Technology and Health Care 24 (2016) 31–42
DOI 10.3233/THC-151071
IOS Press

Machine learning models in breast cancer survival prediction

Mitra Montazeri[a,b], Mohadeseh Montazeri[c,d], Mahdieh Montazeri[e,*] and Amin Beigzadeh[f]
[a]*Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran*
[b]*Computer Engineering Department, Shahid Bahonar University, Kerman, Iran*
[c]*Social Determinants of Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran*
[d]*Department of Computer, Technical and Vocational University, Kerman, Iran*
[e]*Research Center for Modeling in Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran*
[f]*Health Services Management Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran*

Received 31 March 2015
Accepted 20 October 2015

Abstract.
BACKGROUND: Breast cancer is one of the most common cancers with a high mortality rate among women. With the early diagnosis of breast cancer survival will increase from 56% to more than 86%. Therefore, an accurate and reliable system is necessary for the early diagnosis of this cancer. The proposed model is the combination of rules and different machine learning techniques. Machine learning models can help physicians to reduce the number of false decisions. They try to exploit patterns and relationships among a large number of cases and predict the outcome of a disease using historical cases stored in datasets.
OBJECTIVE: The objective of this study is to propose a rule-based classification method with machine learning techniques for the prediction of different types of Breast cancer survival.
METHODS: We use a dataset with eight attributes that include the records of 900 patients in which 876 patients (97.3%) and 24 (2.7%) patients were females and males respectively. Naive Bayes (NB), Trees Random Forest (TRF), 1-Nearest Neighbor (1NN), AdaBoost (AD), Support Vector Machine (SVM), RBF Network (RBFN), and Multilayer Perceptron (MLP) machine learning techniques with 10-cross fold technique were used with the proposed model for the prediction of breast cancer survival. The performance of machine learning techniques were evaluated with accuracy, precision, sensitivity, specificity, and area under ROC curve.
RESULTS: Out of 900 patients, 803 patients and 97 patients were alive and dead, respectively. In this study, Trees Random Forest (TRF) technique showed better results in comparison to other techniques (NB, 1NN, AD, SVM and RBFN, MLP). The accuracy, sensitivity and the area under ROC curve of TRF are 96%, 96%, 93%, respectively. However, 1NN machine learning technique provided poor performance (accuracy 91%, sensitivity 91% and area under ROC curve 78%).
CONCLUSIONS: This study demonstrates that Trees Random Forest model (TRF) which is a rule-based classification model was the best model with the highest level of accuracy. Therefore, this model is recommended as a useful tool for breast cancer survival prediction as well as medical decision making.

Keywords: Breast cancer survival prediction, classification, machine learning models

*Corresponding author: Mahdieh Montazeri, Research Center for Modeling in Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran. Tel./Fax: +98 31325406; E-mail: montazeri@kmu.ac.ir.

1. Introduction

Breast cancer is one of the most common cancers with a high mortality rate among women. As reported by WHO, it is the 2nd most common cause of cancer death among women [1]. In the United States, an individual woman has a 1-in-8 chance of developing breast cancer [2]. With early diagnosis and detection of breast cancer (up to five years after cell division) survival will increase from 56% to more than 86% [3]. Uncontrolled growth of cells in breast tissues is the main cause of breast cancer [4]. The abnormal growth of cells can be benign or malignant. Benign tumors are non-intensive while malignant tumors are cancerous and can spread to other parts of the body. Thus, correct diagnosis of tumor for treatment should be considered [5]. 5%–10% of cancers are due to an abnormality that is inherited from the parents and about 90% of breast cancers are due to genetic abnormalities that happen as a result of the aging process [6]. An expert physician faces with complex issues in the diagnosis of breast cancer. But modern medical diagnoses are based on data obtained through clinical observation or other tests. In this regard, they can help physicians to diagnose breast cancer. Therefore, an accurate and reliable system is necessary for the early diagnosis of benign or malignant tumors [7].

One example of these models is machine learning models which are increasingly used in the field of science [8–10]. The main objective of these models is to determine effective variables and the relationship between them [11]. These models can also be used for prediction and estimation [12]. Machine learning models can be defined as a process to design a model that is learned through experience and to improve its performance. These models are a field of artificial intelligence and are an active field of research in different sciences. Also, in medical, machine learning techniques are widely used for the diagnosis of breast cancer and also to differentiate between benign and malignant cancers [13].

