International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 6 Issue 7 ǁ July. 2018 ǁ PP.80-101

Diagnosis of Breast Cancer Using Cwv-Bann-Svm

Madhusmita Mohanty, Pradip Kumar Sahoo, Ipsita Samal, Sulochana Nanda
Department of Electronics and Communication Engineering, NM Institute of Engineering and Technology,Bhubaneswar , Odisha
Department of Electronics and Communication Engineering, Raajdhani Engineering College,Bhubaneswar,Odisha
Department of Electronics and Communication Engineering,Aryan Institute of Engineering and Technology Bhubnaeswar , Odisha
Department of Electronics and Communication Engineering,Capital Engineering College,Bhubaneswar,Odisha

ABSTRACT
This paper presents a new data mining technique for an accurate prediction of breast cancer (BC), which is one of the major mortality causes among women around the globe. The main objective of our study is to expand an automatic expert system (ES) to provide an accurate diagnosis of BC. Both, Support Vector Machines (SVMs) and Artificial Neural Networks (ANNs) were applied to analyze BC data. The well-known Wisconsin Breast Cancer Dataset (WBCD), available in the UCI repository, was examined in our study. We first tested the SVM algorithm using various values of the C, e and c parameters. As a result of the first experiment, we were able to observe that the adjustment of these regularization parameters can greatly improve the performance of the traditional SVM algorithm applied for BC detection. The high-est obtained accuracy at the first step was 99.71%. Then, we performed a new BC detection approach based on two ensemble learning techniques: the confidence-weighted voting method and the boosting ensemble technique. Our model, called CWV-BANNSVM, combines boosting ANNs (BANN) and two SVMs, using optimal parameters selected during the first experiment. The performance of the applied methods was evaluated using several popular metrics, such as specificity, sensitivity, precision, FPR, FNR, F_1 score, AUC, Gini and accuracy. The proposed CWV-BANNSVM model was able to improve the per-formance of the traditional machine learning algorithms applied to BC detection, reaching the accuracy of 100%. To overcome the overfitting issue, we determined and used some appropriate parameter values of polynomial SVM. Our comparison with the existing studies dedicated to BC prediction suggests that the proposed CWV-BANN-SVM model provides one of the best prediction performances overall.

Keywords:
Data mining
Machine learning
Ensemble technique
Breast cancer
Support vector machine
Artificial neural network

I. INTRODUCTION

Nowadays, we are facing various diseases and cancers that are common between men and women. However, some of these dis-eases and cancers are much more common for only one gender, such as Breast Cancer (BC) among women (but rarely among men) or Prostate cancer among men. According to [1], BC is one of the main cancers in both developed and developing countries. On the other hand, in both the United States of America (USA) and Asia, BC is the second leading cause of death for women. In another study [2], BC is described as one of the most common causes of death, and the fifth leading cause of death among women. Abdel-Zaher and Eldeib [3] reported that BC includes 18.3% of all cancer types in Egypt. All these studies show the signif-icant importance of the BC cancer research to decline the number of deaths as well as to help patients who suffer from BC. Treatment and surgical costs of BC are very high. Finding appropriate computer-based solutions for its early detection can decrease these costs, which is helpful for both patients and governments.

The presence of an abnormal multiplication of cells in the breast tissue is the main cause of BC [4]. The ammography is one of the most common methods for an early detection of BC. According to the American College of Radiology (ACR), Breast Image Reporting and Data System (BI-RADS) was introduced as a trademark to clas-sify the results obtained by mammograms into four major groups that later on was increased to 6 main groups [5]. Although the mammography gives useful information about BC, based on the BI-RADS method, the biopsy is also required to identify the malig-nant or benign tumor. Data mining and machine

learning methods can be also used to help doctors in their decision making regarding different types of cancers, including BC.

Data mining and machine learning are new and powerful solu-tions computer-based solutions to discover hidden relationships in complex datasets [6]. Indeed, often raw datasets available from dif-ferent sources (e.g., medical science, transportation, news, social media, weather, etc.) have useful information which traditional data classification approaches cannot retrace. In other words, although traditional and manual solutions might reveal some
latent information, their application usully takes longer time and often includes human mistakes. Providing a reliable and trustwor-thy predictive model with the highest precision and accuracy is the main target for data mining and machine learning researchers. The exploration of medical data is a very important issue due to its close relationship to individual's life. Therefore, the proposed mod-els should have the lowest error rates in terms of diagnosis and treatment.

In this paper, an SVM machine learning model with various parameters, such as Regularization parameter (C), Polynomial ker-nel function and Gamma was first applied for an accurate detection of BC. To that end, the Wisconsin Breast Cancer Dataset (WBCD) was considered (see Section 3.3 for more information about this dataset). In our first experiment, we applied SVM with different values of the C, e and c parameters in Polynomial kernel function to obtain appropriate outcomes. This experiment provided us with some appropriate values of the C, e and c parameters. In the second experiment, two SVMs with suitable parameters for Multilayer perceptron (MLP) and Radial basis function (RBF) artificial neural networks were combined by using the Confidence-weighted voting approach. The accuracy of ANN (MLP and RBF) was increased using the boosting technique. This method is called boosting ANN (BANN). Furthermore, 16 records with missing values in our BC dataset were removed and the new prediction model, called CWV-BANN-SVM, was applied on this transformed BC dataset. Our results indicate that in terms of prediction, the proposed methodology could significantly improve the performance of the final model for diagnosing BC. The proposed model reaches the highest possible accuracy of 100% for the Wisconsin Breast Cancer data.

The rest of the paper is organized as follows. In Section 2, we briefly discuss some related work in the field. Section 3 provides information about the machine learning algorithms being applied and the Wisconsin Breast Cancer data. The proposed methodology is presented by Section 4. In Section 5, we report the experimental results obtained using the proposed model. Finally, Section 6 pro-vides some conclusions of this work.

II. RELATED WORK

Nowadays, we can find several studies addressing the issue of an early BC detection using various machine learning and data mining algorithms. Each of the previously proposed methods has its own strengths and weaknesses. Here, we discuss the related work in the field.

Abdel-Zaher and Eldeib [3] used the Levenberg-Marquardt method for training in the BPNN (back propagation neural net-work) approach with the DBN-NN (deep belief network path) algo-rithm. The best accuracy found by DBN-NN was 99.68%. In the paper [7] by Quinlan, the best accuracy of BC prediction was 94.74% through utilizing 10-fold cross validation with the C4.5 algorithm. In [8], a rule induction algorithm based on the approx-imate classification model was applied and the best accuracy found was 94.99%. In another paper by Ster and Dobnikar [9], linear dis-criminant analysis (LDA) and neural networks (NN) were used to deal will the BC prediction problem. Based on their results, the classification accuracy was 96.80% using a hybrid method (LDA

+ NN). Bennett and Blue [10] applied the SVM algorithm to predict BC cases. The best accuracy of SVM they found was 97.20%. A Neuro-Fuzzy method was used by Nauck and Kruse [11] to deal with this challenging problem. According to their results, the accu-racy of 95.06% was obtained when the 10-folds cross-validation approach was applied. In [12], still using the Fuzzy-GA approach, the accuracy of classification model was 97.36%.

Setiono [13], utilizing a feed forward neural network rule extraction, reported the BC prediction accuracy of 98.10%. In this study, 50% of dataset were used for training and the other 50% of data for testing. In 2002, Goodman et al. [14] addressed the prob-lem of BC detection using three machine learning methods, includ-ing Optimized-LVQ, Big-LVQ and AIRS. The best accuracies they found were 96.70%, 96.80% and 97.20%, respectively. Albrecht et al. [15] presented a new method of BC prediction by combining the Logarithmic Simulated Annealing (LSA) algorithm with the per-ceptron approach. Based on their results, the best accuracy for LSA with perceptron was 98.80%. In [16], the Supervised Fuzzy Cluster-ing (SFC) method was used to study BC data. The best reported accuracy here was 95.57%. Ubeyli [17] used the mixture experts (ME) network structure to predict BC and the highest accuracy obtained was 98.85%, where the dataset was

divided into two main parts as follows: 37% for training and 63% for testing. Polat and Günes [18] used the Least Square Support Vector Machine (LS-SVM) method as well as the 10-fold cross validation approach to obtain the BC prediction accuracy of 98.53%. In [19], the authors applied various machine learning methods, such as probabilistic neural network (PNN), combined neural network (CNN), multilayer perceptron neural network (MLPNN), recurrent neural network (RNN) and SVM, and found that the best BC prediction accuracy of 99.54% was provided by the SVM algorithm.

