Neural Comput & Applic (2014) 24:1163–1177
DOI 10.1007/s00521-012-1324-4

ORIGINAL ARTICLE

Performance analysis of support vector machines classifiers in breast cancer mammography recognition

Ahmad Taher Azar · Shaimaa Ahmed El-Said

Received: 5 May 2012 / Accepted: 17 December 2012 / Published online: 24 January 2013

Abstract Support vector machine (SVM) is a supervised machine learning approach that was recognized as a statistical learning apotheosis for the small-sample database. SVM has shown its excellent learning and generalization ability and has been extensively employed in many areas. This paper presents a performance analysis of six types of SVMs for the diagnosis of the classical Wisconsin breast cancer problem from a statistical point of view. The classification performance of standard SVM (St-SVM) is analyzed and compared with those of the other modified classifiers such as proximal support vector machine (PSVM) classifiers, Lagrangian support vector machines (LSVM), finite Newton method for Lagrangian support vector machine (NSVM), Linear programming support vector machines (LPSVM), and smooth support vector machine (SSVM). The experimental results reveal that these SVM classifiers achieve very fast, simple, and efficient breast cancer diagnosis. The training results indicated that LSVM has the lowest accuracy of 95.6107 %, while St-SVM performed better than other methods for all performance indices (accuracy = 97.71 %) and is closely followed by LPSVM (accuracy = 97.3282). However, in the validation phase, the overall accuracies of LPSVM achieved 97.1429 %, which was superior to LSVM (95.4286 %), SSVM (96.5714 %), PSVM (96 %), NSVM (96.5714 %), and St-SVM (94.86 %). Value of ROC and MCC for LPSVM achieved 0.9938 and 0.9369, respectively, which outperformed other classifiers. The results strongly suggest that LPSVM can aid in the diagnosis of breast cancer.

Keywords Soft computing · Breast cancer diagnosis · Proximal support vector machine (PSVM) · Lagrangian support vector machines (LSVM) · Finite Newton method for Lagrangian support vector machine (NSVM) · Linear programming support vector machines (LPSVM) · Smooth support vector machine (SSVM)

1 Introduction

Worldwide, breast cancer comprises 22.9 % of all cancers in women [5]. In 2008, breast cancer caused 458,503 deaths worldwide (13.7 % of cancer deaths in women) [5]. In the west, earlier research has demonstrated that 1 in 9 women will develop breast cancer in their life, and this risk has been further stratified according to age, with patient up to 25 years, 1 in 15,000; up to age 30, 1 in 1,900; and up to 40, 1 in 200 [35]. Digital mammograms are among the most difficult medical images to be read due to their low contrast and differences in the types of tissues. Important visual clues of breast cancer include preliminary signs of masses and calcification clusters. Also, tumors are of different shapes and some of them have the characteristics of the normal tissue. All these reasons make the decisions that are made on such images more difficult. Unfortunately, in the early stages of breast cancer, these signs are very subtle and varied in appearance, making diagnosis difficult, challenging even for specialists. A false-positive detection may cause an unnecessary biopsy. Statistics show that only

A. T. Azar (✉)
Faculty of Engineering, Misr University for Science and Technology (MUST), 6th of October City, Egypt
e-mail: ahmad_t_azar@yahoo.com; ahmad_t_azar@ieee.org

S. A. El-Said
Electronics and Communications Department, Faculty of Engineering, Zagazig University, Zagazig, Sharkia, Egypt
e-mail: eng.sahmed@windowslive.com

20–30 percentages of breast biopsy cases are proved cancerous. In a false-negative detection, an actual tumor remains undetected that could lead to higher costs or even to the cost of a human life. Here is the trade-off that appears in developing a classification system that could directly affect human life. In addition, the tumors existing are of different types. Early diagnosis requires an accurate and reliable diagnosis procedure that allows physicians to distinguish benign breast tumors from malignant ones. Thus, expert systems and artificial intelligent techniques are increasingly introduced to help improve the diagnostic capability. With the help of these automatic diagnostic systems, the possible errors experts made in the course of diagnosis can be avoided, and the medical data can be examined in shorter time and more detailed as well. Classifying breast cancer features into a discrete set of possible categories is also referred to as supervised learning [36]. Classification is a two-step process. In the first step, a model is learned describing a predetermined set of data classes. The model is constructed by analyzing data described by attributes. Each datum is assumed to belong to a predefined class, as determined by its class label. In the second step, the model is used for classification. First, the predictive accuracy of the model (or classifier) is estimated. There are several ways to estimate the accuracy. The most common method is often referred to as holdout approach which estimates the predictive accuracy of a model by measuring its accuracy on a set of examples (test data set) that it is not allowed to access when constructing the model. The accuracy of a model on a given test set is the percentage of test set examples that correctly classified by the model. A preferred method for accuracy estimation is to use cross-validation. In k-fold cross-validation, the available data are partitioned into k separate sets of approximately equal size. The cross-validation procedure involves k iterations in which the learning method is given $k - 1$ of the subsets to use as training data and is tested on the set left out. Each iteration leaves out different subsets so that each is used as the test set exactly once. The cross-validation accuracy of the given algorithm is simply the average of the accuracy measurement from the individual folds.

Support vector machines are an attractive approach to data classification. The main contribution of this paper is to exploit the maximum generalization capability of six Support vector machine (SVM) tools and apply them to the breast cancer diagnosis to distinguish benign breast tumor from malignant one. These classifiers are used to achieve very fast, simple, and efficient breast cancer diagnosis. They are actually much related. The different methods are small adaptations of proximal support vector machine (PSVM).

The rest of this paper is organized as follows. Section 2 surveys related work. Section 3 provides subjects and methods that are used in this paper. Section 4 presents brief description for each SVM tool and its modeling methodology. Section 4 reports the results of experimental evaluations and comparisons of the proposed six linear SVM classifiers. Section 6 concludes the study and discusses directions for future research.

2 Related work

Artificial intelligent techniques have been investigated to diagnose the disease of breast cancer with high classification accuracies. Liu et al. [30] compared SVM with k-means cluster and neural network to diagnose breast cancer, and it was noted that the SVM exhibited the better whole performance. The classification results obtained by other studies for breast cancer data set were presented in Table 1. As noticed from these studies, SVM has been used to diagnose the breast cancer and achieved the highest classification accuracy among the available artificial intelligent methods in literature. However, from authors' opinion, the SVM approach has not received the attention it deserves in the breast cancer diagnosis literature when compared with other research fields despite its great potential. SVM as a relatively new machine learning technique was first introduced by Vapnik [57]. It aims to minimize the upper bound of the generalization error based on the structural risk minimization (SRM) principal that is known to have high generalization performance. The training of SVM is equivalent to solving a linear constrained quadratic programming problem [11]. Thus, it is unlikely to be trapped in the local optimum [13, 52]. In addition to the good properties of SVM, it has found its application in a wide variety of fields including handwritten digit recognition [7, 12, 47, 48], object recognition [4], speaker identification [46], face detection in images [39], text categorization [25], and so forth. When using SVM for tackling practical problems, there are two issues have to be taken into considerations [11]. First, the appropriate kernel parameter setting plays a significant role in designing an effective SVM model. The first parameter, penalty parameter C, determines the trade-off between the fitting error minimization and model complexity. The second parameter, gamma (γ or d) of the kernel function, defines the nonlinear mapping from the input space to some high-dimensional feature space. Second, determining the optimal input feature subset also influences the performance of the SVM model in great part.

3 Subjects and methods

In this paper, the medical data related to breast cancer are considered. This database was obtained from the university of Wisconsin hospital, Madison from Dr. William H.

