational diagnostic tools and artificial intelligence techniques provide automated procedures for objective judgments by making use of quantitative measures and machine learning techniques. In this paper we propose a Support Vector Machines (SVMs) based classifier in comparison with Bayesian classifiers and Artificial Neural Networks for the prognosis and diagnosis of breast cancer disease. The paper provides the implementation details along with the corresponding results for all the assessed classifiers. Several comparative studies have been carried out concerning both the prognosis and diagnosis problem demonstrating the superiority of the proposed SVM algorithm in terms of sensitivity, specificity and accuracy.

Keywords Breast cancer · Decision support · Diagnosis · Prognosis · Support vector machines · Bayesian classifiers · Artificial neural networks

1 Introduction

According to the American National Cancer Institute, the population of the estimated new breast cancer cases for the 2006 in USA is approximately 214 640, while the estimation of deaths exceeds 41 000 [1]. Comparing to other forms of cancer, these statistics claim that breast cancer occupies the third position of appearance in diagnosed new cases following genital organs' and digestive systems' cancer. As far as the diagnosis of the breast cancer cases and the prognosis of the disease are concerned, the method that can confirm malignancy with high-level sensitivity is the surgical biopsy. Nevertheless, this is considered as a costly operation, which has a negative impact over the patient's psychology as well. Towards these considerations, less invasive machine leaning techniques target to provide the same levels of accuracy, without the negative aspects of surgical biopsy.

In the present research work, a SVM model is implemented for the breast cancer diagnosis and prognosis problem using the Wisconsin Diagnostic Breast Cancer (WDBC) as well as the Wisconsin Prognostic Breast Cancer (WPBC) datasets, which are publicly available at http://ftp.ics.uci.edu/pub/machine-learning-databases/breast-cancer-wisconsin/. These datasets involve measurements taken according the Fine Needle Aspirate (FNA) test [2].

The role of diagnosis is to provide a distinction between the malignant and benign breast masses. In case that a patient is diagnosed with breast cancer, the malignant mass must be excised. After this or a different post-operative procedure, a prediction of the expected course of the disease must be determined. However, prognostic prediction does not belong either on the classic learning paradigms of function approximation or classification. This is due to a patient can be classified as a "recur" case (instance) if the disease is observed, while there is not a threshold point at which the patient can be considered as a "non-recur" case. The data are therefore censored since a time to recur for only a subset of patients is known. For the other patients, the length of time after treatment during which malignant masses are not found is known. This time interval is the disease free survival (DFS) time, which can be reported for an individual

I. Maglogiannis (✉) · E. Zafiropoulos · I. Anagnostopoulos
Department of Information and Communication Systems Engineering, University of the Aegean, 83200 Karlovasi, Samos, Greece
e-mail: imaglo@aegean.gr

E. Zafiropoulos
e-mail: eliaszafir@aegean.gr

I. Anagnostopoulos
e-mail: janag@aegean.gr

patient or for a study population. In particular, the right endpoints of the recurrence time intervals are right censored, as some patients will inevitably change hospital, doctors or die of other unrelated with the cancer causes. The prognosis of the specific time interval is considered a difficult problem since the training data are right censored [2–5].

In this research work, we also study alternative to SVM classification methodologies, such as statistical classifiers; namely, the Naïve Bayes and Bayesian Nets, and an efficient Probabilistic Neural Network. The alternative implementations are compared with the proposed SVM model and the corresponding results are discussed. The remainder of the paper is structured as follows: In Sect. 2, related work on the issue is presented. In Sect. 3, all the details concerning the medical data characteristics and the problem formulation for each dataset in the cases of breast cancer prognosis and diagnosis are presented. The basic principles of the SVM algorithm, the Bayesian classifiers and Artificial Neural Nets, adopted as alternative classification methods, are discussed, along with specific details about their adaptation within the present application. Section 4 presents the proposed implementation for prognosis and diagnosis of the case study datasets and comments all the corresponding results, while Sect. 5 concludes the paper.

2 Related work

During the past years, there has been a significant increase in the level of interest regarding content based data processing and case base reasoning [6]. One major area of accelerated growth is the field of medical informatics where a lot of research is done towards the development of diagnostic tools, designed to support the work of medical professionals. Expert systems and machine learning algorithms are used to provide second opinions in diagnosis of medical data with the integration of knowledge-based strategies.

More specifically in the field of 2D medical imaging, image registration [7], segmentation, and pattern analysis and classification are still in the forefront of interest. In the case of segmentation for instance, knowledge-based approaches [8] are used for tackling problems such as edge detection and region growing [9], and subsequent salient object boundary detection. In addition, deformable models [10, 11] can be employed to track the deformation of organs such as the heart and model several functional organ parameters, while successive scan slices can be used to reconstruct the 3D volume of an internal organ [12]. Image segmentation is followed in the majority of image analysis tools by a feature extraction module, which performs measurements on the pixels that represent a segmented salient object allowing non-visible features to be computed. Color and border based features are then computed and used for classification. Several studies have proven the efficiency of border shape descriptors (similar to those used in this study) for the detection of malignant objects in computer based evaluation methods of several image modalities such as mammography [13], ultrasound [14], x-rays [15], microscopy [16], magnetic resonance imaging (MRI) [17] and skin cancer [18]. Image texture features are also very useful, especially in cases where tissue density is mapped to pixel intensity (i.e. in the cases of ultrasounds [19] or microscopic images of cancer images [20]).

The last stage of a computer based diagnostic tool is classification, which results in decision making. The task involves mainly two phases: training and testing. During the training phase typical feature values are extracted from a sequence of data representing known cases and corresponding classification rules are computed. A wide selection of classification techniques exist in literature varying from the simple statistical paradigm of discriminative analysis to the more sophisticated tools of artificial intelligence like artificial neural networks (ANN). Different data are used for testing the classification process and evaluating the performance of a classifier. The classification techniques utilized in the context of the present study are analyzed thoroughly in this paper later on in Sect. 3.2, while the evaluation methods are described in Sect. 3.3.

Computer based assistive diagnosis of breast cancer preoccupies several research groups. More specifically, the implementation of training algorithms in the prognosis and diagnosis of breast cancer is a research area of great interest, and significant research work has been published in literature in the area of ANNs [2–5]. Many types of artificial neural networks (ANNs) have been implemented in order to deal with the breast cancer problem such as back propagation type ANNs, multi-layer perceptrons, Learning Vector Quantization (LVQ), Radial-Basis Function (RBF) type as well as Self-Organized Maps (SOMs). Apart from neural networks several other methods have been also proposed, namely Support Vector Machines, Bayesian methods and Decision Trees, while recently, swarm intelligence and emergent algorithms are becoming more and more an alternative approach in the area. In [21] the authors compare their work with other 26 different approaches from the literature regarding the diagnosis problem. In the current work SVMs are introduced as an additional method for solving the prognosis breast cancer problem, employing a multi—class approach of ‘one-against-all’ as described in the following section.