In another paper by Akay [20], an SVM-based algorithm with feature selection was used. The obtained accuracy for the proposed method was 99.51% when five BC features were considered. Kara-batak and Ince [21] applied the Association Rules (AR) and neural network (NN) methods with the 3-fold cross validation to detect BC cases. The classification accuracy of their approach was 97.40%. Paulin and Santhakumaran [22] applied different methods to process BC data, including the Batch Gradient Descent with Momentum (BGDM), Batch Gradient Descent (BGD), Resilient Back Propagation (RBP), Quasi Newton (QN), Levenberg Marquardt (LM) and Conjugate Gradient (CG) methods. According to their results, the Feed Forward Neural Network (FFNN) based on the Back Prop-agation algorithm (BP) using the Levenberg Marquardt (LM) method for training had the best accuracy, of 99.28%, compared to the other methods. Dheeba et al. [23] introduced a new Particle Swarm Optimized Wavelet Neural Network (PSOWNN) method for diagnosing BC as a computer aided diagnosis (CAD) system. The best accuracy found using PSOWNN was 93.67%. In [24], the Naïve Bayesian (NB) method was combined with a weighting approach and the new method named weighted-Naïve Bayesian (W-NB) with the 5-folds cross-validation for training and testing. The BC prediction accuracy found in that study was 98.54%.

In the paper by Nahato et al. [25], the rough set indiscernibility relation algorithm using a back-propagation neural network (RS-BPNN) was carried out and the best prediction accuracy found was 98.60% when Reduct was equal to R1. In another study [26], the K-Nearest Neighbor (KNN), ANN, radial basis function neural network (RBFNN), and the SVM methods were applied to classify BC data. The Independent Component Analysis (ICA), (with 5-fold and 10-fold cross-validations) and partitioning (20%–40%) RBFNN provided the better accuracy, of 90.49%, than the other competing methods. Turabieh and Muhanna [27] used the ANFIS method and a genetic algorithm (GA) to identify the most effective features for BC prediction. The highest prediction accuracy reported in [27] was 71%. Kumari and Singh [28] obtained the highest accuracy, of 99.28%, using the KNN algorithm. In [29], various methods of BC prediction were compared, including the SVM, Decision Tree (C4.5), Naive Bayes (NB) and KNN algorithms. The highest accuracy of 97.13% was obtained using the SVM algorithm. Moreover, the 10-folds cross-validation approach in the WEKA tool was applied as one of the open source data mining tools. Kaushik and Kaur
[5] applied four well-known machine learning algorithms, includ-ing MLP, Random Forest (RF), Random Tree (RT) and Ensemble Classifier (EC). In this study, Ensemble Classifier with the accuracy of 83.50% overpowered the other methods. In addition, the ROC area measured for the EC method in this study was 0.907 out of 1.

In [30], the GA and RF algorithms were applied to detect BC. Here, the GA-based Rotation Forest method applied on 14 BC fea-tures had the best prediction accuracy of 99.48%. Also, a new intelligent model based on a two-phase classification method was presented by Dhull and Gupta [31]. The model was applied on 12 well-known UCI datasets. The obtained outcomes indicated that this two-phase model had the accuracy of 89.70% when the Wisconsin Diagnosis Breast Cancer (WDBC) dataset was consid-ered. Dora et al. [32] proposed a new method called Gauss-Newton representation-based algorithm (GNRBA). It was applied on two types of datasets: the WBCD and WDBC data. The perfect accuracy of 100% was obtained for the WBCD data by the GNRBA algorithm.

III. METHODS AND DATA

This section is divided into 4 major sub-sections, including the description of SVMs, the description of the Polynomial Kernel func-tion, the description of artificial neural networks, and finally, the description of the WBCD dataset.

3.1. Support vector machine algorithm and its linear version

SVM is a well-known binary classification method used in data mining and machine learning [33,34]. Linear SVM can be used as an appropriate solution dealing with linearly separable binary clas-sification problems as it can be observed in the example of Fig. 1 from [35].

Indeed, in the SVM algorithm, the input variables are converted into a high-dimensional linear feature space by using a non-linear transformation. As a result, the optimal decision function is con-structed [36]. Then, by using the kernel function located in the original space, the dot product operation in the higher dimensional feature space is replaced and the regression function is calculated according to Eq. (1):

f ðxÞ ¼ W:Uð x Þ > þb ð1Þ
where W represents the weight vector, Uðx Þ represents the nonlin-ear mapping from the input area to the output area, and b represents the bias. The forecasting function is calculated according to Eq. (2):

$$\text{Rmin} \frac{1}{2k} \|W\|^2 + \sum_{i=1}^{m} C x_i \quad (2)$$

Here the constraint conditions can be calculated as follows (Eq.

(3)):

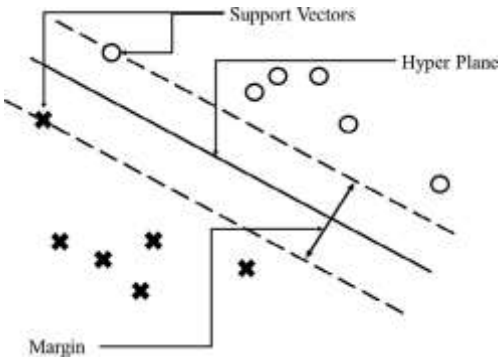

Fig. 1. Solving a linearly separable binary classification problem with the linear SVM algorithm [35].

$$\begin{cases} f(x_i, w, b) - y_i \le \varepsilon + \eta_i, \\ y_i - f(x_i, w, b) \le \varepsilon + \eta_i^*, \\ \eta_i, \eta_i^* \ge 0, i = 1, 2, 3, \ldots, n. \end{cases} \quad (3)$$

where the description of parameters in the Eqs. (2) and (3) are as follows:

C is a penalty factor;
η_i, η_i^* are relaxation factors;

y_i is the class label for the ith training sample;

m is the number of classes of the training samples; l is the number of the training samples.

In the last step, the Lagrange multiplier is introduced and according to the Wolf duality theory, it is changed into an equiva-lent dual problem using Eq. (4):

$$\min \frac{1}{2} (a - a^*)^T Q (a - a^*) + \varepsilon \sum_{i=1}^{l} (a_i + a_i^*) + \sum_{i=1}^{l} y_i (a_i - a_i^*) \quad (4)$$

Eq. (5) gives information about the constraint conditions:
(

$_i$l ¼$_1^ð$ai a_i Þ ¼ 0 5
P ai a_i ¼1 2 3 ...1
ð Þ
0 ; C; I ; ; ;

where a_i are the coefficients for training samples (zero for non-support vectors) and a_i are the coefficients for training samples for the support vector regression (SVR) model. By using the quadra-tic programming, the SVM regression prediction model can be described according to Eq. (6):

l
f x; a_i ; a_iX
¼ a_i a_i k ðx_i; xÞ þ b; ð6Þ
i¼1

where kð x $_i$; xÞ is considered as the kernel function of the SVM algo-rithm [34]. More information about the SVM algorithm can be found in [33–36].

3.2. Polynomial kernel function and regularization parameter in SVM

According to [37], four main kernel SVM functions are as fol-lows: Linear Function (LKF), Polynomial Function (PF), Radial Basis Function (RBF) and Sigmoid Function (SF). In this study, the PF was used as one of the main kernel functions of the SVM algorithm [38]. The polynomial kernel function can be calculated using Eq. (7):

d
k x_i; x_j ¼ cx^T_i :x_j þ r ð7Þ

Gamma (c) is one of the main adjustable parameters in the polynomial kernel function of SVM. Although, this parameter improves the accuracy of the SVM algorithm, its misuse can lead to the overfitting issue. Thus, to deal with this problem, we used the regularization parameter (C) that helps overcome the overfitting problem. We select both c and C at the same time. Indeed, the regularization parameter (C) balances the trade-off between the margin and the accuracy. To monitor the width of the e-insensitive zone, the e parameter can be also used [34,39].