Table 1 Related work for breast cancer diseases diagnosis studies

Reference	Method	Performance analysis
Quinlan [44]	Proposed C4.5 decision tree method	Accuracy = 94.74 %
Hamilton et al. [22]	Proposed a rule induction algorithm based on approximate classification (RIAC) method using tenfold cross-validation (CV)	Accuracy = 94.99 %
Ster and Dobnikar [51]	Proposed linear discriminant analysis (LDA)	Accuracy = 96.8 %
Nauck and Kruse [37]	Proposed a neuro-fuzzy classifier	Accuracy = 95.06 %
Pena-Reyes and Sipper [41]	Proposed a fuzzy-GA system	Accuracy = 97.36 %
Setiono [49]	Developed a classification technique based on a feed forward neural network rule extraction algorithm	Accuracy = 98.10 %
Goodman et al. [20]	Applied three different methods, optimized learning vector quantization (LVQ), big LVQ, and artificial immune recognition system (AIRS)	Accuracy = 96.7 % (for LVQ) accuracy = 96.8 % (for big LVQ) Accuracy = 97.2 % (for AIRS)
Abonyi and Szeifert [1]	Applied a supervised fuzzy clustering technique	Accuracy = 95.57 %
Chang et al. [9]	Compared the performance of SVM with that of a multilayer propagation neural network	Accuracy = 82.40 % (for MLP) Accuracy = 85.6 % (for SVM)
Übeyli [53]	Presented a mixture experts (ME) network structure for breast cancer diagnosis	Accuracy = 98.85 %
Sahan et al. [45]	Proposed a new hybrid method based on fuzzy artificial immune system and k-nn algorithm (Fuzzy-AIS-knn)	Accuracy = 99.14 %
Ubeyli [55]	Proposed a Multilayer perceptron (MLP) neural network using four different methods, combined neural network, probabilistic neural network, recurrent neural network and SVM	Accuracy = 97.36 % (achieved by SVM)
Polat and Gunes [42]	Proposed Least square SVM (LS-SVM)	Accuracy = 98.53 %
Akay [2]	Proposed a SVM-based method combined with F-score method	Accuracy = 99.51 %
Übeyli [54]	Developed adaptive neuro-fuzzy inference system (ANFIS) for breast cancer detection	Accuracy = 99.08 %
Karabatak and Ince [27]	Developed a method combined with association rules and neural networks (AR + NN)	Accuracy = 97.4 %
Huang et al. [24]	Proposed sequential backward selection (SBS) algorithm integrating with back-propagation neural network (BPNN) and Levenberg–Marquardt (LM) (SBS-BPLM), BPNN and particle swarm optimization (PSO) (SBS-BPPSO), respectively	Accuracy = 98.83 % (SBS-BPLM) Accuracy = 97.51 % (SBS-BPPSO)
Cedeño et al. [8]	Proposed artificial metaplasticity multilayer perceptron (AMMLP) algorithm	Accuracy = 99.26 %
Fan et al. [15]	Proposed case-based reasoning approach combined with fuzzy decision tree (CBFDT)	Accuracy = 98.90 %
Chen et al. [10]	Proposed a rough set-based support vector machine classifier (RS_SVM) for breast cancer diagnosis	Highest accuracy = 100 % Average accuracy = 96.87 %
Chen et al. [11]	Proposed a swarm intelligence technique-based support vector machine classifier (PSO_SVM) for breast cancer diagnosis and this system can also be regarded as a promising success with the excellent classification accuracy	Accuracy = 99.3 % (tenfold cross-validation)

Wolberg [31, 56, 60]. It consists of 9 input real variables, 2 output classes and 699 cases, of which 458 are diagnosed as benign, and the remaining 241 are known to be malignant that have been identified on full field digital mammograms. Missing values are handled with the median or mode values for the variable. There are totally 10 attributes (1 class and 9 numeric features), such as clump thickness; uniformity of cell size; uniformity of cell shape; marginal adhesion; single epithelial cell size; bare nuclei; bland chromatin; normal nucleoli; and mitoses. It is obvious that there is a close relationship between these characteristics and identification of benign and malignant cells. The data were rescaled to the range [0–1] because of preventing the data in the greater numeric range dominating those in the smaller range, and the data should be normalized before entered to the classifier. The effect of nine characteristic

parameters on the state of breast cancer and the influence of the involved parameter on the performance of the SVM models are discussed in this study.

4 Support vector machines

Support vector machines are an attractive approach to data modeling. They combine generalization control with a technique to address the curse of dimensionality. Standard SVMs were developed to solve the classification problem, but recently they have been extended to the domain of regression problems [59]. In the literature, the terminology for SVMs can be slightly confusing. The term SVM is typically used to describe classification with support vector methods, and support vector regression is used to describe regression with support vector methods. In this study, the term SVM will refer to classification problem of breast cancer. Standard support vector machines (St-SVMs), which are powerful tools for data classification, classify points by assigning them to one of two disjoint half-spaces. These half-spaces are either in the original input space of the problem for linear classifiers or in a higher dimensional feature space for nonlinear classifiers [34, 58]. Such standard SVMs require the solution of either a quadratic or a linear program that requires considerably longer computational time. In contrast, other modified SVM tools such as PSVM classifiers [19], Lagrangian support vector machines (LSVM) [32], finite Newton method for Lagrangian support vector machine (NSVM) [18], Linear programming support vector machine (LPSVM) [17], and smooth support vector machine (SSVM) [29] do not contain the extensive computational implementation. They are actually much related. The different methods are small adaptations of PSVM. They have comparable test set correctness to that of standard SVM classifiers, but with considerably faster computational time that can be an order of magnitude faster. Their main goal is to choose a model from the hypothesis space, which is closest (with respect to some error measure) to the underlying function in the target space [21]. Classifiers are typically optimized based on some form of empirical risk minimization (ERM) by searching for the hypothesis with the lowest error on the training set. Unfortunately, this approach is doomed to failure without some sort of capacity control [3, 6]. If the data are noisy, which is true of most real world applications, the ERM learner will choose a hypothesis that accurately models the data and the noise. Such a hypothesis will perform badly on unseen data. Note that the SVM is the only algorithm which performs capacity control simultaneously with risk minimization; this is termed structural risk minimization (SRM). Suppose we are given l observations. Each observation consists of a pair: a vector $x_i \in R^n$ and the associated "truth" $y_i \in \{-1, +1\}$. If there is a machine whose task is to learn mapping $x_i \mapsto y_i$. The machine is actually defined by a set of possible mappings $x \mapsto f(x, \alpha)$, where the functions $f(x, \alpha)$ themselves are labeled by the adjustable parameters α [6]. The machine is assumed to be deterministic: for a given input x, and choice of α, it will always give the same output $f(x, \alpha)$. A particular choice of α generates what we will call a "trained machine" [6].

The SVM method is outlined first for the linearly separable case. Kernel functions are then introduced in order to deal with nonlinear decision surfaces. Finally, for noisy data, when complete separation of the two classes may not be desirable, slack variables are introduced. A complete description to the theory of SVMs for pattern recognition is in tutorials by Osuna et al. [39] and Burges [6] on SVMs. The following subsections present a brief description for the standard SVM classifier and its modified versions.

5 Standard support vector machine [34]

The power of SVMs lies in their ability to transform data to a high-dimensional space where the data can be separated using a linear hyperplane. The optimization process for SVM learning therefore begins with the definition of a functional that needs to be optimized in terms of the parameters of a hyperplane. The functional is defined such that it guarantees good classification (if not perfect classification) on the training data and also maximizes the margin (see Fig. 1). St-SVMs classify points by assigning them to one of two disjoint half-spaces [19, 57, 58]. These half-spaces are either in the original input space of the problem for linear classifiers or in a higher dimensional feature space for nonlinear classifiers [19]. The details of mathematical modeling of the standard SVM can be found in [34].

There is an equality constraint present which for large problems necessitates special computational procedures such as sequential minimal optimization (SMO) [43] or $\text{SVM}^{\text{light}}$ [26]. SMO is an improved training algorithm for SVMs. Like other SVM training algorithms, SMO breaks down a large quadratic programming (QP) problem into a series of smaller QP problems. Unlike other algorithms, SMO utilizes the smallest possible QP problems, which are solved quickly and analytically, generally improving its scaling and computation time significantly [43]. One-dimensional optimization problem must be solved in order to determine the locator γ of the separating surface [33]. In order to overcome all these difficulties as well as that of dealing with the necessity of having to essentially invert a very large matrix of the order of $m \times m$, many modified

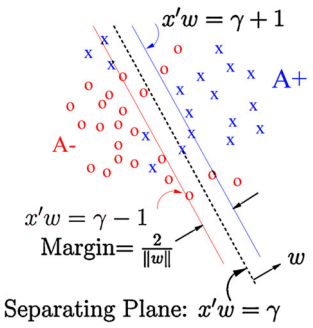

Fig. 1 The standard support vector machine classifier in the w-space of R^n: The approximately bounding planes with a soft (i.e., with some error) margin $2/\|\omega\|$, and the plane that is approximately separating $A+$ from $A-$ [34]

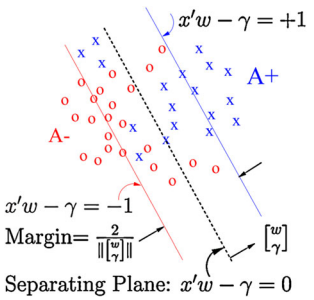

Fig. 2 The proximal support vector machine classifier in the (w, γ) 1-space of R^{n+1}: The planes $x'\omega - \gamma = \pm 1$ around which points of the sets A+ and A− cluster and which are pushed apart by the optimization problem [19]

SVM classifiers are proposed. These modified classifiers are described briefly in the following subsections.