3 Materials and methods

3.1 The Wisconsin breast cancer datasets

The WDBC and WPBC datasets are the results of the efforts made at the University of Wisconsin Hospital for the

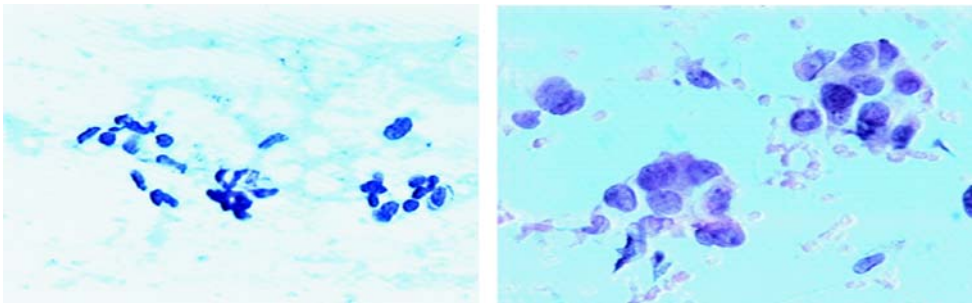

Fig. 1 Images taken using the FNA test: (**a**) Benign, (**b**) Malignant

Table 1 WDBC/WPBC cell nuclei characteristics attributes

Cell nuclei characteristics
1. radius [mean of distances from centre to points on the perimeter],
2. texture [standard deviation of grey-scale values],
3. perimeter,
4. area,
5. smoothness [local variation in radius lengths],
6. compactness [((perimeter)^2/ area) −1],
7. concavity [severity of concave portions of the contour],
8. concave points [number of concave portions of the contour],
9. symmetry,
10. fractal dimension ["coastline approximation" −1].

diagnosis and prognosis of breast tumors solely based on FNA test. This test involves fluid extraction from a breast mass using a small-gauge needle and then visual inspection of the fluid under a microscope. Figure 1 depicts two images, which were taken from fine needle biopsies of breast as appeared in [22].

The WDBC dataset consist of 569 instances (357 benign–212 malignant), where each one represents FNA test measurements for one diagnosis case. For this dataset each instance has 32 attributes, where the first two attributes correspond to a unique identification number and the diagnosis status (benign/malignant). The rest 30 features are computations for ten real-valued features, along with their mean, standard error and the mean of the three largest values ("worst" value) for each cell nucleus respectively. These ten real values, which are depicted at Table 1, are computed from a digitized image of a fine needle aspirate (FNA) of breast tumor, describing characteristics of the cell nuclei present in the image and are recorded with four significant digits.

The WPBC dataset consists of 198 instances (151 non-recur–47 recur), where each one represents follow-up data for one breast cancer case. Each instance has 35 attributes, where the first three attributes correspond to a unique identification number and to the prognosis status (recur/non-recur) following by the recurrence time (Time to Recur—TTR) or the DFS time respectively. Then they follow the above-mentioned 30 features, while the last two attributes are the diameter of the excised tumor (in cm) and the number of positive axillary lymph nodes observed at time of surgery. Four instances were not included in the training/testing set since the Lymph node values were missing. For the addressed problem, both WDBC and WPBC datasets were used in several publications in the medical literature [23–26]. In addition, due to their consistency and robust creation, these datasets are also used for verification purposes over the classification or prediction performance of information systems in other scientific areas [27–29].

3.2 Techniques for automated classification

Several types of popular and diverse approaches, deriving from the areas of Artificial Intelligence and Statistics were introduced to the breast cancer prognosis and diagnosis classification problem; namely Support Vector Machines, Bayesian classifiers and Artificial Neural Networks.

The Support Vector Machines (SVMs) is a novel algorithm for data classification and regression which allows the expansion of the information provided by a training dataset as a linear combination of a subset of the data in the training set (support vectors) [30, 31]. These vectors locate a hypersurface that separates the input data with a very good degree of generalization. The SVM algorithm is a learning machine; therefore it is based on training, testing and performance evaluation, which are common steps in every learning procedure. Training involves optimization of a convex cost function where there are no local minima to complicate the learning process. Testing is based on the model evaluation using the support vectors to classify a test dataset. Performance is based on error rate determination as test dataset size tends to infinity.

The mathematical formulation of the Support Vector Machine algorithm for data classification and regression is presented extensively in literature [29, 31–33]. More specifically, consider the case of:

- a set of N training data points $\{(\mathbf{X}_1, y_1), \ldots, (\mathbf{X}_N, y_N)\}$
- a hyperplane

$$H : y = \mathbf{w} \cdot \mathbf{X} - b = 0 \tag{1}$$

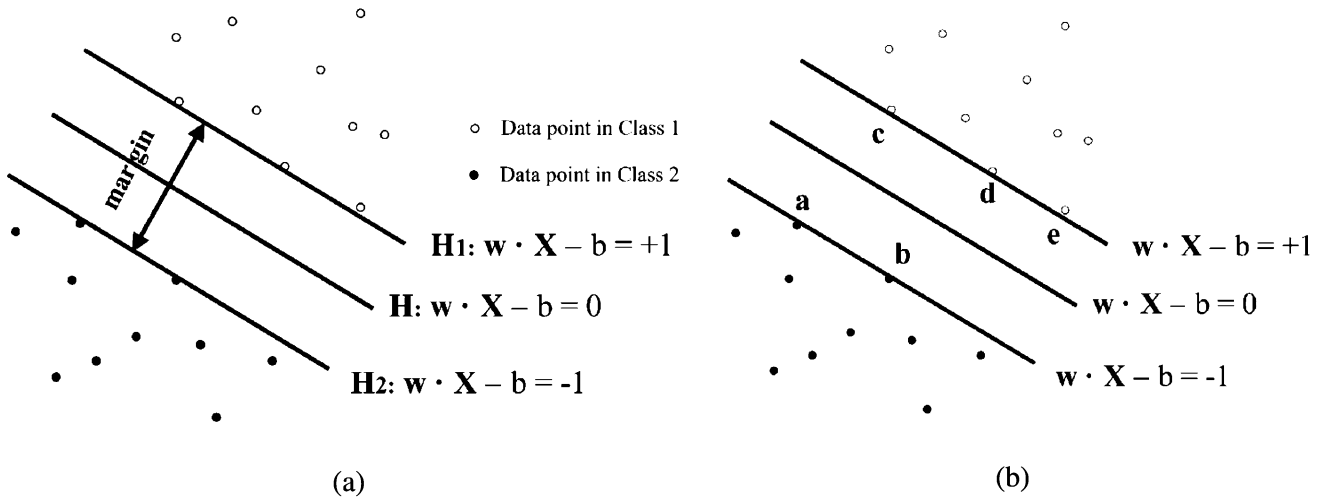

Fig. 2 **a** Decision hyperplanes generated by a linear SVM. **b** Example of support vectors

where $\mathbf{w}$ is normal to the hyperplane, $b/\|\mathbf{w}\|$ the perpendicular distance to the origin and $\|\mathbf{w}\|$ the Euclidean norm of $\mathbf{w}$

- two hyperplanes parallel to H

$$H_1 : y = \mathbf{w}\cdot\mathbf{X} - b = +1, \tag{2}$$

$$H_2 : y = \mathbf{w}\cdot\mathbf{X} - b = -1 \tag{3}$$

with the conditions that there are no data points between H_1 and H_2.