Bahrampour et al. [14] proposed Hidden Markov model to predict the breast cancer mortality. The classification accuracy of this method is 0.939. Wang et al. [15] used Artificial Neural Network (ANN) to predict the patients' five-year survivability. There were five risk factor and 604 patients who were diagnosed and received standard treatment in one hospital during 2000 to 2003 in Taiwan. The accuracy classification and area under the receiver operating characteristic (ROC) curve of their method are 85% and 0.79, respectively. In Ref. [16] proposed three data mining approaches to predict breast cancer survivability. Their proposed data mining methods are artificial neural networks, decision trees and logistic regression with accuracy of 91.2%, 93.6% and 89.2%.

In reference [17] ensemble learning is used to improve classification. In their proposed techniques, three methods namely Bagging, Boosting, and Random subspace were used to detect heart valve disorders. The database used in this method included 91 risk factors. These risk factors were used for feature extraction methods with the audio signal fetal heart rate of 215 cases; 95 and 120 normal and non-normal cases, respectively. In Ref. [18], rapid and adaptive diagnostic systems were presented based on Learning Vector Quantization (LVQ) artificial neural network. In this study, in order to increase the success rate of diagnosis methods and to reduce the decision time, a reinforcement mechanism was proposed for the LVQ network. The parameters of reinforcement learning mechanisms were updated in an adaptive neural network. Their method has been applied both for detecting breast cancer as well as thyroid disorders. In Ref [19], a support vector machine (SVM) was proposed for the diagnosis of urological disorders. An accurate diagnosis of this disease has become a major dilemma. Using information of 381 patients with urological disorders, findings showed that SVM-based method could achieve the classification accuracy at 84.25%. In reference [20], to extract useful information and detect tumors, a combination of tools and a support vector machine (K-SVM) algorithm was presented. The purpose of the K-means algorithm is separately identifying hidden patterns of benign and malignant tumors. Then support vector machine (SVM) is used to classify incoming tumor and the resulting accuracy is 97.38%.

In addition, mobile phones are now widely used to monitor the conditions of the human body. These methods can be effective with the help of automated classification tools for Breast Cancer (active or mobile devices communicate via the web) such as endocrinologists reduction. Classification techniques are automated tools to predict and diagnose medical issues. With the help of machine learning models, we can help the medical community to diagnose and predict diseases [21,22].

Machine learning methods can help physicians to reduce the number of false positive and false Negative decisions [23]. New techniques such as knowledge discovery in databases (KDD) has become a proper research tool for medical researchers who try to identify and exploit patterns and relationships among a large number of variables and predict the outcome of a disease using historical cases stored in datasets [24,25]. In the following, seven machine learning models are presented.

2. Materials and methods

2.1. Data collection

In order to collect data, a database which included the information of 900 patients during 1999–2007 was used. The data were recorded by the Cancer Registry Organization of Kerman Province, in Iran. We focused our study on the information about breast cancer and its associated risk factors among Iranian adult people. The risk factors including Sex (female, male), age (15 to 90 years old), Morphology (Neoplasm malignant, Carcinoma, Infiltrating duct carcinoma, Medullary carcinoma, and Lobular carcinoma), grade (well, moderate and poor), surgery (Yes, No), radiotherapy (Yes, No), chemotherapy (Yes, No) and living status (dead/alive) were used to compare the performance of these two models.

2.2. K-fold cross validation strategy

In this method, the data are partitioned into k subsets. In each $k$iteration, each subset is used for validation and k-1 used for training. This procedure is repeated k times and all data are used exactly k times for training and one for testing. Finally, mean k-time validation result is elected as a final estimate value. In this study, cross-validation of 10-folds is used.

2.3. Naïve Bayes (NB) machine learning models

This model has become recently popular and is increasingly being used [26]. This model is a statistical pattern recognition method which has an accurate assumption about how the data is produced. This model use training samples to estimate the parameters of NB. This classifier is a simple classifier, which assumes all sample features are independent. When the hypothesis is clearly false, NB classify data very well, because the hypothesis of classification are only a sign of the function estimation and the approximation of the function is performed with low accuracy, while the accuracy of the classifier is high [27]. Naïve Bayes learns from training data and the conditional probability of each variable X_k given the class label C [28]. Classification is done by applying Bayes rule to compute the probability of C given the particular instance, $X_1, \ldots X_n$ by the Eq. (1):

$$P(C = c|X_1 = x_1, \ldots, X_n = x_n) \tag{1}$$

Due to the fact that this classifier is based on the assumption that variables are conditionally independent, the posterior probability of the class is formulated by the Eq. (2):

$$P(C = c|X_1 = x_1, \ldots, X_n = x_n) = P(C = c) * \Pi_{X_k} P(X_k = x_k|C = c) \tag{2}$$

