

# A Hybrid Computer-aided-diagnosis System for Prediction of Breast Cancer Recurrence (HPBCR) Using Optimized Ensemble Learning

Mohammad R. Mohebian<sup>a</sup>, Hamid R. Marateb<sup>a,b</sup>, Marjan Mansourian<sup>c,\*</sup>, Miguel Angel Mañanas<sup>b</sup>, Fariborz Mokarian<sup>d,e</sup>

<sup>a</sup> Biomedical Engineering Department, Engineering Faculty, University of Isfahan, Hezar Jerib St., 81746-73441, Isfahan, Iran

<sup>b</sup> Department of Automatic Control, Biomedical Engineering Research Center, Universitat Politècnica de Catalunya, BarcelonaTech (UPC), C. Pau Gargallo, 5, 08028 Barcelona, Spain

<sup>c</sup> Department of Biostatistics and Epidemiology, School of Public Health, Isfahan University of Medical Sciences, Hezar Jerib St., 81745 Isfahan, Iran

<sup>d</sup> Cancer Prevention Research Center, Isfahan University of Medical Sciences, Isfahan, Iran

<sup>e</sup> Department of Internal Medicine, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran

## ARTICLE INFO

### Article history:

Received 30 August 2016

Received in revised form 24 November 2016

Accepted 26 November 2016

Available online 06 December 2016

### Keywords:

Breast cancer

Cancer recurrence

Computer-assisted diagnosis

Machine learning

Prognosis

## ABSTRACT

Cancer is a collection of diseases that involves growing abnormal cells with the potential to invade or spread to the body. Breast cancer is the second leading cause of cancer death among women. A method for 5-year breast cancer recurrence prediction is presented in this manuscript. Clinicopathologic characteristics of 579 breast cancer patients (recurrence prevalence of 19.3%) were analyzed and discriminative features were selected using statistical feature selection methods. They were further refined by Particle Swarm Optimization (PSO) as the inputs of the classification system with ensemble learning (Bagged Decision Tree: BDT). The proper combination of selected categorical features and also the weight (importance) of the selected interval-measurement-scale features were identified by the PSO algorithm. The performance of HPBCR (hybrid predictor of breast cancer recurrence) was assessed using the holdout and 4-fold cross-validation. Three other classifiers namely as supported vector machines, DT, and multilayer perceptron neural network were used for comparison. The selected features were diagnosis age, tumor size, lymph node involvement ratio, number of involved axillary lymph nodes, progesterone receptor expression, having hormone therapy and type of surgery. The minimum sensitivity, specificity, precision and accuracy of HPBCR were 77%, 93%, 95% and 85%, respectively in the entire cross-validation folds and the hold-out test fold. HPBCR outperformed the other tested classifiers. It showed excellent agreement with the gold standard (i.e. the oncologist opinion after blood tumor marker and imaging tests, and tissue biopsy). This algorithm is thus a promising online tool for the prediction of breast cancer recurrence.

© 2016 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Computer-aided diagnosis (CAD) is using computers and software to interpret medical information. The purpose of CAD is to improve the diagnosis accuracy. In fact, CAD is used as a second opinion by the physicians to make the final diagnosis decision [1,2].

*Abbreviations:* CAD, computer-aided diagnosis; DT, decision tree; FH, family history of cancer; HPBCR, the proposed hybrid predictor of breast cancer recurrence; HRT, hormone therapy; I. Node, number of involved axillary lymph nodes; NR, lymph node involvement ratio; T. Node, number of dissected axillary lymph nodes; TS, tumor size; XRT, radiotherapy.

\* Corresponding author.

*E-mail addresses:* [van\\_mohebian@yahoo.com](mailto:van_mohebian@yahoo.com) (M.R. Mohebian), [h.marateb@eng.ui.ac.ir](mailto:h.marateb@eng.ui.ac.ir), [hamid.reza.marateb@upc.edu](mailto:hamid.reza.marateb@upc.edu) (H.R. Marateb), [j\\_mansourian@hlth.mui.ac.ir](mailto:j_mansourian@hlth.mui.ac.ir) (M. Mansourian), [miguel.angel.mananas@upc.edu](mailto:miguel.angel.mananas@upc.edu) (M.A. Mañanas), [mokarian@med.mui.ac.ir](mailto:mokarian@med.mui.ac.ir) (F. Mokarian).

Nowadays, CAD is used in many different fields in medicine including, but not limited to, early detection of breast cancer [3], lung cancer diagnosis [4], arrhythmia detection [5] and dental and maxillofacial lesions diagnosis [6]. Several studies have been reported in the literature focusing on the application of CAD for cancer diagnosis and prognosis [7,8].

Cancer is a collection of diseases that involves growing abnormal cells with the potential to spread to other parts of the body [9]. There are over 200 types of cancer. The most common types of cancer in women are breast, colorectal, lung, and cervical [10].

Cancer is the leading cause of death worldwide, accounting for 8.2 million deaths in 2012 [11]. The most common causes of cancer death are: lung (1.59 million deaths), liver (745,000), stomach (723,000), colorectal (694,000), breast (521,000), and esophageal (400,000) [11]. More than 60% of world's total new annual cancer cases occur in Africa, and Central and South America [12]. The cancer

incidence in Asia is also high [13]. It is expected that annual cancer cases will rise from 14 million in 2012 to 22 million within the next 2 decades [11,12]. Worthwhile, among all types of cancer, breast cancer is the second leading cause of cancer death among women [14].

In Iran, cancer is the third cause of death after coronary heart disease, and accidents [15]. Breast cancer is the leading type of cancer in Iranian females, accounting for 24.6% of all cancers. In Iran, the average women age with breast cancer is 49.6 years [16]. Among all of the provinces in Iran, Isfahan Province is the biggest and most important area located at a desert border in the center of Iran [17]. The rate of cancer is increasing rapidly in Isfahan Province. Cancer control and comprehensive prevention plan is thus necessary [18].

Breast cancer, a complex heterogeneous disease, occurs with a set of different clinical symptoms. It is usually diagnosed using blood tests, MRI, mammography, and CT scan and biopsy. The pathology results from biopsy samples indicate whether the suspicious area is cancerous. Cancer patients, then, undergo a systematic treatment procedure dependent on the cancer stage and additional lab tests such as hormone receptor status [19].

Cancer staging (assigning an ordinal numbers I–IV) is in fact the process to determine the extent to which a cancer has been spread and is based on the following four characteristics: 1) the tumor size, 2) whether the cancer is invasive or non-invasive, 3) the spread of the tumor into the lymph nodes, and 4) the spread of the tumor into other parts of the body (i.e. metastasis). Stage I is an isolated cancer (tumor size  $\leq 20$  mm in greatest dimension) while stage IV is a metastasis cancer. Most cancer deaths are due to cancer that has spread from its primary site to other organs [12,20].

Breast cancer is usually treated with surgery, which may be followed by chemotherapy, radiation, and hormone therapies [21]. When the cancer patient is initially treated, the disease can recur at any time. However, most recurrences happen in the first 5 years after treatment [22]. The recurrence could be local (near the mastectomy scar), regional (spread to nearby lymph nodes) or metastatic (spread to other parts of the body not near the breast). Some of the most common sites of recurrence outside the breast are the lymph nodes, bones, liver, lungs, and brain [23].

An important issue is whether we can optimize the treatments to increase the therapeutic efficacy. In fact, 5-year recurrence-free survival is an important treatment quality measure. In principle, it is possible to predict 5-year cancer recurrence using clinicopathologic characteristics of cancer patients [24]. Such a prediction could be used by doctors to make proper treatment plan to considerably prolong patient life [25].

The prediction systems were used for cancer diagnosis in the literature [7]. However, there are few studies focusing on cancer prognosis (including recurrence or survival analysis). Since the focus of the current study is prediction of cancer recurrence, the literature review on cancer recurrence prediction models is provided. Meanwhile, the table of available methods was provided in the Supplementary material S1.

Zeng suggested a mixture classification model containing a two-layer structure called mixture of rough set and support vector machine (SVM) for breast cancer prognosis with the average accuracy of 91% [26].

The Nottingham prognostic index (NPI) is a prognostic regression model proposed by Galea et al. [27] based on tumor size, histological grade, and lymph node status. The NPI calculation equation is as follows: tumor size (cm)  $\times 0.2$  + histological grade + lymph node point (negative nodes = 1; 1–3 positive nodes = 2;  $\geq 4$  positive nodes = 3). The patients were then classified into the low-risk (NPI point  $< 3.4$ ) and high-risk groups (NPI point  $\geq 3.4$ ) [27,28].

Kim et al. [28] used normalized mutual information index for feature selection and supported vector machines (SVM), Cox-proportional hazard regression model, and artificial neural network classifiers for classification in a sample size of 679 patients (the recurrence prevalence of 28.6%). The following features were used in their prognosis system: local invasion of tumor, number of tumors, number of metastatic

lymph nodes, the histological grade, tumor size, estrogen receptor, and lymphovascular invasion and reached the sensitivity, specificity and area under the curve of 89%, 73% and 0.85, respectively for the best classifier (SVM). Although the statistical power of their system is acceptable (Power = 89%  $>$  80%), the Type I error is beyond the acceptable range ( $\alpha = 0.17 >$  0.05).

Ahmad et al. [29] used the SVM, decision tree, and multilayer perceptron artificial neural network classifiers with feature selection in a sample size of 547 patients (the recurrence prevalence of 21.4%). The predictors were age at diagnosis, menarche and menopause, tumor size, number of involved and dissected axillary lymph nodes, grade and HER2. The best classifier (SVM) had sensitivity, specificity and accuracy of 96%, 91% and 94%, respectively. Despite the high performance of the algorithm proposed by the authors, cases with primary metastasis, a metastasis diagnosed at the time of registry, were not excluded from the analysis. In fact, having metastasis was one of the attributes primarily used by their proposed system. This necessary exclusion has been implemented in similar studies [28,30]. The reason for such an exclusion is that the survival rate of patients with stage IV (metastasis) breast cancer within 5 years is about 10% to 15%. Thus, most of patients with primary metastasis experience cancer recurrence.

