

Hybrid PSO Feature Selection based Association Classification Approach for Breast Cancer Detection

Bilal Sowan^{1*}, Mohammed Eshtay², Keshav Dahal³, Hazem Qattous⁴ and Li Zhang⁵

^{1*}Department of Business Intelligence and Data Analytics,
University of Petra, Amman, Jordan.

²ASAC, LTUC, Amman, Jordan.

³AVCN Research Centre, School of Computing Engineering and
Physical Sciences, University of the West of Scotland, Paisley, UK.

⁴Department of Software Engineering, Princess Sumaya
University of Technology, Amman, Jordan.

⁵Department of Computer Science, Royal Holloway, University of
London, , Surrey TW20 0EX, UK.

*Corresponding author(s). E-mail(s): bilal.sowan@uop.edu.jo;

Contributing authors: m.eshtay@ltuc.com;

keshav.dahal@uws.ac.uk; h.qattous@psut.edu.jo;

li.zhang@rhul.ac.uk;

Abstract

Breast cancer is one of the leading causes of death among women worldwide. Many methods have been proposed for automatic breast cancer diagnosis. One popular technique utilizes a classification-based association called Association Classification (AC). However, most AC algorithms suffer from considerable numbers of generated rules. In addition, irrelevant and redundant features may affect the measures used in the rule evaluation process. As such, they could severely affect the accuracy rates in rule mining. Feature selection identifies the optimal subset of features representing a problem in almost the same context as the original features. Feature selection is a critical preprocessing step for data mining as it tends to increase the prediction speed and accuracy of the classification model and thereby increase performance. In

this research, an ensemble filter feature selection method and a wrapper feature selection algorithm in conjunction with the AC approach are proposed for undertaking breast cancer classification. The proposed approach employs optimal discriminative feature subsets for breast cancer prediction. Specifically, it first utilizes a new bootstrapping search strategy that effectively selects the most optimal feature subset that considers the overall weighted average of the relative frequency-based evaluation criteria function. We employ a Weighted Average of Relative Frequency (WARF) based filter method to compute discriminative features from the ensemble results. The adopted filter algorithms utilize the prioritization ranking technique for selecting a subset of informative features that are used for subsequent AC-based disease classification. Another wrapper feature selection method, namely a hybrid Particle Swarm Optimization (PSO)-WARF filter-based wrapper method, is also proposed for feature selection. Two classification models, i.e., WARF-Predictive Classification Based on Associations (PCBA) and hybrid PSO-WARF-PCBA, are subsequently constructed based on the above filter and wrapper based feature selection methods for breast cancer prediction. The proposed approach of the two models is evaluated using UCI breast cancer datasets. The empirical results indicate that our models achieve impressive performance and outperform a variety of well-known benchmark AC algorithms consistently for breast cancer diagnosis.

Keywords: Association Classification, Feature Selection, Classification, Breast Cancer Prediction

1 Introduction

Breast cancer is one of the most common cancers worldwide [1], including 12% of all new cases and 25% of all female cancers [2]. Breast cancer is a malignant tumour that forms in the cells of the breasts. A large number of ensemble techniques have been widely implemented to diagnose breast cancer with high accuracy rates. However, these are black-box approaches that cannot clarify the causes underlying the diagnosis [3]. As such, not only an accurate classifier is needed to help physicians predict this chronic disease because of the high risks of recurrence [3, 4], but also an effective and interpretable approach is required to diagnose breast cancer patients in order to attain well-performed care and treatment [3]. Classification is the task of categorizing data objects of a training dataset into a given number of class labels based on some characteristics. Classification aims to predict a class label that belongs to a categorical variable [5, 6]. The existing approaches have used association rule mining tasks in the classification process over the last few years with the attempt to explore a new Association Classification (AC) technique [7]. The AC approach is proposed by a combination of two well-known data mining processes, namely association rule mining and classification. The combination is performed for constructing an interpretable classifier [8]. The AC method also improves the

accuracy of the classification process, which also supports many fields of study and application domains, such as medical diagnosis. As mentioned earlier, the classification using association rules is based on integrating classification and association rule mining. Therefore, this integration aims to discover rules that accurately predict a single target variable of class labels. The key strength of association rule mining is that all interesting and strong rules are extracted [4]. For example, in a rule such as $A1, A2 \rightarrow C1$, where $C1$ is the predicted class label, while $A1$ and $A2$ are the values of the attributes. As stated above, this rule can be interpreted as follows: If $A1$ and $A2$, i.e., the values of the attributes occur for any data object together, then this object is classified as $C1$ [4, 8, 9]. In general, association rules mining and AC approaches have extracted all rules globally that satisfy both minimum support and minimum confidence thresholds [10]. Many of the rules are produced by AC approaches but cannot be produced by traditional classification techniques such as C4.5 [11].

Furthermore, many AC algorithms have been proposed in the literature, which achieved impressive results in terms of high accuracy in comparison with the state-of-the-art C4.5 classification model [8], which also achieved scalability, efficiency, effective rule generation, and interpretability (understandable by the end-user). Therefore, the extraction of a full set of classification association rules which contains significant knowledge to assist building a real classification model is required [9]. Indeed, in the medical field, it is believed that physicians have to be involved in the development process of the decision support systems. In other words, they bring expertise and knowledge to build intelligent and expert systems based on data mining techniques. Therefore, in this research, we will shed light upon the specifics of medical datasets for developing a data mining approach for building breast cancer prediction models. For such a purpose, we will address the following research questions:

- How can we select the most significant and informative features and eliminate the redundant features in existing datasets for breast cancer diagnosis?
- How can we build data mining models with reliable performance and sufficient interpretability to physicians? What are the expressed requirements in this regard?

In accordance with these questions, we propose various methods for discriminative feature selection and develop a comprehensive classification model with interpretability for breast cancer diagnosis. Specifically, we propose a new classification model, which combines the advantages of both feature selection methods and a rule-based AC model. Two feature selection methods are developed and offered in the proposed model. The first feature selection variation method is based on the filter method, while the second feature selection method is based on the wrapper method. To be precise, the bootstrapping search strategy, i.e., a Weighted Average of Relative Frequency (WARF) based filter method, is adopted to find the best feature subset in

the first feature selection method. This method calculates the weighted average as an evaluation criterion of all values of the base algorithm's relative frequency. Several filter algorithms are ensembled and aggregated in the base algorithm to compute the importance degrees of the features with respect to all other features for all filtering algorithms. The filtering algorithms used in this research are Correlation-based Feature Selection (CFS), Information Gain (IG), Gain Ratio (GR), Chi-squared (X2), and Relief-F. As a result, the WARF-based filter method is integrated with Association Classification (AC) approach for undertaking breast cancer prediction. Moreover, a hybrid Particle Swarm Optimization (PSO)-WARF wrapper method is also proposed to find the best feature subset in the second feature selection method. The PSO is embedded in the proposed hybrid PSO-WARF method for searching the candidate features. A weighted average of the relative frequency values for all candidate features is computed. Then, it employs the bootstrapping search strategy by calculating the relative frequency value for each candidate that satisfies the weighted average of the relative frequency. The hybrid PSO-WARF method utilizes the PSO algorithm for exploring the best features by embedding several classifiers. The classifiers are Naive Bayes, Random Forest, K-Nearest Neighbor, REPTree, Ensemble Selection, Bagging, Support Vector Machine (SVM), and Classification Based Association (CBA). Therefore, we merge and ensemble candidate features using a base algorithm to find the most optimal feature subset with the weighted average of relative frequency as a criterion function. As a result of this stage, the recommended most optimal feature subset is then used as inputs to the AC approach for breast cancer classification. The proposed approach is distinguished from other approaches in the following aspects. (1) Our approach develops two feature selection methods, i.e., the WARF-based filter method and the hybrid PSO-WARF method, to select the most significant and representative features, which constructs a seed set that determines essential features based on the WARF criterion function. (2) We build two classification models, i.e., WARF-PCBA and hybrid PSO-WARF-PCBA, based on the AC approach. (3) To increase efficiency, our approach eliminates many class association rules that are not required based on the proposed feature selection methods. (4) The proposed approach is used for tackling breast cancer prediction. Evaluated using the widely used breast cancer datasets collected from the University of California at Irvine (UCI) machine learning repository, the experimental results indicate that the proposed classification models achieve superior performance over other well-known AC algorithms.

The main contributions of this research are as follows.

- The proposed approach consists of two main stages, i.e., feature selection for determining the best feature subset, as well as the AC approach for building classification models for breast cancer prediction.
- We propose two methods for feature selection which consider a base algorithm in an ensemble selection mechanism for selecting the best and most

informative feature subsets. The methods are based on their overall weighted average, with relative frequency values used a criterion function. The proposed first method is based on the filter approach, while the second method combines the filter and wrapper approaches and serves as a hybrid ensemble feature selection method. The proposed methods are able to identify optimal feature subsets with competitive performance.

- In the first feature selection method, we search for the best feature subset using the bootstrapping search strategy. It adopts forward selection and backward elimination by employing the weighted average of the relative frequency for each ensembled feature, which is considered an individual threshold of the base algorithm. The bootstrapping search strategy minimizes the empirical loss on the search of the relative frequency process.
- The second method obtains the optimal feature subset using the PSO algorithm. The most important highlight in the ensemble base algorithm is to compute the relative frequency value, which needs to satisfy its overall weighted average as an individual threshold.
- We propose two models, i.e., WARF-PCBA and hybrid PSO-WARF-PCBA, for breast cancer prediction that integrates feature selection method and AC approach. The first model, i.e., WARF-PCBA, has employed the weighted average of the relative frequency values for learning by predictive classification-based association. In the meanwhile, the second model has embedded PSO as a search strategy to compute the weighted average of the relative frequency value features for learning by predictive classification-based association PSO-WARF-PCBA, with the attempt to further enhance performance. This model combines a feature selection method with the AC approach. It is capable of identifying the most discriminative feature subset to inform the subsequent AC-based classification to enhance performance.
- The wrapper approaches are based on selecting the best feature subset tailored for a specific algorithm. In contrast, the filter approaches attempt to quantify the sizes of dataset features, which may discard some important attributes. In comparison, the proposed methods for feature selection aim to be used with any prediction algorithm over any dataset. The research findings in this paper indicate that the proposed ensemble strategy is reliable and effective in selecting the best discriminative features. Specifically, the WARF-PCBA and hybrid PSO-WARF-PCBA models employed two stages to enhance the generalization performance of the classifiers. In the first stage, the PSO algorithm is utilized as a search technique to select the most important attributes among all features in the search space. In the second stage, the proposed models extract rules with one value attribute by employing PCBA, which increases the model efficiency by making the rules generated from data simpler and smaller than those generated by the other methods. Hence, the proposed models better represent the properties of the original dataset, which helps in improving classification accuracy.
- The advantages and micro-effects of the proposed hybrid PSO-WARF-PCBA model for breast cancer detection and prediction are as follows.

Firstly, applying PCBA, we are able to extract rules with one value attribute. Therefore, our model is able to improve runtime efficiency by making the rules created from data simpler and smaller than those generated by the other models. Also, unlike existing AC techniques, the hybrid PSO-WARF-PCBA model extracts rules that have a single feature value in their mechanism. The learning efficiency is enhanced due to accurately identified features' values by the PSO-WARF, which identifies and selects the most significant and discriminative features, thus, improving classification accuracy.

- The empirical results indicate the significance of the proposed approach by selecting the most optimal and significant features as compared with existing methods. As a result, the empirical findings show that our classification models consistently beat a range of baseline models and well-known benchmark AC algorithms for breast cancer diagnosis.

The outline of this paper is as follows. In Section 2, related work is discussed. In Section 3, the proposed methodology embedding feature selection and AC-based classification is demonstrated. Section 4, presents the details of comprehensive evaluation studies. Section 5, summarizes the findings and conclusions of this research.