The classification result is the class with the highest posterior probability Eq. (3):

$$\max_c \Pi_{X_k} P(X_k = x_k | C = c) \tag{3}$$

2.4. Trees Random Forest (TRF) machine learning model

Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [29]. The algorithm, random forest, was developed by Leo Breiman [29], and a "random forest" is a sign of their brand. This is the first so-called "random decision forest" proposed by Tin Kam Ho in 1995 [30]. The avoidance of overtraining has been an important advantage of this approach. This classifier is a technique based on the ensemble learning. Ensemble learning is a model that uses multiple tree models which can be used for classification and regression. Ensemble learning is based on the accuracy and significance of variables.

TRF algorithm is implemented by applying the following steps.

The first phase is training:

- Nominate the number of training samples and classification variable, N and M, respectively. The number of input variables to determine the decision at a node of the tree must be much less than M.
- Production of training data and test data, T, by 10-fold cross validation method.
- Learning T trees by T training set generated in the previous step.
- Estimate the error rate of all the trees by T training samples.
- Iterate algorithm to reduce the error.

The second phase is testing:

- For determining the value of the target variable, voting between the produced trees is performed based on the maximum votes.

2.5. 1-Nearest Neighbor (1NN) machine learning model

K-nearest-neighbor (*KNN*) classifier is one of the most widely used classification algorithms that can be used in a wide variety of applications. *KNN* is a method for statistical pattern recognition to detect different classes of patterns. A similar pattern in the k training samples (k nearest neighbor) is calculated by their distance to the pattern by using weighted Euclidean distance. In this method, by having a feature vector, k closest feature vectors in the training data are found by using standard weighted Euclidean distance. The class of test pattern is identified according to the classes of the k nearest neighbors and the frequency of them. If $k = 1$ (1-NN), the class of pattern will be the nearest sample. This procedure is shown in Fig. 1. In this figure, the class of x_q is positive; in order to the nearest sample to x_q is a sample which belongs to positive class.

2.6. AdaBoost (AD) machine learning model

AD is the abbreviation of Adaptive Boosting and is a machine learning algorithm that was invented by Freund and Robert Shapyr [31]. AD is actually a meta-algorithm to improve the performance along with other learning algorithms.

AdaBoost belong to a particular method of training a boosted classifier which is a classifier in the form of Eq. (4):

$$F_T = \Sigma_{t=1}^{T} f_t(x), \tag{4}$$

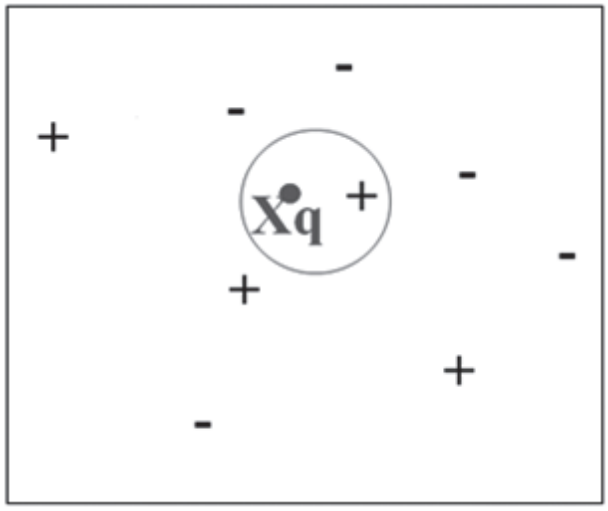

Fig. 1. Examples of the performance of the 1NN algorithm.

where each f_t is a weak learner that give an object as input and get back a real valued result indicating the class of the sample. The sign of the weak learner output identifies the predicted sample class and the value gives the confidence in that classification. Similarly, the T-layer classifier will be positive if the sample is believed to be in the positive class and negative otherwise.

Each weak learner produces an output, hypothesis $h(x_i)$, for each sample in the training set. At each iteration, t, a weak learner is selected and assigned a coefficient such that the sum training error of the resulting t-stage boost classifier is minimized (Eq. (5)).

$$E_t = \Sigma_i E[F_{t-1}(x_i) + \alpha_t h(x_i)], \tag{5}$$

Here $F_{t-1(x)}$ is the boosted classifier that has been built up to the previous stage of training, $E(F)$ is some error function and $f_t(x) = \alpha_t h(x)$ is the weak learner that is being considered for addition to the final classifier.