3.3. Artificial neural networks

The Artificial Neural Networks (ANN) algorithms (sometimes named Neural Networks) is a well-known machine learning method to deal with complex pattern-oriented issues of both time-series and categorization types of data [40]. Basically, the nonparametric version of neural networks allows the methods to be developed without having any previous information or knowledge on the distribution of the data population or probable interaction effects among different variables. An ANN algorithm includes three main layers including input layer, hidden layer(s), and output layer. ANNs try to prepare a mapping between the input layer (input space) and the output layer (optimal space) by understanding intrinsic relationships between different features [41]. The hidden layer(s) investigates the received information from the input layer and then transfers it to the output layer. Even though there are several types of ANNs, this study uses only two of them: the MLP and RBF networks.

3.3.1. Multilayer perceptron (MLP) network

An MLP-based network is one of commonly-used network architectures. The structure of MLP includes one input layer, one or more hidden layers, as well as only one output layer. An MLP network can be trained using different algorithms, such as back-propagation algorithm. This training method decreases the error rate by using weights and bias adjustments. In this study, an MLP neural network with two hidden layers was used. Since the BC dataset we considered had nine features and two classes, the MLP has an input layer with 9 neurons and an output layer with 2 neurons. The used training structure of the MLP classifier for detection of BC was 9-9-9-2. An MLP neural network with an input layer, n hidden layers and one output layer is shown in Fig. 2.

3.3.2. Radial basis function (RBF) network

The RBF neural network is another widely-used architecture of neural networks algorithms [43]. It uses the RBFs as activation functions of an ANN algorithm. RBFs are very effective to deal with different problems such as multi-dimensional problems [44]. Unlike the MLP architecture, RBFs generally have only one hidden layer. A schematic structure of an RBF neural network is shown in Fig. 3.

3.4. Wisconsin breast cancer dataset

As discussed before, BC is one of the major reasons of death of women around the world. Though, finding some accurate diagnos-ing and treatment solutions (using both clinical and computer-based methods) is absolutely indispensable. To deal with this mat-ter, the WBCD dataset from the University of California, Irvine (UCI) Machine Learning Repository was selected. This dataset has been collected by Dr. William H. Wolberg at the University of Wisconsin-Madison Hospitals from 1989 to 1991 [45]. This dataset includes 699 records divided into two classes (class 2 for benign BC cases (negative class) and class 4 for malignant BC cases (positive class)). Ten features (i.e., attributes or variables) were used to describe the records. Note that the first feature (sample code num-ber: id number), which does not give useful information for BC pre-

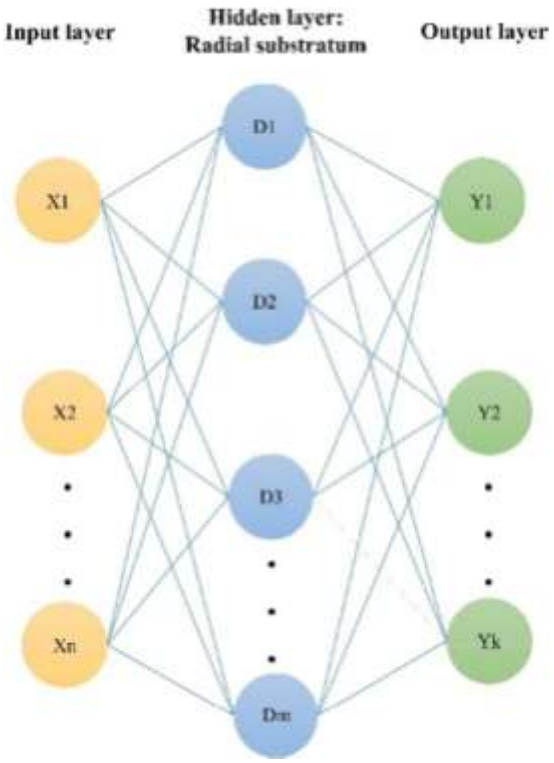

Fig. 3. A schematic structure of an RBF neural network [44].

diction, was removed from the dataset. In total, 458 records belong to class 2 and 241 records belong to class 4. Some detailed informa-tion about the WBCD dataset is presented in Table 1.

IV. PROPOSED CWV-BANN-SVM MODEL

In this section, we describe our new methodology which incor-porates different machine learning algorithms and applies them to BC data. The proposed model, called CWV-BANN-SVM, includes the SVM and boosting artificial neural network (BANN) algorithms. Moreover, the model consists of boosting and confidence-weighted voting ensemble techniques. A valuable point is that our model includes two voting ensemble techniques which help improve the performance of the final model. In addition, we iden-tify the most important attributes of our BC dataset and the opti-mal values of some important parameters of our SVM algorithm.

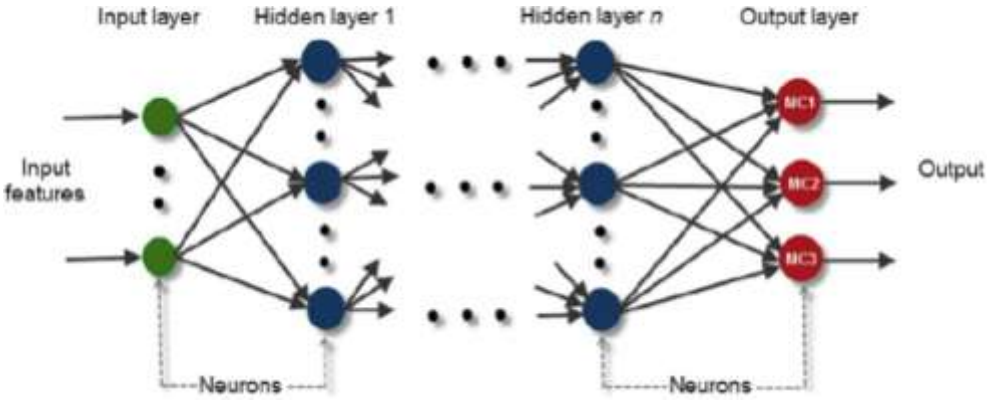

Fig. 2. A schematic view of an MLP neural network [42].

Table 1
Description of the attributes (features or variables) of the WBCD dataset.

Feature Number	Features description	Values of attributes	Mean	Standard error of mean	Standard deviation	Variance
1	Clump thickness	1–10	4.418	0.107	2.816	7.928
2	Uniformity of cell size	1–10	3.134	0.115	3.051	9.311
3	Uniformity of cell shape	1–10	3.207	0.112	2.972	8.832
4	Marginal adhesion	1–10	2.807	0.108	2.855	8.153
5	Single epithelial cell size	1–10	3.216	0.084	2.214	4.903
6	Bare nuclei	1–10	3.464	0.138	3.641	13.255
7	Bland chromatin	1–10	3.438	0.092	2.438	5.946
8	Normal nucleoli	1–10	2.867	0.115	3.054	9.325
9	Mitoses	1–10	1.589	0.065	1.715	2.941

In other words, the current work tries to improve the results found in previous studies dedicated to BC prediction.

Indeed, the main part of our study is related to the second experiment in which all the methods are combined using the Confidence-Weighted Voting (CWV) technique. The block diagram of the proposed CWV-BANN-SVM model is shown in Fig. 4. More details about the CWV-BANN-SVM model are given below (see Section 4.1).

4.1. Model description

The methods used in the first experiment was carried out using IBM SPSS Modeler 14.2 on a PC computer equipped with an Intel Core i7 processor and 8 GB of RAM running under Windows 8.1

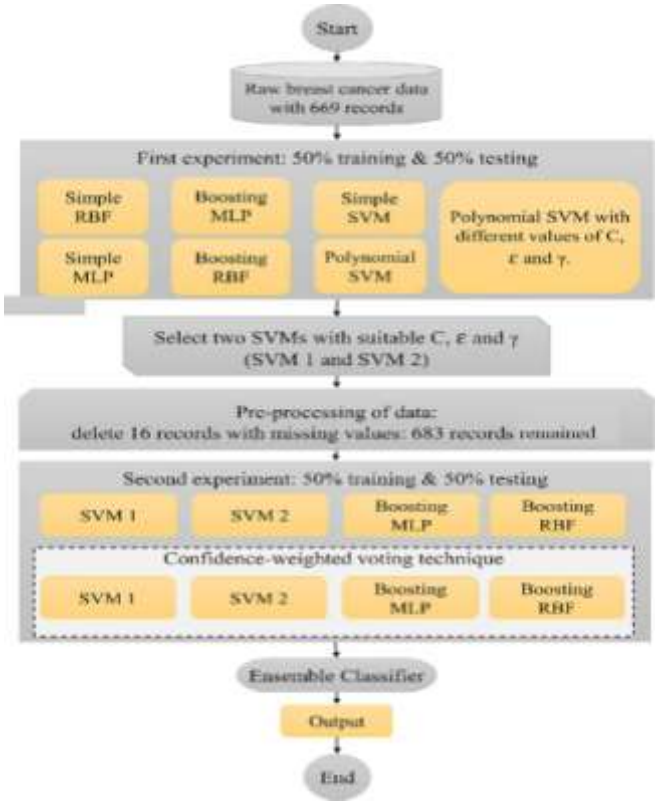

Fig. 4. A schematic view of the proposed CWV-BANN-SVM model.