The discriminative prognosis predictors vary regionally in different studies [27–29]. Also, such prognosis tools have been recently used to optimize the treatment protocol. An example is including the cancer recurrence risk to breast cancer treatment guidelines by the American Society of Clinical Oncology (ASCO) and the National Comprehensive Cancer Network (NCCN). Thus, there is a need to design prognosis tools in regions that are different based on cancer epidemiology. Meanwhile, it is necessary to validate the cancer prognosis system using extensive diagnosis validation criteria. Therefore, the aim of this study was to *reliably* and *accurately* predict breast cancer recurrence using pathological and demographics features of the patients. In the proposed CAD system, HPBCR (hybrid predictor of breast cancer recurrence), a hybrid technique including statistical features selection, meta-heuristic population-based optimization and ensemble learning were used to predict breast cancer recurrence in the first 5 years after the diagnosis.

The rest of the paper is organized as follows: in the next section, information about the experimental protocol and the pattern recognition methods used in this study is presented. Section 3 provides the results of HPBCR and its comparison with the state-of-the-art. The discussion is provided in Section 4 and finally, the conclusions are summarized in Section 5.

## 2. Materials and Methods

### 2.1. The Cohort Dataset

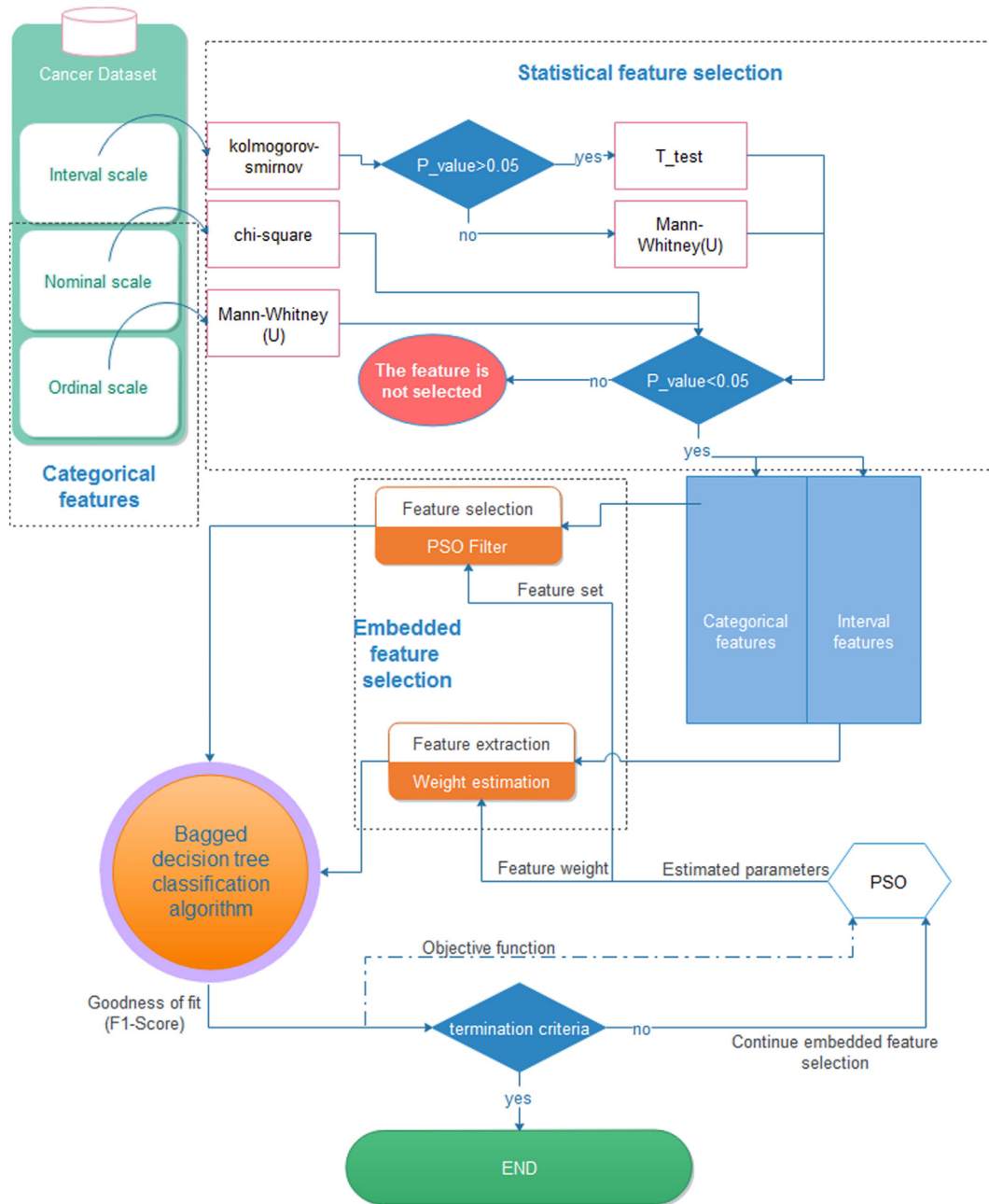
In this study we used a 16-year registry cohort database (1998–2014) on 1085 women who diagnosed with breast cancer in Isfahan Sayed-o-Shohada cancer research center. History of clinical conditions and therapy of patients was continued until death or lost to follow up. Characteristics of variables and tumor have been recorded by interview and reported as pathology results. Time to recurrence was based on the physicians' opinion. The following information was extracted from each patient: age at diagnosis of breast cancer, lymph node involvement ratio (NR) defined as ratio of involved to dissected lymph nodes [31], age of menarche, number of pregnancy (No. Preg), primary tumor size (TS), cellular marker for proliferation (Ki67), number of involved (I. Node) and dissected (T. Node) axillary lymph nodes, number of chemotherapies (No. Chemo) and categorical covariates of family history of cancer (FH: negative/positive), having more than one tumor in the breast (multifocal: negative/positive), estrogen receptor expression (ER: negative (also known as absent)/positive), progesterone receptor expression (PR: negative (also known as absent)/positive), tumor protein 53 (p53: negative/positive), type of surgery (MRM: modified radical mastectomy,

BCS: breast-conserving surgery, Mast: mastectomy), epidermal growth factor receptor-2 (Her2: negative/positive (also known as overexpressed)), Cathepsin-D protein status (Cathepsin: negative/positive), hormone therapy (HRT: negative/positive), radiotherapy (XRT: negative/positive) and the histological grade of tumor (1 to 4). Moreover, five main molecular subtypes of breast cancer were extracted as shown to affect recurrence prognosis in the literature. They were calculated using ER, PR, HER2 and Ki-67 (the cutoff value of 14% to define 'low' and 'high') status. These subtypes are: luminal A (ER and/or PR-positive/HER2 negative/low Ki-67), luminal B (ER- and/or PR-positive/HER2-negative/high Ki-67), HER2-positive luminal B (ER- and/or PR-positive/HER2 positive/any Ki-67), non-luminal HER2-positive

(ER and PR negative/HER2 positive) and triple negative (ER and PR negative/HER2-negative) [32]. The brief description of the above features used in our study was provided in the Supplementary material S2.

Eighty nine subjects were excluded from the analysis because of primary metastasis, a metastasis diagnosed at the time of registry [30]. A number of 579 breast cancer patients with complete information were involved in our analysis. All subjects gave informed consent to the experimental procedure. The experimental protocol was approved by the Isfahan University of Medical Sciences Panel on Medical Human Subjects and conformed to the Declaration of Helsinki.

The structure of HPBCR is depicted in Fig. 1. First, the statistical features' selection was used to identify significant discriminative



**Fig. 1.** The structure of the proposed prognosis system (HPBCR). Other classifiers such as SVM and MLP could be used instead of BDT. The pseudo-code of HPBCR is provided in the Supplementary material S3. The input features in HPBCR were: diagnosis age, nodal ratio, menarche age, the number of pregnancy, tumor size, Ki67, the number of involved and dissected nodes, as interval features, type of surgery, molecular subtypes, family history of cancer, multifocal tumor, estrogen and progesterone receptor status, p53 mutation, Her2 expression, Cathepsin-D protein status, using hormone therapy, and radiotherapy as nominal features and the tumor grade as an ordinal feature. Briefly, the input features are first selected using statistical feature selection. The selected features were used by Bagged Decision Tree to build the classifier. The optimal feature set and the weight of features are estimated using Particle Swarm Optimization during learning. The algorithm stops if no significant improvement is seen in the objective function or the maximum number of iterations (set to 100 in our study) is reached.

features. Among such features, categorical features (i.e. nominal or ordinal measurement scales [33]) used in the Bagged Decision Tree (BGD) were selected by Particle Swarm Optimization (PSO). The features with the interval measurement scale used by BGD were assigned weights being also tuned by PSO. The fitness function of the PSO was the F-score of the classifier on the training set (Eq. (1)). In the following, the details HPBCR are discussed. The pseudo-code of the HPBCR was presented in the Supplementary material S3.

$$fitness = FScore_{train} \quad (1)$$

## 2.2. Statistical Feature Selection

Feature selection (FS), the process of selecting a subset of relevant features, is divided into wrapper, embedded and filter methods. Filter method selects subsets of variables as a pre-processing step, independently from the chosen classifier while the other two methods, are used during machine learning procedure [34]. In wrapper methods, the selection of a features set is formulated as a search problem, in which different feature combinations are evaluated. Embedded methods, on the other hand, learn which set best contribute to the model accuracy while the model is being created. The former method searches the entire feature subsets while in the later one, search is guided by learning process which is prone to less over-fitting [35].

In fact, Filter FS also known as Statistical FS (SFS) applies a statistical measure to assign a score to each feature [36]. Such features are then ranked and selected based on the resulting score. Univariate SFS was first used in our analysis. Accordingly, the normality of the features with interval measurement scale was assessed using the Kolmogorov–Smirnov (KS) test [37]. Then, independent-sample t-test was used to identify statistically discriminative normally-distributed features. Meanwhile, Wilcoxon Mann–Whitney test was used to distinguish discriminative features with ordinal or non-normal interval measurement scales [38] (Fig. 1).