In this algorithm, classification in each new stage is based on samples incorrectly classified. AD is sensitive to the data noise and outliers, but over fitting is more superior to other learning algorithms. The base classifier used in this algorithm is random classification (50%). With more repetitions, a random classification is better and thus the performance of the algorithm is improved. Although the error of base classifier is high, the overall performance of the model improves with a negative coefficient. In each AD algorithm iteration, $t = 1, \ldots, T$, add a weak classifier. In each call, based on the samples, the weights are updated. In each iteration, the weights of incorrectly classified samples are increased and weights of correctly classified are decreased. Therefore, a new classifier focuses on the samples that are difficultly learned. The AD used in this paper is M1 type [32].

2.7. Support Vector Machine (SVM) machine learning model

A SVM is a supervised learning method [33] that is used for classification [34] and regression [35]. This algorithm is a relatively new approach and has a good performance in recent years. SVM classifier is based on linear classifiers and in data which separated by line is attempted to select a line which is

marginally more confidence. The goal is solving the Eq. (6) for x data point and it finds the optimal line for this data.

$$W^T x + b = 0 \tag{6}$$

Where W is data point and b is the bias weight

Before separating data by line, in order to the model is able to classify the data with high complexity, data is mapped to a higher dimension by function [36]. The high dimension problem can be solved by Binary Lagrange's theorem [37]. In this method, a simple function called a kernel function that is the Vector multiplication of phi function is used. Different kernel functions, including exponential kernels, polynomial, and Sigmoid can be used [38].

2.8. RBF Network (RBFN)

A radial basis function network is a kind of artificial neural network that the activation functions are radial basis functions. A linear combination of radial basis functions of the inputs and neuron parameters are used as the output of the network. One of the applications of the Radial basis function networks is classification. They were first formulated in 1988 by Broomhead and Lowe, both researchers at the Royal Signals and Radar Establishment [39–41].

RBFN typically have three layers: an input layer, a hidden layer with a non-linear RBF activation function and a linear output layer. The input can be modeled as a vector of real numbers. The output of the network is a scalar function of the input vector, $\varphi : R^n \to R$, and is given by Eq. (7):

$$\varphi(x) = \Sigma_{i=1}^{N} \alpha_i \rho(||x - c_i||) \tag{7}$$

where N is the number of neurons in the hidden layer, is the center vector for neuron i, and is the weight of neuron i in the linear output neuron.

2.9. Multilayer Perceptron (MLP)

This Classifier uses back-propagation to classify samples. MP can be created either by hand or by an algorithm. During the training time, the network can also be modified. In this network, all the nodes are sigmoid except for when the class is numeric in which case the output nodes become unthresholded linear units [42].

In MLP, some neurons use a nonlinear activation function which was developed to model the frequency of action potentials, or firing, of biological neurons in the brain. The two main activation functions used in this study are both sigmoids, and are described by Eqs (8) and (9):

$$y(v_i) = \tanh(v_i), \tag{8}$$

$$y(v_i) = (1 + e^{-v_i})^{-1}, \tag{9}$$

The first function is a hyperbolic tangent which ranges from -1 to 1, and the second, the logistic function, is similar in shape but ranges from 0 to 1. Here is the output of the i^{th} *node* (neuron) and is the weighted sum of the input synapses.

Table 1
Risk factors of breast cancer dataset

Variable	Levels	Frequency (%)
Sex	Female	876
	Male	24
Age	15–30	61
	30–45	337
	45–60	354
	60–75	128
	75–90	20
Morphology	Neoplasm malignant	391
	Carcinoma	28
	Infiltrating duct carcinoma	454
	Medullary carcinoma	16
	Lobular carcinoma	11
Grade	Well	61
	Moderate	791
	Poor	48
Surgery	Yes	446
	No	454
Radiotherapy	Yes	329
	No	571
Chemotherapy	Yes	254
	No	646
Status	Alive	803
	Died	97

3. Result

The proposed machine learning models are used to Breast Cancer survival prediction. In the proposed method, the seven models of machine learning NB, TRF 1NN, AD, SVM, RBFN and MLP were applied in the prediction of breast cancer survival.

In this study, Weka software was used [43]. The data set included 900 patient records in which 803 patients were alive and 97 patients were dead. Thus, the target variable is divided into two groups: alive or dead. The risk factors used in this study are shown in Table 1.

3.1. Configurations of the proposed models

Suppose that the numbers of positive and negative examples are P and N, respectively, the following definition can be stated:

$FP =$ Alive people incorrectly identified as dead.
$TP =$ Dead people correctly diagnosed as dead.
$TN =$ Alive people correctly identified as dead.
$FN =$ Dead people incorrectly identified as alive.