2 Related Works

Many data mining techniques have achieved promising results, but only some of them can construct interpretable classifiers that are important in many fields, such as healthcare [12, 13] or gene expression [14]. In general, these techniques employ different knowledge discovery methodologies based on rules or trees that can obtain highly interpretable models. Using the Decision Tree (DT), a small change in input data will suffer from an adaptability problem that significantly affects the constructed model [15]. Other techniques do not consider the full dataset for extracting rules, but small samples of data and, hence, the classifier could not be representative and comprehensive of the whole cases. Many rules generated by the traditional classification techniques are hard to understand since these techniques are based on domain-independent biases and heuristics for generating a small set of rules to build a classifier. In contrast, the AC techniques build a classifier (i.e., a classification model) from training data by extracting a set of strong classification association rules [15]. AC can also create a reliable classifier because it is able to generate a set of rules that were not formed by other classifiers. Furthermore, unlike neural networks and other probabilistic approaches, which build classification models that are difficult to understand [16], the rules generated in AC are interpretable and easy to understand, and may also be modified manually by the end users. Many empirical studies [15] indicated that AC algorithms were usually able to construct better classifiers concerning a high accuracy rate than other classification models based on DT [17] and rule induction [18]. In the last decade, many AC algorithms have been proposed, such as Classification

Based Association (CBA) [11, 19], Classification based on Predicted Association Rule (CPAR) [20], Multi-class Classification using Association Rule (MCAR) [11], Multi-class, Multi-label Associative Classification (MMAC) [21], and Lazy Associative Classification (LAC) [22]. In the AC technique, the training process consists of extracting hidden knowledge, mainly using the discovered association rule to build a classifier after sorting and pruning uninteresting and redundant rules [9]. CBA is the well-known algorithm, the first AC algorithm proposed for classification. It operates based on two steps, i.e. a set of association rules is produced via the Apriori algorithm in the first step, and the extracted rules are then ranked in the second step to build a classifier model [8]. The CMAR algorithm [23] was proposed to generate multiple rules to predict unseen instances and deal with a complex data structure to speed up the runtime training process. These proposed algorithms work reasonably well in terms of accuracy and efficiency. However, CPAR [20] has been proposed, which is based on a combination of traditional rule induction and AC techniques, that has both accuracy and runtime improvements while, in part, ignoring interpretability. Many other works have proposed classification models in different application fields.

Kumar et al. [12] developed a cloud and IoT-based mobile healthcare application to monitor, predict and diagnose severe diseases. This application is based on a new classification algorithm called fuzzy rule-based neural classifier for diagnosing and estimating diabetes conditions and its severity. The proposed classifier was applied on medical data sets collected from the UCI repository and the real health records obtained from different hospitals by medical sensors to predict who has been severely affected by diabetes. The experimental findings indicated that the proposed classifier outperformed other frameworks for disease prediction. Lakshmanaprabu et al. [24] presented an IoT with a cloud-based clinical decision support system framework to predict Chronic Kidney Disease (CKD). The framework was employed a Deep Neural Network (DNN) classifier for the prediction of CKD. The dataset was collected from patients utilizing the medical sensors. Furthermore, a PSO-based feature selection method was applied to enhance the classification performance, confirming that the feature selection method significantly improved the classification results. Ahmed, H. et al. [25] introduced a real-time heart disease prediction model. The relief and univariate feature selection algorithms were used to select the most significant subset of features from the dataset. The experiments were conducted on four classification techniques, namely DT, SVM, Random Forest, and Logistic Regression. The experimental outcomes showed that the Random Forest classifier outperformed the other three techniques. Alwidian et al. [4] proposed a weighted classification based on the association rules algorithm called WCBA, based on the AC technique. The WCBA is applied to classify breast cancer data set, intended to improve the accuracy results by employing a statistical measure. An expert determined the importance of the dataset features by allocating weights for each feature. The experimental results demonstrated that WCBA outperforms the other AC

algorithms in most cases. The feature selection method reduces dataset dimensions by finding the best subset of features by searching for all the possible cases representing the dataset. Even if all features are useful for its original task, they may not be used to learn the prediction model. Therefore, feature selection has to be performed on the training data that aims to perform knowledge discovery. This process is comparatively easy to understand since some features may become redundant and irrelevant, which can lead to extracting useless and inaccurate knowledge. Moreover, irrelevant or redundant features cause unreliable and noisy data, making the extracted knowledge more difficult during the training [26]. Gao et al. [27] experimentally have shown that, in some cases, the model after the feature selection process can perform almost good as the model with all features, even if 85% of the features have been removed. Several techniques have been proposed in previous work for building an accurate classifier, some of them based on rules [15], DT [17], or SVM [28]. Various attempts have been proposed to develop an AC-based classification model [11, 19, 22].

According to several studies, it has been noticed that AC has many advantages as compared with other classification techniques such as DT, rule induction, and many others. The AC can extract all hidden patterns in the form of a set of rules from training data discarded by other classification techniques [9]. The advantages of employing AC allow updating and adjusting any particular rule without affecting the full set of rules. In contrast with the DT technique, any modification process for a rule requires reshaping the complete tree [29]. The AC model is developed using the best rules that are learned and produced from training data [19]. As mentioned earlier, the AC approach integrates both association rule mining and classification tasks. This approach is more accurate than traditional techniques [20, 29]. Additionally, AC approaches produce rules that are easy to understand by end user. They could even be updated manually, unlike probabilistic methods such as neural networks, which generate classification models that are difficult to interpret [30]. Data quality is one of the main factors that affect the success of the data mining task. The data may include irrelevant or redundant features, or noisy and unreliable data. Therefore, knowledge discovery during training is more complicated. As such, feature selection is necessary for identifying a subset of features and removing the irrelevant and insignificant features [31]. Feature selection is the process of finding the most appropriate informative and significant feature subset for a particular dataset to improve classification performance. The feature subset can help classify an original problem in the same manner as the original feature sets. Feature selection is often an essential task that considers a preprocessing phase for data mining techniques since it increases the efficiency of the prediction methods in terms of speed and accuracy. Also, feature selection enhances a deeper understanding of the original problem, where the data often includes noisy, complex, in a large-dimensional feature space. A large amount of data is needed for most data mining techniques such as prediction and classification. Some feature selection methods

have been performed based on selecting a subset of features by assigning a score to each feature, which ranks the features instead of determining features subset only [31, 32]. One of the most important tasks in developing and building an efficient classification model is the selection of the most significant and useful features and eliminating the remaining insignificant ones. In other words, it reduces the original set of features into a smaller one [1].

The evaluation of feature selection methods should consider the stability, performance, and efficiency when building a classification model with a small set of features. Stability is defined as the robustness of a subset of features generated by a feature selection method, which are smaller in size but represent the original problem effectively. The frequently used empirical evaluation measure involves the performance's accuracy. The efficiency is achieved by eliminating redundant and irrelevant features, in the meanwhile enhancing classifier performance [33].

Breast cancer is a malignant tumour occurring from breast cells. Even though some of the risk factors (aging, hereditary risk factors, family history, menstrual cycles, nulliparity, obesity) increase the possibility of developing breast cancer in women. The causes of breast cancer and how many factors induce cells to change into malignant tumors are unknown. Many studies are performed to find out the fundamental reasons, and scientists are significantly improving their understanding of how such changes in DNA that can affect healthy breast cells to turn into cancerous ones. However, breast cancer is the leading cause of death of women in most countries [1].

Many feature selection methods have been proposed, separated into three categories: filter, wrapper, or embedded methods. The filter methods select features from the dataset without any learning interaction. The wrapper methods evaluate selected features by using a learning algorithm. The embedded methods combine the feature selection step and classifier construction stage. In the filter methods, the relevance of features is evaluated by exploring them only at the underlying properties of the data. A feature relevance score is measured, and the low-scoring features are eliminated in most cases. Subsequently, this subset of features is provided as input for the classification algorithm. Many feature selection methods exist for scoring features such as information gain, gini index, chi-squared, relief, and correlation-based feature selection. The filter methods are fast, computationally simple, scalable for large datasets, as well as operating independently of the classification process. In wrapper methods, the model hypothesis search is embedded into the feature selection process, which interacts with the classification algorithm. The selected features are evaluated in the training and test stages of a specific classifier that is attempting to adapt to a provided dataset. The wrappers also evaluate selected features by using accuracy estimates provided by the actual target learning algorithm. In addition, in the third type of feature selection methods, the embedded methods, the feature selection process is performed during classifier construction. Thus, embedded methods are specific to a given learning algorithm, and like

the wrapper methods, they include the interaction with the classifier [31, 32]. Some popular filter-based feature ranking methods are introduced below.

- **Information Gain** is one of the simplest and fastest feature ranking methods. It is used to measure the quality of the dataset attributes. It quantifies the relevance between features and target classes, based on information entropy. It is widely adopted in machine learning fields for feature ranking and selection. Information entropy is one of the several methods for assessing diversity, impurity, and uncertainty of features for predicting the value of the target class label. Hence, information gain is used to reduce data high dimensionality to assess the effectiveness of data features in the classification or the prediction model. The information gain of the data can be calculated as defined in Eq. (1) [31]:

$$H(C) = - \sum_{i=1}^m P(C_i) \log_2 P(C_i) \quad (1)$$

where $H(C)$ is the entropy of a binary class, m is a class label, $P(C_i)$ denotes the probability that a class label $C \in m$. Entropy measures the uncertainty of each feature f in the selection process. Eqs. (2) and (3) calculates the conditional entropy of feature f and class c .

$$\begin{aligned} H(C | F) &= \sum_{f \in F} P_f(F) H(C | F = f) \\ &= - \sum_{f \in F} P_f(F) \sum_{c \in C} p(c | f) \log_2 p_c(c | f) \end{aligned} \quad (2)$$

$$= - \sum_{f \in F} \sum_{c \in C} P_{fc}(c | f) \log_2 p_c(c | f) \quad (3)$$

$$IG(c | f) = H(c) - H(c | f)$$

where $IG(c | f)$ is the information gain, $H(c)$ is class entropy and $H(c | f)$ is a conditional class entropy given the value of feature f .

When a class distribution is homogeneous, the misclassification error and entropy are both high. The entropy of minimum value is achieved when all the data samples belong to the same class. Thus, the smaller impurity value, the highest skewed class distribution. Information gain with a higher value means better decision-making, which is an excellent measure to discriminate the relevance of feature for a classification task. Not all features are contributed and considered equally in building a model and achieving good results. The features are ranked in information gain to depict the importance of features in building a model and attaining satisfying results [34].

- **Chi-squared (X2)** is a numerical test that measures the dependency between two features in a dataset and assesses their independence. It is a statistical test used for a categorical or nominal dataset. Chi-squared selects

the best features related to the target class after scoring them based on the class and extracting the top-ranking ones [35].

Besides the introduction of feature selection methods, we also introduce the AC method below. Typically, an AC algorithm works in three main steps. The first step is to find a hidden relationship (correlation) among attributes and class labels attribute in training data to generate Class Association Rules (CARs) in the form of IF-Then. In the second step, rank and prune irrelevant and redundant CARs, where the ranking process sorts CARs depending on certain thresholds such as confidence and support values. In the third step, it builds a classification model based on the CARs, which can be evaluated to measure the effectiveness of the constructed classifier in predicting the class label of unseen testing data. The evaluation is accomplished using the accuracy or error rate of the classifier [9]. For the classification task, AC is formally defined as follows:

Definition: given training dataset D , which comprising n distinct attributes (features) $A = \{A_1, A_2, \dots, A_n\}$, where C is a set of class labels including several distinct values. A set of transactions $TID = \{t_1, t_2, \dots, t_m\}$ is formed from D and each transaction t_j comprises a subset of attributes $\{A_k, \dots, A_l\}$, $1 \leq k, l \leq n$ and a specific value for the class C . As a result, building a classifier from D is defined as follows: let A be the set of attribute values, and C is the set of class labels. The classification task could be formally described for predicting (classifying) the class value of a testing instance from testing data S , such as $S = \{S_1, S_2, \dots, S_l\}$, where $\forall s \in S : S \subseteq A$. This task of mapping a set of input attributes to an output attribute is conducted using the classifiers (rules) as $A \rightarrow C$ [30].

As mentioned above, the formal definition of the AC problem represented in the form of “IF-Then” as follows: $A_i \rightarrow C_j$, where A_i is an itemset (attribute value set) called the antecedent, while C_j is a class label called the consequent. Let D be a training dataset of n distinct attributes $A = \{A_1, A_2, \dots, A_n\}$, and C is a set of class labels including several distinct values. Attributes can be either categorical or continuous. In the case of categorical attributes, all possible values are transformed into a set of positive integers, while in the case of continuous attributes, a discretization method is applied. The classification process utilizes the association rules as stated in a rule that $A_i \rightarrow C_j$. A set of transactions $TID = \{t_1, t_2, \dots, t_m\}$ includes a subset of attributes [7]. Each TID denoted as a training object can be described as a set of attribute A_i and values V_i , joined up a class denoted by C_j . The item can be described as an attribute A_i and value V_i , where (A_i, V_i) is an item. An itemset is a set of items covered in a training data object. A rule itemset is a form $\langle itemset, C_j \rangle$, where C_j is a class label. Therefore, The CARs comprise all possible rules that are both frequent and accurate [8]. The CAR is formally expressed and presented as follows: $(A_1, V_{11}) \wedge (A_i, V_{il}) \wedge \dots \wedge (A_n, V_{nm}) \rightarrow C_j$ [4]. The support count (actual occurrence and frequency) of the itemset. To find the support count of rule itemset that belongs to a specific class label C_j that contains

itemset and C_j divided by $|TID|$ (i.e., $support = \frac{itemset \cup C_j}{|TID|}$), where $|TID|$ is the number of transactions). The frequent rule itemset that satisfies $minconf$ is accurate (i.e., $confidence \geq minconf$, $confidence = \frac{support(itemset \cup C_j)}{support(itemset)}$), which is called Class Association Rules, i.e. CARs.