The above situation is illustrated in Fig. 2(a). If d_+ (d_-) is the shortest distance from the separating hyperplane H to the closest positive (negative) data point where the hyperplanes H_1 (H_2) is located, then the distance between the hyperplanes H_1 and H_2 is $d_+ + d_-$. Since $d_+ = d_- = 1/\|\mathbf{w}\|$, then the margin equals $2/\|\mathbf{w}\|$. The problem is to find the pair of hyperplanes that give the maximum margin:

$$\min_{\mathbf{w},b} \frac{1}{2}\mathbf{w}^T\mathbf{w} \quad \text{s.t.} \quad y_i(\mathbf{w}\cdot\mathbf{X} - b) \geq 1. \tag{4}$$

The parameters $\mathbf{w}$, b control the function and are called weight vector and bias respectively. The optimization problem presented in (4) can be stated in a convex, quadratic problem in $(\mathbf{w}, b)$ in a convex set. Using the Lagrangian formulation, the constraints will be replaced by constraints on the Lagrange multipliers themselves. Additionally in this reformulation, as a consequence the training data will only appear in the form of dot product between data vectors. Introducing Lagrangian multipliers $\alpha_1, \ldots, \alpha_N \geq 0$, a Lagrangian function for the optimization problem can be defined:

$$L_P(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_i (a_i y_i(\mathbf{w}\cdot\mathbf{X} - b) - a_i). \tag{5}$$

Using the Wolfe dual formulation and the constraints of the Lagrangian optimization problem [30], the parameters α_i can be calculated and the parameters $\mathbf{w}$, b which specify the separating hyperplane can be calculated using the following equations:

$$\mathbf{w} = \sum_i a_i y_i \mathbf{X}_i, \tag{6}$$

$$a_i(y_i(\mathbf{w}\cdot\mathbf{X}_i + b) - 1) = 0 \quad \forall i. \tag{7}$$

According to (6), the parameters α_i that are not equal to zero correspond to data $\mathbf{X}_i$, y_i that are the support vectors (Fig. 2(b)).

If the surface separating the two classes is not linear, the data points can be transformed to another high dimensional feature space where the problem is linearly separable. If the transformation to the high dimensional space is $\Phi()$ then the Lagrangian function can be expressed as:

$$L_D = \sum_i a_i - \frac{1}{2}\sum_{i,j} a_i a_j y_i y_j \Phi(\mathbf{X}_i)\cdot\Phi(\mathbf{X}_j). \tag{8}$$

The dot product $\Phi(\boldsymbol{X}_i)\ \Phi(\boldsymbol{X}_j)$ in that high dimensional space defines a kernel function $k(\mathbf{X}_i, \mathbf{X}_j)$ and therefore it is not necessary to be explicit about the transformation $\Phi()$ as long as it is known that the kernel function corresponds to a dot product in some high dimensional feature space [30]. This case is presented in Fig. 3.

With a suitable kernel, SVM can separate in the feature space the data that in the original input space was non-separable. There are many kernel functions that can be used, for example:

$$k(X_i, X_j) = e^{-\|X_i - X_j\|^2/2\sigma^2}$$

(Gaussian radial basis function kernel), (9)

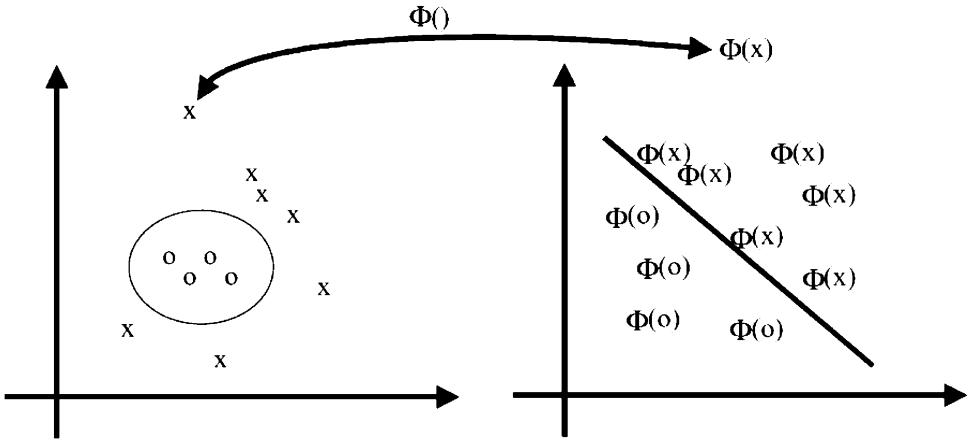

Fig. 3 A non-linear separating region transformed into a linear one

$$k(X_i, X_j) = (\mathbf{X}_i \cdot \mathbf{X}_j + m)^p$$

(the polynomial kernel). (10)

A kernel function has a good performance if the support vectors that are calculated by using the corresponding transformation are few and the classification of the test data is successful. In the proposed automated diagnostic system, we experimented with both Gaussian radial basis function kernel and the polynomial kernel.

To sum up, in order to separate a data set, a train data set (**X, Y**) is selected, the optimization problem is solved and the parameters α_i, **w**, b are calculated. Then, a given data vector **X** of the initial data set is classified according to the value of $\text{sgn}(\mathbf{w}\cdot\mathbf{X}^* + b)$. The performance of the support vectors calculated is tested using a test data set derived from the initial data set.

In the present case study WDBC data should be classified in two categories, benign or malignant breast masses, which is a typical computational requirement for developing an SVM model. The calculated support vectors provide a framework for categorizing each breast mass, based on a simple sign calculation of a linear mathematical expression. On the other hand, the prognosis problem is more complicated, since it requires the estimation of the time-to-recur (TTR) or disease-free (DFS) time, given a set of time intervals. Therefore, prognosis of TTR or DFS can be dealt as a multi-class classification of WPBC data and the SVM model is modified accordingly. Several attempts have been published in literature concerning SVM algorithms for multi-category classification, where the strategy of "one-against-all" seem to perform very well comparing to "one-to-one" and max-win voting [34, 35]. The "one-against-all" strategy is based on constructing m binary classifiers for categorizing data into m classes. Each classifier i is dealing with whether each case in the dataset 'belongs' or not to the i-th class. On the other hand, "one-against-all" strategy is based on constructing $m(m-1)/2$ classifiers in order to categorize the data into the m classes. In this case, each classifier categorize each case in the dataset between two classes out of the m, that is to say $m(m-1)/2$ total combinations of classes. Each case is appointed to the class that won the maximum votes among the $m(m-1)/2$ classifications (max-win-voting). In the proposed SVM model for the WPBC case, the "one-against-all" strategy is adopted, by integrating 4 SVM two-class modules.

An alternative approach for this classification problem might be Bayesian classifiers, which obtain the posterior probability of each class, C_i, using Bayes rule. The Naive Bayes classifier (NBC) makes the simplifying assumption that the attributes, A, are independent given a specific class, so the likelihood can be obtained by the product of the individual conditional probabilities of each attribute given the class. Thus, the posterior probability, $P(C_i|A_1, \ldots, A_n)$, is given by:

$$P(C_i|A_1, \ldots, A_n) = P(C_i)P(A_1|C_i)\ldots P(A_n|C_i)/P(A). \quad (11)$$

The assumption that attributes are independent (given the class) in real life certainly is a simplistic one. But despite the disparaging name, Naïve Bayes works very well when tested on actual datasets, particularly when combined with attribute selection procedures that eliminate redundant and hence non-independent, attributes. According to Naïve Bayes methodology, the probability density functions of the dataset's attributes are approximated as a normal distribution based on training data. Then the corresponding mean and variance values for each distribution and consequently the prior probabilities of belonging in each time interval are calculated.