5.1 Proximal support vector machine classifiers (PSVM) [19]

Instead of standard SVM that classifies points by assigning them to one of two disjoint half-spaces, points are classified by assigning them to the closer of two parallel planes (in input or feature space) that are pushed apart as far as possible. This proximal support vector machine classifiers (PSVM) formulation [19], which can also be interpreted as regularized least squares and considered in the much more general context of regularized networks, leads to an extremely fast and simple algorithm for generating a linear or nonlinear classifier that merely requires the solution of a single system of linear equations. In contrast, standard SVMs solve a quadratic or a linear program that requires considerably longer computational time [19]. Furthermore, PSVM formulation leads to a strongly convex objective function which is not always the case in [14]. Strong convexity plays a key role in the simple proximal code as well the very fast computational times obtained. In order to overcome the SVM classifications difficulties, Fung and Mangasarian [19] have proposed a simple modification to the standard SVM formulation. This modification, even though very simple, changes the nature of optimization problem significantly. In fact, it turns out that we can write an explicit exact solution to the problem in terms of the problem data, whereas it is impossible to do that in the previous formulations because of their combinatorial nature. Geometrically in Fig. 2, the planes $x'\omega - \gamma = \pm 1$ are not bounding planes anymore, but can be thought of as "proximal" planes, around which the points of each class are clustered and which are pushed as far apart as possible by the term $(\acute{\omega}\omega + \gamma^2)$ in the objective function which is nothing other than the reciprocal of the 2-norm distance squared between the two planes in the (ω, γ) space of R^{n+1} [19].

5.2 Smooth support vector machine (SSVM) [29]

Smoothing methods are used to generate and solve an unconstrained smooth reformulation of the support vector machine for pattern classification using a completely arbitrary kernel [29]. A fast Newton–Armijo algorithm for solving the SSVM converges globally and quadratically. In order to overcome the SVM classifications difficulties, Lee and Mangasarian [29] have proposed simple but critical modifications to the standard SVM formulation.

5.3 Lagrangian support vector machine (LSVM)

Mangasarian and Musicant [32] have proposed an implicit Lagrangian for the dual of a simple reformulation of the standard quadratic program of a linear support vector machine. It is used for a highly effective iterative scheme. This leads to the minimization of an unconstrained differentiable convex function in a space of dimensionality equal to the number of classified points. LSVM requires the inversion at the outset of a single matrix of the order of the much smaller dimensionality of the original input space plus one [32].

5.4 Linear programming support vector machines (LPSVM)

Fung and Mangasarian [17] have presented a fast method for solving a fundamental classification problem of data mining with a pronounced feature selection property for linear classifiers. When nonlinear kernels are used, the algorithm performs feature selection in a high-dimensional space of the dual variable, which results in a nonlinear kernel classifier that depends on a small number of kernel functions. This makes the method a very good choice for classification when feature selection or a fast nonlinear kernel classifier is required, as in the case of online decision making such as fraud or intrusion detection. The LPSVM algorithm requires only a linear equation solver,

which makes it simple, fast and easily accessible. In addition, LPSVM can be applied very effectively to classification problems in very large dimensional input spaces, which is often the case in the analysis of gene expression microarray data. LPSVM can also be used effectively for classifying large data sets in smaller dimensional input space.

5.5 Finite Newton method for Lagrangian support vector machine classification (NSVM)

In order to handle problems with very large dimensional input spaces, a fast finite Newton method is used [18] for finding the unconstrained unique global minimum solution of the implicit Lagrangian associated with the classification problem. The solution is obtained by solving a system of linear equations a finite number of times. Fung and Mangasarian [18] have proposed modifications to the standard SVM formulation in order to overcome the SVM classifications difficulties. Twice the reciprocal squared of the margin is used to solve the classification problem instead of measuring the margin in the $(n + 1)$ dimensional space that induces strong convexity and has little or no effect in general on the problem as was shown in [33].

6 Results and discussions

In this section, the performance analysis and comparison of the six linear SVM classifiers presented in Sect. 5 are demonstrated. The simulations were performed by using an Intel (R) Core (TM) i3 CPU 530–2.93 GHz personal computer and a Microsoft Windows 7 64-bit operating system. The core of the SVM calculations was implemented by using the MATLAB software package (MATLAB version 7.0) (www.mathworks.com).

6.1 Setting model parameters

The accuracy of an SVM model is largely dependent on the selection of the model parameters. The parameters that should be optimized for standard SVM include penalty parameter C and the kernel function parameters such as the gamma (γ) for the radial basis function (RBF) kernel. The penalty parameter C is a positive constant parameter used to control the trade-off between complexity and classification accuracy. The grid search approach [23] is an alternative to finding the best C and gamma when using the radial basis function (RBF) kernel function. In the grid search approach, pairs of (C, γ) are tried and the one with the best cross-validation accuracy is chosen. After identifying a "better" region on the grid, a finer grid search on that region can be conducted. To get good generalization ability, grid search approach uses a validation process to decide parameters. Generally, the search gets slower as the value of the C parameter gets larger, so it is best to restrict it to a reasonable range. In this study, the optimal values found of epsilon, C, and gamma were 0.001, 0.1, and 0.001, respectively.

6.2 Performance analysis

The performance of each classifier was evaluated by using performance indices such as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, and F-Measure. Sensitivity is also called recall rate in some fields and measures the proportion of actual positives which are correctly identified, while specificity measures the proportion of negatives which are correctly identified. The positive predictive value or precision rate is the proportion of positive test results that are true positives (such as correct diagnoses). It is a critical measure of the performance of a diagnostic method, as it reflects the probability that a positive test reflects the underlying condition being tested for. Its value does, however, depend on the prevalence of the outcome of interest, which may be unknown for a particular target population. The negative predictive value (NPV) is a summary statistic used to describe the performance of a diagnostic testing procedure. It is defined as the proportion of subjects with a negative test result who are correctly diagnosed. A high NPV for a given test means that when the test yields a negative result, it is most likely correct in its assessment. In the familiar context of medical testing, a high NPV means that the test only rarely misclassifies a sick person as being healthy. Some of the main formulations are defined as follows:

6.2.1 *Classification accuracy*

The classification accuracy is a common method that is used in the pattern recognition applications. The classification accuracy for the experiment is taken as the ratio of the number of samples correctly classified to the total number of samples.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \times 100\,\% \tag{1}$$

6.2.2 *Sensitivity and specificity*

Sensitivity and *specificity* are statistical measures of the performance of a classification test. *Sensitivity* measures the proportion of actual positives which are correctly identified as such (e.g., the percentage of sick people who are correctly identified as having the condition). *Specificity* measures the proportion of negatives which are correctly identified (e.g., the percentage of healthy people who are

correctly identified as not having the condition). For sensitivity and specificity analysis, we use the following expressions:

$$\text{Senstivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\,\% \tag{2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \times 100\,\% \tag{3}$$

6.2.3 *Positive predictive value (PPV), negative predictive value (NPV)*

Positive predictive value is the proportion of positive test results that are true positives (such as correct diagnoses). It is a critical measure of the performance of a diagnostic method, as it reflects the probability that a positive test reflects the underlying condition being tested for. The *negative predictive value* (NPV) is a summary statistic used to describe the performance of a diagnostic testing procedure. It is defined as the proportion of subjects with a negative test result who are correctly diagnosed. A high NPV for a given test means that when the test yields a negative result, it is most likely correct in its assessment.

$$\text{PPR} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\,\% \tag{4}$$

$$\text{NPR} = \frac{\text{TN}}{\text{TN} + \text{FN}} \times 100\,\% \tag{5}$$

In Eqs. (1–5), TP is the number of true positives (benign breast tumor), FN is the number of false negatives (malignant breast tumor), TN is the number of true negatives, and FP is the number of false positives. They are defined as a confusion matrix. Receiver operating characteristic, ROC, curves are also used to evaluate the performance of a diagnostic test [28]. This method consists of a lot of information for comprehensibility and improving classifiers performance. The ROC curve is defined as sensitivity and specificity which are the number of true-positive decision over the number of actually positive cases and the number of true negative decisions over the number of actually negative cases, respectively. This method leads to the measurement of diagnostic test performance [40]. The advantages of ROC analysis are the robust description of the network's predictive ability and an easy way to change the existence network based on differential cost of misclassification and varying prior probabilities of class occurrences. However, it requires visual inspection because the best classifiers are hard to recognize when the curves are mixed.