3 The Proposed WARF-PCBA and PSO-WARF-PCBA models

The section describes the proposed models for efficient breast cancer prediction. A new AC approach is proposed for breast cancer detection and prediction, including a preprocessing comprising feature selection and classification models. The proposed approach consists of two phases. The first phase is used to collect the necessary breast cancer dataset from the UCI machine learning repository, and then the proposed two feature selection methods, i.e. WARF-PCBA and PSO-WARF-PCBA are applied. The second phase is responsible for breast cancer prediction with diagnosis and recurrences. In this phase, two classification models are constructed based on the AC approach, which uses the most significant and informative features in breast cancer datasets identified and selected by the proposed WARF and PSO-WARF feature selection methods. As a result, accurate and efficient medical classifier models are built. The proposed models are presented in Figs. 1 and 2.

3.1 The proposed methodology for feature selection

We present the two proposed feature selection methods in this subsection. The first feature selection method uses the bootstrapping method to extract the best feature subset. It performs optimal feature selection by searching all the possible candidate attributes and calculating the weighted average for all features' relative frequency-based evaluation criteria. The identified best feature subset is then used as inputs of the AC approach.

- WARF-based filter method:

The first proposed first feature selection method is a weighted average of relative frequency (WARF) based filter method applied to the entire feature space. It is regarded as a filter method owing to searching the most significant features based on weight and importance. It is accomplished by ensembling and aggregating several well-known feature selection and ranking algorithms such as CFS, IG, GR, X2, and Relief-F.

Algorithm 1 presents the pseudo-code of the proposed WARF-based filter method for feature selection. The procedures of the WARF method, which selects the features based on bootstrapping search strategy, are used as a selection mechanism that aims to find occurrence times. Thus, the subset features were chosen from the selected and highest-ranked features generated by the aggregated base algorithm and should be satisfied that each feature must have the same ranking for all adopted algorithms. Then, the features are divided

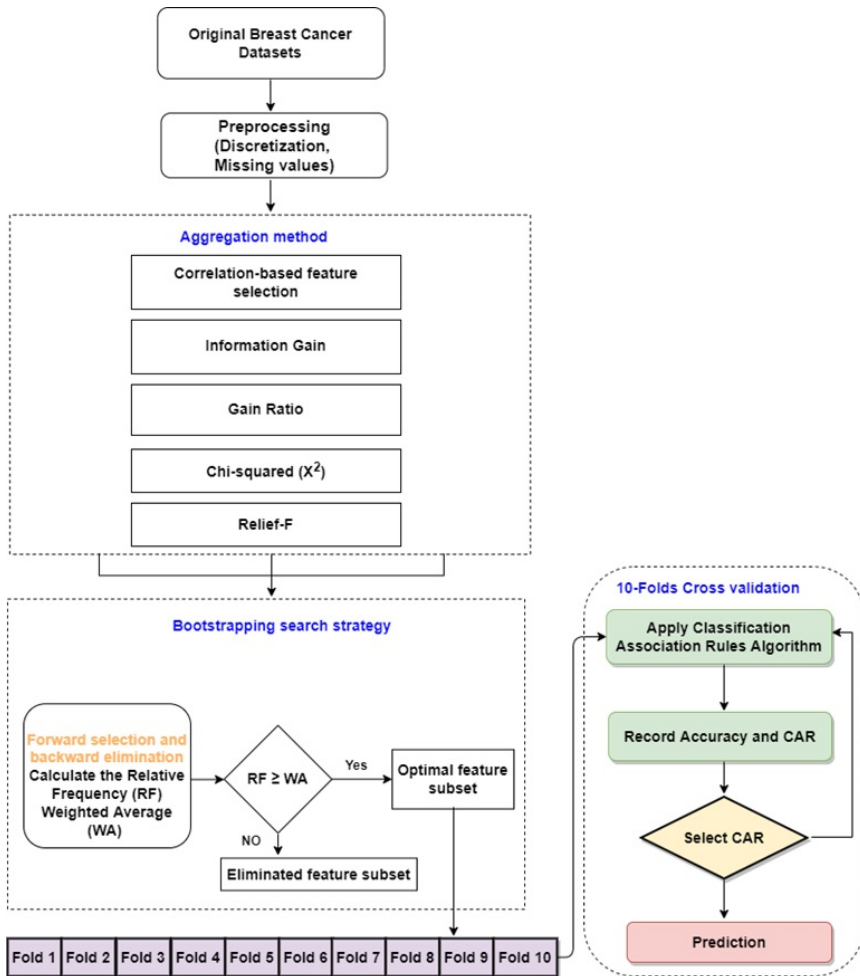


Fig. 1 The proposed first WARF-PCBA model.

into two groups in the proposed WARF-based filter method. The first group refers to the selected candidate features (called forward selection) of high-rank value, while the second group denotes the eliminated features (called backward elimination) of low-rank value. In both parts, the relative frequency is computed for each aggregated feature to calculate the feature's weight and occurrence times among all other features. The proposed WARF-based filter method employs the weighted average to filter out the features and find the optimal feature subset. Feature selection is accomplished based on the relative frequency value of each selected feature which must satisfy the weighted average of relative frequency for all features. As a result, the most important features with the highest weight and influence scores are extracted and selected, and the lowest influential features are excluded.

The steps of the WARF-based filter method can be summarized as

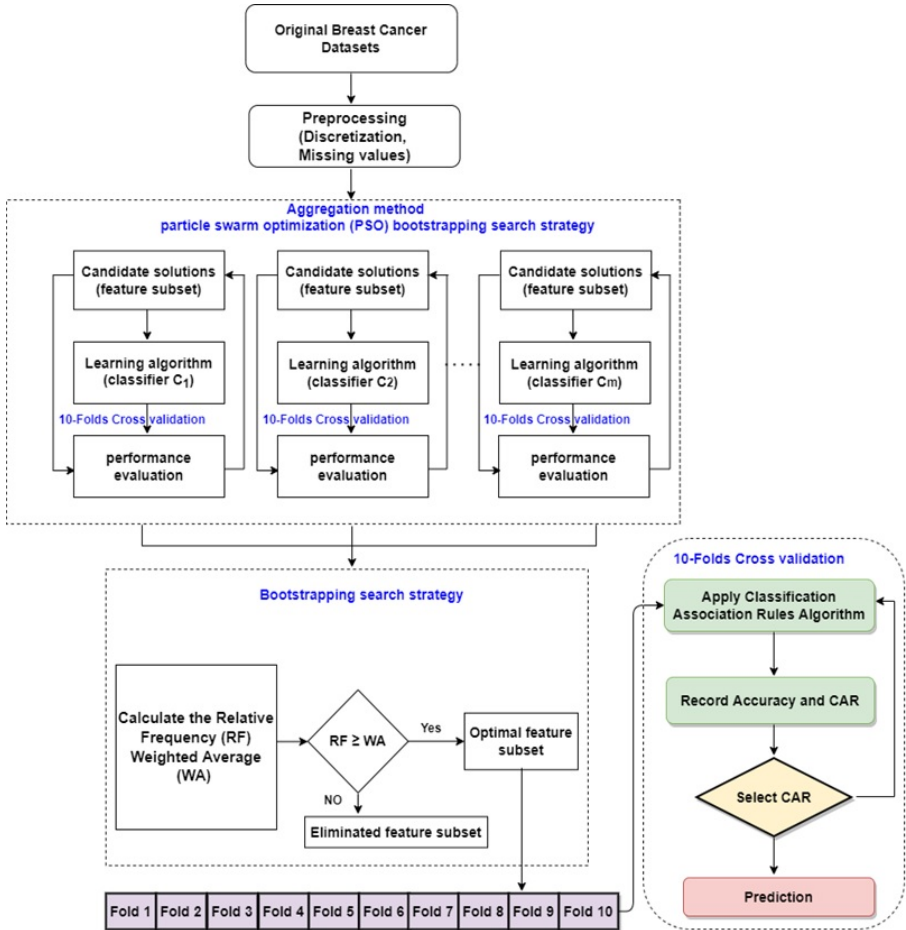


Fig. 2 The proposed second hybrid PSO-WARF-PCBA model.

follows:

Step 1: Several filter algorithms are ensembled and aggregated in a base algorithm. The result of this step is the ranked and selected features in each algorithm based on their importance scores. The higher the scores are, the more important the features are.

Step 2: The features are ranked and divided into selected, and eliminated candidate features based on the median of n as a threshold for features dividing, where n is the number of features in the respective dataset.

Step 3: The relative frequency is calculated for each candidate feature and used to calculate the feature weight. Such a feature weight defined in Eq. (4) refers to importance score of a feature against those of other features as a weighting parameter:

$$RelativeFrequency = \frac{FeatureOccurance_{ij}}{\sum_{i=1}^n FeatureOccurance} \quad (4)$$

where, *FeatureOccurance* is the number of times the feature occurred among all algorithms. $\sum_{i=1}^n FeatureOccurance$ is the total and overall frequencies of all features among all algorithms.

Step 4: The weighted average *WA* is calculated to select the optimal feature subset as indicated in Eq. (5):

$$WA = \frac{\sum_{i=1}^n RelativeFreq * Importance}{\sum_{i=1}^n Feature Importance} \quad (5)$$

where, *WA* is the weighted average that satisfies the following condition, $RelativeFrequency \geq FeatureImportance$ is calculated as $\frac{1}{n}$, where *n* is the number of features.

The proposed WARF-based filter method selects the optimal feature subset by satisfying the relative frequency's overall weighted average evaluation function. The forward selection and backward elimination search strategies are divided based on the median ranked score of features as a search division threshold. In the forward selection search strategy, the relative frequency is calculated for all candidate selected features, and then the weighted average is computed for all features of the highest weight and frequency. While in the backward elimination feature search strategy, the relative frequency is calculated for all eliminated features, and then the weighted average is computed for all features of the lowest weight and frequency.

Definition: Let a dataset *D* consisting of a number of attributes or features $F = \{F_1, F_2, \dots, F_n\}$ and $MD = \{MD_1, MD_2, \dots, MD_m\}$ be a filter algorithm for selecting and ranking features. Let $RankedF_{ij} = \{F_j \mid 1 \leq j \leq n\}$ for a method $MD_i \in MD, 1 \leq i \leq m$. The ensembled and aggregation method generates the selected features *RankedF_j* with their importance scores calculated by the *MD_i* algorithm. Then apply the bootstrapping search strategy using the forward pass selection and backward pass elimination. In the case of the forward pass selection with respect to the feature selection, the Candidate Feature Selection *CFS* search strategy is applied by dividing *RankedF_{ij}* based on the median. It is calculated as the average of the ranked score based on a threshold, where the minimum value of median or average is less than or equal (or less than) ranked score. While, in case of the backward pass selection, the Candidate Feature Elimination *CFE* search strategy is applied by dividing *RankedF* based on the median (i.e. the average of the ranked score) as a threshold, where the min of median or average is greater than or equal to the ranked score. As a result, the optimal selected features subset *SF* is obtained through $SF = \{SF_j \mid CFS_{ij}.RelativeFreq_{ij} \geq WA\}$ or $SF = \{SF_j \mid (CFE_{ij}.RelativeFreq_{ij} \geq WA) - RankedF\}$.

- Hybrid PSO-WARF filter-based wrapper method:

The second feature selection method employs the bootstrapping search strategy to find the best feature subset. It does this by searching all the possible

Algorithm 1 The proposed WARF-based filter method

Input: all dataset features F_n , feature ranking method MD_m .