operating system. The methods in the second experiment were carried out using IBM SPSS Modeler 18 a PC computer equipped with an Intel Core i5 processor and 8 GB of RAM running under Windows 8.1 operating system. In our study, the WBCD dataset was divided in two groups: training and testing sets - 50 percent of the original records were included in each of these sets [13,15]. In the first step, we executed the simple SVM algorithm (without any other parameters). In second step, we used the expert item in the SVM node to improve the prediction performance. In this step, we tested several values of the C and c parameters. As mentioned earlier, to overcome the overfitting issue, an appropri-ate C parameter should be determined and used. For the Polyno-mial SVM model, we used the default parameters of IBM SPSS Modeler, which were: C = 10, e = 0.1 and c = 1. To provide an opti-mal classifier, more experiments were conducted using various values of the parameter C.

Indeed, based on the existing knowledge about appropriate val-ues of these parameters, SVMs can show better performance. For example, the parameters C and e were tested, as discussed in [34,39,46–48], in order to optimize the performance of the SVM method. As explained in these papers, choosing the best parameter values in SVM is very important, For this purpose, we tested vari-ous values of the parameters C, e and c to achieve an acceptable performance in terms of different evaluation metrics, such as specificity, sensitivity, precision, accuracy, F_1 etc., and to decrease the overfitting issue.

Fig. 4 represents the main steps of the proposed methodology. As shown in Fig. 4, in the first step, the classical methods are applied to the original WBCD dataset with 699 records. The first experiment consists of two well-known machine learning tech-niques: SVM with the polynomial kernel function and the ANN using MLP and RBF. Here, seven prediction methods were tested. They were as follows: simple MLP, simple RBF, simple SVM, poly-nomial SVM (with default parameters), boosting MLP, boosting RBF and also polynomial SVM with different values of parameters. As a result, the best SVMs (two polynomial SVM methods with the best parameter values), boosting MLP and boosting RBF were cho-sen for the next implementation stage. It should be mentionned that in this step, the performance of MLP and RBF was improved using boosting techniques as an ensemble approach. Finally, the methods providing the best performances were selected for the next step.

In the second experiment, a pre-processing approach was used to deal with missing data in the WBCD dataset. Hence, 16 records with missing values were removed and the rest of the study was conducted with 683 records. As in the first experiment, 50% of data were used for training and another 50% for testing the proposed model. In the first step of the second experiment simple MLP, sim-ple RBF, boosting MLP and boosting RBF

were applied to the trans-formed BC data. Then, all boosting MLP, boosting RBF and two SVMs, chosen in the first experiment, were combined using the Confidence-Weighted Voting (CWV) ensemble approach. The final model was named the CWV-BANN-SVM ensemble learning model.

V. RESULTS

This section discusses the results obtained without and with the proposed methodology. First, a brief discussion on the evaluation metrics will be presented. As mentioned above, our study include two main experiments. In the first experiment, the results of clas-sical methods will be reported and discussed. Then, in the second experiment, we present and discuss the results obtained using the proposed methodology.

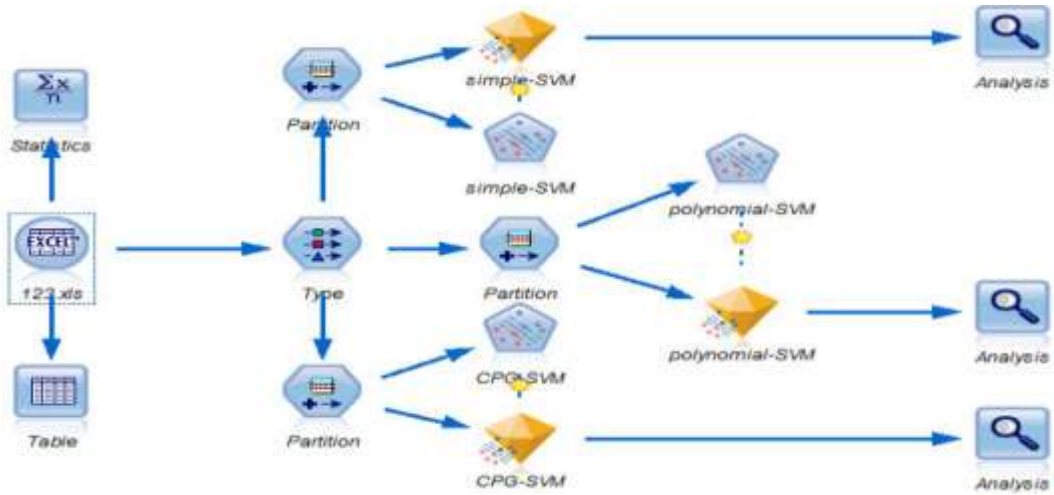

Fig. 5. Proposed system for diagnosis of BC used in the first experiment (SVM).

5.1. Evaluation metrics

The comparison of the performance methods by evaluation metrics is one of the most important steps in data mining. By uti-lizing these metrics, we can analyze the performance of the pro-posed system and compare it with other methods. For this regard, seven metrics have been considered, which are specificity, sensitivity, precision, FPR, FNR, F_1 and accuracy, computed using the True Negative (TN), False Positive (FP), False Negative (FN)

Table 2

Confusion Matrix in training step.

Algorithm	TP	FP	FN	TN	SUM
Simple SVM	120	6	5	222	353
polynomial-SVM	120	6	2	225	353
CPG-SVM (1)	126	0	0	227	353
CPG-SVM (2)	126	0	0	227	353
CPG-SVM (3)	125	1	0	227	353
CPG-SVM (4)	118	8	3	224	353
CPG-SVM (5)	120	6	3	224	353
CPG-SVM (6)	124	2	0	227	353
CPG-SVM (7)	124	2	0	227	353
CPG-SVM (8)	126	0	0	227	353
CPG-SVM (9)	125	1	0	227	353
CPG-SVM (10)	125	1	0	227	353

Table 3
Confusion Matrix in testing step.

Algorithm	TP	FP	FN	TN	SUM

Simple SVM	112	3	6	225	346
polynomial-SVM	113	2	1	230	346
CPG-SVM (1)	114	1	0	231	346
CPG-SVM (2)	114	1	0	231	346
CPG-SVM (3)	114	1	0	231	346
CPG-SVM (4)	113	2	1	230	346
CPG-SVM (5)	113	2	1	230	346
CPG-SVM (6)	113	2	0	231	346
CPG-SVM (7)	113	2	0	231	346
CPG-SVM (8)	114	1	0	231	346
CPG-SVM (9)	114	1	0	231	346
CPG-SVM (10)	114	1	0	231	346

Table 4
Evaluation metrics of SVM for training step (%).

Algorithm	C	e	c	Specificity	Sensitivity	Precision	FPR	FNR	F_1	Accuracy (training)
Simple SVM	–	–	–	97.36	96.00	95.23	2.64	4.0	95.61	96.88
polynomial-SVM	10	0.1	1	97.40	98.36	95.23	2.60	1.64	96.77	97.73
CPG-SVM (1)	50	0.05	2.5	100	100	100	0.0	0.0	100	100
CPG-SVM (2)	50	0.1	2.5	100	100	100	0.0	0.0	100	100
CPG-SVM (3)	10	0.05	2.5	99.56	100	99.20	0.44	0.0	99.60	99.72
CPG-SVM (4)	10	0.05	0.5	96.55	97.52	93.65	3.45	2.48	95.54	96.88
CPG-SVM (5)	50	0.05	0.5	97.39	97.56	95.23	2.61	2.44	96.38	97.45
CPG-SVM (6)	50	0.1	1	99.12	100	98.41	0.88	0.0	99.20	99.43
CPG-SVM (7)	100	0.1	1	99.12	100	98.41	0.88	0.0	99.20	99.43
CPG-SVM (8)	50	0.05	2	100	100	100	0.0	0.0	100	100
CPG-SVM (9)	50	0.05	1.5	99.56	100	99.20	0.44	0.0	99.60	99.72
CPG-SVM (10)	100	0.1	1.5	99.56	100	99.20	0.44	0.0	99.60	99.72

Table 5
Evaluation metrics of SVM for testing step (%).