6.3 Training phase of classifiers

The classification process starts by obtaining a data set (input–output data pairs) and dividing it into a training set and validation data set. Hold out cross-validation was used for better reliability of test results [16]. Different lengths of the cross-validation data set ranging from one-tenth to one-third of the window size were examined. Apparently, choosing one-third of the original data leads to short data set for the training process that may cause difficulty to reach the error goal. Choosing 20 % of window size leads to weakness in detecting the features of the expected data set in prediction stage thereby leading to relatively short data set for the cross-validation procedure. Therefore, it was decided to select the length of data set for cross-validation utilized in our study to be 25 % of the original data set. The classification results of the training phase obtained from the six SVM classifiers are displayed in Table 2 by using a confusion matrix. In a confusion matrix, each cell contains the raw number of exemplars classified for the corresponding combination of desired and actual classifier outputs where TP and TN are the number of samples which are correctly identified as positives or negatives by the classifier in the test set, respectively, and FN and FP represent the numbers of samples corresponding to those cases as they are mistakenly classified as benign or malignant, respectively. Considering imbalanced positive and negative samples in the data sets, another appropriate quantity for evaluating the classification accuracy of imbalanced positive and negative samples is the Matthews correlation coefficient MCC, which is given as follows [61]:

$$\text{MCC} = \frac{\text{TP.TN} - \text{FN.FP}}{\sqrt{(\text{TP} + \text{FN})(\text{TP} + \text{FP})(\text{TN} + \text{FN})(\text{TN} + \text{FP})}} \tag{6}$$

Obviously, the scope of the MCC is within the range of $[-1, 1]$. The larger the MCC value, the better the classifier performance.

The performance analyses of the training phase by SVM classifiers are given in Table 2 and also represented graphically in Fig. 3. The results illustrated the overall accuracies of LPSVM, LSVM, SSVM, PSVM, NSVM, and St-SVM achieved 97.33 % with 333 correct classifications, 95.6107 % with 327 correct classifications, 96.76 % with 331 correct classifications, 96.1832 % with 331 correct classifications, 96.9466 % with 332 correct classifications, and 97.71 % with 333 correct classifications, respectively. The results indicated that LSVM has the lowest accuracy, while St-SVM performed better than other methods for all performance indices and is closely followed by LPSVM. Value of ROC and MCC for St-SVM achieved 0.9951 and 0.9504, respectively, which is superior to those of other classifiers. ROC curve for detecting benign breast tumors using St-SVM classifier is shown in Fig. 4.

Besides classification accuracy, the amounts of times needed for classifier construction and for classification are

Table 2 Classification comparison of SVM methods during training phase for breast cancer diagnosis

Category	LPSVM		LSVM		SSVM	
	Benign	Malignant	Benign	Malignant	Benign	Malignant
Benign	333 (TP)	12 (FN)	327 (TP)	16 (FN)	331 (TP)	13 (FN)
Malignant	2 (FP)	177 (TN)	8 (FP)	173 (TN)	4 (FP)	176 (TN)
Total	335	189	335	189	335	189
Category	PSVM		NSVM		St-SVM	
	Benign	Malignant	Benign	Malignant	Benign	Malignant
Benign	331 (TP)	16 (FN)	332 (TP)	13 (FN)	333 (TP)	10 (FN)
Malignant	4 (FP)	173 (TN)	3 (FP)	176 (TN)	2 (FP)	179 (TN)
Total	335	189	336	188	335	189

"Benign" and "Malignant" in the column headings indicate histologic findings

TP true positive, *TN* true negative, *FN* false negative, *FP* false positive

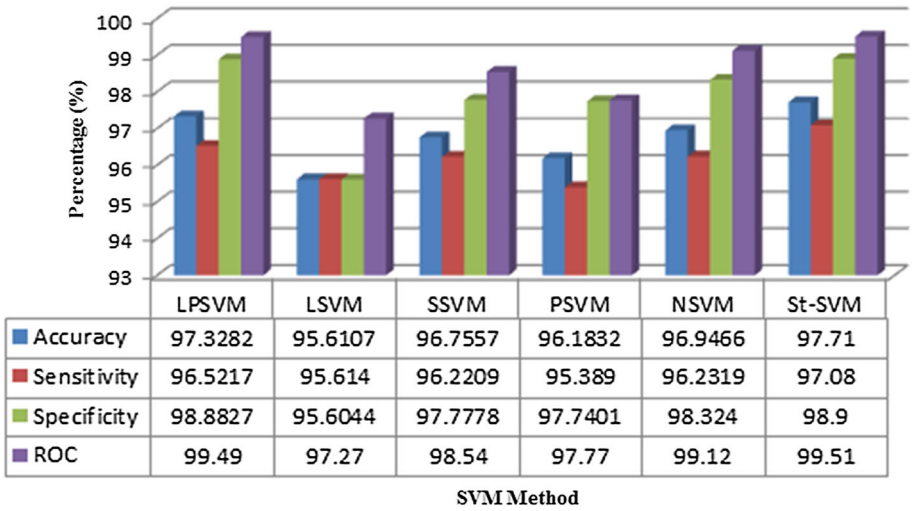

Fig. 3 Performance comparison of SVM methods during training phase

also an important factor for consideration. In this regard, computational expenses were compared between all methods of SVM (see Table 3). Results showed that the training time of LSVM was much longer than the other methods. LPSVM was significantly faster than other types of SVM. Overall, the time cost difference mainly came from the training phase as the testing time for all methods was short and practically feasible.

As shown in Fig. 4, ROC graphs are two-dimensional graphs in which true positive rate (TPR) is plotted on the Y axis and false positive rate (FPR) is plotted on the X axis. The lower left point (0, 0) represents the strategy of never issuing a positive classification; such a classifier commits no false-positive errors but also gains no true positives. The opposite strategy, of unconditionally issuing positive classifications, is represented by the upper right point (1, 1). The point (0, 1) represents perfect classification. Informally, one point in ROC space is better than another if it is to the northwest (TPR is higher, FPR is lower, or both) of the first. Classifiers appearing on the left-hand side of an ROC graph, near the X axis, may be thought of as conservative: they make positive classifications only with strong evidence so they make few false-positive errors, but they often have low true-positive rates as well. Classifiers appearing on the left-hand side of an ROC graph, near the X axis, may be thought of as conservative: they make positive classifications only with strong evidence so they make few false-positive errors, but they often have low true-positive rates as well. Classifiers on the upper right-hand side of an ROC graph may be thought of as liberal: they make positive classifications with weak evidence so they classify nearly all positives correctly, but they often have high false-positive rates.