Output: The optimal selected feature subset SF .

```

1: procedure WARF
2:   for each method  $MD_i$  do // the feature ranking method  $i$ , where
    $i \in MD$  and  $1 \leq i \leq m$ .
3:      $RankedF_{ij} \leftarrow Rank$  the features  $F_j$  in each method  $MD_i$ .
4:   end for
5:   for each method  $MD$  do
6:      $RF_j \leftarrow Rank$  the features  $F_i$  in each method  $MD_j$ .
7:   end for
8:   for each  $RankedF_{ij}$  do // the ranked feature  $j$  by method  $MD_i$ ,
   where  $j \in F$  and  $1 \leq j \leq n$ .
9:     Calculate median, average, and standard deviation for the ranked
   score of all algorithms.
10:    Apply bootstrapping search strategies: forward selection (candidate
   features selection  $CFS_{ij}$ ) and backward elimination (candidate features
   elimination  $CFE_{ij}$ ) by dividing  $RankedF_{ij}$  into based on the median,
   average of the ranked score as a threshold.
11:    if candidate features selection  $CFS$  then
12:      Forward pass selection of min (median, average) of the ranked
   score where min is less than or equal (or less than) ranked score.
13:      Calculate the relative frequency  $RelativeFreq$  for all  $CFS$ ,
    $RelativeFreq$ .
14:      Calculate the weighted average  $WA$  of  $RelativeFreq$ ,  $WA$ .
15:       $SF \leftarrow$  contains the optimal selected feature subset that is
    $RelativeFreq_{ij}$  greater than or equal to  $WA$ .
16:    end if
17:    if candidate features elimination  $CFE$  then
18:      Backward pass elimination of min (median, average) of the
   ranked score where min is greater than or equal (or greater than) ranked
   score.
19:      Calculate the relative frequency  $RelativeFreq$  for all  $CFE$ ,
    $RelativeFreq$ .
20:      Calculate the weighted average  $WA$  of  $RelativeFreq$ ,  $WA$ .
21:       $EF \leftarrow$  contains the eliminated feature subset that is
    $RelativeFreq_{ij}$  greater than or equal to  $WA$ .
22:       $SF \leftarrow RankedF - EF$ .
23:    end if
24:  end for
25: end procedure

```

candidate features extracted in the hybrid PSO-WARF filter-based wrapper method. It aggregates the wrapper method that embeds several classifiers by ensembling candidate features in a base algorithm. After that, the best feature subset is selected by utilizing the weighted average for all features' relative frequency-based evaluation criteria. The identified most optimal feature subset is then used as inputs to the AC approach for breast cancer classification. The proposed hybrid method integrates the bootstrapping search strategy with a PSO-based wrapper method to identify optimal features in the overall feature space. In particular, the PSO-based search process is considered as a generalized wrapper-filter method for searching the most significant features based on weighting and importance scores. The hybrid PSO-WARF method utilizes the PSO algorithm for identifying the most optimal features by embedding several classifiers (such as Naive Bayes, Random Forest, K-Nearest Neighbor, REPTree, Ensemble Selection, Bagging, SVM, CBA). Therefore, the candidate features are merge and ensembled in a base algorithm in order to identify the best feature subset using the weighted average of relative frequency as a selection criterion.

Algorithm 2 presents the pseudo-code of the proposed hybrid PSO-WARF filter-based wrapper method for feature selection. This hybrid PSO-WARF wrapper method selects the most optimal features based on the PSO algorithm by involving each classification model. Serving as a base algorithm, this method also incrementally and iteratively performs the ensemble aggregation of the selected features of all classifiers. The bootstrapping search strategy is applied as a selection mechanism that aims to find occurrence times. Thus, the optimal features were chosen from the highest-ranked features generated by the aggregated base algorithm and should satisfy that each feature must have the same ranking for all adopted algorithms. The relative frequency is computed for all the aggregated features in the base algorithm to calculate the feature's weight and occurrence times among all other features. The proposed PSO-WARF filter-based wrapper method employed the weighted average to filter out the features and extract the optimal subset features. The feature selection is accomplished based on the relative frequency value of each selected feature, which must satisfy the weighted average of relative frequency for all features. As a result, the higher weighting scores and more important the features are. **The steps of the hybrid PSO-WARF filter-based wrapper method can be summarized as follows:**

Step 1: The wrapper approach employs several classifiers ensembled and aggregated in a base algorithm. The results of this step are the selected features in each wrapper algorithm incorporating PSO and a classification method.

Step 2: In the base algorithm, the relative frequency is calculated for each selected feature to obtain its weight and importance score against those of other features as a weighting parameter. This process is defined in Eq. (4).

Step 3: the weighted average WA is calculated which is used to select the optimal features as defined in Eq. (5).

The proposed PSO-WARF filter-based wrapper method selects the most optimal feature subset by satisfying the relative frequency's overall weighted average evaluation function. The PSO-bootstrapping search strategy is utilized based on a wrapper approach mechanism for searching significant features. In the bootstrapping search strategy, the relative frequency is calculated for all candidate selected features in the base algorithm, and then the weighted average is computed for all features with the highest weight and frequency scores.

Definition: Let a dataset D consisting of a number of attributes or features $F = \{F_1, F_2, \dots, F_n\}$ and $Clr = \{Clr_1, Clr_2, \dots, Clr_m\}$ be a classifier employed in the wrapper algorithm. Let $ClrSF_{ij}$ be the selected feature for a particular classifier Clr in combination with PSO-based search strategy, where $Clr_i \in Clr$ and $1 \leq i \leq m$. The ensembled and aggregation base algorithm then applies the bootstrapping search strategy. As a result, the identified optimal feature subset SF is obtained through $SF = \{SF_j \mid ClrSF_{ij}.RelativeFreq_{ij} \geq WA\}$.

3.2 The proposed AC models

Finding the most significant risk factor features in medical datasets is one of the key challenges in data mining tasks. A useful feature selection method is required for improving the classification performance by removing irrelevant and redundant features from high-dimensional datasets. The proposed methodology is able to achieve accurate and compact classification results. The accuracy can be improved by using the most important and relevant features for building the classifier. While the compactness can be achieved by eliminating redundant features, which in turn will affect the generated numbers of Class Association Rules optimistically. In this study, the proposed WARF and PSO-WARF methods are able to reduce the feature space and then identify the most optimal feature subset to aid building more accurate prediction models. The proposed classification models, namely WARF-PCBA and hybrid PSO-WARF-PCBA, as shown in Algorithm 3, are based on integrating well-known algorithms CBA and predictive apriori.

The proposed models are based on three main tasks as follows: In the first task, ensemble of rule generation refers to a set of rules that have been generated using the integration of CBA-RG and predictive apriori algorithms [8, 36]. In the second task, a rule pruning method is applied as shown in Algorithm 4, and in the final task, the prediction performance is evaluated as demonstrated in Algorithm 5. The advantages of the proposed models are as follows. They act as an ensemble approach that integrates the feature selection method with the AC approach. Thus, the proposed models comprise high accuracy with impressive performance and interpretability. The prediction accuracy metrics refers to the well-known statistical measures presented in the following subsection, while the number of generated CARs defines the term of prediction performance. As a result, the proposed models improve the decision-making process with respect to breast cancer diagnosis.

Due to the fact that PSO is one of the important evolutionary computation techniques with compact and robust exploratory capabilities, it has proven to be useful in resolving problems involving supervised wrapper-based feature selection. PSO searches for a feature subset as significant features, similar to how a group of flying birds as a swarm travels over time to find the optimal landing spot for all groups to arrive at once. PSO attempts to find an optimal solution based on an iterative search process [37]. The particle swarm in PSO represents a multidimensional search vector in finding the best feature subsets in multidimensional velocity. After each iteration, the particle's velocity is updated based on the particle's personal and global best positions, which are defined by the objective function [38]. The objective function aims to select features with high relevancy to class labels and eliminate redundant attributes that have no significance to class labels. The use of the filter and wrapper approaches is based on the assessment of evaluation criteria. The assessment is based on integrating both strategies as a hybrid approach. Thus, the evaluation process takes into account both filter and wrapper metrics. The main advantage of using PSO is appropriately for the domain where there is a structure in how candidate features interact highly, with the intention that quick convergence does not result in local optima [39]. PSO is one of the best search methods for feature selection because of its great efficiency in tackling high dimensional data. In PSO, the particle swarm has a memory that allows it to keep track of the information about the potential optimal solutions. At the same time, it explores the problem space by using different particles to search for alternative optimal positions in the solution space. PSO is a powerful, yet simple search method that can find the most optimal solution in an iterative process. It is easy to implement, requires minimal parameters, and has almost proven to find the optimal global solution for a variety of optimization problems. Furthermore, the PSO algorithm uses less memory and runtime than those of other feature selection methods and does not require sophisticated mathematical operations. Moreover, the PSO algorithm can handle binary and discrete datasets [40]. Concerning the AC approach, the results of employing such an approach are distinguishable and can be categorized into two important points. The first point is that the extracted rules are simpler, smaller, and easier for end users to understand and interpret. The second point is that, in terms of classification performance, the AC approaches outperform conventional data mining techniques [41].

The micro-effects of the PSO-based feature selection against those of filter and other search methods were discussed below. As an example, Tan et al. [42] proposed a PSO variant for feature selection with respect to skin cancer detection. Their PSO model followed local and global best and worst leader signals to identify the most important features while avoiding less optimal regions. In comparison with filter-based methods (e.g. ReliefF) without the consideration of feature interaction, their model selected comparatively smaller but more effective feature subsets. For example, their selected features included clinically important attributes such as shape (e.g. border irregularity, asymmetry, and

compactness) and colour (e.g. variance of image darkness, entropy, skewness, correlation, and mean of RGB) features with respect to melanoma diagnosis, while some of such features were not ranked high enough or regarded as crucial enough to be selected by the filter- and other wrapper-based methods, therefore affecting their model performance.

Srisukkhram et al. [38] exploited a Bare-Bones PSO (BBPSO) based feature selection method for leukaemia diagnosis. Their model adopted hybrid leaders for the search of optimal features. Clinically important morphological features such as the cytoplasm and nucleus areas, the ratio between nucleus and cytoplasm areas, and texture changes related to open or close chromatin, were selected consistently by their model, while some of such features were often completely discarded or not co-existence in the feature subsets extracted by other search methods. A similar observation was also obtained through the studies of Xie et al. [43]. They employed another PSO variant integrated with random distributions for swarm leader enhancement and hybrid signals for feature selection. Their model identified morphologically important features such as form factor, compactness, perimeter, and eccentricity for blood cancer classification, as well as crucial features such as maximum heart rate, serum cholesterol, chest pain type, and ST depression for coronary heart disease identification. In contrast, such sophisticated feature interactions were not well captured by other existing search methods. They either excessively eliminated some of the above crucial features or selected a significant number of redundant features with limited discriminative capabilities. Such micro-effects of PSO-based feature selection were also exploited in detail in other existing review studies [39].

3.3 Performance measure

The most popular model evaluation method in data mining and machine learning is the ten-fold cross-validation, which concerns resampling selection. The cross-validation technique estimates the prediction accuracy in evaluating the performance of the classifier model. The prediction accuracy is calculated for each fold. As a result, the cross-validation accuracy rate is the average of the results of all the ten folds obtained on the validation sets [44].

The prediction accuracy is defined as a total number of all correct predictions divided by the number of all samples N in a dataset D . For a more formal definition; the prediction accuracy can be defined as follows [45]:

$$Accuracy(N_{ACC}) = \frac{1}{N} \sum_{i=1}^N W(\hat{y}_i, y_i) \quad (6)$$

Algorithm 2 The proposed hybrid PSO-WARF filter-based wrapper method.

Input: all dataset features F_n , wrapper method employing m classifiers Clr .