Bayesian Nets are a set of methods for probabilistic calculations in most problems characterized by uncertainty. A simple Bayesian Network for medical diagnosis is depicted in Fig. 4, where the dependence among the disease and the n symptoms is demonstrated.

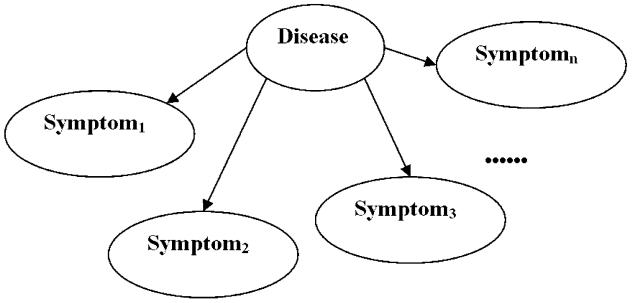

Fig. 4 A simple Bayesian Network for medical diagnosis

A Bayesian Net classifier is characterized as a theoretically well-founded way of representing probability distributions concisely and comprehensibly in a graphical manner. It consists of a set of variables and a set of directed edges between variables, representing their "cause-effect" relationships. The variables-nodes are the data attributes and class and the arcs, together with conditional probability tables represent their interdependencies graphically and numerically.

The way to construct a learning algorithm for Bayesian classifiers is to define two components: a function for evaluating a given network based on the data and a method for searching through the space of possible networks. Several algorithms have been presented in literature for the aforementioned problems such as K2, TAN (Tree Augmented Naïve Bayes), AD (All Dimensions) Trees [37]. In the present work, the K2 learning algorithm has been selected for the Naïve Bayes model learning as well as for the Bayesian Net model, since it converges faster than the rest of learning methods. The Bayesian Net model involves the definition of the belief network structure as well as the conditional probability tables. In this way, a belief network is constructed where the class attribute (time interval for WPBC or diagnosis result for WDBC) is considered as 'parent', with directed arcs connecting it with all the other attributes considered as 'childs'. All the probability tables of this belief network are populated via the learning algorithm and the training data [36].

Finally, artificial neural networks (ANNs) originate from engineering and two of its main areas of application are classification and decision problems. An ANN is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in order to solve specific problems. ANNs, like people, learn by example. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This learning process also occurs in neural networks, which are also chosen for computational reasons, since once trained, they operate very fast.

In this work, we have tested radial basis ANNs, namely Probabilistic and Generalized Regression neural networks [37, 38] applying in parallel all the necessary amendments and modifications for the problems of breast cancer diagnosis and prognosis. For the diagnosis problem we propose a Probabilistic Neural Network (PNN), since this kind of networks present high-generalization ability and do not require large amount of training data [39]. The neural classifier decides whether the input instance corresponds to a benign or a malignant case. The topology for this case consists of three layers (31-568-2). The input layer consists of 31 nodes, which correspond to the diagnosis status, followed by the 30 calculated values (mean, standard error and "worst" value) from the digitized image of each instance. The second layer is the pattern layer, which organizes the training set in such a way, that an individual processing element represents each normalized input vector, consisting of 568 nodes. This amount of nodes corresponds to the total amount of the dataset instances excluding the one left for testing each training epoch, according to the jackknife test method. Finally, the network has an output layer consisting of 2 nodes, representing the decision upon malignancy or not.

For the prognosis problem the neural network belongs to the Generalized Regression radial basis neural networks type (GRNN) and it was chosen since this type of neural networks has the capability of dealing with sparse and non-stationary data where non-linear relationships exist among the input and the output. An additional reason for selecting this type of neural classifier is due to the small amount of instances (194 instances), which the Wisconsin Breast Cancer dataset consists of for the prognosis problem. The role of this neural network is to calculate a time interval that corresponds to a possible right end-point of the patient's disease-free survival (DFS) time.

The topology of the prognosis neural network is 14-193-4-4 and it is depicted in Fig. 5. The input layer consists of 14 nodes, where each node correspond to the prognosis status, the TTR or the DFS time, the ten cell nuclei characteristics attributes of Table 1, the diameter of the excised tumor and the number of positive axillary lymph nodes observed at time of surgery. Four instances were not included in the training/testing set since their lymph node values were missing and thus, the second layer (middle layer) consists of 193 nodes (194 nodes except for the one needed for testing purposes). Finally, the division layer consists of 4 nodes that feed a same amount of processing elements in the output layer, representing the classified time intervals that correspond to a possible right-end of the DFS time.

3.3 Evaluation techniques for automated classifiers

The final phase of learning machines' development involves evaluation, where the efficiency of the proposed model

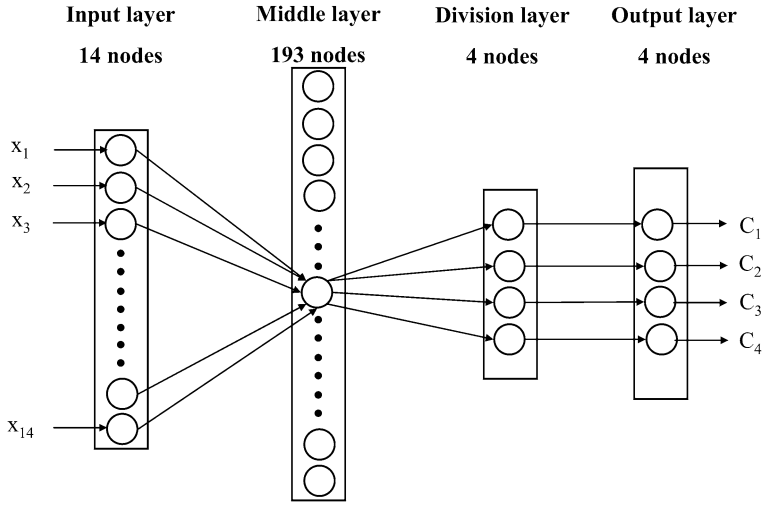

Fig. 5 Topology of the prognosis neural network

is calculated using a predefined procedure. The dominant methods presented in literature are n-fold cross validation and leave one out (jacknife) procedures.

N-fold cross-validation is a widely applied procedure mainly used in situations where the total amount of data is limited in order to provide a sufficient amount of data for training and separately testing the developed model. According to the proposed procedure, the total amount of data is randomly separated in N equal subsets; the $N-1$ subsets are used as training set for the model and the remaining set is used a test set. This is repeated N times and the total error rate is the average for all N test sets. The most widely applied values for parameter N is 3, 10 and 20. In the developed SVM and Bayesian Network model, the 10-fold cross validation has been selected as a validation procedure.