6.4 Validation phase of classifiers

Once the model structure and parameters have been identified, it is necessary to validate the quality of the resulting

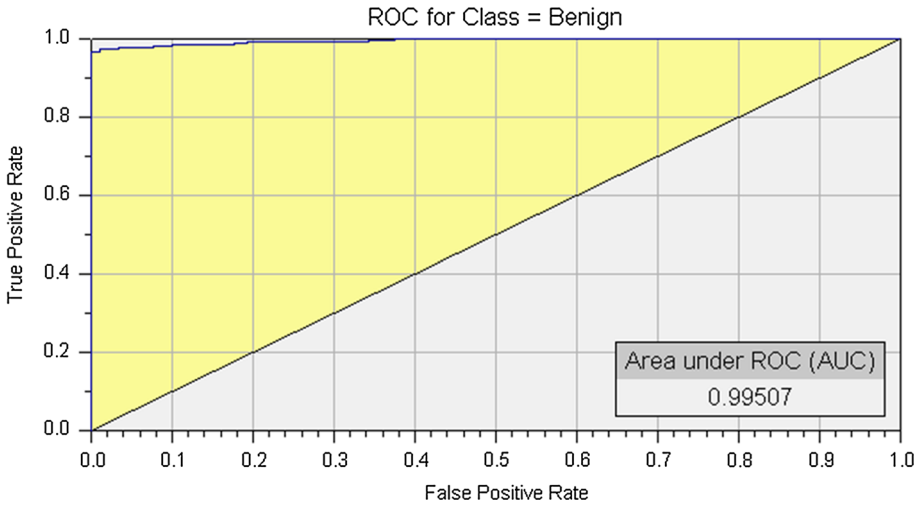

Fig. 4 ROC curves for detecting benign breast tumors using St-SVM classifier during training Phase

Table 3 Performance indices for training phase of SVM classifiers

Performance index	LPSVM	LSVM	SSVM	PSVM	NSVM	St-SVM
Accuracy (%)	97.3282	95.6107	96.7557	96.1832	96.9466	97.71
Sensitivity (Recall) (%)	96.5217	95.6140	96.2209	95.3890	96.2319	97.08
Specificity (%)	98.8827	95.6044	97.7778	97.7401	98.3240	98.90
Geometric mean of sensitivity and specificity (%)	97.7022	95.6092	96.9994	96.5646	97.2780	97.99
Positive predictive value (PPV) (Precision) (%)	99.4030	97.6119	98.8060	98.8060	99.1045	99.40
Negative predictive value (NPV) (%)	93.6508	92.0635	93.1217	91.5344	93.1217	94.71
Geometric mean of PPV and NPV (%)	96.5269	94.8377	95.9638	95.1702	96.1131	97.03
F-measure	0.9794	0.9660	0.9750	0.9707	0.9765	0.9823
Area under ROC curve (AUC)	0.9949	0.9727	0.9854	0.9777	0.9912	0.9951
MCC	0.9422	0.9044	0.9296	0.9172	0.9338	0.9504
Time (Sec)	0.0085	0.025	0.012	0.013	0.011	0.069

Table 4 Classification comparison of SVM methods during validation phase for breast cancer diagnosis

Category	LPSVM		LSVM		SSVM	
	Benign	Malignant	Benign	Malignant	Benign	Malignant
Benign	112 (TP)	2 (FN)	111 (TP)	4 (FN)	113 (TP)	4 (FN)
Malignant	3 (FP)	58 (TN)	4 (FP)	56 (TN)	2 (FP)	56 (TN)
Total	115	60	115	60	115	60
Category	PSVM		NSVM		St-SVM	
	Benign	Malignant	Benign	Malignant	Benign	Malignant
Benign	111 (TP)	3 (FN)	113 (TP)	4 (FN)	110 (TP)	4 (FN)
Malignant	4 (FP)	57 (TN)	2 (FP)	56 (TN)	5 (FP)	56 (TN)
Total	115	60	115	60	115	60

"Benign" and "Malignant" in the column headings indicate histologic findings. *TP* true positive, *TN* true negative, *FN* false negative, *FP* false positive

model. In principle, the model validation should not only validate the accuracy of the model, but also verify whether the model can be easily interpreted to give a better understanding of the modeled process. It is therefore important to combine data-driven validation, aiming at checking the accuracy and robustness of the model, with

Table 5 Performance indices for validation phase of SVM classifiers

Performance index	LPSVM	LSVM	SSVM	PSVM	NSVM	St-SVM
Accuracy (%)	97.1429	95.4286	96.5714	96	96.5714	94.86
Sensitivity (Recall) (%)	98.2456	96.5217	96.5812	97.3684	96.5812	95.65
Specificity (%)	95.0820	93.3333	96.5517	93.4426	96.5517	93.33
Geometric mean of sensitivity and specificity (%)	96.6638	94.9275	96.5665	95.4055	96.5665	94.49
Positive predictive value (PPV) (precision) (%)	97.3913	96.5217	98.2609	96.5217	98.2609	96.49
Negative predictive value (NPV) (%)	96.6667	93.3333	93.3333	95	93.3333	91.80
Geometric mean of PPV and NPV (%)	97.0290	94.9275	95.7971	95.7609	95.7971	94.12
F-Measure	0.9782	0.9652	0.9741	0.9694	0.9741	0.9607
Area under ROC curve (AUC)	0.9938	0.9718	0.9835	0.9775	0.9835	0.9661
MCC	0.9369	0.8986	0.9236	0.9117	0.9236	0.8864

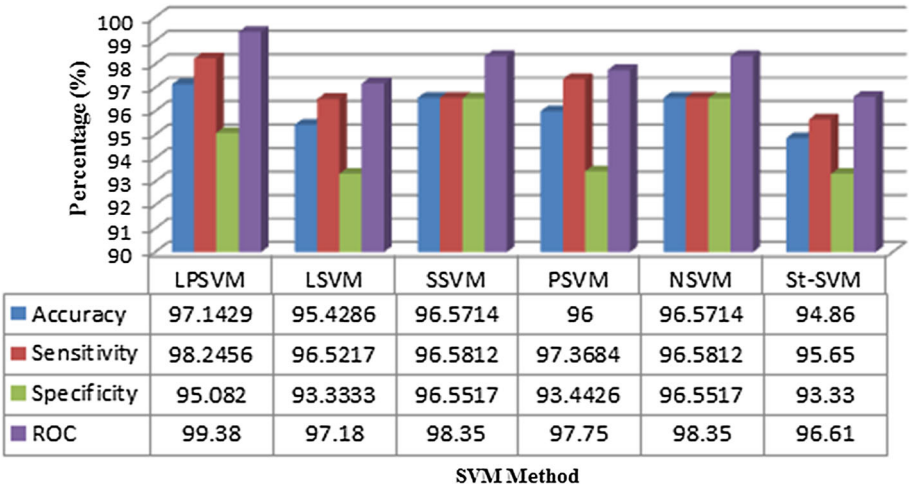

Fig. 5 Performance comparison of SVM methods during validation phase

more subjective validation, concerning the interpretability of the model. There will usually be a challenge between flexibility and interpretability, the outcome of which will depend on their relative importance for a given application. While it is evident that numerous cross-validation methods exist, the choice of the suitable cross-validation method to be employed in the SVM is based on a trade-off between maximizing method accuracy, stability, and minimizing the operation time. In this research, fourfold cross-validation method is adopted for SVM because of its accuracy and possible implementation. The classification and performance results of the validation phase by SVM methods are given in Tables 4 and 5.

It can be seen from Tables 4 and 5 that the overall accuracies of LPSVM achieved 97.1429 %, which is superior to LSVM (95.4286 %), SSVM (96.5714 %), PSVM (96 %), NSVM (96.5714 %), and St-SVM (94.86 %). Value of ROC and MCC for LPSVM achieved 0.9938 and 0.9369, respectively. SSVM and NSVM gave the same results in detecting breast cancer for all performance indices as shown in Fig. 5. However, St-SVM performed better than other methods in the training phase. It has achieved the lowest accuracy in the validation phase. It is illustrated that LPSVM has excellent performance to distinguish breast cancer tumors.

6.5 Agreement comparison between standard SVM and other methods

Statistical analysis was performed using NCSS 2007 [38]. All the methods (LPSVM, LSVM, SSVM, PSVM, and NSVM) were compared against St-SVM which was taken as a reference method. Comparison of means between the six types of SVM was made using the paired-sample t test. The p level reported with a t test represents the probability of error involved in accepting the hypothesis about the existence of a difference between the means of two groups, and it is the significance level for the statistical test. If the p value is less than α, say 0.05, the null hypothesis is rejected. If the p value is greater than α, the null hypothesis is accepted. Median algebraic difference (median Δ) between two methods is used an index of systematic bias of one method with respect to the other. Logistic regression analysis was used to assess the correlation between SVM

classifiers. The most common measure of correlation is the Pearson's product-moment correlation (called Pearson's correlation for short). Pearson's correlation (R) reflects the degree of linear relationship between two variables. The squared correlation coefficient is called the coefficient of determination. The coefficient of determination is one of the best means for evaluating the strength of a relationship that represents the fraction of the variation in one variable that may be explained by the other variable.