Output: The optimal selected feature subset SF .

```

procedure HYBRID PSO-WARF
2:   for each classifier  $Clr_i$  do // the wrapper method employing classi-
      fiers, where  $i \in Clr$  and  $1 \leq i \leq m$ .
      for each fold 1 to 10 do // apply 10-fold cross-validation.
4:     Initialize particles population randomly a group of a candidate
      solution (candidate feature subset solution).
      for each  $t = 1$  to  $MaxGenerations$  do // number of genera-
      tions.
6:       for each particle  $i$  do
          fitness value = Apply  $Clr_i$  using features extracted from
          particle  $i$ .
8:         if  $val > pBest_i$  then
             $pBest_i = val$ .
10:        end if
        end for
12:      Choose the particle having the best fitness value as the
       $gBest_t$ .
        for each particle  $i$  do
14:           $v = v + c1 * rand * (pBest_i - p) + c2 * rand * (gBest_t - p)$ .
             $p = p + v$ .
16:        end for
        end for
18:       $F = gBest$ .
        end for
20:       $ClrSF_i \leftarrow$  Selected features from  $F$  for classifier  $Clr_i$ .
        end for
22:      for each  $ClrSF_i$  do // the selected feature  $j$  by wrapper employing
      classifier  $Clr_i$ , where  $j \in F$  and  $1 \leq j \leq n$ .
        Calculate the relative frequency  $RelativeFreq$  for all  $CFS$ ,
         $RelativeFreq$ .
24:      Calculate the weighted average  $WA$  of  $RelativeFreq$ ,  $WA$ .
         $SF \leftarrow$  contains the optimal selected feature subset that is
         $RelativeFreq_i$  greater than or equal to  $WA$ .
26:      end for
end procedure

```

The 0-1 Win $W(\hat{y}_i, y_i)$ is defined as:

$$W(\hat{y}_i, y_i) = \begin{cases} 0 & \text{if } \hat{y}_i \neq y_i \\ 1 & \text{if } \hat{y}_i = y_i \end{cases} \quad (7)$$

Algorithm 3 PCBA classifier

Input: Dataset D , Desired number of class association rules n **Output:** PCBA classifier // predictive classification based association.

```

procedure PCBA
    FSD=WARF( $D$ ) OR PSO-WARF( $D$ ) // feature selection for dataset
     $D$ .
3:   PCBA(FSD).
    for  $i = 1$  to  $K$  fold ( $K$  is 10-fold) do
        Divide  $FSD$  into training  $Tr$  and testing  $Ts$  (10-fold cross-
        validation).
6:    $GR =$  empty rule set.
         $GR =$  RuleGeneration( $Tr$ ). // Ensemble of RuleGen [36] and CBA-
        RG algorithm [8].
         $StrongR =$  RulePruning( $GR, Tr$ ) // GR is a set of generated rules
         $R$ .
9:    $AccFold =$  Prediction( $StrongR, Ts$ ) // StrongR is set of strong
        rules.
    end for
    Average Accuracy of  $AccFold$  // average accuracy of all folds.
12: end procedure

```

Algorithm 4 Rule pruning method

Input: GR, Tr // GR is a set of generated rules, and Tr is training dataset.**Output:** Classifier (Clr) // a set of best predictive rules.

```

procedure RULEPRUNING( $GR, Tr$ )
     $GR' =$  sort( $GR$ ) // sort all generated rules based on their predictive
    accuracy equation in [36] and rule length (less number of attributes has
    the highest priority) in descending order.
     $PCBA(FSD)$ 
4:    $Folds = 10$ 
    for each rule  $r_i \in GR'$  do
        Find all instance cases  $ic$  in  $Tr$  that satisfies the conditions of  $r_i$ 
        if  $ic_j$  satisfies the conditions of  $r_i$  then
8:           Insert the rule  $r_i$  at the end of classifier  $Clr$ .
           Delete all instance cases in  $Tr$  covered by  $r_i$ 
        end if
        if all instance cases in  $Tr$  cannot correctly be covered by  $r_i$  then
12:           Delete  $r_i$  from  $GR'$ 
        end if
    end for
    Insert the majority class as the default class to the  $Clr$ .
16: end procedure

```

Algorithm 5 Prediction method

Input: $StrongR, Ts$ // StrongR is a set of strong rules, and Ts is testing dataset.

Output: Prediction Accuracy

```

procedure PREDICTION( $StrongR, Ts$ )
   $Clr = StrongR$  // Clr is the classifier
   $PCBA(FSD)$ 
  for each test case  $ts_i$  in  $Ts$  do
5:   for each ranked rule  $r_i$  in  $Clr$  do
     if rule  $r_i$  matches  $ts_i$  condition then
       Assign the class of  $r_i$  to  $ts_i$ .
     end if
   end for
10:  if all rule in  $Clr$  does not matches  $ts_i$  condition then
     Assign the default class to  $ts_i$ 
   end if
  end for
  Compute  $accuracy$ .
15: end procedure

```

where y_i is the i^{th} actual class label and \hat{y}_i is the i^{th} predicted class label respectively.

$$e = 1 - Accuracy(N_{ACC}) \quad (8)$$

where e is the prediction error, i.e. the loss over N samples in a dataset D , which is computed as the expected value in the range of $[0, 1]$.

Accuracy is the most widely used measurement for evaluating the performance of prediction and classification models. The accuracy rate can be calculated based on the confusion matrix as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (9)$$

where, TP is True Positive, FP is False Positive, FN is False Negative, and TN is True Negative.

3.4 The breast cancer datasets

The datasets used in our experimental studies are obtained from the University of California at Irvine (UCI) machine learning repository [46]. Two breast cancer datasets are used in the experiments. The first dataset is breast cancer recurrences, and the second one is a breast cancer diagnosis. The breast cancer recurrences dataset includes ten attributes and 286 instances, as shown in Table 1, whereas the breast cancer diagnosis dataset consists of ten features and 699 cases, as shown in Table 2. The same datasets are used in numerous other research efforts to improve breast cancer prediction.

Table 1 The features of the breast cancer recurrences dataset.

No.	Name of feature	Possible Values	Number of Values
1	Age	f1 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 8089, 90-99.	9
2	Menopause	f2 lt40, ge40, premeno.	3
3	Tumor-size	f3 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59.	12
4	Inv-nodes	f4 0-2, 3-5, 6-8, 9-11, 12-14, 15-17, 18-20, 21-23, 2426, 27-29, 30-32, 33-35, 36-39.	13
5	Node-caps	f5 yes, no.	2
6	Deg-malig	f6 1, 2, 3.	3
7	Breast	f7 left, right.	2
8	Breast-quad	f8 left-up, left-low, right-up, right-low, central.	5
9	Irradiat	f9 yes, no.	2
10	Class	f10 No-recurrence-events, Recurrence-events	2

Table 2 The features of the breast cancer diagnosis dataset.

No.	Name of feature	Domain
1	Clump Thickness	f1 1-10
2	Cell Size Uniformity	f2 1-10
3	Cell Shape Uniformity	f3 1-10
4	Marginal Adhesion	f4 1-10
5	Single Epithelial Cell Size	f5 1-10
6	Bare Nuclei	f6 1-10
7	Bland Chromatin	f7 1-10
8	Normal Nucleoli	f8 1-10
9	Mitoses	f9 1-10
10	Class	f10 Benign and Malignant

4 Experimental Results and Analysis

To test model efficiency, a comprehensive evaluation is performed. The adopted experimental studies consist mainly of three parts; the experiments were conducted to validate the breast cancer recurrences. Similarly, the experiments were also performed to validate breast cancer diagnosis. Furthermore, extensive experiments were conducted to select the most significant and informative features. The proposed approach was compared with the state-of-the-art algorithm and with several AC algorithms, namely, CMAR, MCAR, FACA, ECBA, WCBA, and CBA.

Feature importance plays an essential role in deciding to what extent the features play a role in the predictive power for the classification model. In this respect, the sensitivity analysis for selecting features is conducted based on removing features at each iteration. The features generated by the proposed WARF method were utilized to investigate the importance of the selected features. In other words, the proposed WARF method generates a new subset of features. In the first part of the comparison.

Table 3 illustrates the results obtained using Algorithm 1, and the experiments show the relative frequency of forwarding selection and backward elimination of the breast cancer recurrences dataset when the weighted average is equal to 0.11 as a threshold for the selection or elimination of features. The selected features by the proposed WARF method is then used in subsequent experiments for classification purposes. The empirical results indicate that the yielded classification models using the selected features have produced impressive and consistent performance.

Table 3 The relative frequency of forward selection and backward elimination of the breast cancer recurrences dataset (Weighted Average=0.11).

Forward pass selection		Backward pass elimination	
Candidate features selection (selected feature=4)	Candidate features selection (selected feature=5)	Candidate features elimination (eliminated features=4)	Candidate features elimination (eliminated features=5)
f1 0.05	f1 0.04	f1 0.2	f1 0.16
f2 0.05	f2 0.04	f2 0.2	f2 0.16
f3 0.15	f3 0.16	f3 0.05	f3 0.08
f4 0.2	f4 0.16	f4 0.05	f4 0.04
f5 0.2	f5 0.16	f5 0.05	f5 0.04
f6 0.2	f6 0.2	f6 0	f6 0.04
f7 0.0	f7 0.04	f7 0.2	f7 0.2
f8 0.05	f8 0.04	f8 0.2	f8 0.16
f9 0.1	f9 0.16	f9 0.05	f9 0.12
Selected features: f3, f4, f5, f6	Selected features: f3, f4, f5, f6, f9	Eliminated features: f1, f2, f7, f8	Eliminated features: f1, f2, f7, f8, f9

Fig 3 shows the performance results of the accuracy evaluation measure of the WARF-PCBA model with respect to the breast cancer recurrences dataset. As indicated in Fig 3, the experimental results show that the WARF-PCBA model produced better results with an accuracy rate of 76.2%, when the number of selected features is 4 on the breast cancer recurrences dataset, in comparison with the accuracy rate of 70.3% using the full feature set. The accuracy rate of WARF-PCBA is 76.2% for the breast cancer recurrences dataset by removing age, menopause, breast, breast-quad, and irradiat features. The results were based on a sensitivity analysis of the different models where the selected features are then used to confirm the effectiveness of the proposed WARF method and WARF-PCBA model.

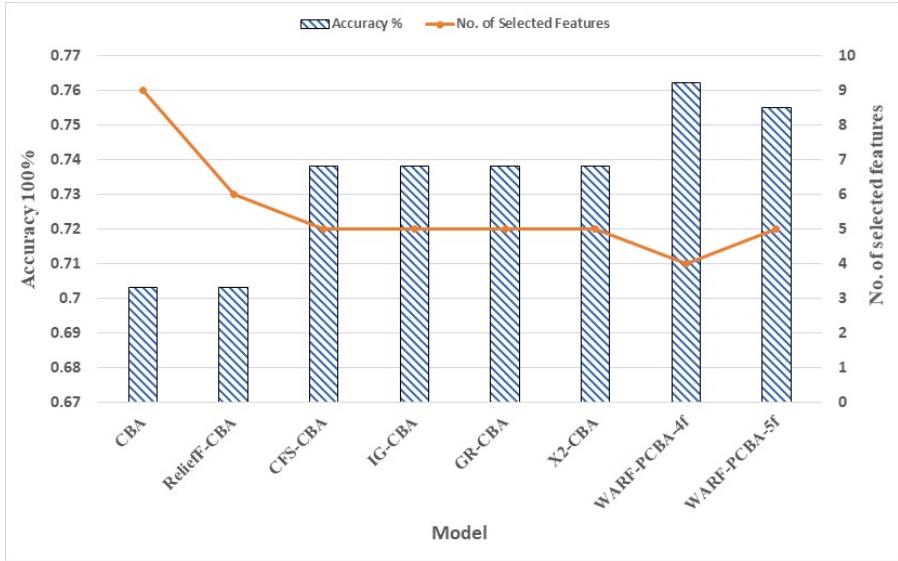


Fig. 3 Comparison of WARF-PCBA model in terms of selected features and accuracy of the breast cancer recurrences dataset.

The results in Table 4 are produced using Algorithm 1, where the experiments show the relative frequency of forwarding selection and backward elimination of the breast cancer diagnosis dataset when the weighted average is equal to 0.11 as a threshold for the selection or elimination features. Also, the sensitivity analysis is then performed to show the effectiveness of the selected features. The empirical results indicate that the yielded classification models using the selected features have produced impressive and consistent performance.

Fig 4 shows the performance results of the accuracy evaluation measure of the WARF-PCBA model with respect to the breast cancer diagnosis dataset. As indicated in Fig 4, the experimental results show that the WARF-PCBA model produced a better accuracy rate of 96.3%, when the number of selected features is 3, on the breast cancer diagnosis dataset, in comparison with the accuracy rate of 95.7% using the full feature set. The accuracy rate of WARF-PCBA is 96.3% for the breast cancer diagnosis dataset by removing clump thickness, marginal adhesion, single epithelial cell size, bland chromatin, normal nucleoli, and mitoses features. The results were based on a sensitivity analysis of the different models where the selected features are used to confirm the effectiveness of the proposed WARF method and WARF-PCBA model.

In the second part of the comparison, the proposed WARF-PCBA model, i.e., the integration of feature selection and AC, was also compared to CBA, ReliefF-CBA, CFS-CBA, IG-CBA, GR-CBA, X2-CBA models, all of which are evaluated using both the breast cancer recurrences and breast cancer diagnosis datasets. As illustrated in Figs. 3 and 4, the proposed WARF-PCBA model

Table 4 The relative frequency of forwarding selection and backward elimination of the breast cancer diagnosis dataset (Weighted Average=0.11).