Leave-one-out cross validation (LOOCV) is a K-fold cross validation, where K equals to N, which is the number of data points in the set. In other words, this means that the function approximator is trained on all the data except for one point and a prediction is made for that point, for N distinct times. Then the average error is computed and used to evaluate the model. Additionally, the quality of a prediction can be evaluated examined using the LOOCV technique in which, each instance is singled out in turn, as a test instance with the remaining instances used to train the neural network.

Through this test method a main arising advantage is that the data can be used for training, while none has to be held back in a separate test set and thus, the possible bias introduced by relying on any one particular division into test and train components is avoided. LOOCV often works well for estimating generalization error for continuous error functions such as the mean squared error. However, it may perform poorly for discontinuous error functions such as the number of misclassified cases.

Finally the jackknife method resembles to the LOOCV method since both involve omitting each training case in turn and retraining the network on the remaining subset. However, while LOOCV is used to estimate generalization error, the jackknife is used to estimate the bias of a statistic. Thus, in the latter method, one can calculate some statistics of interest in each subset of the data. The average of these subset statistics is then compared with the corresponding statistic computed from the entire sample in order to estimate the potential bias of the whole dataset. In addition, the jackknife method can be used to estimate the bias of the training error and consequently to estimate the generalization error.

4 The proposed implementation

4.1 Classification of the WPBC patient data based on the disease free or recurrence time (prognosis)

The WPBC instances were divided over four classes, namely C_1, C_2, C_3 and C_4, according to the value of the recurrence or the DFS time. In other words, C_1 corresponds to the instances, in which the DFS time or the recurrence time was between 1 and 12 months, while C_2, C_3 and C_4 correspond to intervals between 1–3 years, 3–6 years and more than 6 years. Table 2 depicts the amount of the WPBC dataset instances in respect to the above-mentioned categorization. The first column indicates the time interval class, while the second and the third columns present the amount of instances, when the tumor recurred (N_R) and the amount of instances when the tumor did not recur (N_N).

Based on the categorization of the WPBC instances in the four intervals depicted in Table 2, the SVM algorithm has been applied for the corresponding two-class classification problem of each time interval, based on the "one-against-all" strategy. The attributes used were the ones depicted in Table 1, together with the "tumor size" and "lymph node status" features found only within the WPBC dataset. Several

kernel functions were tried in order to find the least complex function that results in low number of support vectors comparing to the training set and exhibits satisfactory performance in correct data classification. The 10-fold cross validation method has been applied and the top results are depicted in Table 3.

In order to develop an SVM-model we experimented with both polynomial and Gaussian radial base functions, as they are presented in (9) and (10). The polynomial kernel function resulted in a model that failed to classify correctly the majority of the WPBC data. The improvement of the accuracy when parameter p is set to 3 does not continue for p set to 4 or 5, not to mention that in all trials of a polynomial kernel function the correct cases recognized to belong in each interval (True Positive) were very few. However, the Gaussian Radial Base function (RBF) SVM model performed excellently, reaching the top performance with $\sigma = 1$ with accuracy varying from 96.91% to 94.84% for the four time intervals. The SVM algorithm was implemented in Matlab using a Pentium PC at 2.6 GHz with 512 MB RAM. The execution time for calculating the support vectors with different kernel functions varied for 11 to 14 seconds.

On the other hand, the Naïve Bayes and the Bayesian Nets classifiers have not the same performance in the WPBC dataset, as it can be observed through the results presented in Table 3. In all time intervals the achieved accuracy was lower than the respective SVM model, i.e. its accuracy did not exceed 80%, and the True Positive cases were very few in every time interval. The Bayesian classifiers were developed in Weka [36, 40] and the calculations in each model didn't need more than 6 seconds using a Pentium PC at 2.6 GHz with 512 MB RAM.

Regarding the Artificial Neural Net classifier, its accuracy for each time interval was high, overpassed 90% of the WPBC dataset, while in the case of the C3 time interval the correctly classified instances approximate 94,5%. However, the ANNs achieved performance was still lower than the corresponding SVM model of $\sigma = 1$ in all classes (i.e. time intervals).

The performance of the implemented classifiers can be further evaluated using the sensitivity and specificity indices. Assuming that if an instance belongs to the time interval it is classified positive, otherwise it is classified negative, the sensitivity and specificity indices can be defined as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \tag{12}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{13}$$

Table 2 WPBC instances according to the categorized interval time and prognosis status

Class	Interval time	N_R	N_N	Total
C_1	Less than 1 year	20	23	43
C_2	1 year–3 years	14	34	48
C_3	3 years–6 years	7	48	55
C_4	More than 6 years	5	43	48
Total				194

Table 3 Results of the SVM algorithm using alternative kernel functions, Bayesian classifiers and Artificial Neural Nets for the classification of WPBC cases in each time interval

		SVM (Kernel functions)					Bayes Net	Naïve Bayes	ANNs
		Polynomial		Gaussian RBF					
		$p = 2$	$p = 3$	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$			
C1	No of SVs	74	91	104	93	104			
	Errors	152	42	10	27	39	42	56	13
	Accuracy (%)	21.65	78.35	94.84	86.08	79.90	78.35	71.13	93.33
C2	No of SVs	77	106	122	103	112			
	Errors	46	46	9	20	31	46	60	14
	Accuracy (%)	76.29	76.29	95.36	89.69	84.02	76.29	69.07	92.80
C3	No of SVs	192	97	108	99	109			
	Errors	55	55	6	31	43	55	68	11
	Accuracy (%)	71.65	71.65	96.91	84.02	77.84	71.65	64.95	94.44
C4	No of SVs	177	78	100	93	95			
	Errors	143	51	9	19	33	60	56	19
	Accuracy (%)	26.29	73.71	95.36	90.21	82.99	69.07	71.13	90.22

Table 4 Sensitivity and specificity indexes for the SVM algorithm implementation using the Gaussian radial base function with $\sigma = 1$ as kernel function

	Time intervals							
	C1		C2		C3		C4	
Positive	43		48		55		48	
Negative	151		146		139		146	
Classifier	*ANN*	*SVM*	*ANN*	*SVM*	*ANN*	*SVM*	*ANN*	*SVM*
True Positive Classified	35	42	38	46	47	52	41	46
True Negative Classified	140	142	142	139	136	136	134	139
False Positive Classified	11	9	4	7	3	3	12	7
False Negative Classified	8	1	10	2	8	3	7	2
SENSITIVITY (%)	81.40	97.67	79.17	95.83	85.45	94.55	85.42	95.83
SPECIFICITY (%)	92.72	94.04	97.26	95.21	97.24	97.84	91.28	95.21