Consequently, 10 to 12 formulas are represented:

$$\text{True positive rate } = \frac{TP}{P} \tag{10}$$

$$\text{False positive rate} = \frac{FP}{N} \tag{11}$$

Table 2
A comparison between seven machine learning models, NB, TRF, 1NN, AD, SVM, RBFN and MLP (%) by 10-fold CV validation strategy

No.	Machine learning model	Classification accuracy	TP rate	FP rate	Precision	Sensitivity	Specificity	ROC area
1	NB	0.95	0.95	0.333	0.94	0.95	0.99	**0.94**
2	TRF	**0.96**	**0.96**	**0.241**	**0.96**	**0.96**	0.98	0.93
3	1NN	0.91	0.92	0.346	0.91	0.92	0.95	0.78
4	AD	0.94	0.94	0.324	0.94	0.94	0.98	0.94
5	SVM	0.94	0.95	0.333	0.94	0.95	0.98	0.81
6	RBFN	0.95	0.95	0.323	0.95	0.95	**0.99**	0.93
7	MLP	0.95	0.96	0.25	0.95	0.96	0.98	0.93

$$\text{Classification accuracy } = \frac{TP + TN}{P + N} \tag{12}$$

3.2. Performance of the proposed model

In this study, seven models of machine learning NB, TRF, 1NN, AD, SVM, RBFN and MLP are used to Breast cancer survival prediction. In this study, in MLP model, we set the learning rate to 0.3, momentum to 0.2, training time to 500, and validation threshold to 20. MLP uses back-propagation to classify instances and the nodes in this network are all sigmoid (except for when the class is numeric in which case the output nodes become unthresholded linear units). In NB classifire use a default precision of 0.1 for numeric attributes when Classifier is build with zero training instances. In TRF model the maximum depth of TRF is set to 0 and the 10 number of trees to be generated. 1NN model uses normalized Euclidean distance to find the training instance closest to the given test instance, and predicts the same class as this training instance. If multiple instances have the same (smallest) distance to the test instance, the first one found is used. In AD model the base classifier is Decision Stump [44]. The number of iterations to be performed is set 10 and weight threshold for weight pruning is set 100. In SVM model the complexity parameter is set to 1 and the epsilon for round-off error (shouldn't be changed) is set to 1.0E-12. The kernel function is used in this study is the polynomial kernel [45]. The tolerance parameter (shouldn't be changed) is set to 0.0010. RBFN model use tha value of 0.1 to minimize the standard deviation for the clusters. The number of clusters for K-Means to generate and the Ridge value for the logistic or linear regression is set to 2 and 1.0E-8, respectively.

These models of machine learning have been implemented and the results of comparison are summarized in Tables 2 and 3. As can be seen from this table, these models are compared in terms of specificity, sensitivity, precision, accuracy, and the ROC curve. In each evaluation the best value is highlighted. As you can see in these tables, TRF and MLP machine learning model are the best model in the most evaluation merits in Tables 2 and 3, respectively.

Also, ROC curve is plotted in Fig. 3 for seven machine learning models. As can be seen from this figure, NB model and AD are in a higher position than the other machine learning models, and therefore have more area under the curve. In the ROC curve values from 0 to 0.5 represents random classification and 0.5 to 1 indicates that the model has a general diagnostic ability.

4. Discussion and conclusions

Breast cancer is a kind of cancer with a high mortality rate among women and it is the most common cause of death in women. With the early diagnosis and detection of breast cancer survival will increase

Table 3
A comparison between seven machine learning models, NB, TRF, 1NN, AD, SVM, RBFN and MLP (%) by training-testing validation strategy

No.	Machine learning model	Classification accuracy	TP rate	FP rate	Precision	Sensitivity	Specificity	ROC area
1	NB	0.95	0.64	0.014	0.849	0.64	0.986	0.95
2	TRF	**0.97**	0.8	0.011	0.895	0.8	0.988	**0.98**
3	1NN	0.81	**1**	0.213	0.362	**1**	0.787	0.89
4	AD	0.94	0.68	0.032	0.717	0.68	0.967	0.95
5	SVM	0.95	0.65	0.011	0.875	0.65	0.988	0.82
6	RBFN	0.95	0.67	0.019	0.813	0.67	0.981	0.95
7	MLP	**0.97**	0.79	**0.01**	**0.905**	0.79	**0.990**	0.96