The agreement comparisons between the standard SVM and other methods are shown in Table 6. The results showed the training and testing accuracies for the fourfold of cross-validation. The statistical analysis demonstrated that there was a significant difference ($p < 0.05$) between St-SVM and all other SVM types during the training phase except for LPSVM method ($p = 0.1486$). In the validation phase, there was a statistically significant difference ($p < 0.05$) between St-SVM and all other SVM types except for LSVM ($p = 0.6672$) and NSVM ($p = 0.1762$). The biggest absolute difference in training phase from St-SVM mean values was obtained using LSVM (|bias| = 2.21) and PSVM (|bias| = 1.5275), while in the validation phase, the biggest difference was obtained using LPSVM (|bias| = 2.283), SSVM (|bias| = 1.9025), and NSVM (|bias| = 1.7075). The least absolute difference from St-SVM mean values was found to be 0.3889 % using LPSVM method (|bias| of 0.38) during the training phase. However, in the validation phase, the least absolute difference was found to be 0.5956 % using LSVM method (|bias| = 0.565).

Box plot accuracy comparisons between standard SVM and other SVM methods during training and validation phases are shown in Figs. 6 and 7, respectively. A box-and-whisker plot is a way of summarizing a set of data measured on an interval scale. It is often used in exploratory data analysis. It is a type of graph which is used to show the shape of the distribution, its central value, and variability. The range of the vertical scale is from the minimum to the maximum value of the selected column, or, to the highest or lowest of the displayed reference points. The box stretches from the lower hinge (defined as the 25th percentile) to the upper hinge (the 75th percentile) and therefore contains the middle half of the scores in the distribution. The median is shown as a line across the box. Therefore, 1/4 of the distribution is between this line and the top of the box, and 1/4 of the distribution is between this line and the bottom of the box. The length of the box, that is, the distance between the 25th and 75th percentiles, is known as the interquartile range. If any whisker is more than 1.5 times as long as the length of the box, then we have evidence of outliers.

6.6 Agreement comparison between SVM classifiers

Accuracy agreement comparisons between LPSVM, LSVM, SSVM, PSVM, and NSVM were also obtained. As shown in Table 7, the statistical analysis demonstrated significant differences in the training phase ($p < 0.05$) between LPSVM–LSVM ($p = 0.006$), LPSVM–PSVM ($p = 0.0021$), LSVM–SSVM ($p = 0.0239$), LSVM–NSVM ($p = 0.016$), SSVM-PSVM ($p = 0.0143$), and PSVM–NSVM (0.009). In the validation phase, no statistically significant differences were found between all SVM classifiers except for LPSVM and PSVM ($p = 0.0397$).

Correlation between standard SVM and other methods during training and validation phases was studied by linear regression analysis and analysis of variance (ANOVA) as shown in Table 8. The regression analysis revealed that there was a moderate negative correlation between St-SVM and LSVM in the training phase ($R = -0.6781$, $R^2 = 0.4599$). In the validation phase, moderate positive correlations were found between St-SVM and SSVM

Table 6 Agreement comparison between the standard SVM as a reference method and other types

Method	Accuracy agreement comparisons between the standard SVM and other methods							
	Training phase				Testing phase			
	Mean ± SD	Mean bias	Median difference Δ	*P* value	Mean ± SD	Mean Bias	Median difference Δ	*P* value
LPSVM	97.71 ± 0.18 versus 97.33 ± 0.42	−0.38	−0.23	0.1486	94.86 ± 0.63 vs. 97.14 ± 0.81	2.283	−0.94	0.0044*
LSVM	97.71 ± 0.18 versus 95.61 ± 0.70	−2.21	−2.17	0.0012*	94.86 ± 0.63 vs. 95.43 ± 2.42	0.565	1.38	0.6672
SSVM	97.71 ± 0.18 versus 96.76 ± 0.31	−0.948	−1.04	0.0018*	94.86 ± 0.63 versus 96.76 ± 0.88	1.9025	2.225	0.0125*
PSVM	97.71 ± 0.18 versus 96.18 ± 0.15	−1.5275	−1.485	0.00001*	94.86 ± 0.63 versus 96.02 ± 0.26	1.1625	1.465	0.0143*
NSVM	97.71 ± 0.18 versus 96.946 ± 0.38	−0.765	−0.555	0.0106*	94.86 ± 0.63 versus 96.57 ± 2.14	1.7075	1.655	0.1762

SD standard deviation; * statistically significant $p < 0.05$

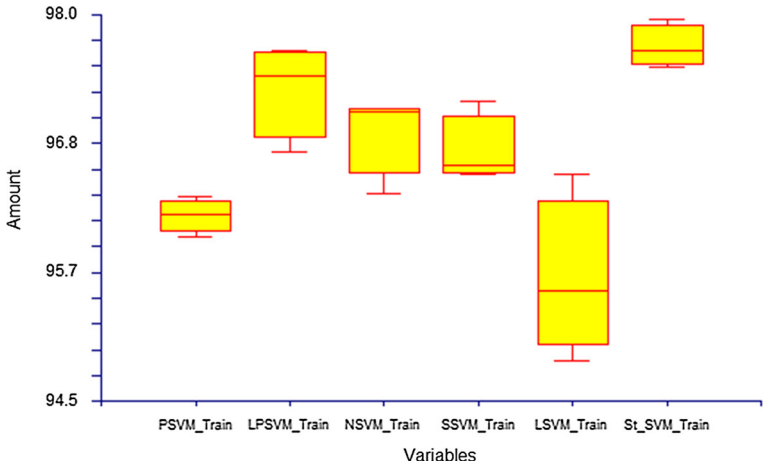

Fig. 6 Box plot accuracy comparison between standard SVM and other methods during training phase

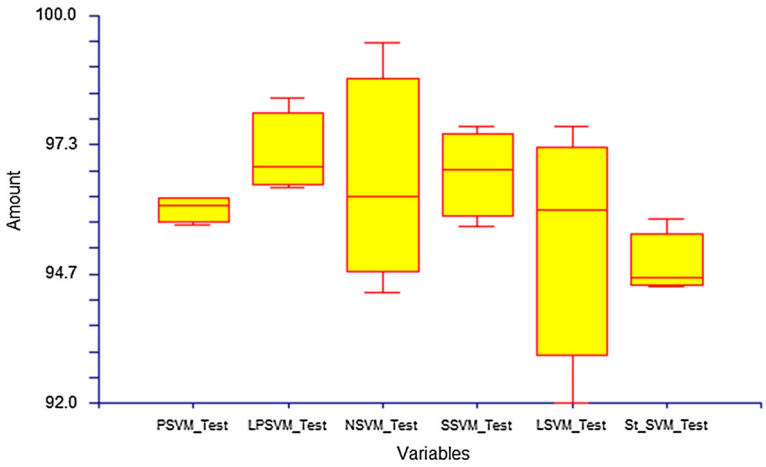

Fig. 7 Box plot accuracy comparison between standard SVM and other methods during validation phase

Table 7 Paired-sample *t* test for accuracy agreement comparison between SVM classifiers

Method	Accuracy agreement comparisons between SVM classifiers									
	Training phase					Validation phase				
	LPSVM	LSVM	SSVM	PSVM	NSVM	LPSVM	LSVM	SSVM	PSVM	NSVM
LPSVM	–	0.006*	0.0720	0.0021*	0.2219	–	0.2271	0.5495	0.0397*	0.6329
	–	$T = 4.19$	$T = 2.18$	$T = -5.14$	$T = 1.36$	–	$T = 1.35$	$T = 0.63$	$T = -2.62$	$T = 0.50$
LSVM	0.006*	–	0.0239*	0.1629	0.016*	0.2271	–	0.3389	0.6409	0.506
	$T = 4.19$	–	$T = 3.00$	$T = 1.59$	$T = 3.34$	$T = 1.35$	–	$T = 1.04$	$T = 0.49$	$T = 0.71$
SSVM	0.0720	0.0239*	–	0.0143*	0.4806	0.5495	0.3389	–	0.1583	0.8715
	$T = 2.18$	$T = 3.00$	–	$T = -3.41$	$T = 0.75$	$T = 0.63$	$T = 1.04$	–	$T = -1.61$	$T = -0.17$
PSVM	0.0021*	0.1629	0.0143*	–	0.009*	0.0397*	0.6409	0.1583	–	0.6308
	$T = -5.14$	$T = 1.59$	$T = -3.41$	–	$T = -3.76$	$T = -2.62$	$T = 0.49$	$T = -1.61$	–	$T = -0.51$
NSVM	0.2219	0.016*	0.4806	0.009*	–	0.6329	0.506	0.8715	0.6308	–
	$T = 1.36$	$T = 3.34$	$T = 0.75$	$T = -3.76$	–	$T = 0.50$	$T = 0.71$	$T = -0.17$	$T = -0.51$	–

SD standard deviation; * statistically significant $p < 0.05$

($R = 0.6368$, $R^2 = 0.4055$). Otherwise, no correlations were found between standard SVM and other methods during training and validation phases.