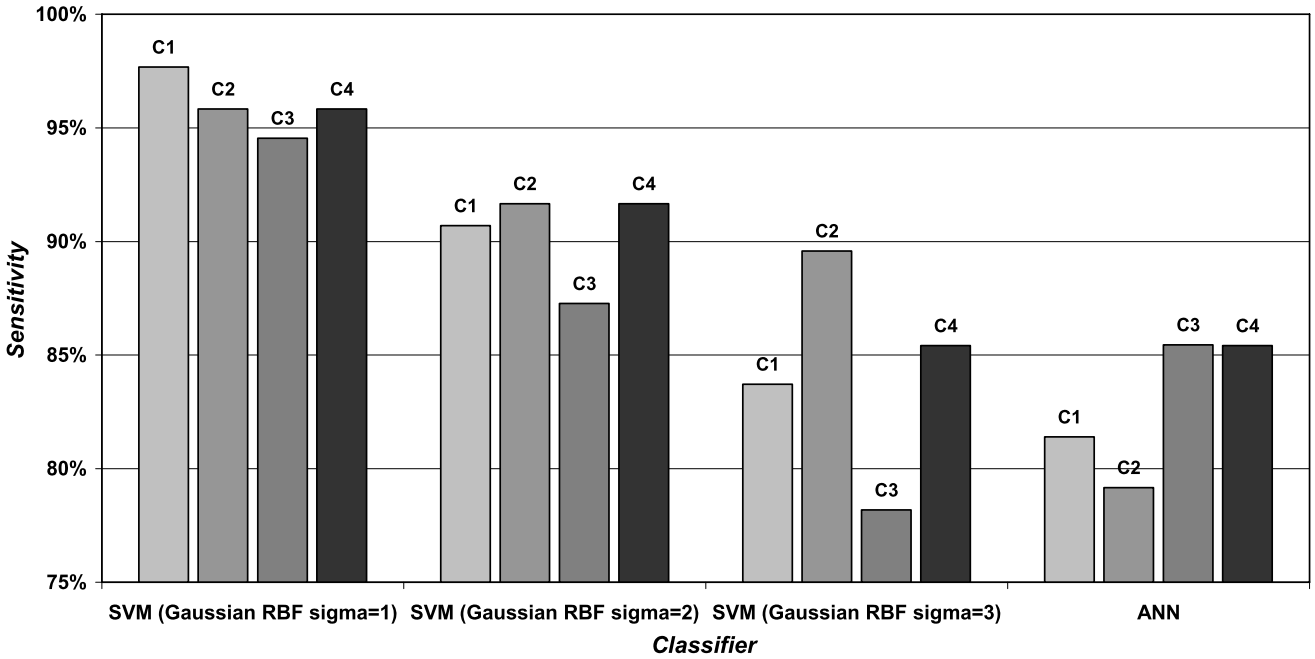

Fig. 6 Sensitivity indices for the best SVM learning machine models using alternative kernel functions and the ANNs for the classification of WPBC datasets according to the recurrence or the disease-free time survival (DFS) time (sigma stands for σ parameter)

where: *TP* (*TN*) = Number of True Positive (True Negative) classified instances; i.e. instances that the learning machine classifies correctly. *FP* (*FN*) = Number of False Positive (False Negative) classified instances; i.e. the learning machine labels the instance as positive (negative) while it is negative (positive).

These indices are presented in Table 4, corresponding to the best performed SVM classifier and the Artificial Neural net model. In Figs. 6 and 7, the indices of sensitivity, specificity and accuracy are presented graphically for the rest of the SVM classifiers and ANNs. In these figures, the specificity and sensitivity indices for the Gaussian RBF ($\sigma = 1$) are higher comparing to other classifiers in all cases of classification in the four time intervals. This fact together with the high total accuracy depicted in Table 3 indicates that the proposed SVM learning algorithm with the Gaussian RBF ($\sigma = 1$) kernel function is a superior classifier for the breast cancer prognosis problem, particularly in what concerns the sensitivity index, which is the most important.

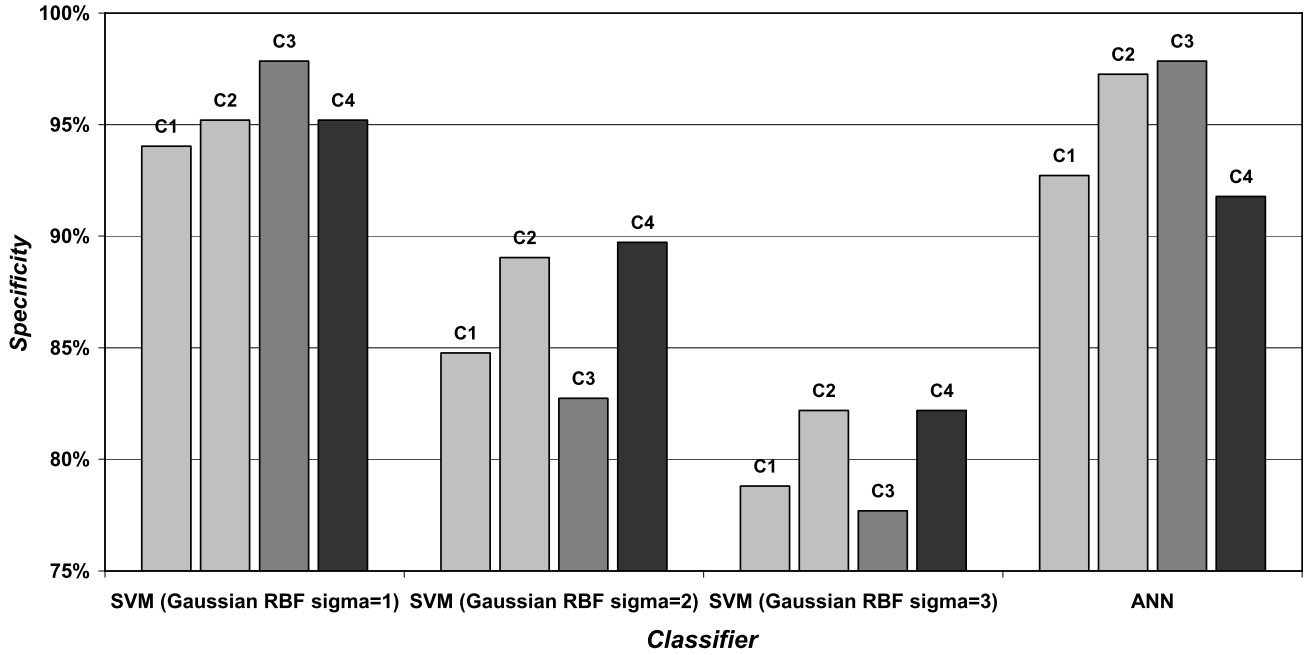

Fig. 7 Specificity indices for the SVM learning machine using alternative kernel functions and ANNs for the classification according to the recurrence or the disease-free time survival (DFS) time (sigma stands for σ parameter)

Table 5 Results of the SVM algorithm for the diagnosis of benign/malignant breast cancer instances for the WDBC dataset

Classifiers	No of SVs	Errors	Accuracy (%)
SVM Gaussian RBF ($\sigma = 3$)	243	41	92.79
SVM Gaussian RBF ($\sigma = 2$)	192	37	93.50
SVM Gaussian RBF ($\sigma = 1$)	139	30	94.73
SVM Gaussian RBF ($\sigma = 0.8$)	131	21	96.31
SVM Gaussian RBF ($\sigma = 0.7$)	126	14	97.54
SVM Gaussian RBF ($\sigma = 0.6$)	123	20	96.49
ANN		11	97.90
Bayes Net		41	92.80
Naïve Bayes		49	91.39

4.2 Automated diagnosis of breast cancer based on the wdbc patient data

In this section we discuss the problem of automated diagnosis of benign vs. malignant breast cancer instances in the case of the WDBC patient data. New classifiers based on SVMs, Bayesian classifiers and Artificial Neural Networks have been developed in order to procure a reliable automated system of breast cancer diagnosis. Several SVM models with different kernel functions were tried in order to find the best classifier that would be characterized by high performance indices (in terms of accuracy, specificity, sensitivity).