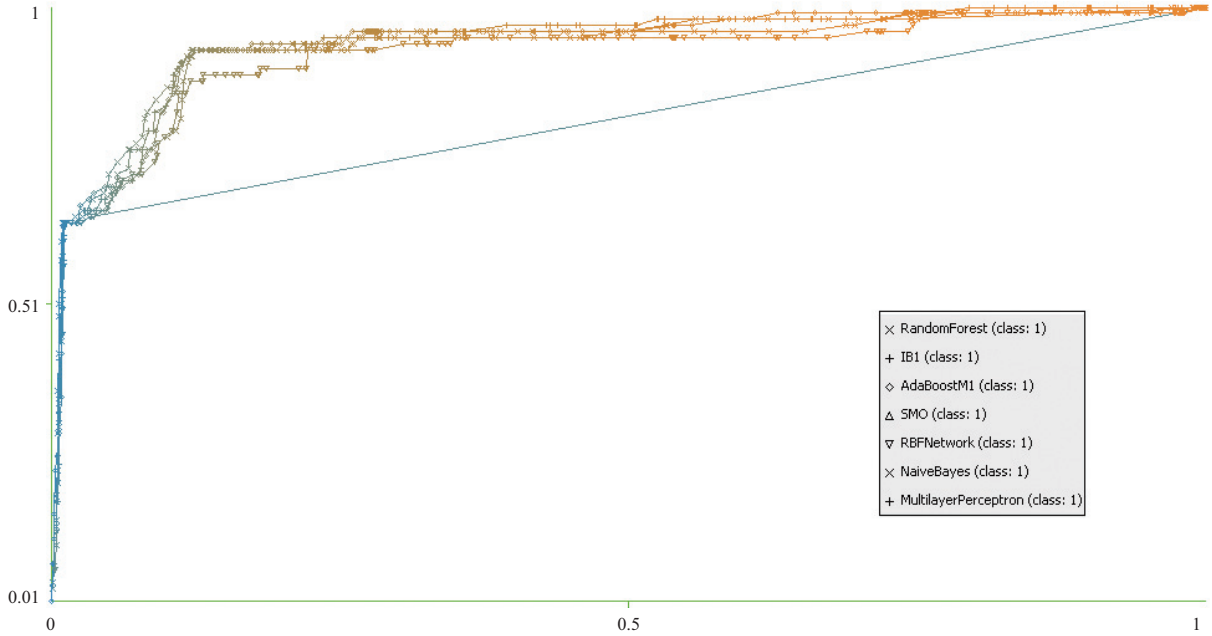

Fig. 2. Receiver operating characteristic curves (ROC) for the seven machine learning models. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/THC-151071)

significantly. Automatic classification tools can be used as a diagnostic tool to reduce the workload of doctors. This study evaluates the selected classification algorithm for predicting breast cancer survival. The data from the Cancer Registry Organization of Kerman Province, in Iran, were considered for model comparison. The classifiers used in this study included NB, TRF, 1NN, AD, SVM, RBFN and MLP. These algorithms were evaluated based on the criteria of sensitivity, specificity, accuracy, ROC curve and precision. This study is related to health research and it is particularly important in resource allocation therapy for high-risk patients.

The result of comparison shows that the performance of the TRF model was the best. Therefore, TRF model can be proposed for breast cancer survival. This method is fast to in application and can be used for feature selection (finding efficient risk factors) alone. This method resists over training and ability to handle data without preprocessing, e.g. data does not need to be rescaled, transformed, or modified. During the evaluation, TRF has the following significant performances:

– Natural handling of “mixed” type data.

- Handling of missing values.
- Robustness to outlier in input space.
- Insensitive to monotone transformation of inputs.
- Computational scalability.
- Ability to deal with irrelevant inputs.

Comparing TRF with SVM, RBFN and MLP, it is true that they have about the same accuracy. However, TRF is more interpretable due to these facts:

- Feature importance can be estimated during training for little additional computation.
- Plotting of sample proximities.
- Visualization of output decision trees.

TRF readily handles larger numbers of predictors and faster to train. It has fewer parameters and cross validation is unnecessary (it generates an internal unbiased estimate of the generalization error (test error) as the forest building progresses).

Also, in comparison to AD, both derive many benefits from ensembling and both can be applied to ensembling decision trees. However,

- AD performs an exhaustive search for the best predictor to split on; TRF searches only a small subset.
- AD may be more difficult to model and requires more attention to parameter tuning than TRF.
- On very large training sets, AD can become slow with many predictors, while TRF which selects only a subset of predictors for each split can handle significantly larger problems before slowing.
- TRF will not overfit the data. AD can overfit the data (though means can be implemented to lower the risk of it).
- If parallel hardware is available, (e.g. multiple cores), TRF embarrassingly parallel with the need for shared memory as all trees are independent.