Chi-square test was used to check whether a nominal-measurement-scale feature was discriminative. The discriminative features were selected and further refined as the inputs of the classification system (BDT), next discussed.

## 2.3. Bagged Decision Tree

Decision tree builds classification models in the form of a tree structure. It divides a dataset into small subsets to incrementally develop an associated decision tree with decision and leaf nodes [39]. It uses entropy to calculate the homogeneity of samples to build the tree [40]. We used the statistical classifier C4.5 decision tree [41] with pruning in HPBCR. C4.5 sorts the data at every node of the tree to determine the best splitting attribute. When the tree is constructed, pruning is carried out from the leaves to the root. The redundant sub-trees are removed and thus the prediction error decreases. C4.5 handles both continuous and discrete attributes and is thus suitable for our application. Moreover, the online implementation of the resulting tree is simple.

Decision tree has been applied to many fields; for example in prediction of breast cancer relapse after surgery [42], classification mass spectrometry of blood serum samples from people with and without lung cancer [43] and gene expression data for cancer classification [44]. In fact, decision trees are reliable and effective decision making techniques used in different areas of medical decision making [45]. Meanwhile, in evidence-based medicine, it was used for clinical decision analysis [46].

Bagging is a method for improving results of machine learning classification algorithms. This method was formulated by Breiman and its name was deduced from the phrase “bootstrap aggregating”. The details of bagging could be found in [47,48]. Bagging is usually applied to decision tree methods, though it can be used with other classifiers. In summary bagged tree, resamples the training dataset several times

(bootstrapping), and builds a decision tree model from each; then aggregate these models together for a final classifier. This method was used for breast cancer classification [49], gene expression data analysis [50] and prostate cancer classification [51]. In our study, the number of trees was set to 10 for bagging.

The proper combination of categorical selected features and also the weight (importance) of the selected interval-measurement-scale features were identified using a meta-heuristics population-based stochastic optimization method entitled as Particle Swarm Optimization (PSO) (Fig. 1), discussed in the next chapter.

## 2.4. Particle Swarm Optimization (PSO)

PSO is an evolutionary computational method inspired by flocking birds [52], applied in many different areas, including manufacturing [53], civil engineering [54], optimum design [55], computational neuroscience [56] and breast cancer detection [57]. PSO is initialized with a population of random solutions known as particles, growing over generations to find optimal solutions. In a population, each particle has a velocity (i.e. rate of change in solutions) to enable them to fly through the problem space. Therefore, each particle is represented by a position and a velocity [58]. The current velocity of a particle depends on its previous velocity and on the distances of the particle from the personal (cognitive term) and neighborhood (social term) best positions (Eq. (2)).

$$v_i^{n+1} = wv_i^n + c_1r_1^n(p_i^n - x_i^n) + c_2r_2^n(p_g^n - x_i^n) \quad (2)$$

where  $w$  is the inertia weight;  $c_1$ ,  $c_2$  are the cognitive and social acceleration coefficients;  $i$  is the particle number,  $r_1$  and  $r_2$  are independent random numbers whose elements are uniformly distributed in  $[0, 1]$ ;  $n$  denotes the iteration number,  $p_i$  is the personal best of the  $i$ th particle and  $p_g$  is the neighborhood best.

Each particle records its best position (personal best) so far and the best position achieved in the group (neighborhood best) among all personal bests. Then, the new position of a particle is determined based on its previous position information and its current velocity (Eq. (3)).

$$x_i^{n+1} = x_i^n + v_i^{n+1} \quad (3)$$

The objective is to find the particle that best maximizes the target fitness function (Eq. (1)). Moreover, the neighborhood of each particle is the entire swarm in our method (star topology).

In HPBCR, the acceleration coefficients  $c_1$  and  $c_2$  were set to 2. A large value of inertia weight supports global search (“exploration”), while a small value favors local search (“exploitation”). A suitable strategy is to first motivate exploration, and then exploitation [56]. Thus, the inertia coefficient was set to 1.00 at the first iteration and was linearly decreased with the damping coefficient of 0.99 at each iteration, and the number of particles was set to 20 and 100 iterations were performed. Moreover, in order to increase the convergence possibility, velocity clamping ( $V_{max} = 4$ ) was used.

PSO algorithm was used to filter categorical features and to estimate the weights of interval features (Fig. 1). In categorical features, the position of the particles was rounded to 0 and 1 (Binary PSO) but for interval features, such values were real numbers between zero and one. The pseudo-code of the embedded feature selection based on PSO was provided in the Supplementary material S3 (section B).

## 2.5. State-of-the-art

In our study, other classification methods namely as MLP, SVM and DT were used for comparison. Such methods were proposed in similar studies in the literature [28,29]. The selected features using SFS were used as the input to these classifiers.

MLP is a feed-forward artificial neural network model that maps sets of input data onto a set of appropriate outputs. It consists of multiple fully connected layers of nodes in which learning happens by changing connection weights through back propagation. It is a modification of the standard linear perceptron and can be used for classification of data that are not linearly separable [59]. In this study the MLP with 1 hidden layer with 10 neurons and sigmoid activation function was used. Such an activation function was reported by Isa et al. to be suitable for the prediction of breast cancer [60]. Meanwhile, the sensitivity analysis was performed on the number of neurons. Cross-validation on the training set (70% of data) was performed. When increasing the number of neurons up to 10, the prediction accuracy increased and did not improve considerably afterwards. Thus, ten neurons were used as to avoid over-fitting problem.

SVM constructs a hyper plane or set of hyper planes in a high or infinite-dimensional space, which can be used for classification [61]. The SVM algorithm has been widely applied in the biological and many other fields [62–64]. In our study, the radial basis function (RBF) kernels were used. The soft-margin parameter and the radius of the RBF kernel should be set properly, because inappropriate parameter settings result in poor classification. The method proposed by Wu and Wang [65] was used to set the soft-margin and RBF kernel parameters.

2.6. Validation

2.6.1. The Performance Indices for Each Classifier

The performance of the classifiers was assessed using the holdout method, an approach to out-of-sample evaluation, where the dataset was split into two mutually exclusive sets (70% training and 30% test sets). The classifiers were then trained on the first set and tested on the other set [66]. Additionally, 4-fold cross-validation was used to assess the performance of the best classifier to overcome a possible biased error estimate [67]. A comprehensive list of performance indices [68–72] was selected and reported for each classifier to identify different statistical errors which is crucial in medical diagnosis especially in cancer studies. The performance measures of the classifiers are listed in Table 1, along with their definitions. Discriminant power (DP) was characterized as “poor” when  $DP < 1$ , “limited” when  $DP < 2$ , “fair” when  $DP < 3$  and “good” in other cases [72,73]. Among those measures AUC ROC is a global measure of diagnosis accuracy. ROC (receiver operating characteristic) curve is a plot of true positive rate (i.e. sensitivity) versus false positive rate (i.e. 1-specificity) of the diagnosis system as its discrimination threshold is varied. AUC (area under roc curve) is an effective and combined measure of the validity of the diagnosis test, which is also known as balanced accuracy defined the average of the sensitivity and specificity of the diagnosis system [72]. Its ranges for excellent, very good and good diagnosis accuracy are (0.9–1.0), (0.8–0.9) and (0.7–0.8) respectively [74].

Along with such indices, the agreement rate between the output of the classifier and the gold standard was assessed using Cohen’s kappa coefficient [75]. Kappas over 0.75 was characterized as excellent, 0.40 to 0.75 as fair to good, and below 0.40 as poor [76]. A diagnosis system was considered as reliable if (1) its sensitivity and specificity were at

least 80% (the minimum statistical power of 80%) and 95% (the maximum Type I error of 0.05), respectively [77,78], and (2) its false discovery rate ( $FDR = 1 - precision$ ) is as low as 5% [79]. Moreover, as a general rule, diagnosis tests with positive likelihood ratio ( $LR^+$ ) greater than 10 or negative likelihood ratio ( $LR^-$ ) less than 0.1 are preferred [80]. The application of sensitivity and specificity is highly dependent on the prevalence of the disease. A conservative method could be fulfilling both LR conditions (i.e. DOR greater than 100). An illustration of the validation criteria was presented in the Supplementary material S4.

2.6.2. Comparison Between Different Classifiers

McNemar’s test, also known as the Gillick test, was used to compare different classifiers to identify whether one classifier statistically significantly outperforms the other or not [67,81]. It was originally used to analyze retrospective case–control studies, where each case is matched to a particular control or it can be used to analyze experimental studies, where two treatments are given to match subjects [82].

To compare two classifiers A and B, z statistics is estimated as:

$$z = \frac{|x - y| - 1}{\sqrt{x + y}} \tag{4}$$

where x is the number of samples misclassified by A but not B and y is number of samples misclassified by B but not A. The null hypothesis (that the classifiers have the same error) can be rejected (with probability of incorrect rejection of 0.05) if  $|z| > 1.96$  [67,83]. In such a case, a classifier with higher DP value was reported as superior to the other one.

All data processing was performed off-line using Matlab version 8.6 (The MathWorks Inc., Natick, MA, USA). All statistical analysis and calculations were performed using the SPSS statistical package, version 18.0 (SPSS Inc., Chicago, IL, USA).

2.7. Online Web-based HPBCR Implementation

Ten extracted rules were obtained from BDT whose feature weights were optimized using PSO. The prognosis system was written in C# language and used in ASP.net5 which is an open-source framework from Microsoft Corporation. The prediction model is freely available in the website (<http://www.Prognosis.ir>). A snapshot of the developed online program is shown in the Supplementary material S5.