The equation of the straight line relating St-SVM and LSVM in the training phase using fourfold cross-validation is estimated as

Table 8 Regression analyses between between the standard SVM and other methods

	Accuracy agreement regression analysis between the standard SVM and other methods									
	Training phase					Testing phase				
	LPSVM	LSVM	SSVM	PSVM	NSVM	LPSVM	LSVM	SSVM	PSVM	NSVM
R	−0.0445	−0.6781	−0.4982	0.4012	0.2994	−0.4812	0.1585	0.6368	0.4614	−0.2972
R^2	0.002	0.4599	0.2482	0.1610	0.0897	0.2316	0.0251	0.4055	0.2129	0.0883
RMSE	0.5152	0.6327	0.3251	0.1662	0.4404	0.8739	2.9263	0.8309	0.2867	2.4993
Intercept	107.397	352.03	178.78	64.205	36.245	156.27	37.519	12.162	77.64	192.46
Slope	−0.1030	−2.6242	−0.8394	0.3273	0.6212	−0.6234	0.6104	0.8918	0.1938	−0.0883
F-ratio	0.004	1.7029	0.6602	0.3838	0.1970	0.6027	0.0515	1.3643	0.5410	0.1938
p value	0.9555	0.3219	0.5018	0.5988	0.7006	0.5188	0.8415	0.3632	0.5386	0.7028

$$\mathrm{LSVM}_{\mathrm{train}} = 352.03 - (2.624 * \ \mathrm{St} - \mathrm{SVM}_{\mathrm{train}}) \quad (7)$$

The y-intercept, the estimated value of $\mathrm{LSVM}_{\mathrm{train}}$ when St-SVM is zero, is 352.03 with a standard error of 196.4954. The slope, the estimated change in $\mathrm{SVM}_{\mathrm{train}}$ per unit change in St-SVM, is −2.624 with a standard error of 2.0110.

Similarly, the equation of the straight line relating St-SVM and SSVM in the testing phase using fourfold cross-validation is estimated as

$$\mathrm{SSVM}_{\mathrm{test}} = 12.162 + (0.8918 * \ \mathrm{St} - \mathrm{SVM}_{\mathrm{test}}) \quad (8)$$

The y-intercept, the estimated value of $\mathrm{LSVM}_{\mathrm{train}}$ when St-SVM is zero, is 12.162 with a standard error of 72.4313. The slope, the estimated change in $\mathrm{SVM}_{\mathrm{train}}$ per unit change in St-SVM, is 0.8918 with a standard error of 0.7735.

Since the support vector machines demonstrate a good accuracy in classification, the type of SVM is very important for diagnosis and knowledge discovery. Compared with standard SVM, computational results on publicly available data sets indicate that the modified SVM classifiers have comparable test set correctness to that of standard SVM classifiers and with considerably faster computational time that can be an order of magnitude faster.

7 Conclusions and future works

Breast cancer is a malignant tumor that develops when cells in the breast tissue divide and grow without the normal controls on cell death and cell division. Studies on breast cancer have been approached for many years from different angles of importance. The role of a surgeon alone is adequate for the breast cancer that is initially at a local and regional disease process. However, when the breast cancer is a systematic disease, then the involvement of systematic treatments is needed besides the surgery alone to cure it and there will be a role for earlier detection of breast cancer. With the improvements in decision support systems, expert systems, and machine learning, the effects of these tools are entering to more application domains and medical field is one of them. Decision making in medical field can be a trouble sometimes. Classification systems that are used in medical decision making provide medical data to be examined in shorter time and more detailed. According to the statistical data for breast cancer in the world, this disease is among the most prevalent cancer types. The prediction and classification of this disease have been a challenging research problem for many researchers. In this paper, six linear SVM classifiers are used to achieve very fast, simple, and efficient breast cancer diagnosis. Experiments were conducted on the WBCD data set to diagnose breast cancer in a fully automatic manner using SVM classifiers. The classification performance of St-SVM is compared with those of the other modified classifiers such as LPSVM, LSVM, SSVM, PSVM, and NSVM. The training results indicated that LSVM has the lowest accuracy of 95.6107, while St-SVM performed better than other methods for all performance indices (accuracy = 97.71 %) and is closely followed by LPSVM (accuracy = 97.3282). These results revealed that classifier and kernel function selection are necessary to get the best results. However, in the validation phase, the overall accuracies of LPSVM achieved 97.1429 %, which was superior to LSVM (95.4286 %), SSVM (96.5714 %), PSVM (96 %), NSVM (96.5714 %), and St-SVM (94.86 %). Value of ROC and MCC for LPSVM achieved 0.9938 and 0.9369, respectively, which outperformed other classifiers. The results strongly suggest that LPSVM can aid in the diagnosis of breast cancer. LPSVM classifier can be very helpful to the physicians for their final decision on their patients. By using such an efficient tool, they can make very accurate decisions. The study suggests that all other SVM classifiers are very efficient to be a potential practical methodology for clinical assistant breast cancer diagnosis by providing

the physicians with the immediate second opinion and is also possible to help the inexperienced physicians avoid misdiagnosis and then reduce the mortality rate of breast cancer. The practical obstacle of the SVM-based (as well as neural networks) classification model is its black-box nature. A possible solution for this issue is the use of SVM rule extraction techniques or the use of hybrid-SVM model combined with other more interpretable models. These issues remain to be solved in future research. It is hoped that more interesting results will follow on further exploration of data. Although developed method is built as an offline diagnosing system, it can be rebuilt as an online diagnosing system in the future. In addition, the future investigation will pay much attention to evaluating the SVMs in other medical diagnosis problems and improving the performance of SVMs using high-performance computing techniques.

Acknowledgments I would like to highly appreciate and gratefully acknowledge, Phillip H. Sherrod [50], software developer and consultant on predictive modeling, for his support and consultation during modeling process.