Forward pass selection		Backward pass elimination	
Candidate features selection (selected feature=4)	Candidate features selection (selected feature=5)	Candidate features elimination (eliminated features=4)	Candidate features elimination (eliminated features=5)
f1 0.05	f1 0.08	f1 0.15	f1 0.16
f2 0.25	f2 0.2	f2 0	f2 0
f3 0.25	f3 0.2	f3 0	f3 0
f4 0	f4 0	f4 0.25	f4 0.2
f5 0.05	f5 0.12	f5 0.1	f5 0.16
f6 0.2	f6 0.2	f6 0	f6 0.04
f7 0.1	f7 0.12	f7 0.1	f7 0.12
f8 0.1	f8 0.08	f8 0.15	f8 0.12
f9 0	f9 0	f9 0.25	f9 0.2
Selected features: f2, f3, f6	Selected features: f2, f3, f5, f6, f7	Eliminated features: f1, f4, f8, f9	Eliminated features: f1, f4, f5, f7, f8, f9

produced better performance in terms of accuracy as compared with those of other baseline models.

The proposed WARF-PCBA model, i.e., the integration of feature selection and AC, was also used to compare with CBA, ReliefF-CBA, CFS-CBA, IG-CBA, GR-CBA, and X2-CBA models. These methods are all evaluated using both breast cancer recurrences and breast cancer diagnosis datasets. As shown in Fig 5, the proposed WARF-PCBA model has produced better performance in terms of accuracy as compared with those of all the baseline models.

In the third part of the comparison, Fig 6 shows that the hybrid PSO-WARF-PCBA model produced a better result with an accuracy rate of 75.9% on the breast cancer recurrences dataset in comparison with those of other models. Fig 6 shows the comparison between the proposed hybrid PSO-WARF-PCBA model and baseline models in terms of selected features and the classification evaluation measure for the breast cancer recurrences dataset. The results were based on a sensitivity analysis of the different models and selected features to confirm the effectiveness of the proposed PSO-WARF method and hybrid PSO-WARF-PCBA model. Fig 6 indicates that the graph showing the high accuracy score is produced where the number of features is three. The accuracy rate of the hybrid PSO-WARF-PCBA model is 75.9% for the breast cancer recurrences dataset by removing age, menopause, tumor-size, breast, breast-quad, and irradiat features.

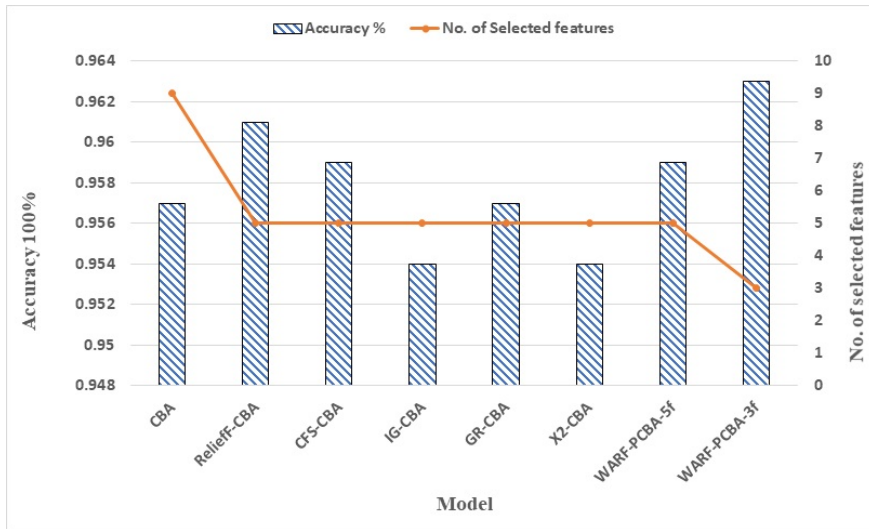


Fig. 4 Comparison of WARF-PCBA model in terms of selected features and accuracy of the breast cancer diagnosis dataset.

Furthermore, as shown in Fig 7, the hybrid PSO-WARF-PCBA model produced a better result with an accuracy rate of 97.3% on the breast cancer diagnosis dataset as compared with those of other models. Fig 7 illustrates the comparison between the proposed hybrid PSO-WARF-PCBA model and baseline models in terms of selected features and the classification evaluation measure with respect to the breast cancer recurrences dataset. The results were based on a sensitivity analysis of the different models where the selected features are used to test the effectiveness of the proposed PSO-WARF method and PSO-WARF-PCBA model. Fig 7 implies that the graph showing the high accuracy score is produced where the number of features is five. The accuracy rate of hybrid PSO-WARF-PCBA model is 97.3% for the breast cancer diagnosis dataset by removing `cell_size_uniformity`, `marginal_adhesion`, `single_epi_cell_size`, and `mitoses` features.

The proposed hybrid PSO-WARF-PCBA model, i.e., the integration of feature selection and AC, was also used to compare with Wrapper-Naive Bayes, Wrapper-Random Forest, Wrapper-K-Nearest Neighbor, Wrapper-REPTree, Wrapper-EnsembleSelection, Wrapper-Bagging, Wrapper-SVM, and Wrapper-CBA models. These models are all evaluated using both breast cancer recurrences and breast cancer diagnosis datasets. As shown in Fig 8, the proposed hybrid PSO-WARF-PCBA model has produced better performance in terms of accuracy as compared with those of all the benchmark models.

Figs. 9 and 10 show part of the classification rules generated by the hybrid PSO-WARF-PCBA model for breast cancer recurrences and diagnosis datasets. It is worth mentioning that these classification rules comprised the most significant features selected by the proposed PSO-WARF method.

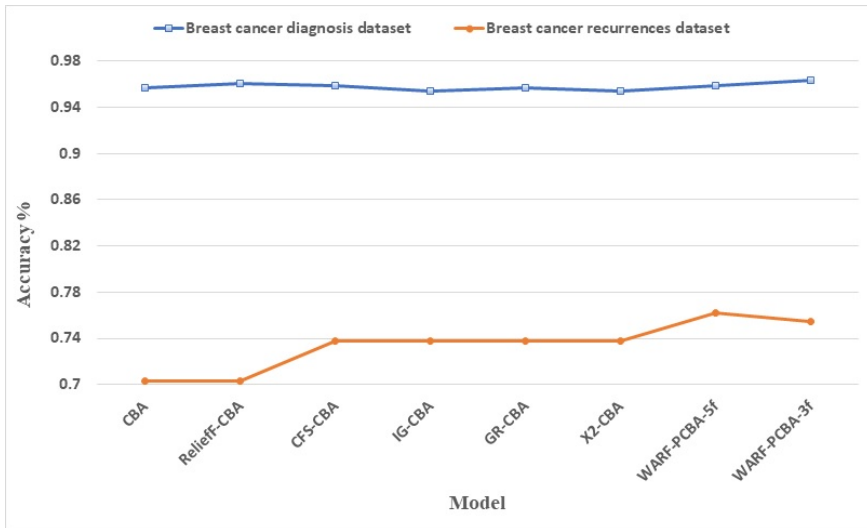


Fig. 5 The predictive accuracy rates of the proposed and all baseline models (CBA, Relief-CBA, CFS-CBA, IG-CBA, GR-CBA, X2-CBA, WARF-PCBA-4f, WARF-PCBA-5f) on both breast cancer recurrences and breast cancer diagnosis datasets.

The Wilcoxon rank sum test results shown in Tables 5 and 6 are used to compare the statistical significance between the proposed models, i.e., WARF-PCBA and hybrid PSO-WARF-PCBA, and other baseline models, for the breast cancer recurrences and diagnosis datasets, respectively. For the breast cancer recurrences dataset using WARF-PCBA, the comparison reveals that the accuracy produced by WARF-PCBA is statistically significantly better than the following baseline models: CBA, Relief-CBA, CFS-CBA, IG-CBA, GR-CBA, and X2-CBA. For the breast cancer diagnosis dataset using WARF-PCBA, the comparison reveals that the accuracy produced by WARF-PCBA is statistically significantly better than the following models: CBA, CFS-CBA, IG-CBA, GR-CBA, and X2-CBA. The only exception is for Relief-CBA, which achieves similar results as those of WARF-PCBA.

On the other hand, for the breast cancer recurrences dataset using hybrid PSO-WARF-PCBA, the comparison reveals that the accuracy produced by hybrid PSO-WARF-PCBA is statistically significantly better than the following models: Wrapper-Naive Bayes, Wrapper-Random Forest, Wrapper-K-Nearest Neighbor, Wrapper-REPTree, Wrapper-SVM, and Wrapper-CBA. At the same time, there is no difference in the prediction by hybrid PSO-WARF-PCBA and other baseline models such as Wrapper-EnsembleSelection and Wrapper-Bagging. For the breast cancer diagnosis dataset using hybrid PSO-WARF-PCBA. The comparison reveals that the accuracy produced by the hybrid PSO-WARF-PCBA model is statistically significantly better than the following models: Wrapper-Naive Bayes, Wrapper-Random Forest, Wrapper-K-Nearest Neighbor, Wrapper-REPTree,

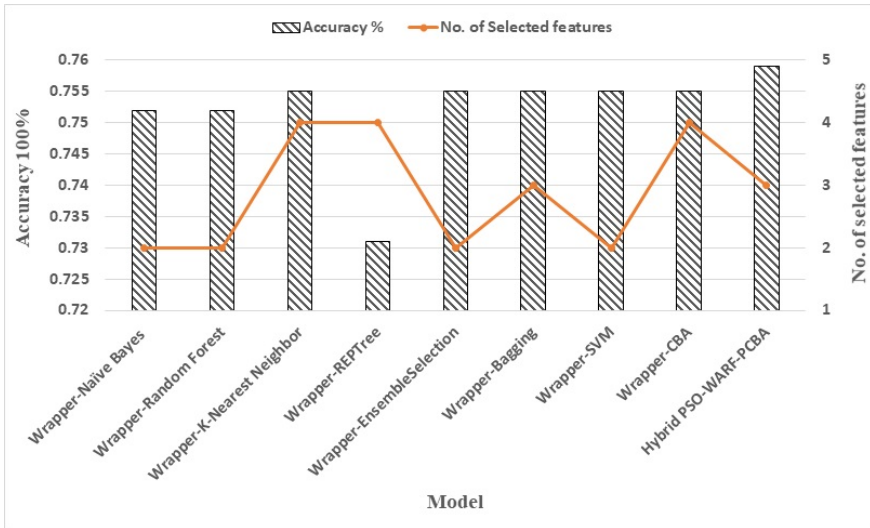


Fig. 6 Comparison of hybrid PSO-WARF-PCBA model in terms of selected features and accuracy of the breast cancer recurrences dataset.

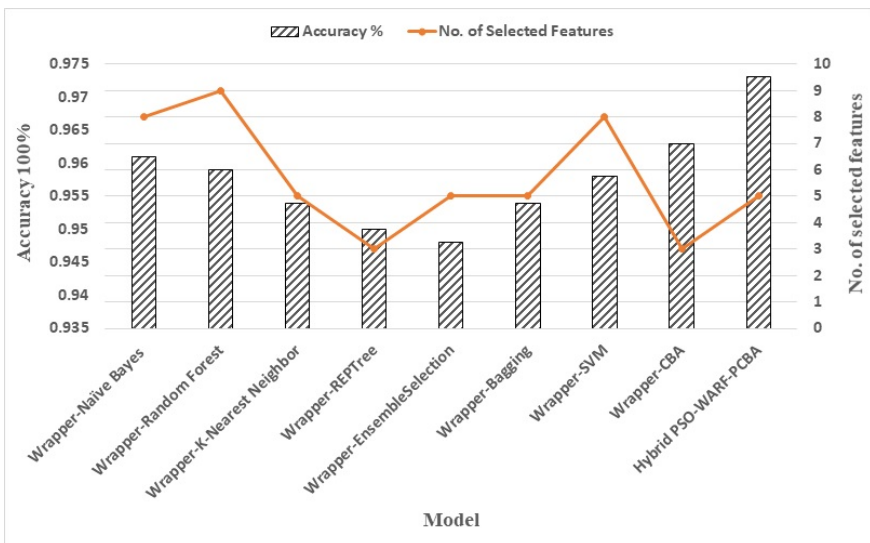


Fig. 7 Comparison of the hybrid PSO-WARF-PCBA model in terms of selected features and accuracy for the breast cancer diagnosis dataset.