3. Results

The average age of the participants was  $46.8 \pm 10.0$  years. Among the number of 579 patients participated in our study, 19.3% had cancer recurrence during five years after diagnosis. The demographic and pathological characteristics of the participants, grouped by their classification with/without recurrence, are depicted in Tables 2, 3 and 4 for the features with interval, nominal and ordinal measurement scales, respectively. In fact for the features with binary, nominal with more than two categories, ordinal and interval measurement scales, their existence, category, rank and value is important in the diagnosis system. Among the interval measurement scale features, only age at diagnosis

**Table 1**  
The classification performance measures used in our study.

$Se = \frac{TP}{TP+FN}$	$Sp = \frac{TN}{TN+FP}$	$Acc = \frac{TP+TN}{TP+TN+FP+FN}$
$Pr = \frac{TP}{TP+FP}$	$Alpha = 1 - Sp$	$Beta = 1 - Se$
$F\text{-score} = \frac{2 \times Pr \times Se}{Pr + Se}$	$AUC = \frac{1}{2} \times (Sp + Se)$	$LR+ = \frac{Se}{1 - Sp}$
$LR- = \frac{1 - Se}{Sp}$	$DOR = \frac{LR+}{LR-}$	
$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}}$		
$DP = \left(\frac{\sqrt{3}}{\pi}\right) \times \log(DOR)$		

True positive (TP): subjects with cancer recurrence, correctly identified; false positive (FP): subjects without recurrence, incorrectly identified; true negative (TN): subjects without recurrence, correctly identified; false negative (FN): subjects with recurrence, incorrectly identified; Se: sensitivity = Power; Sp: specificity; Acc: Accuracy; Pr: precision; AUC: area under the receiver operating characteristic (ROC) curve; LR: likelihood ratio; DOR: diagnosis odds ratio; MCC: Matthews correlation coefficient; DP: discriminant power.

**Table 2**  
Comparison of clinical and biochemical features (with interval measurement scale) of included subjects with/without cancer recurrence in  $\mu \pm \sigma$  [min, max].

Variable	With recurrence	Without recurrence
Age* (year)	45.4 $\pm$ 10.1 (ND) [27.0, 68.0]	47.2 $\pm$ 9.9 [24, 75.0]
NR*	0.47 $\pm$ 0.36 [0.0, 1.0]	0.24 $\pm$ 0.32 [0.0, 1.0]
Menarche	13.3 $\pm$ 1.3 [11.0, 17.0]	13.4 $\pm$ 1.4 [10.0, 17.0]
No. Preg	3.5 $\pm$ 1.9 [0.0, 10.0]	3.7 $\pm$ 2.1 [0.0, 10.0]
TS* (cm)	4.0 $\pm$ 1.7 (ND) [1.0, 8.5]	3.7 $\pm$ 1.8 [0.5, 11.0]
Ki67*	21.1 $\pm$ 12.6 [6.0, 74.0]	18.0 $\pm$ 11.8 [1.0, 86.0]
I. Node*	6.6 $\pm$ 7.3 [0.0, 39.0]	3.0 $\pm$ 4.9 [0.0, 36.0]
T. Node*	12.8 $\pm$ 6.1 [1.0, 39.0]	11.4 $\pm$ 5.4 [1.0, 36.0]
No. Chemo	7.7 $\pm$ 0.9 [4.0, 10.0]	7.4 $\pm$ 1.3 [1.0, 9.0]

Age (age at diagnosis); NR (lymph node involvement ratio); menarche (menarche age); No. Preg (number of pregnancy); TS (tumor size); Ki67 (Ki67 proliferation marker); I. Node: number of involved auxiliary lymph nodes; T. Node: number of dissected auxiliary lymph nodes; No. Chemo: number of chemotherapy; ND: normally distributed. The sample size was 579 and the recurrence prevalence was 19.3%.

\* Statistically significant features ( $P < 0.05$ ).

and TS were normally distributed in the recurrence group. Thus, due to the rejection of class dependence normality, Mann–Whitney test was used for SFS in interval features. Selected discriminative features using

**Table 3**  
Comparison of clinical and biochemical features (with nominal measurement scale) of included subjects with/without cancer recurrence in percentage.

Variable	With recurrence	Without recurrence
FH	P 18 N: 82	P: 22 N: 78
Multifocal	33 67	27 73
ER	50 50	57 43
PR*	50 50	61 39
P53	42 58	34 66
Surgery*	MRM: 51 BCS: 14 Mast: 35	56 24 20
Her2	58 42	58 42
Cathepsin	93 7	95 5
HRT*	55 43	70 30
XRT	99 1	98 2
Subtypes*	LA: 17 LB: 16 HLB: 26 NLH: 31 3N: 10	14 14 26 32 14

FH (family history of cancer); multifocal (having more than one tumor in the breast); ER (estrogen receptor); PR (progesterone receptor); P53 (tumor protein 53); surgery: type of surgery (MRM: modified radical mastectomy, BCS: breast-conserving surgery, Mast: mastectomy); Her2 (epidermal growth factor receptor-2); Cathepsin (Cathepsin-D); HRT (hormone therapy); XRT (radiotherapy); subtypes: cancer molecular subtypes (LA: luminal A, LB: luminal B; HLB: HER2-positive luminal B, NLH: non-luminal Her2, 3N: triple negative). The characteristics were shown as positive and negative percentages (i.e. relative frequent table) for binary variables, respectively and the mode was underlined. The sample size was 579 and the recurrence prevalence was 19.3%.

\* Statistically significant features ( $P < 0.05$ ).

**Table 4**  
Comparison of the tumor "histological grade" feature (with ordinal measurement scales) of included subjects with/without cancer recurrence.

Grade categories	With recurrence	Without recurrence
1	22	10
2	<u>44</u>	<u>59</u>
3	33	29
4	1	2

The characteristics were shown as percentages (i.e. relative frequent table), and the mode (the most frequent item) was underlined in each category (recurrent or non-recurrent patient groups). Tumor grade was not statistically significant in subjects with/without recurrence. The sample size was 579 and the recurrence prevalence was 19.3%.

SFS were: age at diagnosis, NR, TS, I. Node, T. Node, Ki67, PR, HRT, surgery and cancer subtypes.

In our study, discretization was used on the following attribute intervals: age at diagnosis, NR and TS. In machine learning, discretization refers to the procedure of partitioning continuous attributes to discretized variables as to improve the classification performance [84,85]. We used a set of cut points taken from the literature, indicating the diagnosis properties of individual attributes as below to create categorical variables.

Age was categorized (Age.cat) as less than 40 years and more than 40 years [86]. NR was categorized (NR.cat) in  $\leq 0.25$  and  $> 0.25$  [87,88]. TS was also categorized (TS.cat) as  $\leq 2$  (T1), (2,5) (T2) and  $> 5$  (cm) (T3) [89–91].

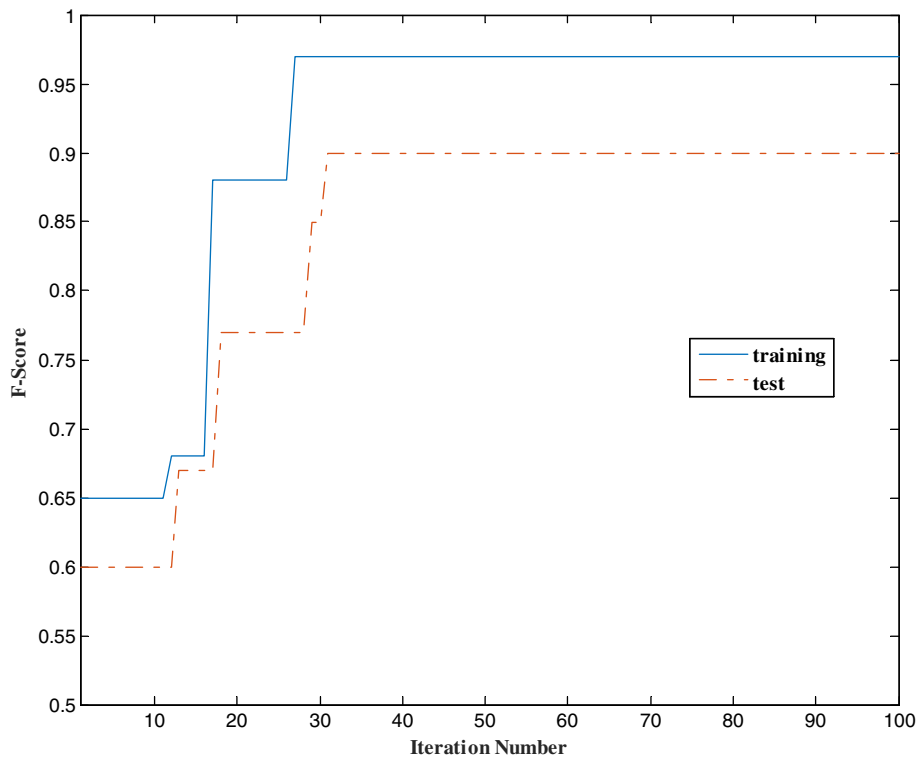
The value of the fitness function (F-score on the training set) and the test set performance (F-score on the test set) at some iteration were shown in Fig. 2. The following categorical features were kept after running PSO and BDT: PR, HRT, Surgery, age.cat, TS.cat and NR.cat. Among the features with interval measurement scale, Ki67 and T. Node were not selected by BDT. In fact BDT uses entropy principle to choose the best discriminative features [92]. The estimated weight of the selected interval-scale feature I. Node was 0.74.

The performance of the classifiers on the test set (when trained on the training set) using the hold-out validation method was shown in Table 5. The performance of the classifiers was further assessed using 4-fold cross validation (Table 6). The confusion matrix of HPBCR on the test set was shown in Table 7. The performance of HPBCR was statistically significantly higher than that of three other classifiers (McNemar's test;  $P < 0.05$ ). HPBCR showed limited discriminant power ( $DP = 1.28$ ), very good diagnosis accuracy ( $AUC = 0.90$ ), and excellent agreement with the gold standard ( $Kappa = 0.83$ ). The minimum statistical power and maximum Type I error ( $\alpha$ ) were 77% and 0.07, respectively in HPBCR (Table 6). Overall, the average statistical power and Type I error of HPBCR was 80% and 0.04, respectively (Table 6). The training time of HPBCR, SVM, DT and MLP was  $412.10 \pm 14.50$  (s),  $66.20 \pm 11.75$  (s),  $52.00 \pm 4.50$  (s) and  $132.45 \pm 15.50$  (s), while the validation time was  $0.62 \pm 0.31$  (s),  $0.49 \pm 0.24$  (s),  $0.39 \pm 0.33$  (s) and  $0.49 \pm 0.31$  (s), respectively. The average running time was the average of 10 runs over 405 and 174 subjects in the training and test sets respectively (hold-out 70%) on an Intel Core-i7 2 GHz CPU with 4 GB of RAM.