In order to develop an SVM-model we experimented with both polynomial and Gaussian radial base functions, as they are presented in (9) and (10). The SVM models, as well as the rest of the previously described classifiers, have been developed in the same platforms and needed similar computational time. The diagnosis of breast cancer is a typical two-category classification; therefore a single SVM model was developed for different kernel function, rather than many SVM models "one-against-all" in the multi-category classification during prognosis. All the top results of the performed trials are presented in Table 5. The development of SVM models using polynomial kernel functions resulted in low performance classification and the corresponding results are not included in Table 5. Additionally, the accuracy, sensitivity and specificity indices for all the classifiers with significant results are presented comparatively in Fig. 8.

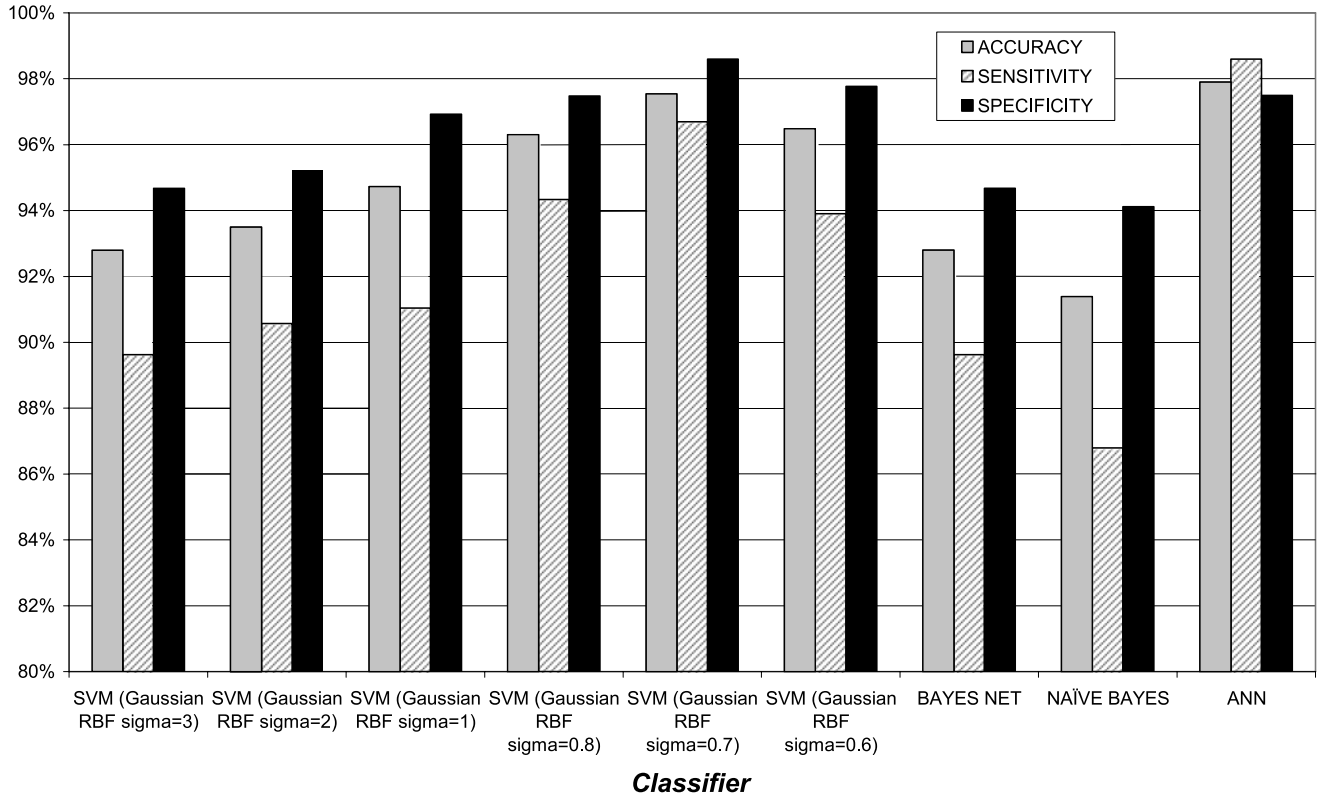

Fig. 8 Sensitivity, specificity, and accuracy for the top performance kernel functions in the SVM learning machine, the Naïve Bayes and Bayesian Nets methods as well as for the ANNs (sigma stands for σ parameter)

Regarding the diagnosis problem, a SVM with a Gaussian RBF kernel function classifier and the probabilistic ANN exhibited comparable excellent performance, reaching an accuracy rate of 98%, even higher than in the prognosis problem. The achieved sensitivity and specificity were high as well, surpassing for both classifiers 96%. The rest of the SVM models, with Gaussian RBF kernel function and σ parameter of 3, 2, 1, 0.8, 0.6, performed significant results, since all of them approximated and exceeded 93% of accuracy and acceptable sensitivity and specificity indices. The Bayesian classifiers had lower performance than the previously mentioned classifiers, since their accuracy was just above 90%.

5 Discussion and conclusions

This paper is focused on the implementation of SVM based classification algorithms comparatively with alternative methodologies (Naïve Bayes, Bayesian Nets and Artificial Neural Nets) for the diagnosis and prognosis of breast cancer. The main objective of the paper is to test the ability of our proposed SVM and ANN architectures for the Wisconsin prognosis and diagnosis breast cancer problem in comparison to other classifiers.

Initially, the suggested classification methods are implemented for the prognosis problem based on Wisconsin Prognostic Breast Cancer datasets. The examined dataset was divided in four classes, namely C1, C2, C3, and C4. The anticipated outcome of our methodology was to predict a class that corresponds to a possible tumor recurrence in four time intervals, which their range varies. This is due to the fact that the range of the predicted interval elongates for larger DFS times and the predicted probabilities for the right-end-point of the patient's DFS time is in lower levels in respect to the actual ones. As a result, the learning set over the categorized intervals could not be equally divided, while the first time interval to be predicted was shorter (in month duration), following by the second, the third and finally the forth one. Thus, the results in respect to sensitivity and specificity were expected to differ. The optimized SVM algorithm performed excellently, exhibiting high values of accuracy (up to 96.91%), specificity (up 97,67%) and sensitivity (up to 97,84%), superior than the alternative approaches for all classes.

Similarly at a second stage, another optimized SVM classifier, was implemented for automated diagnosis of the Wisconsin Diagnostic Breast Cancer datasets and the accuracy was approximately 97%, while sensitivity and specificity

indices were also satisfactory. In the specific diagnosis problem the probabilistic ANN had similar performance.