Algorithm	C	e	c	Specificity	Sensitivity	Precision	FPR	FNR	F_1	Accuracy (testing)
Simple SVM	–	–	–	98.68	94.91	97.39	1.32	5.09	96.13	97.40
polynomial-SVM	10	0.1	1	99.13	99.12	98.26	0.87	0.88	98.68	99.13
CPG-SVM (1)	50	0.05	2.5	99.56	100	99.13	0.44	0.0	99.56	99.71
CPG-SVM (2)	50	0.1	2.5	99.56	100	99.13	0.44	0.0	99.56	99.71
CPG-SVM (3)	10	0.05	2.5	99.56	100	99.13	0.44	0.0	99.56	99.71
CPG-SVM (4)	10	0.05	0.5	99.13	99.12	98.26	0.87	0.88	98.68	99.13
CPG-SVM (5)	50	0.05	0.5	99.13	99.12	98.26	0.87	0.88	98.68	99.13
CPG-SVM (6)	50	0.1	1	99.14	100	98.26	0.86	0.0	99.12	99.42
CPG-SVM (7)	100	0.1	1	99.14	100	98.26	0.86	0.0	99.12	99.43

CPG-SVM (8)	50	0.05	2	99.56	100	99.13	0.44	0.0	99.56	99.71
CPG-SVM (9)	50	0.05	1.5	99.56	100	99.13	0.44	0.0	99.56	99.71
CPG-SVM (10)	100	0.1	1.5	99.56	100	99.13	0.44	0.0	99.56	99.71

Table 6

Coefficients showing the importance of each attribute.

Sensitivity ¼ TPR ¼ TP = TP þ FN ð9Þ

No.	Features name	Coefficients for attributes
1	Bare nuclei	0.28
2	Uniformity of cell shape	0.23
3	Normal nucleoli	0.14
4	Uniformity of cell size	0.12
5	Bland chromatin	0.10
6	Mitoses	0.07
7	Clump thickness	0.05
8	Single epithelial cell size	0.02
9	Marginal adhesion	0.00

and True Positive (TP) rates. The Confusion Matrix was created at all steps. More information about each of these terms are pre-sented below:

TPs are the true positives,

FNs are the false negatives,

FPs are the false positives,

TNs are the true negatives.

As mentioned above, seven metrics have been used to compute the performance of various algorithms. They can be calculated using Eqs. (8)–(14) [48]:

Specificity ¼ TNR ¼ TN = TN þ FP ð8Þ

Table 7

Evaluation metrics of ANN for training step (%).

Precisio
n ¼ TP = TPþFP ð10Þ

FPR ¼ FP=FP þ TN ¼
1 TNR ð11Þ

FNR ¼ FN=FN þ TP
¼ 1 TPR ð12Þ

F_1 ¼ 2TP = ð2TP þ FP þ FNÞ ð13Þ

Accurac
y ¼ TPþTN = TPþTNþFP þFN ð14Þ

5.2. First experiment

This step discusses the obtained results using the SVM algo-rithm with different parameters. As we mentionned above, in our first experiment, the whole WBCD dataset with 699 records was used [45]. The proposed system for diagnosis of BC is illustrated in Fig. 5. The Confusion Matrices are shown in Tables 2 and 3, and the obtained results for the simple SVM and CPG-SVM algo-rithms for both training and testing stages are given in Tables 4 and 5, respectively.

According to Table 3, it can be observed that in training stage, in the best situation, 114 malignant records were detected as malig-nant and were classified as TP, whereas in the best case, 231 benign records were detected as benign and were classified as TN. More-

Algorithm	Specificity	Sensitivity	Precision	FPR	FNR	F_1	Accuracy (training)
Simple MLP	97.96	97.19	95.41	2.04	2.81	96.29	97.73
Simple RBF	96.74	93.45	92.59	3.26	6.55	93.02	95.75
MLP + boosting	100	100	100	0	0	100	100
RBF + boosting	98.78	96.26	97.16	1.22	3.74	96.71	98.02

Table 8

Evaluation metrics of ANN for testing step (%).

Algorithm	Specificity	Sensitivity	Precision	FPR	FNR	F_1	Accuracy (testing)
Simple MLP	94.81	97.01	92.19	5.19	2.99	94.54	95.66
Simple RBF	96.22	93.28	93.98	3.78	6.72	93.63	95.09
MLP + boosting	99.69	95.52	94.81	3.31	4.48	95.16	96.24
RBF + boosting	96.22	92.53	93.93	3.78	7.47	93.23	94.80

Table 9
Evaluation metrics of the SVMs and the ANNs for training step (%).

Algorithm	Specificity	Sensitivity	Precision	FPR	FNR	$^{F}1$	Accuracy (training)
SVM 1	100	99.20	100	0.0	0.80	99.60	99.71
SVM 2	100	99.20	100	0.0	0.80	99.60	99.71
Simple MLP	96.36	99.20	93.98	3.64	0.80	96.52	97.40
Simple RBF	97.72	92.06	95.86	2.28	7.94	93.92	95.66
MLP + boosting	100	100	100	0.0	0.0	100	100
RBF + boosting	97.72	94.44	95.96	2.28	5.56	95.20	96.53

Table 10
Evaluation metrics of the SVMs and the ANNs for testing step (%).

Algorithm	Specificity	Sensitivity	Precision	FPR	FNR	$^{F}1$	Accuracy (testing)
SVM 1	100	98.23	100	0.0	1.77	99.10	99.41
SVM 2	100	99.11	100	0.0	0.89	99.55	99.70
Simple MLP	97.32	96.46	94.78	2.68	3.54	95.61	97.03
Simple RBF	97.32	94.69	96.69	2.68	5.31	94.69	96.44
MLP + boosting	98.21	95.57	96.42	1.79	4.43	96.00	97.33
RBF + boosting	96.87	95.57	93.91	3.13	4.43	94.73	96.44

over, in the best situation, 0 benign records were classified as malignant and only 1 malignant record was classified in the benign class. Therefore, in the best cases of the first experiment, the FP and the FN values of the confusion matrix were 0 and 1, respectively. The calculated specificity, sensitivity, precision, FPR, FNR, F_1 and accuracy values are illustrated in Tables 4 and 5, for both training and testing steps, respectively. As noted earlier, the values of the parameters C, e and c should be selected carefully to deal with the overfitting issue. Since the default value for c was 1, we then changed this parameter slightly. As a result, the process of choos-ing the appropriate parameters enhanced the performance of SVM and, at the same time, decreased the probability of overfitting.

According to Table 5, it is obvious that the highest accuracies were obtained in CPG-SVM 1, 2, 3, 8, 9 and 10 with 99.71%. It can be also seen that various tested values of the C parameter were 10, 50 and 100, for the e parameter these values were 0.05 and 0.1, and for the c parameter, they were 0.5, 1, 1.5, 2 and 2.5. We obtained the best prediction accuracy (99.71%) when the values of the parameters C, e and c, were 50, 0.05 and 1.5, respectively. Although, by reducing the value of c from 2 to 1.5, the accuracy of training step has been decreased from 100% to 99.72%. However, by reducing the value of the parameter c, we can deal with overfit-ting challenge, and the accuracy of testing step remains unchanged. Our experiments show that the parameters C and c are more important and have greater impact on the accuracy than the parameter e. As a result, the two best SVMs were chosen for our nest experiments. According to the results presented in Table 5, we select the best values for these parameters as follows: C = 50 and

Table 11

Final results (confusion matrix) for diagnosis of BC using the CWV-BANN-SVM model in training step.

'Partition' = 1_Training	Benign	Malignant
Benign	220	0
Malignant	0	126

Table 12

Final results (confusion matrix) for diagnosis of BC using the CWV-BANN-SVM model in testing step.

'Partition' = 2_Testing	Benign	Malignant
Benign	224	0
Malignant	0	113

100, e = 0.05 (the lowest value), and c = 1.5 (the lowest value).
Hence, we will consider two SVMs with the following parameters:
SVM 1: C = 50, e = 0.05, and c = 1.5,
SVM 2: C = 100, e = 0.05, and c = 1.5.