## 4. Discussion

### 4.1. The Risk Factors of the Cancer Recurrence

In our study, the following risk factors were selected by HPBCR to predict breast cancer recurrence: Age at diagnosis, TS, NR, I. Node, PR, HRT and type of surgery. Approximately 7% of women with breast cancer are diagnosed before the age of 40 years. It has been shown in the literature that younger age is not only an independent predictor of adverse outcome in survival rate [86], but it is also a risk factor in recurrence metastasis [89,93]. Bock et al. showed a hazard ratio of 2.8 for local recurrence in patients less than 35 years compared to those above



**Fig. 2.** The value of the fitness function (F-Score of the proposed classifier (HPBCR) on the training set-solid line) and the F-Score on the test set (dash-dot line) during optimization procedure. The termination criterion was only the maximum number of iterations (i.e. 100) in this plot.

50 years [94]. Bharat et al. estimated the risk of breast cancer recurrence for women diagnosed below the age of 40 to be 1.53 times higher than in those diagnosed above 40 years [95]. In our study, the first and second age quartile for patients without recurrence was 41 and 47 years compared with 38 and 43 years for those with recurrence. Meanwhile, patients who were less than 40 years old had 33.5% higher hazard of recurrent metastasis in comparison with more than 40 year old patients [89]. In fact, the odds of recurrence in the young (<40 years) and old (>40 years) population in our study was 0.43 and 0.19, respectively. Such a trend is in agreement with the literature. Meanwhile, in recurrent cancer patients, the odds ratio (OR) associated with age  $\leq 40$  compared with age  $> 40$  was 1.8, indicating that discretized age was a good feature for the developed prognosis system.

TS is an independent prognostic factor, with distant recurrence rates increasing with larger tumor size [96]. It has been in fact an important recurrence prognosis factor in several studies [97–100]. It has been shown in the literature that the median time to the development of metastatic disease shortens as tumor size increases [101–103]. In our study, the recurrence odds for  $TS \leq 2$  (T1),  $2 < TS \leq 5$  (T2), and  $TS > 5$  (T3) was 0.13, 0.27 and 0.29, respectively showing that increased tumor size is associated with higher probability of cancer recurrence.

NR was shown to be a prognosis factor for cancer recurrence in the literature [100]. Voordeckers et al. showed that NR is one of the main prognosis risk factors for cancer recurrence [104]. Elkhodary et al.

showed that the hazard ratio of breast cancer recurrence risk increases when NR increases [105]. Tazhibi et al. indicated that  $NR (\leq 0.25$  vs.  $> 0.25)$  is associated with hazard of distance recurrence [89]. In our study, the recurrence odds for the  $NR \leq 0.25$ , and  $NR > 0.25$  was 0.14, and 0.43, respectively showing that increased NR is associated with higher probability of cancer recurrence. Meanwhile, in recurrent cancer patients, the OR associated with  $NR > 0.25$  compared with  $NR \leq 0.25$  was 1.6, indicating that discretized NR was a good feature for the developed prognosis system.

It was shown in the literature that I. Node is an important predictive factor for recurrence risk [106,107]. T. Node was also considered a predictor for cancer recurrence [108]. Although SFS selected this feature, it was not used in BDT. One explanation could be that in our dataset, I. Node and T. Node were moderately correlated ( $r = 0.56$ ;  $P < 0.001$ ). Also, NR defined as the ratio between I. Node to T. Node, was also used by the classifier. Thus, adding T. Node could not increase the amount of information in the prognosis system. In fact, reducing the number of correlated features improves the classification accuracy [109]. Higher Ki67 expression, on the other hand, was shown to be a prognosis indicator of increased cancer recurrence in the literature [110]. In our study, the second quartile of Ki67 was 14.3 and 17.4 in non-recurrent and recurrent cancer patient, respectively. Although SFS selected this feature, it was not used in BDT either. In our study, Higher Ki67 index tumors showed larger size (TS), and more lymph node

**Table 5**

The holdout performance estimate (%) of the selected classifiers.

Method	Se %	Sp %	Acc %	Pr %	F-score %	Alpha	Beta	AUC	MCC	DOR	DP	Kappa
HPBCR	81	98	90	97	89	0.02	0.19	0.90	0.81	208.9	1.28	0.83
SVM	67	88	78	84	75	0.12	0.33	0.77	0.57	14.9	0.65	0.73
Decision tree	75	78	77	79	77	0.22	0.25	0.76	0.58	10.6	0.57	0.72
MLP	69	81	76	79	73	0.19	0.31	0.75	0.52	10.63	0.57	0.71

Se: sensitivity, Sp: specificity, Acc: accuracy, Pr: precision, AUC: area under the curve, MCC: Matthews correlation coefficient, DOR: diagnostics odds ratio; DP: discriminant power; SVM: supported vector machines; MLP: multilayer perceptron artificial neural network. Seventy percent of the data was used for training and the rest for validation.

**Table 6**  
The performance of the classifiers in percent based on 4-fold cross-validation (mean  $\pm$  SD) [min, max].

Method	Sensitivity	Specificity	Accuracy	Precision
Proposed method	80.0 $\pm$ 0.3 [77.0, 80.5]	96.1 $\pm$ 1.0 [93.0, 95.0]	89.2 $\pm$ 0.6 [85.0, 89.0]	96.5 $\pm$ 1.0 [95.0, 98.2]
SVM	73.0 $\pm$ 1.0 [66.0, 74.0]	85.3 $\pm$ 1.2 [80.0, 89.0]	77.6 $\pm$ 0.4 [76.1, 78.9]	82.1 $\pm$ 0.3 [80.0, 84.8]
Decision tree	74.0 $\pm$ 0.9 [71.8, 76.0]	77.0 $\pm$ 0.8 [75.1, 78.5]	77.1 $\pm$ 0.2 [76.6, 77.5]	78.1 $\pm$ 0.6 [77.3, 80.2]
MLP	67.5 $\pm$ 0.2 [66.2, 67.9]	84.3 $\pm$ 1.0 [80.8, 85.0]	76.0 $\pm$ .0.3 [75.0, 76.8]	84.2 $\pm$ 1.8 [79.9, 86.0]

SVM: supported vector machines; MLP: multilayer perceptron artificial neural network.

involvement (I. Node), which is in agreement with what was reported in the literature [111]. This could be the justification that the final prognosis system did not use Ki67 expression.

PR expression was shown to be an independent recurrence prognostic variable in breast cancer in the literature [112–114]. Purdie et al. showed that absent PR expression was significantly associated with poor cancer recurrence prognosis [112]. In our study, the recurrence odds for the PR + and PR – was 0.20, and 0.31, respectively which is in agreement with what was mentioned in the literature. Meanwhile, in non-recurrent cancer patients, such odds associated with PR + compared with PR – was 1.6, indicating that PR was a distinctive feature for the developed prognosis system.

It was shown that HRT reduces the risk of breast cancer recurrence in the literature [115]. In our study, the recurrence odds for the HRT + and HRT – was 0.20, and 0.35, respectively which is in agreement with what was mentioned in the literature. Meanwhile, in non-recurrent cancer patients, the recurrence odds associated with HRT + compared with HRT – was 2.4, indicating that HRT was a suitable predictor for the developed prognosis system. It was proposed in the literature that the impact of type of surgery on the breast cancer recurrence might be dependent on age [116]. In our study, the non-recurrence odds for young and old patients underwent MRM surgery was 2.8 and 5.4, respectively. Such odds for the BCS surgery was 4.8 and 8.3. This could in principal support that the performance of the surgery is age dependent.

#### 4.2. The Properties and Performance of HPBCR

In our study, among 1085 breast cancer patients, only 579 subjects with complete information and without primary metastasis were included in the prediction. Thus, subjects with missing data were excluded from the analysis. Missing data can introduce bias and reduce efficiency [117]. It is possible to use multiple imputation to handle such incomplete data. Although multiple imputation has potential to improve the validity of medical research, it requires the user to model the missing variable distribution. The validity of the imputed results highly depends on such modeling and underlying assumptions [118].

**Table 7**  
The overall confusion matrix of HPBCR on the test set.

Total population		Condition (as determined by "gold standard")	
		Condition positive	Condition negative
Test outcome	Test outcome positive	True positive 51	False positive (Type I error) 2
	Test outcome negative	False negative (Type II error) 11	True negative 110

Gold standard: recurrent cancer; test outcome: the decision of the proposed diagnosis system (HPBCR). The classifier was trained on the training set (70% of the samples) and its performance was shown on the test set.

In our algorithm, SFS was first used to identify the discriminative features. Such a selection was then refined by PSO and also the importance of the interval features, represented by the estimated weights, was identified. Overall, the embedded FS was used in our study where FS is performed as part of the learning procedure. PSO is a meta-heuristics population-based stochastic optimization method. In principle, it is possible to use any other method in this category that could deal with continuous (feature weights) and discrete (feature selection) problems. We also tested genetic algorithm (GA) [119,120]. In multiple runs, both PSO and GA converged to an identical optimum solution. However, GA required about 150 iterations in comparison with 50 iterations in PSO (Fig. 2), for converging.

An alternative FS method is using multivariate FS namely as multiple logistic regression (MLR), taking into account the relationship between different features. MLR, known as feature vector machine in machine learning, can be used to select statistically significant features [84,121]. MLR was also tested in our study. However, due to the lack of fit of the MLR model, univariate SFS was used.