Future work on the area aims at the evaluation of the proposed SVM based classifier in other breast cancer datasets. In particular, authors intent to continue their research over the Surveillance Epidemiology and End Results (SEER) breast cancer dataset from the US National Cancer Institute (http://www.seer.cancer.gov). The SEER Breast cancer data consist of more than 400 000 records, where every record corresponds to a breast cancer case. However, the SEER dataset contains a large number of cases with relatively few coarsely-measured features and a high percentage of missing values (in contrast with the Wisconsin Breast Cancer Datasets), thus requiring a thorough effort on cleaning and preparing the data for predictive modeling.

An interesting issue left also as future work is the deployment of an efficient voting classification system that would combine the effectiveness of the best performed algorithms in order to improve the automated diagnosis potential. In general the results of this study are quite encouraging for the future. It is our belief that the testing methods presented in this paper have significant evidence to warrant trials with important clinical outcomes.

References

1. Estimated new cancer cases and deaths for 2005. http://seer.cancer.gov/cgi-bin/csr
2. Mangasarian et al (1995) Breast cancer diagnosis and prognosis via linear programming. Oper Res 43(4):570–577
3. Burke HB, Goodman PH et al (1997) Artificial neural networks improve the accuracy of cancer survival prediction. Cancer 79:857–862
4. Choong PL, de Silva CJS et al (1996) Entropy maximization networks, an application to breast cancer prognosis. IEEE Trans Neural Netw 7(3):568–577
5. Street WN (1998) A neural network model for prognostic prediction. In: Proceedings of the fifteenth international conference on machine learning, Madison, WI. Kaufmann, Los Altos
6. Perner P, Holt A, Richter M (2005) Image processing in case-based reasoning. Knowl Eng Rev 20(3):311–314
7. Maintz J, Viergever M (1998) A survey of medical image registration. Med Image Anal 2(1):1–16
8. Tsotsos J (1985) Knowledge organization and its role in representation and interpretation for time-varying data: the ALVEN system. Comput Intell 1(1):16–32
9. Martelli A (1976) An application of heuristic search methods to edge and contour detection. Commun ACM 19:73–83
10. McInerney T, Terzopoulos D (1996) Deformable models in medical image analysis: a survey. Med Image Anal 1(2):91–108
11. Metaxas D, Terzopoulos D (1991) In: Medioni G, Horn B (eds) Constrained deformable superquadrics and nonrigid motion tracking, computer vision and pattern Recognition. IEEE Computer Society, Los Alamitos, pp 337–343
12. Staib L, Duncan J (1996) Model-based deformable surface finding for medical images. IEEE Trans Med Imaging 78(5):720–731
13. Bocchi L, Nori J (2007) Shape analysis of microcalcifications using Radon transform. Med Eng Phys 29(6):691–698
14. Amores J, Radeva P (2005) Registration and retrieval of highly elastic bodies using contextual information. Pattern Recognit Lett 26(11):1720–1731
15. Antani Sameer, Lee DJ, Long LR, Thoma GR, (2004) Evaluation of shape similarity measurement methods for spine X-ray images. J Vis Commun Image Represent 15(3):285–302
16. Zapater V, Martinez-Costa L, Ayala G (2002) Classifying human endothelial cells based on individual granulometric size distributions. Image Vis Comput 20(11):783–791
17. Martens H, Thybo AK, Andersen HJ, Karlsson AH, Donstrup S, Stodkilde-Jorgensen H, Martens M (2002) Sensory analysis for magnetic resonance-image analysis: using human perception and cognition to segment and assess the interior of potatoes. Lebensm Wiss Technol 35(1):70–79
18. Maglogiannis I, Pavlopoulos S, Koutsouris D (2005) An integrated computer supported acquisition, handling and characterization system for pigmented skin lesions in dermatological images. IEEE Trans Inf Technol Biomed 9(1):86–98
19. Morris D (1988) An evaluation of the use of texture measures for tissue characterization of ultrasound images of in vivo human placenta. Ultrasound Med Biol 14(1):387–395
20. Doukas C, Maglogiannis I, Chatzioannou A, Papapetropoulos C (2006) Automated angiogenesis quantification through advanced image processing techniques. In: Proceedings of the 28th IEEE engineering in medicine and biology conference, EMBC 2006, pp 2345–2348
21. Anagnostopoulos, Anagnostopoulos C, Vergados D, Rouskas A, Kormentzas G (2006) The Wisconsin breast cancer problem: diagnosis and TTR/DFS time prognosis using probabilistic and generalised regression information classifiers. Oncol Rep, special issue Computational analysis and decision support systems in oncology 15:975–982
22. Wolberg WH, Street WN, Heisey DM, Mangasarian OL (1995) Computer-derived nuclear features distinguish malignant from benign breast cytology. Hum Pathol 26:792–796
23. Tourassi GD, Markey MK, Lo JY, Floyd CE Jr (2001) A neural network approach to breast cancer diagnosis as a constraint satisfaction problem. Med Phys 28:804–811
24. Wolberg WH, Street WN, Heisey DM, Mangasarian OL (1995) Computer-derived nuclear features distinguish malignant from benign breast cytology. Hum Pathol 26:792–796
25. Wolberg WH, Street WN, Mangasarian OL (1994) Machine learning techniques to diagnose breast cancer from fine-needle aspirates. Cancer Lett 77:163–171
26. Wolberg WH, Street WN, Mangasarian OL (1995) Image analysis and machine learning applied to breast cancer diagnosis and prognosis. Anal Quant Cytol Histol 17(2):77–87
27. Hoya T, Chambers JA (2001) Heuristic pattern correction scheme using adaptively trained generalized regression neural networks. IEEE Trans Neural Netw 12(1):91–100
28. Kaban A, Girolami M (2000) Initialized and guided EM-clustering of sparse binary data with application to text based documents. In: 15th international conference on pattern recognition, vol 2, September 2000, pp 744–747
29. Anagnostopoulos I, Anagnostopoulos C, Loumos V, Kayafas E (2004) Classifying web pages employing a probabilistic neural network classifier. IEE Proc Softw 151(3):139–150
30. Burges C. A tutorial on support vector machines for pattern recognition. http://www.kernel-machines.org/
31. Schölkopf B. Statistical learning and kernel methods. http://research.Microsoft.com/~bsc
32. Campbell C. Kernel methods: a survey of current techniques. http://www.kernel-machines.org/
33. Maglogiannis IG, Zafiropoulos EP (2004) Characterization of digital medical images utilizing support vector machines. BMC Med Inform Decis Mak 4:4

34. Duan KB, Keerthi SS (2005) Which is the best multiclass SVM method? In: Oza NC et al (eds) An empirical study, MCS 2005. Lecture notes in computer science, vol 3541, pp 278–285
35. Lee Y, Lee C-K (2003) Classification of multiple cancer types by multicategory support vector machines using gene expression data. Bioinformatics 19(9):1132–1139
36. Witten IH, Frank E (2005) Data mining: practical machine learning tools and techniques, 2nd ed. Kaufmann, Los Altos
37. Specht DF (1990) Probabilistic neural networks. Neural Netw 3(1), 109–118
38. Specht DF (1996) Probabilistic neural networks and general regression neural networks. In: Chen CH (ed) Fuzzy logic and neural network handbook. McGraw-Hill, New York
39. Masters T (1995) Advanced algorithms for neural networks. Wiley, New York
40. University of Waikato. Weka software. Available at http://www.cs.waikato.ac.nz/