Finally, the CPG-SVM method has been applied with C = 50, e = 0.05 and c = 1.5 to identify the predictor importance for all BC attributes, as reported in Table 6, which presents information about the coefficients for each attribute.

According to Table 6, it can be seen that the feature Bare nuclei, Uniformity of cell shape, Normal nucleoli, Uniformity of cell size and Bland chromatin were the most important features for diag-nosing BC, whereas the feature Single epithelial cell size and Mar-ginal adhesion were the least important features for diagnosing BC. The importance of all features is shown using the coefficients fea-tures in Table 8. For instance, the coefficient value of Bare nuclei is 0.28, which is the most important feature for diagnosing BC. The coefficient value of Marginal adhesion has the lowest importance among all features with the value of 0. These findings could be very helpful for physicians working BC. In other words, considering the most important features can improve the quality of BC detection, decreasing the error rate of BC detection.

Furthermore, at this step we assessed the performance of sim-ple MLP, simple RBF, boosting MLP and boosting RBF on the dataset with 699 records. The detailed results provided by these algo-rithms are illustrated in Tables 7 and 8.

As Table 7 shows, boosting MLP has the best performance in training step when the original data with 699 records are consid-ered. Furthermore, Table 8 indicates that once again boosting MLP has the best performance in testing step in terms of speci-ficity, precision, FPR, F_1, and accuracy.

5.3. Second experiment

In this experiment, the pre-processing technique is used to improve the performance of the model. In other words, the pre-processing removes unnecessary data and/or data with missing values. In our first experiment, the original dataset with 699 records were used, while in the second experiment, 16 records with missing values were removed from it. Thus, we assess the performance of the two selected SVMs (SVM 1: C = 50, e = 0.05,

Table 13

A summary of the introduced model per each algorithm in ensemble model. and c = 1.5, and SVM 2: C = 100, e = 0.05, and c = 1.5) on the reduced dataset including 683 records. Moreover, the perfor-mances of MLP and two hidden layers and RBF with one hidden layer will be also compared. As previously, we will apply the boost-ing technique with artificial neural network (with both MLP and RBF) to increase the accuracy of methods. More details are pre-sented in Tables 9 and 10.

Table 9 shows that boosting MLP has the best performance compared to the other methods in the training step, whereas Table 10 reveals that SVM 2 has the best performance in the testing step. As discussed in [49], even though some ensemble techniques can enhance the performance of machine learning algorithms, they may also deteriorate their performance. Moreover, due to valuable performance of ANN on Parkinson disease in [50], this study applies the same method for BC as well. Moreover, our first exper-

Model (algorithm)	Max Profit	Max Profit occurs in (%)	Lift (top 30%)	Overall accuracy (%)	No. features	AUC
Neural Network 1	1195	34	2.858	100	9	1.0
SVM 2	1195	34	2.858	99.707	9	1.0
SVM 1	1185	34	2.858	99.561	9	1.0
Neural Network 2	1110	36	2.786	97.365	9	0.995

Table 14

Evaluation of the CWV-BANN-SVM model.

Phase	Specificity	Sensitivity	Precision	FPR	FNR	$^{F}1$	Accuracy	AUC	Gini
Training	100	100	100	0.0	0.0	100	100	1.0	1.0
Testing	100	100	100	0.0	0.0	100	100	1.0	1.0

Table 15

Comparison of different methods applied for BC prediction.

No.	Study	Year	Algorithm	Accuracy (%)	
1	Quinlan [7]	1996	C4.5	94.74	(10-CV)
2	Hamilton et al. [8]	1996	RIAC	94.99	(10-CV)
3	Ster and Dobnikar [9]	1996	LDA + NN	96.80	(10-CV)
4	Bennett and Blue [10]	1998	SVM	97.20	(5-CV)
5	Nauck and Kruse [11]	1999	NEFCLASS	95.06	(10-CV)
6	Pena-Reyes and Sipper [12]	1999	Fuzzy-GA	97.36	(fuzzy diagnostic system with two rules) (train dataset: 50% – test dataset: 50%)
7	Setiono [13]	2000	Neuro-Rule	98.10	
8	Goodman et al.	2002	Optimized-LVQ	96.70	(10-CV)

	[14]				
	Goodman et al.				
9	[14]	2002	Big-LVQ	96.80	(10-CV)
	Goodman et al.				
10	[14]	2002	AIRS	97.20	(10-CV)
					(train dataset:
	Albrecht et al.				50% – test
11	[15]	2002	LSA with perceptron method	98.80	dataset: 50%)
	Abonyi and				
12	Szeifert [16]	2003	SFC	95.57	(10-CV)
					(train dataset:
					37% – test
13	Ubeyli [17]	2005	ME	98.85	dataset: 63%)
	Polat and Günes				
14	[18]	2007	LS-SVM	98.53	(10-CV)
					(train dataset:
					37% – test
15	Ubeyli [19]	2007	SVM	99.54	dataset: 63%)
					99.51(train dataset:
					80% – test dataset:
16	Akay [20]	2009	F-score + SVM	20%)	
	Karabatak and				
17	Ince [21]	2009	AR+NN	97.40	(3-CV)
	Paulin and				(train dataset:
	Santhakumaran				70% – test
18	[22]	2011	FFNN + BP + LM	99.28	dataset: 30%)
	Dheeba et al.				
19	[23]	2014	PSOWNN	93.67	
20	Karabatak [24]	2015	W-NB	98.54	(5-CV)
					(train dataset:
	Nahato et al.				80% – test
21	[25]	2015	RS-BPNN	98.60	dataset: 20%)
22	Mert et al. [26]	2015	RBFNN	90.49	(10-CV)
	Turabieh and				71 (train dataset: 75%
23	Muhanna [27]	2015	GA + ANFIS	– test dataset: 25%)	
					(train dataset:
					54.90% – test
	Abdel-Zaher				dataset:
24	and Eldeib [3]	2016	DBN-NN	99.68	45.10%)
25	Asri et al. [29]	2016	SVM	97.13	(10-CV)
					(train dataset:
	Kaushik and				70% – test
26	Kaur [5]	2016	EC	83.50	dataset: 30%)
	Alicˇkovicˊ and				
27	Subasi [30]	2017	Rotation Forest + GA	99.48	(10-CV)
					100 (train dataset: 70%
28	Dora et al. [32]	2017	GNRBA	– test dataset: 30%)	
	Kumari and				
29	Singh [28]	2018	KNN	99.28	(10-CV)
	Nguyen et al.				
30	[51]	2015	GSAM + Wavelets	97.40	(5-CV)
	The current				100 (train dataset: 50%
31	study	2019	CWV + BANN + SVM	– test dataset: 50%)	

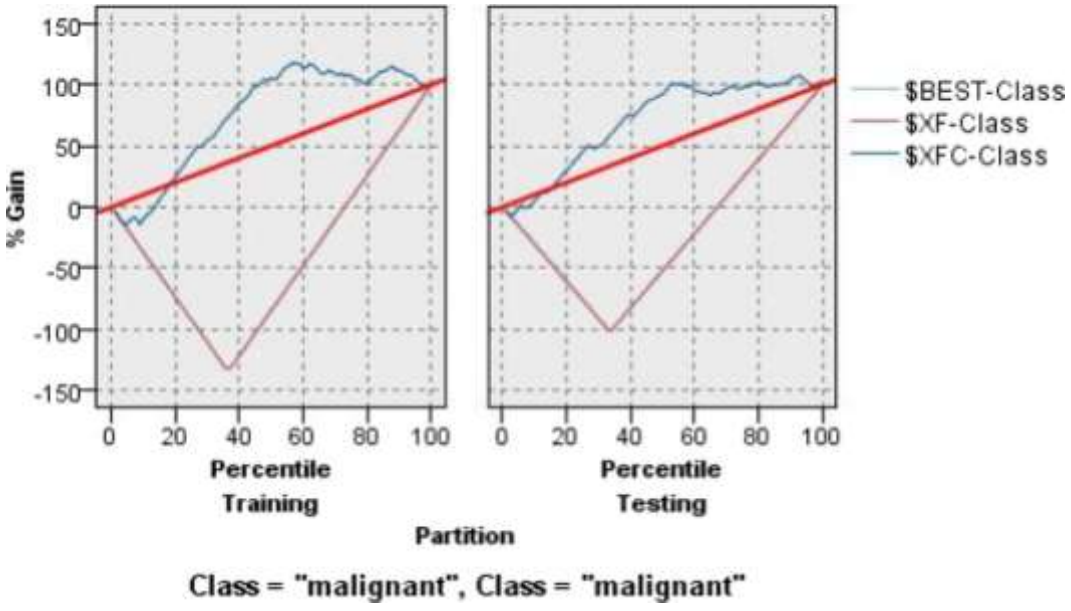

Fig. 6. Gain chart for the CWV-BANN-SVM model.