In HPBCR, F-score was used as the fitness function. When the dataset is imbalanced, for example in our case where the recurrence prevalence was 19.3, overall accuracy is not a suitable fitness measure. In fact, choosing the wrong objective function while developing a classification model can introduce bias towards majority [122]. Meanwhile HPBCR could be used for other mixed data where each instance in a data set is described by more than one type of measurement scale attribute [33]. In this case, different attributes must be treated differently. In our algorithm, different attributes were manipulated distinctly.

The average sensitivity and specificity of HPBCR are estimated as 80% and 96% (Table 6). Also, considering the maximum Types I ( $\alpha = 0.07$ ) and II errors ( $\beta = 0.23$ ), maximum false discovery rate (FDR = 5%), guarding against testing hypotheses suggested by the data (Type III errors [123]) done by cross-validation, very good diagnosis accuracy (AUC = 0.90), and finally the "excellent agreement" between the gold standard and the outcome of the system (Kappa = 0.83), HPBCR is promising for clinical prognosis tests. HPBCR significantly outperformed the other systems namely as MLP, DT and SVM (McNemar's test;  $P < 0.05$ ). The other three tested classifiers showed poor DP, good diagnosis accuracy and good agreement with the gold standard (Table 5). The Types I and II errors of the proposed algorithm were better than those obtained by the other classifiers (Tables 5 and 6). In HPBCR the average Type I error was less than 0.05 while the average statistical power was higher than 80% in the hold-out (Table 5) and cross validation (Table 6). However, the maximum Type I error (0.07) and minimum power (77%) in the folds used for cross validation were beyond the acceptable clinical ranges and must be improved. The variation of the performance indices in different folds was low (Table 6), showing consistency in the results of HPBCR.

The performance of HPBCR was further compared with the state-of-the-art. Kim et al. [28] used hold-out validation method (70%: training set; 30% test set) and reported the performance of the algorithm on the test set as the AUC of 0.85, sensitivity of 89%, specificity of 73%, precision of 75% (FDR = 25%) and accuracy of 85%. In a similar setting (Table 5), HPBCR had AUC of 0.90, sensitivity of 81%, specificity of 98%,



**Table 8**

The holdout performance estimate (%) of the prognosis system with modifications.

Scenario	Se %	Sp %	Acc %	Pr %	F-score %	Alpha	Beta	AUC	MCC	DOR	DP	Kappa
Baseline	81	98	90	97	89	0.02	0.19	0.90	0.81	208.9	1.28	0.83
1	78	81	78	93	86	0.18	0.21	0.79	0.51	15.1	0.65	0.78
2	81	98	90	96	89	0.02	0.19	0.90	0.81	208.9	1.28	0.83
3	80	96	88	97	87	0.04	0.2	0.88	0.79	116.9	1.14	0.81

Baseline: original developed HPBCR; Scenario 1: excluding PSO from the algorithm, 2: using SVM classifier instead of BDT, 3: using MLP classifier instead of BDT, Se: sensitivity, Sp: specificity, Acc: accuracy, Pr: precision, AUC: area under the curve, MCC: Matthews correlation coefficient, DOR: diagnostics odds ratio; DP: discriminant power; SVM: supported vector machines; MLP: multilayer perceptron artificial neural network. Seventy percent of the data was used for training and the rest for validation.

precision of 97% (FDR = 3%) and accuracy of 90%. Our algorithm outperformed the other prediction method in terms of AUC, FDR and accuracy.

NPI method proposed by Galea et al. [27] was implemented and tested in our dataset. The value of the performance measures AUC, sensitivity, specificity, precision, and accuracy were 0.51, 100%, 2%, 25% and 26%, respectively. This method is not suitable in our study because of high Type I error ( $\alpha = 0.98$ ).

#### 4.3. Additional Analysis

The contribution of BDT and PSO to the overall classification performance was also considered. PSO was excluded from the algorithm flow-chart and the resulting hold-out performance was reported in Table 8 (Scenario 1), compared with what obtained in the HPBCR (baseline). HPBCR significantly outperformed the modified prognosis system (McNemar's test;  $P < 0.05$ ). Particularly, discriminant power is deteriorated when excluding PSO in the developed algorithm. Moreover, the modified system is not clinically reliable (Power = 78%, Type I error = 0.09). PSO is in fact used for embedded feature selection and is a necessary section of the program.

Moreover, SVM (Scenario 2) and MLP (Scenario 3) classifiers were used instead of BDT whose features were tuned using PSO. The performance of these two modified prognosis systems was comparable to what obtained in HPBCR (Table 8; McNemar's test;  $P > 0.05$ ). Thus, BDT could be substituted by SVM and MLP. However, it was shown in the literature that BDT is as good as SVM and MLP [124]. Meanwhile, bagging avoids over-fitting. Using BDT, it was possible to implement HPBCR online efficiently (Section 2.7).

However, the performance of SVM or MLP when PSO is used for embedded feature selection (Table 8) was significantly better than that of SVM or MLP without using PSO (Table 5). This is the reason that hybrid feature selection performed by PSO is essential in the proposed prognosis system. Also, the training time of SVM + PSO, and MLP + PSO was  $415.77 \pm 10.50$  (s) and  $710.02 \pm 22.55$  (s), while the validation time was  $0.50 \pm 0.27$ , and  $0.50 \pm 0.33$ , respectively (hold-out 70%). Since at each iteration, feature sets are calculated for each 20 particle, the training time is more than cases where PSO is not used.

#### 4.4. Final Considerations

This study was a validation of the prognostic performance of HPBCR. The limitation of the current study is that it was a retrospective, single-institution study. The results might have been biased on the basis of available specimen and patterns of referral to the hospital. The sample size must be increased as to improve the statistical power in our prognosis system [125]. The output of the classification system is not fuzzy. It will be useful to add risk assessment to the algorithm and report the risk of recurrence which is the focus of our future activity.

### 5. Conclusions

We proposed a computer-aided prognosis system to identify breast cancer recurrence within 5 years after diagnosis. This method is

accurate and precise and could be possibly used by clinicians to make proper treatment plan.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.csbj.2016.11.004>.

### Acknowledgements

This work was supported by the People Programme (Marie Curie Actions) of the European Union Seventh Framework Programme (FP7/2007–2013) under REA grant agreement no. 600388 (TECNIOspring Programme), from the Agency for Business Competitiveness of the Government of Catalonia, ACCIÓ and from Spanish Ministry of Economy and Competitiveness- Spain (project DPI2014-59049-R). This work was supported in part by the University of Isfahan and Isfahan University of Medical Sciences.

### References

- Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph* 2007;31:198–211.
- Castellino RA. Computer aided detection (CAD): an overview. *Cancer Imaging* 2005;5:17–9. <http://dx.doi.org/10.1102/1470-7330.2005.0018>.
- Rangayyan RM, Ayres FJ, Desautels JL. A review of computer-aided diagnosis of breast cancer: toward the detection of subtle signs. *J Franklin Inst* 2007;344:312–48.
- Kim JH. Computer-aided diagnosis for lung cancer. *J Lung Cancer* 2004;3:67–70.
- Oweis R, Hijazi L. A computer-aided ECG diagnostic tool. *Comput Methods Programs Biomed* 2006;81:279–84.
- Katsumata A, Fujita H. Progress of computer-aided detection/diagnosis (CAD) in dentistry CAD in dentistry. *Jpn Dent Sci Rev* 2014;50:63–8.
- Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J* 2015;13:8–17.
- Hadjiiski L, Sahiner B, Chan HP. Advances in computer-aided diagnosis for breast cancer. *Curr Opin Obstet Gynecol* 2006;18:64–70. <http://dx.doi.org/10.1097/01.gco.0000192965.29449.da>.
- Anand P, Kunnumakara AB, Sundaram C, Harikumar KB, Tharakan ST, Lai OS, et al. Cancer is a preventable disease that requires major lifestyle changes. *Pharm Res* 2008;25:2097–116. <http://dx.doi.org/10.1007/s11095-008-9661-9>.
- Kuper HBP, Adami HO. Tobacco use and cancer causation: association by tumour type. *J Intern Med* 2002;252:206–24.
- Wild BWSaCP. World cancer report 2014. World Health Organization; 2014.
- Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *Cancer J Clin* 2014;64:9–29.
- Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* 2015;65:87–108. <http://dx.doi.org/10.3322/caac.21262>.
- Thomsen AKJ. Chemoprevention of breast cancer. *Am J Health Syst Pharm* 2008.
- Mousavi SM, Guoya MM, Ramazani R, Davanlou M, Hajsadeghi N, Seddighi Z. Cancer incidence and mortality in Iran. *Oxf J Med Health Ann Oncol* 2009;20:556–63.
- Jazayeri SB, Saadat S, Ramezani R, Kaviani A. Incidence of primary breast cancer in Iran: ten-year national cancer registry data report. *Cancer Epidemiol* 2015;39:519–27.
- Maracy MR, Moradpour F, Hosseini SM, Tirani M. Cancer incidence and prevalence in Isfahan: application of mortality data to estimates and projects for the period 2001–2015. *Int J Prev Med* 2012;3:867–74.
- Mokarian F, Ramezani MA, Heydari K, Tabatabaiean M, Tavazohi H. Epidemiology and trend of cancer in Isfahan 2005–2010. *J Res Med Sci* 2011;16:1228–33.
- Dunnwald LK, Rossing MA, Li CI. Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast Cancer Res* 2007;9:R6. <http://dx.doi.org/10.1186/bcr1639>.
- de Martel CFJ, Franceschi S, et al. Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *Lancet Oncol* 2012;13:607–15.
- Lin R, Tripuraneni P. Radiation therapy in early-stage invasive breast cancer. *Indian J Surg Oncol* 2011;2:101–11. <http://dx.doi.org/10.1007/s13193-011-0048-8>.