Wrapper-EnsembleSelection, Wrapper-Bagging, Wrapper-SVM, and Wrapper-CBA. Most of the p values obtained from statistical analysis were less than 0.001, confirming that our proposed models significantly differed from other classification models. It is worth mentioning that the proposed models produce satisfactory results than those of other models in the comparisons mentioned above.

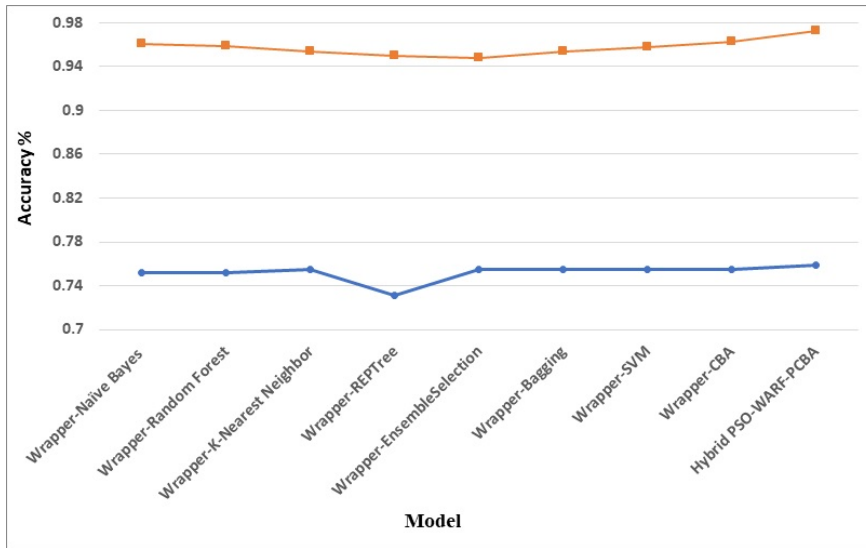


Fig. 8 The predictive accuracy rates of the proposed and benchmark models (i.e., Hybrid-PSO-WARF-PCBA, Wrapper-Naive Bayes, Wrapper-Random Forest, Wrapper-K-Nearest Neighbor, Wrapper-REPTree, Wrapper-EnsembleSelection, Wrapper-Bagging, Wrapper-SVM, and Wrapper-CBA) on both breast cancer recurrences and breast cancer diagnosis datasets.

- | |
|---|
| 1) Deg_Malig=2 → Class=no_recurrence_events |
| 2) Inv_Nodes=(0-2) → Class=no_recurrence_events |
| 3) Node_Caps=no → Class=no_recurrence_events |
| 4) Inv_Nodes=(3-5), Deg_Malig=1 → Class=no_recurrence_events |
| 5) Deg_Malig=1, Inv_Nodes=(0-2) → Class=no_recurrence_events |
| 6) Inv_Nodes=(12-14), Node-Caps=yes → Class=recurrence_events |
| 7) Inv_Nodes=(6-8), Deg_Malig=3 → Class=recurrence_events |
| 8) Node_Caps=yes, Deg_Malig=3 → Class=recurrence_events |

Fig. 9 Part of the classification rules generated by the hybrid PSO-WARF-PCBA model for the breast cancer recurrences dataset.

Furthermore, as shown in Figs. 8 and 11 and Tables 5 and 6, the proposed hybrid PSO-WARF-PCBA model performs better and significantly surpasses those obtained using the original subset features without feature selection compared to other models. From a clinical standpoint, essential diagnostic features (characteristics) for the breast cancer recurrences dataset include tumor-size, inv-nodes, node-caps, deg-malig, and irradiat, which are identified by our proposed model. The importance of these factors on breast cancer diagnosis from a clinical perspective is as follows: tumor-size and the amount of axillary lymph node involvement are two of the most important prognostic factors for breast cancer [47]. A breast cancer's stage can be determined by the size of the tumor, whether it has spread to nearby lymph nodes, whether it has reached distant

1) Cell_Shape_Uniformity=(1.9-2.8) → Class=benign
2) Cell_Shape_Uniformity=(7.3-8.2) → Class=malignant
3) Cell_Size_Uniformity=(4.6-5.5) → Class=malignant
4) Cell_Size_Uniformity=(6.4-7.3) → Class=malignant
5) Cell_Size_Uniformity=(7.3-8.2) → Class=malignant
6) Bare_Nuclei=(5.5-6.4) → Class=malignant
7) Bare_Nuclei=(8.2-9.1) → Class=malignant
8) Bare_Nuclei=(1.9-2.8), Cell_Size_Uniformity=(1.9-2.8) → Class=benign

Fig. 10 Part of the classification rules generated by the hybrid PSO-WARF-PCBA model for the breast cancer diagnosis dataset.

Table 5 Wilcoxon rank sum test of the proposed model and other models for breast cancer recurrences dataset.

WARF-PCBA versus	<i>p</i> -value	Significance	Hybrid PSO-WARF-PCBA versus	<i>p</i> -value	Significance
CBA	<0.00001	Yes	Wrapper-Naive Bayes	0.0346	Yes
ReliefF-CBA	<0.00001	Yes	Wrapper-Random Forest	0.0286	Yes
CFS-CBA	0.0399	Yes	Wrapper-K-Nearest Neighbor	0.0497	Yes
IG-CBA	0.0062	Yes	Wrapper-REPTree	0.0064	Yes
GR-CBA	0.0019	Yes	Wrapper-EnsembleSelection	0.116	No
X2-CBA	0.0104	Yes	Wrapper-Bagging	0.249	No
			Wrapper-SVM	0.0407	Yes
			Wrapper-CBA	0.0492	Yes

Table 6 Wilcoxon rank sum test of the proposed model and other models for breast cancer diagnosis dataset.

WARF-PCBA versus	<i>p</i> -value	Significance	Hybrid PSO-WARF-PCBA versus	<i>p</i> -value	Significance
CBA	0.0312	Yes	Wrapper-Naive Bayes	0.0001	Yes
ReliefF-CBA	0.0897	No	Wrapper-Random Forest	0.00035	Yes
CFS-CBA	0.0457	Yes	Wrapper-K-Nearest Neighbor	0.000048	Yes
IG-CBA	0.0039	Yes	Wrapper-REPTree	0.00002	Yes
GR-CBA	0.0303	Yes	Wrapper-EnsembleSelection	0.000015	Yes
X2-CBA	0.0049	Yes	Wrapper-Bagging	0.000046	Yes
			Wrapper-SVM	0.000143	Yes
			Wrapper-CBA	0.00062	Yes

regions of the body, and what cancer's biomarkers are [48]. Finding biomarkers to anticipate and track a tumor's response to radiation therapy is crucial for developing individualized radiotherapy treatment plans [49]. Breast cancer develops when abnormal cells in the breast expand and divide out of control, resulting in a mass of tissue known as a tumor. Feeling a lump in the breast, noticing a change in breast size, or noticing changes to the breast's skin can all indicate breast cancer. Since the important factor of deg-malig is to aid in the early identification of breast cancer for early-stage treatment, a

higher number means a higher level of malignancy. The recommendation is to use mammograms for diagnosing and screening the breast in the early stages. Mammography aims to detect breast cancer in its earliest stages by detecting unusual lumps or microcalcifications [50]. The innovative technique of irradiating regional lymph nodes, notably axillary lymph nodes, is one of the critical aspects. The significance of accelerated partial breast irradiation and the solution to the question of which patient population will benefit from partial breast radiation therapy are also crucial [51]. Such aforementioned important factors are selected by the proposed model in most test cases, which in turn help build an effective classifier to tackle breast cancer identification.

The same investigation has been conducted for the breast cancer diagnosis dataset, where the clinically important features such as cell size uniformity, cell shape uniformity, single epithelial cell size, bare nuclei, and bland chromatin features [4], are selected by the proposed hybrid PSO-WARF-PCBA model. As a result, our model is able to capture complex feature interaction and micro-effects of the clinically crucial feature subsets to inform breast cancer diagnosis. As indicated in Figs. 8 and 11, it can be inferred that the best results are obtained when applying the proposed model for both recurrences and breast cancer diagnosis datasets. It should be noted that, the experimental findings confirmed selecting the most significant features by the proposed model. In other words, the learning performance of the proposed hybrid PSO-WARF-PCBA model has been improved, owing to the selection of significant features that clinically participate in detecting breast cancer. As a result, the classification performance is greatly enhanced due to the selected discriminative feature representations as well as the elimination of insignificant attributes. Accordingly, a suitable picture of breast cancer diagnosis leads a physician for better understanding and interpreting of the results.

In contrast, the baseline method such as ReliefF only selects partial of the above clinically significant attributes, i.e., inv-nodes in the breast cancer recurrences dataset, and discards other crucial features such as tumor-size, node-caps, deg-malig, and irradiat. Therefore such inefficient selected feature subsets affect its model performance. Similarly, the baseline method GR also only regards and ranks the following features as important for the breast cancer recurrences dataset, i.e., node-caps, inv-nodes, deg-malig, irradiat, tumor-size, age, breast-quad, breast, and menopause, and overlooks characteristics with respect to other key factors, e.g., tumor-size and irradiat.

Since from clinical perspective, factors such as the size of the tumor, whether or not it has spread to lymph nodes or other distant parts of the body, are closely associated with the staging of the disease condition, as indicated in the above analysis, such crucial factors are often discarded or overlooked by the baseline methods, but well captured by the proposed model. Therefore, the above observation well indicates the micro-effects of the proposed PSO-based feature selection against those of the existing baseline methods.

In the fourth part of the comparison, the proposed models, namely WARF-PCBA and hybrid PSO-WARF-PCBA, were applied in the breast cancer

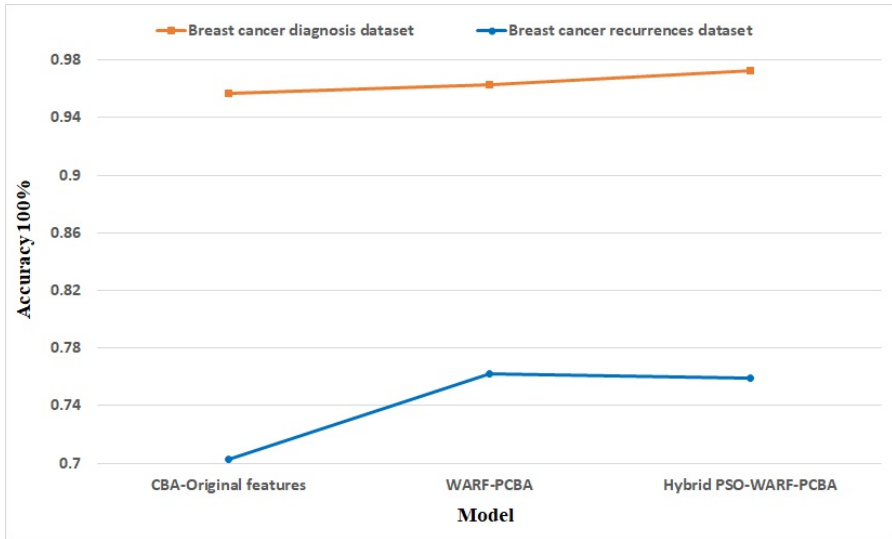


Fig. 11 The predictive accuracy rates of the proposed and CBA-Original features models (CBA-Original features, WARF-PCBA, Hybrid PSO-WARF-PCBA) on both breast cancer recurrences and breast cancer diagnosis datasets.

recurrences and breast cancer diagnosis datasets. The results of the proposed models were also compared with those of state-of-the-art AC algorithms. As depicted in Table 7, the 10-fold cross-validation performances for the proposed models outperformed various other AC algorithms with accuracy rates of 75.5% and 75.9%, respectively with respect to the breast cancer recurrences dataset. Meanwhile, for the breast cancer diagnosis dataset, the proposed WARF-PCBA and hybrid PSO-WARF-PCBA models outperformed other AC algorithms with an accuracy rates of 96.3% and 97.3%, respectively. These results indicate that the proposed models perform better than the other AC algorithms, particularly WCBA, since this algorithm eliminates some features based on an expert. In contrast, in the proposed models, the proposed feature selection methods, namely WARF and PSO-WARF, are used for feature selection. The resulting WARF-PCBA and hybrid PSO-WARF-PCBA models achieve outstanding results for the two classification tasks. The reason behind this is that the proposed models identify the essential features that have the most decisive influence on breast cancer classification while eliminating irrelevant features from the highest-ranked rule list generated during the rule pruning process.