References

[1] Marklund, L. and L. Hammarstedt, *Impact of HPV in oropharyngeal cancer.* Journal of oncology, 2010. **2011**.
[2] Bellaachia, A. and E. Guven, *Predicting breast cancer survivability using data mining techniques.* Age, 2006. **58**(13): p. 10-110.
[3] Paulin, F. and A. Santhakumaran, *Extracting rules from feed forward neural networks for diagnosing breast cancer.* Artificial Intelligent Systems and Machine Learning, 2009. **1**(4): p. 143-146.
[4] Liu, X., et al., *Global and Local Structure Preservation for Feature Selection.* Neural Networks and Learning Systems, IEEE Transactions on, 2013. **PP**(99): p. 1-1.
[5] Hota, H., *Diagnosis of breast cancer using intelligent techniques.* Int J Emerg Sci Eng (IJESE), 2013. **1**(3): p. 45-53.
[6] Reif, M. and F. Shafait, *Efficient feature size reduction via predictive forward selection.* Pattern Recognition, 2014. **47**(4): p. 1664-1673.
[7] Gayathri, B., C. Sumathi, and T. Santhanam, *BREAST CANCER DIAGNOSIS USING MACHINE LEARNING ALGORITHMS–ASurvey.* International Journal of Distributed and Parallel Systems, 2013. **4**(3): p. 105.
[8] Montazeri, M. and H. Nezamabadi-pour. *Automatic extraction of eye field from a gray intensity image using intensity filtering and hybrid projection function*. in *Communications, Computing and Control Applications (CCCA), 2011 International Conference on*. 2011. IEEE.
[9] Montazeri, M., H. Nezamabadi-pour, and M. Montazeri, *Automatically Eye Detection with Different Gray Intensity Image Conditions.* Computer Technology and Application, 2012. **3**: p. 525-532.
[10] Mitra, M., et al. *Cooperating of Local Searches based Hyperheuristic Approach for Solving Traveling Salesman Problem*. in *IJCCI (ECTA-FCTA)*. 2011.
[11] Montazeri, M., H. Naji, and M. Montazeri. *A Novel Memetic Feature Selection Algorithm*. in *5th International Conference on Information and Knowledge Technology (IKT 2013)*. 2013.