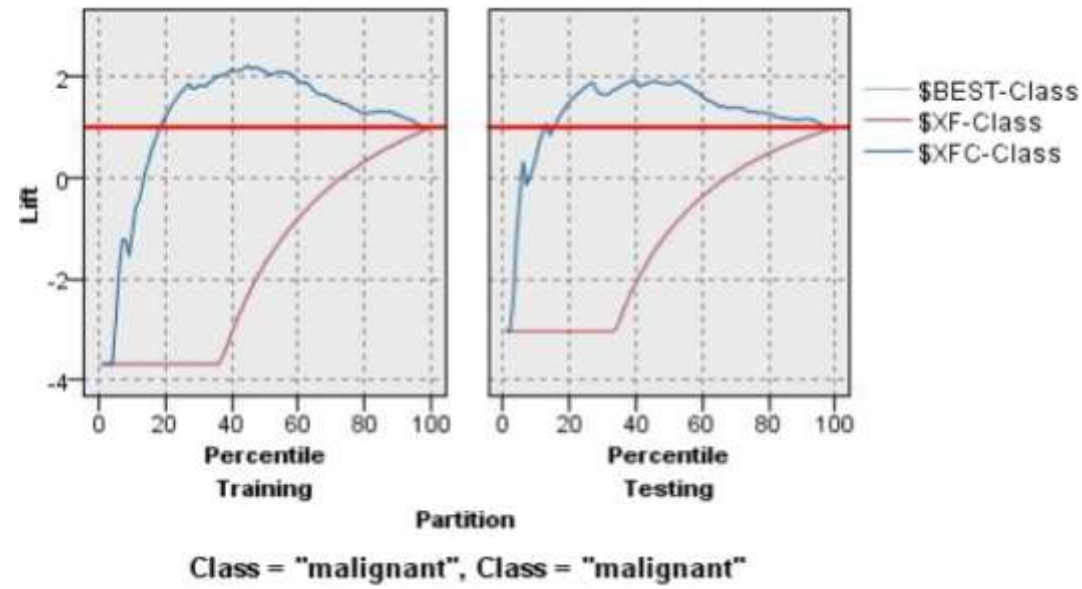

Fig. 7. Lift chart for the CWV-BANN-SVM model.

iment showed that the parameters C, e and c are very important and should be selected carefully when the polynomial SVM algo-rithm is applied.

Overal, we applied two types of ANNs: MLP with two hidden layers and RBF with one hidden layer. Thus, our final model included four algorithms: SVM 1. SVM 2, MLP, and RBF. We cus-tomized the number of units in hidden layers for MLP (Neural Net-work 1) and RBF (Neural Network 2). Since, our dataset had 9 input features, both hidden layers of MLP (hidden layer 1 and hidden layer 2) also had 9 units (neurons). Therefore, the final architecture of MLP was set to 9-9-9-2. In addition, the hidden layer related to RBF also had 9 units (neurons) and the final architecture of RBF was set to 9-9-2. The results obtained by using the CWV-BANN-SVM method are shown in Tables 11 and 12.

Summary of the proposed model regarding the performance of each algorithm is given in Table 13.

Table 9 shows that the Neural Network 1 (MLP) algorithm with boosting technique was quite influential on the final result of the model such that the final model provided the best values of the selected metrics. It should be noted that even though Neural Net-work 1 (MLP) had 100% accuracy, this result was obtained when our new CWV-BANN-SVM model was applied. We showed that the performance of Neural Network 1 (MLP) with boosting tech-nique was not the same when it was applied without our model. In other words, the introduced model could improve the perfor-mance of Neural Network 1 (MLP), compared to the original Neural Network 1 (MLP) applied in the previous experiment. The results provided by the final CWV-BANN-SVM model are summarized in Table 14.

According to Table 14, it can be observed that the proposed CWV-BANN-SVM model improved the detection performance in terms of all metrics in training and testing steps.

5.4. Discussion

For more clarity, we compared the results obtained using the proposed methodology with those presented in several existing studies related to BC prediction. Table 15 provides a comparaison of the accuracy of our model and those proposed previously. We can see that that the proposed CWV-BANN-SVM model with 100% accuracy has the highest performance for diagnosis of BC, compared to the existing studies (except the study by Dora et al. [32]).

However, we would argue that our model uses less training samples, compared to Dora et al. [32]. This means that the pro-posed CWV-BANN-SVM model is very flexible and needs less records for the training step. Moreover, the testing set in the cur-rent study is bigger than in [32]. Thus, we can conclude that the

CWV-BANN-SVM model is quite effective for an early diagnosis of BC. Graphical representations of Gain, Lift, Profit, Response, ROC Curve (ROC), and Return on investment (ROI) are illustrated in Figs. 6–11, respectively.

Figs. 6–11 indicate the performance of the CWV-BANN-SVM model when for the class ‘‘Malignant”. As indicated, CWV-BANN-SVM has the highest performance in terms of different factors. Indeed, having the best performance for a clinical decision support system is a key issue which have a significant impact for decision making process. In our study, the WBCD dataset was divided into two groups: training and testing, with 50 percent of data desig-nated for training and another 50 percent of data for testing [13,15]. It is worth noting that the k-fold cross validation can be also used. Additionally, our work mainly concentrates on the opti-mization of classical algorithms using two ensemble techniques.

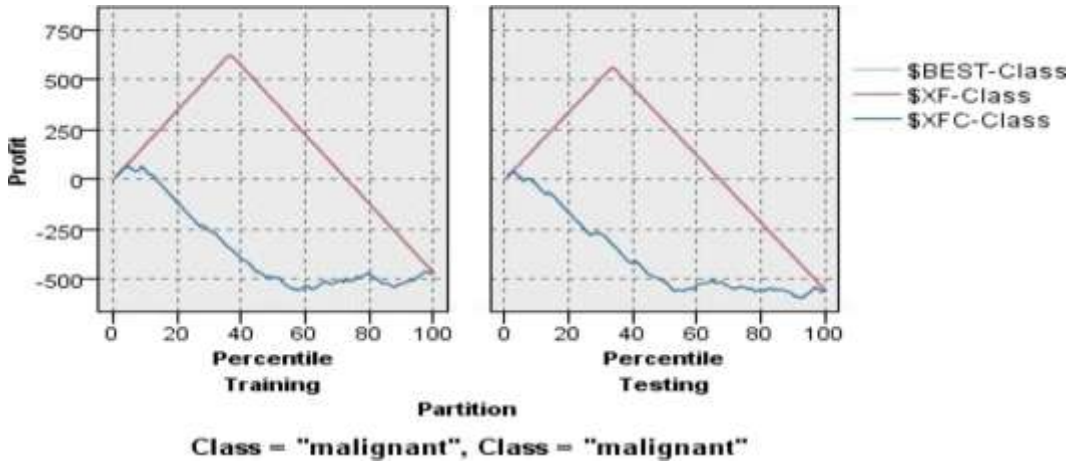

Fig. 8. Profit chart for the CWV-BANN-SVM model.