- [22] Saphner T, Tormey DC, Gray R. Annual hazard rates of recurrence for breast cancer after primary therapy. *J Clin Oncol* 1996;14:2738–46.
- [23] Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011;61:69–90. <http://dx.doi.org/10.3322/caac.20107>.
- [24] Gonzalez-Angulo AMLJ, Broglio KR, Meric-Bernstam F, et al. High risk of recurrence for patients with breast cancer who have human epidermal growth factor receptor 2-positive, node-negative tumors 1 cm or smaller. *J Clin Oncol* 2009.
- [25] Weeks JC, Cook EF, O'Day SJ, Peterson LM, Wenger N, Reding D, et al. Relationship between cancer patients' predictions of prognosis and their treatment preferences. *JAMA* 1998;279:1709–14.
- [26] Zeng TLJ. Mixture classification model based on clinical markers for breast cancer prognosis. *Artif Intell Med* 2010;48:129–37.
- [27] Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham prognostic index in primary breast cancer. *Breast Cancer Res Treat* 1992;22:207–19.
- [28] Kim W, Kim KS, Lee JE, Noh D-Y, Kim S-W, Jung YS, et al. Development of novel breast cancer recurrence prediction model using support vector machine. *J Breast Cancer* 2012;15:230–8.
- [29] Ahmad I, Eshlaghy A, Poorebrahimi A, Ebrahimi M, Razavi A. Using three machine learning techniques for predicting breast cancer recurrence. *J Health Med Inform* 2013;4:124.
- [30] Jafari-Koshki T, Mansourian M, Mokarian F. Exploring factors related to metastasis free survival in breast cancer patients using Bayesian cure models. *Asian Pac J Cancer Prev* 2014;15:9673–8.
- [31] Wishart GC, Bajdik CD, Dicks E, Provenzano E, Schmidt MK, Sherman M, et al. PREDICT Plus: development and validation of a prognostic model for early breast cancer that includes HER2. *Br J Cancer* 2012;107:800–7. <http://dx.doi.org/10.1038/bjc.2012.338>.
- [32] Shim HJ, Kim SH, Kang BJ, Choi BG, Kim HS, Cha ES, et al. Breast cancer recurrence according to molecular subtype. *Asian Pac J Cancer Prev* 2014;15:5539–44.
- [33] Marateb HR, Mansourian M, Adibi P, Farina D. Manipulating measurement scales in medical statistical analysis and data mining: a review of methodologies. *J Res Med Sci* 2014;19:47–56.
- [34] Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res* 2003;3:1157–82.
- [35] Saeyns Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics* 2007;23:2507–17. <http://dx.doi.org/10.1093/bioinformatics/btm344>.
- [36] Nilsson R. Statistical feature selection: with applications in life science. SE58183 Linköping, Sweden: Linköping University, The Institute of Technology; 2007, 191 <http://www.diva-portal.org/smash/get/diva2:17577/FULLTEXT01.pdf>.
- [37] Destercke S, Strauss O. Kolmogorov–Smirnov test for interval data. Springer; 2014 416–25.
- [38] Hilton JF. The appropriateness of the Wilcoxon test in ordinal data. *Stat Med* 1996; 15:631–45.
- [39] Quinlan JR. Induction of decision trees. *Mach Learn* 1986;1:81–106.
- [40] Quinlan JR. C4.5: programs for machine learning. Elsevier; 2014.
- [41] Salzberg SL. C4.5: programs for machine learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. *Mach Learn* 1994;16:235–40.
- [42] Jerez-Aragón JM, Gómez-Ruiz JA, Ramos-Jiménez G, Muñoz-Pérez J, Alba-Conejo E. A combined neural network and decision trees model for prognosis of breast cancer relapse. *Artif Intell Med* 2003;27:45–63.
- [43] Markey MK, Tourassi GD, Floyd CE. Decision tree classification of proteins identified by mass spectrometry of blood serum samples from people with and without lung cancer. *Proteomics* 2003;3:1678–9.
- [44] Tan AC, Gilbert D. Ensemble machine learning on gene expression data for cancer classification; 2003.
- [45] Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision trees: an overview and their use in medicine. *J Med Syst* 2002;26:445–63.
- [46] Bae J-M. The clinical decision analysis using decision tree. *Epidemiol Health* 2014; 36.
- [47] Breiman L. Bagging predictors. *Mach Learn* 1996;24:123–40.
- [48] Machová K, Barcak F, Bednár P. A bagging method using decision trees in the role of base classifiers. *Acta Polytech Hung* 2006;3:121–32.
- [49] Lavanya D, Rani KU. Ensemble decision tree classifier for breast cancer data. *Int J Inf Technol Converg Serv* 2012;2:17.
- [50] Valentini G, Muselli M, Ruffino F. Bagged ensembles of support vector machines for gene expression data analysis. *IEEE*; 2003 1844–9.
- [51] Qu Y, Adam B-L, Yasui Y, Ward MD, Cazares LH, Schellhammer PF, et al. Boosted decision tree analysis of surface-enhanced laser desorption/ionization mass spectral serum profiles discriminates prostate cancer from noncancer patients. *Clin Chem* 2002;48:1835–43.
- [52] Eberhart RC, Kennedy J. A new optimizer using particle swarm theory. Proceedings of the 6th IEEE International Symposium on Micro Machine and Human Science (MHS '95); October 1995. p. 39–43. <http://dx.doi.org/10.1109/MHS.1995.494215> Nagoya, Japan.
- [53] Sha D, Lin H-H. A multi-objective PSO for job-shop scheduling problems. *Expert Syst Appl* 2010;37:1065–70.
- [54] Gopalakrishnan K. Particle swarm optimization in civil infrastructure systems: state-of-the-art review. Metaheuristic applications in structures and infrastructures. London (UK): Elsevier; 2013. p. 49–76.
- [55] Durán O, Pérez L, Batocchio A. Optimization of modular structures using particle swarm optimization. *Expert Syst Appl* 2012;39:3507–15.
- [56] Marateb HR, McGill KC. Resolving superimposed MUAPs using particle swarm optimization. *IEEE Trans Biomed Eng* 2009;56:916.
- [57] Modiri A, Kiasaleh K. Permittivity estimation for breast cancer detection using particle swarm optimization algorithm. *IEEE*; 2011 1359–62.
- [58] Gopalakrishnan K, Ceylan H, Attoh-Okine NO. Intelligent and soft computing in infrastructure systems engineering: recent advances. Springer; 2009.
- [59] Wasserman PD, Schwartz T. Neural networks. II. What are they and why is everybody so interested in them now? *IEEE Expert* 1988;3:10–5.
- [60] Isa I, Saad Z, Omar S, Osman M, Ahmad K, Sakim HM. Suitable MLP network activation functions for breast cancer and thyroid disease detection. *IEEE*; 2010 39–44.
- [61] Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;20:273–97.
- [62] Akay MF. Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Syst Appl* 2009;36:3240–7.
- [63] Widodo A, Yang B-S. Support vector machine in machine condition monitoring and fault diagnosis. *Mech Syst Signal Process* 2007;21:2560–74.
- [64] George G, Raj, VC. Review on feature selection techniques and the impact of SVM for cancer classification using gene expression profile. (arXiv preprint arXiv:1109.1062 2011).
- [65] Wu K-P, Wang S-D. Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space. *Pattern Recognit* 2009;42:710–7.
- [66] Sammut C, Webb GI. Encyclopedia of machine learning. Springer Science & Business Media; 2011.
- [67] Webb AR. Statistical pattern recognition. John Wiley & Sons; 2011.
- [68] Metz CE. Basic principles of ROC analysis. Elsevier; 1978 283–98.
- [69] Lalkhen AG, Mccluskey A. Clinical tests: sensitivity and specificity. *Contin Educ Anaesth Crit Care Pain* 2008;8:221–3.
- [70] Huang J, Ling CX. Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans Knowl Data Eng* 2005;17:299–310.
- [71] Glas AS, Lijmer JG, Prins MH, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003;56:1129–35.
- [72] Sokolova M, Japkowicz N, Szpakowicz S. Performance accuracy, f-score and ROC: a family of discriminant measures for performance evaluation. Hobart, Australia: Springer-Verlag; 2006 1015–21.
- [73] Mert A, Kilic N, Bilgili E, Akan A. Breast cancer detection with reduced feature set. *Comput Math Methods Med* 2015;2015, 265138. <http://dx.doi.org/10.1155/2015/265138>.
- [74] Simundic A. Measures of diagnostic accuracy: basic definitions. *eJIFCC* 2008; 19 (4). 2013.
- [75] Smeeton NC. Early history of the kappa statistic. *JSTOR* 1985.
- [76] Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions. John Wiley & Sons; 2013.
- [77] Banerjee A, Chitnis UB, Jadhav SL, Bhawalkar JS, Chaudhury S. Hypothesis testing, type I and type II errors. *Ind Psychiatry J* 2009;18:127–31. <http://dx.doi.org/10.4103/0972-6748.62274>.
- [78] Ellis PD. The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results. Cambridge; New York: Cambridge University Press; 2010.
- [79] Colquhoun D. An investigation of the false discovery rate and the misinterpretation of p-values. *Open Sci* 2014;1:140216.
- [80] Guilherme HMO, Nesbitt GC, Murphy JG, Habermann TM. Mayo Clinic medical manual and Mayo Clinic internal medicine review. 7th ed. MN, USA: CRC Press Rochester; 2007.
- [81] Dietterich TG. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Comput* 1998;10:1895–923.
- [82] Eliasziw M, Donner A. Application of the McNemar test to non-independent matched pair data. *Stat Med* 1991;10:1981–91.
- [83] Demšar J. Statistical comparisons of classifiers over multiple data sets. *J Mach Learn Res* 2006;7:1–30.
- [84] Marateb HR, Goudarzi S. A noninvasive method for coronary artery diseases diagnosis using a clinically-interpretable fuzzy rule-based system. *J Res Med Sci* 2015;20:214–23.
- [85] Huan L, Setiono R. Feature selection via discretization. *IEEE Trans Knowl Data Eng* 1997;9:642–5. <http://dx.doi.org/10.1109/69.617056>.
- [86] Anders CK, Johnson R, Litton J, Phillips M, Bleyer A. Breast cancer before age 40 years. *Semin Oncol* 2009;36:237–49. <http://dx.doi.org/10.1053/j.seminoncol.2009.03.001>.
- [87] Keam B, Im SA, Kim HJ, Oh DY, Kim JH, Lee SH, et al. Clinical significance of axillary nodal ratio in stage II/III breast cancer treated with neoadjuvant chemotherapy. *Breast Cancer Res Treat* 2009;116:153–60. <http://dx.doi.org/10.1007/s10549-008-0160-9>.
- [88] Truong PT, Berthelet E, Lee J, Kader HA, Olivetto IA. The prognostic significance of the percentage of positive/dissected axillary lymph nodes in breast cancer recurrence and survival in patients with one to three positive axillary lymph nodes. *Cancer* 2005;103:2006–14. <http://dx.doi.org/10.1002/cncr.20969>.
- [89] Tazhibi M, Fayaz M, Mokarian F. Detection of prognostic factors in metastatic breast cancer. *J Res Med Sci* 2013;18:283–90.
- [90] Narod SA. Tumour size predicts long-term survival among women with lymph node-positive breast cancer. *Curr Oncol* 2012;19:5. <http://dx.doi.org/10.3747/co.19.1043>.
- [91] Saadatmand S, Bretveld R, Siesling S, Tilanus-Linthorst MMA. Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173,797 patients. *BMJ Br Med J* 2015;351.
- [92] Zadrozny B, Elkan C. Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers. *Citeseer*; 2001 609–16.
- [93] Assi HA, Khoury KE, Dbouk H, Khalil LE, Mouhieddine TH, El Saghir NS. Epidemiology and prognosis of breast cancer in young women. *J Thorac Dis* 2013;5(Suppl. 1):S2–8. <http://dx.doi.org/10.3978/j.issn.2072-1439.2013.05.24>.
- [94] de Bock GH, van der Hage JA, Putter H, Bonnema J, Bartelink H, van de Velde CJ. Isolated loco-regional recurrence of breast cancer is more common in young