References

1. Abonyi J, Szeifert F (2003) Supervised fuzzy clustering for the identification of fuzzy classifiers. Pattern Recognit Lett 24(14): 2195–2207
2. Akay MF (2009) Support vector machines combined with feature selection for breast cancer diagnosis. Expert Syst Appl 36(2): 3240–3247
3. Bishop C (1997) Neural networks for pattern recognition. Clarendon Press, Oxford
4. Blanz V, Scholkopf B, Bulthoff H et al (1996) Comparison of view–based object recognition algorithms using realistic 3d models. In: von der Malsburg C, von Seelen W, Vorbruggen JC, Sendhoff B (eds) Artificial Neural Networks—ICANN'96, Springer Lecture Notes in Computer Science, Berlin, vol 1112, pp 251–256
5. Boyle P, Levin B (2008) World Cancer report 2008. International Agency for Research on Cancer, Lyon
6. Burges CJC (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Discov 2(2):121–167
7. Burges CJC, Scholkopf B (1997) Improving the accuracy and speed of support vector learning machines. In: Mozer M, Jordan M, Petsche T (eds) Advances in neural information processing systems 9. MIT Press, Cambridge, pp 375–381
8. Cedeño AM, Domíngueza JQ, Andina D (2011) WBCD breast cancer database classification applying artificial metaplasticity neural network. Expert Syst Appl 38(8):9573–9579
9. Chang RF, Wu WJ, Moon WK et al (2003) Support vector machines for diagnosis of breast tumors on US images. Acad Radiol 10(2):189–197
10. Chen HL, Yanga B, Liua J, Liu DY (2011) A support vector machine classifier with rough set based feature selection for breast cancer diagnosis. Expert Syst Appl 38(7):9014–9022
11. Chen HL, Yang B, Wang G et al (2011) Support vector machine based diagnostic system for breast cancer using swarm intelligence. J Med Syst. doi:10.1007/s10916-011-9723-0
12. Cortes C, Vapnik V (1995) Support vector network. Mach Learn 20:273–297
13. Cristianini N, Taylor JS (2000) An introduction to support Vector Machines: and other kernel-based learning methods. Cambridge University Press, Cambridge
14. Evgeniou T, Pontil M, Poggio T (2000) Regularization networks and support vector machines. In: Bartlett P, Scholkopf B, Schurmans D, Smola AJ (eds) Advances in large margin classifiers. MIT Press, Cambridge, pp 171–203
15. Fan CY, Changb PC, Linb JJ, Hsieh JC (2011) A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. Appl Soft Comput 11(1):632–644
16. Francois D, Rossi F, Wertz V, Verleysen M (2007) Resampling methods for parameter-free and robust feature selection with mutual information. Neurocomputing 70:1276–1288
17. Fung G, Mangasarian OL (2004) A feature selection Newton method for support vector machine classification. Comput Optim Appl 28(2):185–202
18. Fung G, Mangasarian OL (2003) Finite {N}ewton method for {L}agrangian support vector machine classification. Neurocomputing 55(1–2):39–55
19. Fung G, Mangasarian OL (2001) Proximal support vector machine classifiers. Proceedings of KDD'01 seventh ACM SIGKDD international conference on Knowledge Discovery and Data Mining, San Francisco, pp 77–86. ISBN: 1-58113-391-X. doi:10.1145/502512.502527
20. Goodman D, Boggess L, Watkins A (2002) Artificial immune system classification of multiple-class problems. In: Dagli CH, Buczak AL, Ghosh J, Ersoy O, Kercel SW (eds) Intell Eng Syst Artif Neural Net, vol 12, pp 179–184
21. Gunn SR (1998) Support vector machines for classification and regression. Technical Report, Faculty of Engineering, University of Southampton
22. Hamilton HJ, Shan N, Cerone N (1996) RIAC: a rule induction algorithm based on approximate classification. Technical Report CS 96-06, University of Regina. ISBN 0-7731-0321-X
23. Hsu CW, Chang CC, Lin CJ (2003) A practical guide to support vector classification. Technical Report, Department of Computer Science and Information Engineering, National Taiwan University
24. Huang ML, Hung YH, Chen WY (2010) Neural network classifier with entropy based feature selection on breast cancer diagnosis. J Med Syst 34(5):865–873
25. Joachims T, Nedellec C, Rouveirol C (1998) Text categorization with support vector machines: learning with many relevant. Springer, Springer-Verlag GmbH, Berlin
26. Joachims T (1998) SVM light. http://svmlight.joachims.org/
27. Karabatak M, Ince MC (2009) An expert system for detection of breast cancer based on association rules and neural network. Exp Syst Appl 36(2, Part 2):3465–3469
28. Kerekes J (2008) Receiver operating characteristic curve confidence intervals and regions. IEEE Geosci Remote Sens Lett 5(2):251–255
29. Lee YJ, Mangasarian OL (2001) {SSVM}: a smooth support vector machine. Comput Optim Appl 20:5–22
30. Liu HX, Zhang RS, Luan F et al (2003) Diagnosing breast cancer based on support vector machines. J Chem Inf Comput Sci 43(3):900–907
31. Mangasarian OL, Setiono R, Wolberg WH (1990) Pattern recognition via linear programming: theory and application to medical diagnosis. Proceedings of the workshop on large-scale numerical optimization, SIAM, Philadelphia, pp 22–31
32. Mangasarian OL, Musicant DR (2000) Lagrangian Support Vector Machine Classification. Tec. Report, Data Mining Institute, Computer Sciences Department, University of Wisconsin
33. Mangasarian OL, Musicant DR (1999) Successive overrelaxation for support vector machines. IEEE Trans Neural Networks 10:1032–1037

34. Mangasarian OL (2000) Generalized support vector machines. In: Smola A, Bartlett P, Scholkopf B, Schuurmans D (eds) Advances in large margin classifiers. MIT Press, Cambridge, pp 135–146
35. McAree B, O'Donnell ME, Spence A et al (2010) Breast cancer in women under 40 years of age: a series of 57 cases from Northern Ireland. Breast 19(2):97–104
36. Mitchell T (1997) Machine learning. The McGraw-Hill Companies, Inc., New York
37. Nauck D, Kruse R (1999) Obtaining interpretable fuzzy classification rules from medical data. Artif Intell Med 16(2):149–169
38. NCSS (2012) Statistical and power analysis software. http://www.ncss.com. Accessed in April 2012
39. Osuna E, Freund R, Girosit F (1997) Training support vector machines: an application to face detection. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun 17–19, pp 130–136
40. Park SH, Goo JM, Jo CH (2004) Receiver operating characteristic (ROC) curve: practical review for radiologists. Korean J Radiol 5(1):11–18
41. Pena-Reyes CA, Sipper M (1999) A fuzzy-genetic approach to breast cancer diagnosis. Artif Intell Med 17(2):131–155
42. Polat K, Gunes S (2007) Breast cancer diagnosis using least square support vector machine. Digit Signal Process 17(4):694–701
43. Platt J (1998) Sequential minimal optimization: A fast algorithm for training support vector machines. Technical Report MSR-TR-98-14
44. Quinlan J (1996) Improved use of continuous attributes in C4. 5. J Artif Intell Res 4:77–90
45. Sahan S, Polat K, Kodaz H, Günes S (2007) A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis. Comput Biol Med 37(3): 415–423
46. Schmidt M (1996) Identifying speaker with support vector networks. Interface'96 Proceedings, Sydney
47. Scholkopf B, Burges C, Vapnik V (1995) Extracting support data for a given task. In: Fayyad UM, Uthurusamy R (eds) Proceedings, first international conference on knowledge discovery & data mining. AAAI Press, Menlo Park
48. Scholkopf B, Burges C, Vapnik V (1996) Incorporating invariances in support vector learning machines. In: von der Malsburg C, von Seelen W, Vorbruggen JC, Sendhoff B (eds) Artificial neural networks- ICANN'96, vol 1112. Springer Lecture Notes in Computer Science, Berlin, pp 47–52
49. Setiono R (2000) Generating concise and accurate classification rules for breast cancer diagnosis. Artif Intell Med 18(3):205–219
50. Sherrod PH (2011) DTREG predictive modeling software. www.dtreg.com
51. Ster B, Dobnikar A (1996) Neural networks in medical diagnosis: comparison with other methods. Proceedings of the international conference on engineering applications of neural networks, pp 427–430
52. Taylor JS, Cristianini N (2004) Kernel methods for pattern analysis. Cambridge University Press, Cambridge
53. Übeyli ED (2005) A mixture of experts network structure for breast cancer diagnosis. J Med Syst 29(5):569–579
54. Übeyli ED (2009) Adaptive neuro-fuzzy inference systems for automatic detection of breast cancer. J Med Syst 33(5):353–358
55. Ubeyli ED (2007) Implementing automated diagnostic systems for breast cancer detection. Expert Syst Appl 33(4):1054–1062
56. UCI (2012) Machine learning repository. http://archive.ics.uci.edu/ml/index.html. Accessed on 10 Aug 2012
57. Vapnik VN (1995) The nature of statistical learning theory. Springer, New York
58. Vapnik VN (1999) The nature of statistical learning theory, 2nd edn. New York, Springer
59. Vapnik V, Golowich S, Smola A (1997) Support vector method for function approximation, regression estimation, and signal processing. In: Mozer M, Jordan M, Petsche T (eds) Advances in neural information processing systems 9. Cambridge, MIT Press, pp 281–287
60. Wolberg WH, Mangasarian OL (1990) Multisurface method of pattern separation for medical diagnosis applied to breast cytology. Proc Natl Acad Sci 87:9193–9196
61. Yuan Q, Cai C, Xiao H et al (2007) Diagnosis of breast tumours and evaluation of prognostic risk by using machine learning approaches. Commun Comput Inform Sci 2:1250–1260