5 Conclusion

In this research, we have proposed two feature selection methods, i.e., WARF-based filter and hybrid PSO-WARF methods, as well as two classification models, i.e., WARF-PCBA and hybrid PSO-WARF-PCBA, for breast cancer diagnosis. The proposed filter and wrapper feature selection methods are used

Table 7 Comparison of the proposed WARF-PCBA and hybrid PSO-WARF-PCBA models with the AC algorithms in terms of accuracy applied to both breast cancer datasets.

AC Algorithm	Breast cancer recurrences dataset	Breast cancer diagnosis dataset
CBA	0.703	0.957
CMAR	0.6279	0.912
MCAR	0.6744	0.941
FACA	0.6861	0.951
ECBA	0.6744	0.936
WCBA	0.7326	0.974
WARF-PCBA	0.755	0.963
Hybrid PSO-WARF-PCBA	0.759	0.973

to extract the most informative features in each dataset. The resulting classification models based on the combination of the AC approach and classification techniques are subsequently used for breast cancer prediction. In particular, the AC approach is able to find a correlation or association between attributes, while the classification technique is able to predict the class label.

Specifically, the proposed WARF-PCBA and hybrid PSO-WARF-PCBA classification models are developed based on the feature selection methods with the attempt to degrade the generation of a large number of class association rules. In other words, the WARF-PCBA and hybrid PSO-WARF-PCBA models employed the selected features yielded by the WARF and hybrid PSO-WARF methods.

Evaluated using two UCI data sets, the results demonstrated an enhancement in prediction accuracy using WARF-PCBA and hybrid PSO-WARF-PCBA models. Owing to the elimination of irrelevant and contradictory features, our proposed models outperformed a variety of baseline methods and AC techniques, and can be used as an alternative, reliable, and competitive methods for breast cancer diagnosis. Our study also indicates that the class association rule-based technique can be used as an emerging paradigm for classification tasks in other domains.

In future work, the efficiency of the proposed models will be further investigated using other UCI datasets with respect to classification and regression tasks. Moreover, different search algorithms and hyper-parameter optimization will also be explored to further enhance performance.

Acknowledgments. This work is supported by Deanship of Scientific Research and Graduate Studies at University of Petra, Amman, Jordan.

Conflict of Interest. The authors declare no conflict of interest.

Data Availability Statement. The datasets were taken from UCI public repository at <https://archive.ics.uci.edu/ml/datasets.php>

References

- [1] Aličković, E., Subasi, A.: Breast cancer diagnosis using ga feature selection

- and rotation forest. *Neural Computing and Applications* **28**(4), 753–763 (2017)
- [2] Biblowit, M.J.: Resources: Breast Cancer Statistics And Resources (2022). <https://www.bcrf.org/breast-cancer-statistics-and-resources> Accessed 2022-02-1
- [3] Wang, S., Wang, Y., Wang, D., Yin, Y., Wang, Y., Jin, Y.: An improved random forest-based rule extraction method for breast cancer diagnosis. *Applied Soft Computing* **86**, 105941 (2020)
- [4] Alwidian, J., Hammo, B.H., Obeid, N.: Wcba: Weighted classification based on association rules algorithm for breast cancer disease. *Applied Soft Computing* **62**, 536–549 (2018)
- [5] Sowan, B.: A comparative analysis of exam timetable using data mining techniques. *IJCSNS* **17**(1), 73 (2017)
- [6] Sowan, B., Qattous, H.: A data mining of supervised learning approach based on k-means clustering (2017)
- [7] Sowan, B., Dahal, K., Hossain, M.A., Zhang, L., Spencer, L.: Fuzzy association rule mining approaches for enhancing prediction performance. *Expert Systems with Applications* **40**(17), 6928–6937 (2013)
- [8] Liu, B., Hsu, W., Ma, Y., *et al.*: Integrating classification and association rule mining. In: *Kdd*, vol. 98, pp. 80–86 (1998)
- [9] Abdelhamid, N., Thabtah, F.: Associative classification approaches: review and comparison. *Journal of Information & Knowledge Management* **13**(03), 1450027 (2014)
- [10] Sowan, B.I., Dahal, K.P., Hossain, A.M., Alam, M.S.: Diversification of fuzzy association rules to improve prediction accuracy. In: *International Conference on Fuzzy Systems*, pp. 1–8 (2010). IEEE
- [11] Thabtah, F., Cowling, P., Peng, Y.: Mcar: multi-class classification based on association rule. In: *The 3rd ACS/IEEE International Conference on Computer Systems and Applications*, 2005., p. 33 (2005). IEEE
- [12] Kumar, P.M., Lokesh, S., Varatharajan, R., Babu, G.C., Parthasarathy, P.: Cloud and iot based disease prediction and diagnosis system for health-care using fuzzy neural classifier. *Future Generation Computer Systems* **86**, 527–534 (2018)
- [13] Venkatesan, C., Karthigaikumar, P., Paul, A., Satheeskumaran, S., Kumar, R.: Ecg signal preprocessing and svm classifier-based abnormality

- detection in remote healthcare applications. *IEEE Access* **6**, 9767–9773 (2018)
- [14] Kim, S.G., Theera-Ampornpunt, N., Fang, C.-H., Harwani, M., Grama, A., Chaterji, S.: Opening up the blackbox: an interpretable deep neural network-based classifier for cell-type specific enhancer predictions. *BMC systems biology* **10**(2), 243–258 (2016)
- [15] Han, J., Kamber, M., Pei, J.: Data mining concepts and techniques third edition. The Morgan Kaufmann Series in Data Management Systems **5**(4), 83–124 (2011)
- [16] Fong, R.C., Vedaldi, A.: Interpretable explanations of black boxes by meaningful perturbation. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3429–3437 (2017)
- [17] Quinlan, J.: "c4. 5: Programs for machine learning", morgan kaufmann publishers, san mateo, ca (1993)
- [18] Jensen, D.D., Cohen, P.R.: Multiple comparisons in induction algorithms. *Machine Learning* **38**(3), 309–338 (2000)
- [19] Chien, Y.-W.C., Chen, Y.-L.: Mining associative classification rules with stock trading data—a ga-based method. *Knowledge-Based Systems* **23**(6), 605–614 (2010)
- [20] Yin, X., Han, J.: Cpar: Classification based on predictive association rules. In: Proceedings of the 2003 SIAM International Conference on Data Mining, pp. 331–335 (2003). SIAM
- [21] Thabtah, F.A., Cowling, P., Peng, Y.: Mmac: A new multi-class, multi-label associative classification approach. In: Fourth IEEE International Conference on Data Mining (ICDM'04), pp. 217–224 (2004). IEEE
- [22] Veloso, A., Meira, W., Zaki, M.J.: Lazy associative classification. In: Sixth International Conference on Data Mining (ICDM'06), pp. 645–654 (2006). IEEE
- [23] Li, W., Han, J., Pei, J.: Cmar: Accurate and efficient classification based on multiple class-association rules. In: Proceedings 2001 IEEE International Conference on Data Mining, pp. 369–376 (2001). IEEE
- [24] Lakshmanaprabu, S., Mohanty, S.N., Krishnamoorthy, S., Uthayakumar, J., Shankar, K., *et al.*: Online clinical decision support system using optimal deep neural networks. *Applied Soft Computing* **81**, 105487 (2019)

- [25] Ahmed, H., Younis, E.M., Hendawi, A., Ali, A.A.: Heart disease identification from patients' social posts, machine learning solution on spark. *Future Generation Computer Systems* **111**, 714–722 (2020)
- [26] Shao, Y., Liu, B., Wang, S., Li, G.: A novel software defect prediction based on atomic class-association rule mining. *Expert Systems with Applications* **114**, 237–254 (2018)
- [27] Gao, K., Khoshgoftaar, T.M., Wang, H., Seliya, N.: Choosing software metrics for defect prediction: an investigation on feature selection techniques. *Software: Practice and Experience* **41**(5), 579–606 (2011)
- [28] Cortes, C., Vapnik, V.: Support-vector networks. *Machine learning* **20**(3), 273–297 (1995)
- [29] Thabtah, F.A.: A review of associative classification mining. *Knowledge Engineering Review* **22**(1), 37–65 (2007)
- [30] Padillo, F., Luna, J.M., Ventura, S.: Evaluating associative classification algorithms for big data. *Big Data Analytics* **4**(1), 1–27 (2019)
- [31] Hall, M.A., Holmes, G.: Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data engineering* **15**(6), 1437–1447 (2003)
- [32] Hira, Z.M., Gillies, D.F.: A review of feature selection and feature extraction methods applied on microarray data. *Advances in bioinformatics* **2015** (2015)
- [33] Kou, G., Yang, P., Peng, Y., Xiao, F., Chen, Y., Alsaadi, F.E.: Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods. *Applied Soft Computing* **86**, 105836 (2020)
- [34] Jadhav, S., He, H., Jenkins, K.: Information gain directed genetic algorithm wrapper feature selection for credit rating. *Applied Soft Computing* **69**, 541–553 (2018)
- [35] Thaseen, I.S., Kumar, C.A.: Intrusion detection model using fusion of chi-square feature selection and multi class svm. *Journal of King Saud University-Computer and Information Sciences* **29**(4), 462–472 (2017)
- [36] Scheffer, T.: Finding association rules that trade support optimally against confidence. In: *European Conference on Principles of Data Mining and Knowledge Discovery*, pp. 424–435 (2001). Springer
- [37] Paul, D., Jain, A., Saha, S., Mathew, J.: Multi-objective pso based online

- feature selection for multi-label classification. *Knowledge-Based Systems* **222**, 106966 (2021)
- [38] Srisukkham, W., Zhang, L., Neoh, S.C., Todryk, S., Lim, C.P.: Intelligent leukaemia diagnosis with bare-bones pso based feature optimization. *Applied soft computing* **56**, 405–419 (2017)
- [39] Xue, B., Zhang, M., Browne, W.N., Yao, X.: A survey on evolutionary computation approaches to feature selection. *IEEE Transactions on Evolutionary Computation* **20**(4), 606–626 (2015)
- [40] Sakri, S.B., Rashid, N.B.A., Zain, Z.M.: Particle swarm optimization feature selection for breast cancer recurrence prediction. *IEEE Access* **6**, 29637–29647 (2018)
- [41] Hadi, W., Al-Radaideh, Q.A., Alhawari, S.: Integrating associative rule-based classification with naïve bayes for text classification. *Applied Soft Computing* **69**, 344–356 (2018)
- [42] Tan, T.Y., Zhang, L., Neoh, S.C., Lim, C.P.: Intelligent skin cancer detection using enhanced particle swarm optimization. *Knowledge-based systems* **158**, 118–135 (2018)
- [43] Xie, H., Zhang, L., Lim, C.P., Yu, Y., Liu, H.: Feature selection using enhanced particle swarm optimisation for classification models. *Sensors* **21**(5), 1816 (2021)
- [44] Berrar, D.: Cross-validation. In: Ranganathan, S., Gribskov, M., Nakai, K., Schönbach, C. (eds.) *Encyclopedia of Bioinformatics and Computational Biology*, pp. 542–545. Academic Press, Oxford (2019)
- [45] Raschka, S.: Model evaluation, model selection, and algorithm selection in machine learning. arXiv preprint arXiv:1811.12808 (2018)
- [46] Dua, D., Graff, C.: UCI Machine Learning Repository, University of California, School of Information and Computer Science, Irvine, CA, 2019 (2019)
- [47] Carter, C.L., Allen, C., Henson, D.E.: Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer* **63**(1), 181–187 (1989)
- [48] Editor in Chief Jyoti D. Patel, F. MD: Breast Cancer: Stages (09/2021). <https://www.cancer.net/cancer-types/breast-cancer/stages#tnm-staging> Accessed 2022-08-14
- [49] Meehan, J., Gray, M., Martínez-Pérez, C., Kay, C., Wills, J.C., Kunkler,

- I.H., Dixon, J.M., Turnbull, A.K.: A novel approach for the discovery of biomarkers of radiotherapy response in breast cancer. *Journal of personalized medicine* **11**(8), 796 (2021)
- [50] Professional, C.C.M.: Breast Cancer (01/21/2022). <https://my.clevelandclinic.org/health/diseases/3986-breast-cancer> Accessed 2022-08-14
- [51] Haussmann, J., Corradini, S., Nestle-Kraemling, C., Bölke, E., Njanang, F.J.D., Tamaskovics, B., Orth, K., Ruckhaeberle, E., Fehm, T., Mohrmann, S., *et al.*: Recent advances in radiotherapy of breast cancer. *Radiation oncology* **15**(1), 1–10 (2020)