- patients and following breast conserving therapy: long-term results of European Organisation for Research and Treatment of Cancer studies. *Eur J Cancer* 2006; 42:351–6. <http://dx.doi.org/10.1016/j.ejca.2005.10.006>.
- [95] Bharat A, Aft RL, Gao F, Margenthaler JA. Patient and tumor characteristics associated with increased mortality in young women (< or =40 years) with breast cancer. *J Surg Oncol* 2009;100:248–51. <http://dx.doi.org/10.1002/jso.21268>.
- [96] Cianfrocca M, Goldstein LJ. Prognostic and predictive factors in early-stage breast cancer. *Oncologist* 2004;9:606–16. <http://dx.doi.org/10.1634/theoncologist.9-6-606>.
- [97] Voogd AC, Nielsen M, Peterse JL, Blichert-Toft M, Bartelink H, Overgaard M, et al. Differences in risk factors for local and distant recurrence after breast-conserving therapy or mastectomy for stage I and II breast cancer: pooled results of two large European randomized trials. *J Clin Oncol Off J Am Soc Clin Oncol* 2001;19:1688–97.
- [98] Arriagada R, Rutqvist LE, Johansson H, Kramar A, Rotstein S. Predicting distant dissemination in patients with early breast cancer. *Acta Oncol* 2008;47:1113–21. <http://dx.doi.org/10.1080/02841860701829661>.
- [99] Silvestrini R, Daidone MG, Di Fronzo G, Morabito A, Valagussa P, Bonadonna G. Prognostic implication of labeling index versus estrogen receptors and tumor size in node-negative breast cancer. *Breast Cancer Res Treat* 1986;7:161–9.
- [100] O'Reilly SM, Camplejohn RS, Barnes DM, Millis RR, Rubens RD, Richards MA. Node-negative breast cancer: prognostic subgroups defined by tumor size and flow cytometry. *J Clin Oncol Off J Am Soc Clin Oncol* 1990;8:2040–6.
- [101] Carter CL, Allen C, Henson DE. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer* 1989;63:181–7.
- [102] Rosen PP, Groshen S, Kinne DW, Norton L. Factors influencing prognosis in node-negative breast carcinoma: analysis of 767 T1N0M0/T2N0M0 patients with long-term follow-up. *J Clin Oncol Off J Am Soc Clin Oncol* 1993;11:2090–100.
- [103] Fisher B, Slack NH, Bross ID. Cancer of the breast: size of neoplasm and prognosis. *Cancer* 1969;24:1071–80.
- [104] Voordeckers M, Vinh-Hung V, Van de Steene J, Lamote J, Storme G. The lymph node ratio as prognostic factor in node-positive breast cancer. *Radiother Oncol* 2004;70:225–30. <http://dx.doi.org/10.1016/j.radonc.2003.10.015>.
- [105] Elkhodary TR, Ebrahim MA, Hatata EE, Niazay NA. Prognostic value of lymph node ratio in node-positive breast cancer in Egyptian patients. *J Egypt Natl Canc Inst* 2014;26:31–5. <http://dx.doi.org/10.1016/j.jnci.2013.10.001>.
- [106] Martin C, Cutuli B, Velten M. Predictive model of axillary lymph node involvement in women with small invasive breast carcinoma: axillary metastases in breast carcinoma. *Cancer* 2002;94:314–22. <http://dx.doi.org/10.1002/cncr.10229>.
- [107] Lin YC, Chen SC, Chang HK, Hsueh S, Tsai CS, Lo YF, et al. Identifying good prognosis group of breast cancer patients with 1–3 positive axillary nodes for adjuvant cyclophosphamide, methotrexate and 5-fluorouracil (CMF) chemotherapy. *Jpn J Clin Oncol* 2005;35:514–9. <http://dx.doi.org/10.1093/jjco/hyi143>.
- [108] Weir L, Speers C, D'Yachkova Y, Olivetto IA. Prognostic significance of the number of axillary lymph nodes removed in patients with node-negative breast cancer. *J Clin Oncol Off J Am Soc Clin Oncol* 2002;20:1793–9.
- [109] Li S, Oh S. Improving feature selection performance using pairwise pre-evaluation. *BMC Bioinformatics* 2016;17:1–13. <http://dx.doi.org/10.1186/s12859-016-1178-3>.
- [110] Yang XQ, Wang FB, Chen C, Peng CW, Zhang JF, Li Y. High Ki-67 expression is a poor prognostic indicator of 5-year recurrence free survival in patients with invasive breast cancer. *Asian Pac J Cancer Prev* 2011;12:3101–5.
- [111] Shokouh TZ, Ezatollah A, Barand P. Interrelationships between Ki67, HER2/neu, p53, ER, and PR status and their associations with tumor grade and lymph node involvement in breast carcinoma subtypes: retrospective-observational analytical study. *Medicine* 2015;94, e1359. <http://dx.doi.org/10.1097/MD.0000000000001359>.
- [112] Purdie CA, Quinlan P, Jordan LB, Ashfield A, Ogston S, Dewar JA, et al. Progesterone receptor expression is an independent prognostic variable in early breast cancer: a population-based study. *Br J Cancer* 2014;110:565–72. <http://dx.doi.org/10.1038/bjc.2013.756>.
- [113] Fernö M, Baldetorp B, Bendahl P-O, Borg Å, Ewers S-B, Olsson H, et al. Recurrence-free survival in breast cancer improved by adjuvant tamoxifen — especially for progesterone receptor positive tumors with a high proliferation. *Breast Cancer Res Treat* 1995;36:23–34. <http://dx.doi.org/10.1007/bf00690181>.
- [114] Fuqua SA, Cui Y, Lee AV, Osborne CK, Horwitz KB. Insights into the role of progesterone receptors in breast cancer. *J Clin Oncol Off J Am Soc Clin Oncol* 2005;23:365:1687–1687–717. <http://dx.doi.org/10.1200/JCO.2005.05.152> [author reply 2–3].
- [115] Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005; 365:1687–717. [http://dx.doi.org/10.1016/s0140-6736\(05\)66544-0](http://dx.doi.org/10.1016/s0140-6736(05)66544-0).
- [116] Cao JQ, Olson RA, Tyldesley SK. Comparison of recurrence and survival rates after breast-conserving therapy and mastectomy in young women with breast cancer. *Curr Oncol* 2013;20:e593–601. <http://dx.doi.org/10.3747/co.20.1543>.
- [117] Barnard J, Meng XL. Applications of multiple imputation in medical studies: from AIDS to NHANES. *Stat Methods Med Res* 1999;8:17–36.
- [118] Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 2009;338. <http://dx.doi.org/10.1136/bmj.b2393>.
- [119] Eberhart RC, Shi Y. Comparison between genetic algorithms and particle swarm optimization. In: Porto VW, Saravanan N, Waagen D, Eiben AE, editors. Evolutionary programming VII: 7th International Conference, EP98 San Diego, California, USA, March 25–27, 1998 proceedings. Berlin Heidelberg, Berlin, Heidelberg: Springer; 1998. p. 611–6.
- [120] Coley DA. An introduction to genetic algorithms for scientists and engineers. Singapore: River Edge, NJ: World Scientific; 2010.
- [121] Marateb HR, Mansourian M, Faghihimani E, Amini M, Farina D. A hybrid intelligent system for diagnosing microalbuminuria in type 2 diabetes patients without having to measure urinary albumin. *Comput Biol Med* 2014;45:34–42. <http://dx.doi.org/10.1016/j.combiomed.2013.11.006>.
- [122] Chawla NV. Data mining for imbalanced datasets: an overview. In: Maimon O, Rokach L, editors. Data mining and knowledge discovery handbook. Boston, MA: Springer US; 2005. p. 853–67.
- [123] Mosteller F. A k-sample slippage test for an extreme population. *Ann Math Stat* 1948;19:58–65.
- [124] Caruana R, Niculescu-Mizil A, Jurafsky D. In: Huang C-R, editor. An empirical comparison of supervised learning algorithms. Pittsburgh, Pennsylvania, USA: ACM; 2006. p. 161–8.
- [125] Rubin A. Statistics for evidence-based practice and evaluation. 3rd ed. Cengage Learning, Belmont, CA: Brooks/Cole; 2013.