[12] Hashemian, A.H., et al., *Comparison of Artificial Neural Networks and Cox Regression Models in Prediction of Kidney Transplant Survival.* International journal of Advanced Biological and Biomedical Research, 2013. **1**(10): p. 1204-1212.
[13] Rajesh, K. and S. Anand, *Analysis of SEER dataset for breast cancer diagnosis using C4. 5 classification algorithm.* International Journal of Advanced Research in Computer and Communication Engineering, 2012. **1**(2): p. 2278-1021.
[14] Bahrampour, A. and M. Montazeri, *Prediction of breast cancer mortality by Hidden Markov model*, 2015, School of Health, Kerman University of Medical Sciences, Kerman, Iran.
[15] Tan-Nai, W., C. Chung-Hao, and C. Hung-Wen. *Predicting post-treatment survivability of patients with breast cancer using Artificial Neural Network methods*. in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. 2013.
[16] Delen, D., G. Walker, and A. Kadam, *Predicting breast cancer survivability: a comparison of three data mining methods.* Artificial intelligence in medicine, 2005. **34**(2): p. 113-127.
[17] Das, R. and A. Sengur, *Evaluation of ensemble methods for diagnosing of valvular heart disease.* Expert Systems with Applications, 2010. **37**(7): p. 5110-5115.
[18] Alkım, E., E. Gürbüz, and E. Kılıç, *A fast and adaptive automated disease diagnosis method with an innovative neural network model.* Neural Networks, 2012. **33**: p. 88-96.
[19] Gil, D. and M. Johnsson, *Using support vector machines in diagnoses of urological dysfunctions.* Expert Systems with Applications, 2010. **37**(6): p. 4713-4718.
[20] Zheng, B., S.W. Yoon, and S.S. Lam, *Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms.* Expert Systems with Applications, 2014. **41**(4): p. 1476-1482.
[21] Montazeri, M., M.S. Baghshah, and A. Enhesari, *Hyper-Heuristic Algorithm for Finding Efficient Features in Diagnose of Lung Cancer Disease.* J. Basic Appl. Sci. Res, 2013. **3**(10): p. 134-140.
[22] Montazeri, M., H.R. Naji, and M. Montazeri, *Memetic Feature Selection Algorithm Based on Efficient Filter Local Search.* J. Basic Appl. Sci. Res., 2013. **3**(10): p. 126-133.
[23] Ramana, B.V., M.S.P. Babu, and N. Venkateswarlu, *A critical study of selected classification algorithms for liver disease diagnosis.* International Journal of Database Management Systems, 2011. **3**(2): p. 101-114.
[24] Montazeri, M. and M. Montazeri, *Machine learning models for predicting the diagnosis of liver disease.* Koomesh, 2014. **16**(1): p. 53-59.
[25] Montazeri, M., et al., *Identifying efficient clinical parameters in diagnose of liver disease.* HealthMED, 2014: p. 1115.
[26] Farid, D.M., et al., *Hybrid decision tree and naive Bayes classifiers for multi-class classification tasks.* Expert Systems with Applications, 2014. **41**(4): p. 1937-1946.
[27] Friedman, N., D. Geiger, and M. Goldszmidt, *Bayesian network classifiers.* Machine learning, 1997. **29**(2-3): p. 131-163.
[28] Dumitru, D., *Prediction of recurrent events in breast cancer using the Naive Bayesian classification.* Annals of the University of Craiova-Mathematics and Computer Science Series, 2009. **36**(2): p. 92-96.
[29] Breiman, L., *Random forests.* Machine learning, 2001. **45**(1): p. 5-32.
[30] Ferenc, G., et al. *A proposed approach to the classification of bearing condition using wavelets and random forests*. in *Embedded Computing (MECO), 2013 2nd Mediterranean Conference on*. 2013. IEEE.
[31] Freund, Y. and R.E. Schapire. *A desicion-theoretic generalization of on-line learning and an application to boosting*. in *Computational learning theory*. 1995. Springer.
[32] Freund, Y. and R.E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting.* Journal of computer and system sciences, 1997. **55**(1): p. 119-139.
[33] Cortes, C. and V. Vapnik, *Support-vector networks.* Machine learning, 1995. **20**(3): p. 273-297.
[34] Raikwal, J. and K. Saxena, *Performance evaluation of SVM and k-nearest neighbor algorithm over medical data set.* International Journal of Computer Applications, 2012. **50**(14).
[35] Abdi, M.J. and D. Giveki, *Automatic detection of erythemato-squamous diseases using PSO–SVM based on association rules.* Engineering Applications of Artificial Intelligence, 2013. **26**(1): p. 603-608.
[36] Aizerman, A., E.M. Braverman, and L. Rozoner, *Theoretical foundations of the potential function method in pattern recognition learning.* Automation and remote control, 1964. **25**: p. 821-837.
[37] Sivapriya, T., A.N.B. Kamal, and V. Thavavel, *Feature Selection for Dementia Classification using Support Vector Machine.* Artificial Intelligent Systems and Machine Learning, 2012. **4**(4): p. 241-247.
[38] Meyer, D., F. Leisch, and K. Hornik, *The support vector machine under test.* Neurocomputing, 2003. **55**(1): p. 169-186.
[39] Broomhead, D.S. and D. Lowe, *Radial basis functions, multi-variable functional interpolation and adaptive networks*, 1988, DTIC Document.
[40] Schwenker, F., H.A. Kestler, and G. Palm, *Three learning phases for radial-basis-function networks.* Neural networks, 2001. **14**(4): p. 439-458.
[41] Vasconcelos, G.C., M.C. Fairhurst, and D.L. Bisset, *Investigating feedforward neural networks with respect to the rejection of spurious patterns.* Pattern Recognition Letters, 1995. **16**(2): p. 207-212.
[42] Hassanien, A.E., et al., *MRI breast cancer diagnosis hybrid approach using adaptive ant-based segmentation and multilayer perceptron neural networks classifier.* Applied Soft Computing, 2014. **14**: p. 62-71.

[43] Hall, M., et al., *The WEKA data mining software: an update.* ACM SIGKDD explorations newsletter, 2009. **11**(1): p. 10-18.
[44] Pang, J., Q. Huang, and S. Jiang, *Multiple instance boost using graph embedding based decision stump for pedestrian detection*, in *Computer Vision-ECCV 2008*. 2008, Springer. p. 541-552.
[45] Burges, C.J., *A tutorial on support vector machines for pattern recognition.* Data mining and knowledge discovery, 1998. **2**(2): p. 121-167.