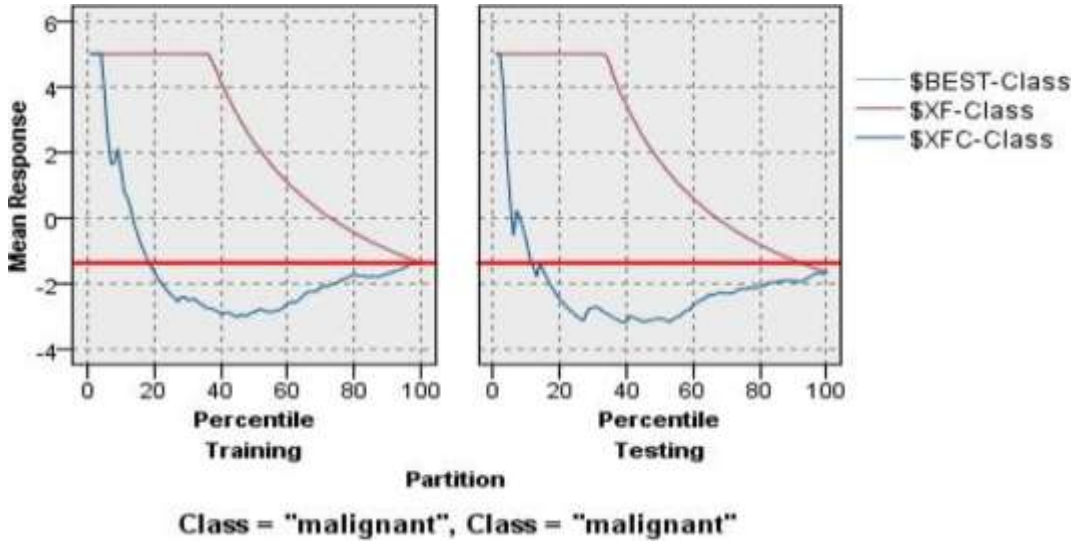

Fig. 9. Response chart for the CWV-BANN-SVM model.

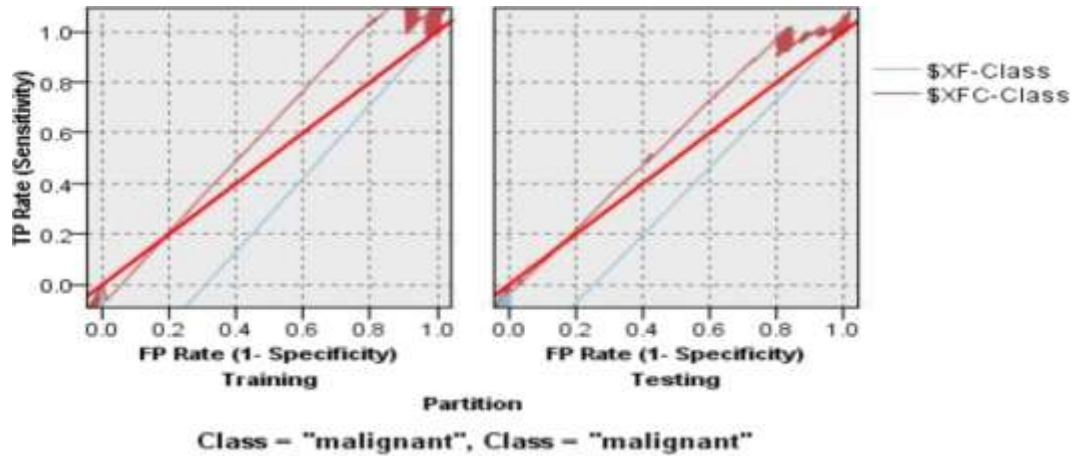

Fig. 10. ROC Curve (ROC) chart for the CWV-BANN-SVM model.

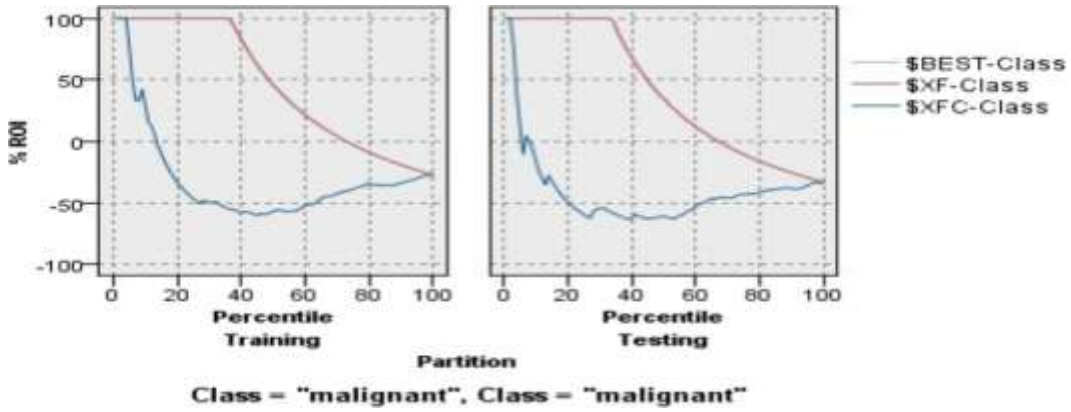

Fig. 11. Return on investment (ROI) chart for the CWV-BANN-SVM model.

Thus, other optimization techniques, such as evolutionary algo-rithms [6,52–59] and different ensemble learning techniques [60] could be also applied for BC prediction purposes.

VI. CONCLUSION

There are different techniques in machine learning and data mining which can be used to process huge amount of medical data. Two well-known methods, SVM and ANN, were used in our study to explore BC data. To this end, the well-known WBCD dataset [45] from the UCI dataset repository was considered. Our study included two main experiments. In the first experiment, the SVM algorithm with three main parameters, C, e and c, was applied with the polynomial kernel function. At this step, a simple SVM model, the polynomial SVM (with default parameters) and several adjusted SVMs with various values of C, e and c were carried out and compared between them. Our results suggest that the right choice of these parameters can greatly enhance the performance of SVM. The optimal values of C, e and c found in the first experi-ment were then used in the following experiments of our study. Moreover, the MLP function with two hidden layers and the RBF function with one hidden layer with boosting technique were also applied. In the second experiment, two SVMs, used with the opti-mal values of C, e and c, were combined with both the boosting multilayer perceptron (BMLP) and boosting radial basis function (BRBF) techniques using confidence-weighted voting ensemble learning algorithm. The new method was called CWV-BANN-SVM. Our final results indicate that the proposed methodology allows us to get 100% accuracy and thus be quite successful in terms of BC prediction. As a conclusion, we would argue that the introduced CWV-BANN-SVM model is worth to be applied for BC prediction in medical research. In the future, it would be interest-ing to test the proposed methodology on different types of medical data. In this work, we used only two ensemble techniques (boost-ing and confidence-weighted voting ensemble). The models incor-porating other types of ensemble techniques should be also investigated. Moreover, some other optimization techniques, such as GA, PSO or ACO, could be applied to find suitable values of SVM parameters for the analysis of BC data.

ACKNOWLEDGEMENT

We would like to express our sincere appreciation to the editor, the associate editor, and the anonymous reviewers for their insightful comments, which have greatly aided us in improving the quality of the paper. Moreover, the authors would like to thank Dr. Mariam Zomorodi-Moghadam from Ferdowsi University of Mashhad for her insightful and valuable comments and for helping to edit the final version of this paper.

REFERENCES

[1] D. Carvalho, P.R. Pinheiro, M.C.D. Pinheiro, A hybrid model to support the early diagnosis of breast cancer, Procedia Comput. Sci. 91 (2016) 927–934.

[2] H.L. Chen, B. Yang, G. Wang, S.J. Wang, J. Liu, D.Y. Liu, Support vector machine based diagnostic system for breast cancer using swarm intelligence, J. Med. Syst. 36 (4) (2012) 2505–2519.

[3] A.M. Abdel-Zaher, A.M. Eldeib, Breast cancer classification using deep belief networks, Expert Syst. Appl. 46 (2016) 139–144.

[4] American Cancer Society, Breast cancer: Facts and figures 2013, ACS, Atlanta.

[5] D. Kaushik, K. Kaur, Application of Data Mining for high accuracy prediction of breast tissue biopsy results, in: 2016 Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC), IEEE, 2016, pp. 40–45.
[6] A.H. Alkeshuosh, M.Z. Moghadam, I. Al Mansoori, M. Abdar, Using PSO Algorithm for Producing Best Rules in Diagnosis of Heart Disease, in: Computer and Applications (ICCA), 2017 International Conference on, IEEE, 2017, pp. 306–311.
[7] J. Quinlan, Improved use of continuous attributes in C4. 5, J. Artif. Intell. Res. 4 (1996) 77–90.
[8] H.J. Hamilton, N. Shan, N. Cercone, RIAC: a rule induction algorithm based on approximate classification, Computer Science Department, University of Regina. International conference on engineering applications of neural networks, 1996.
[9] B. Šter, A. Dobnikar, Neural networks in medical diagnosis: Comparison with other methods, in: International Conference on Engineering Applications of Neural Networks, 1996, pp. 427–430.
[10] K.P. Bennett, J.A. Blue, A support vector machine approach to decision trees, in: Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on, IEEE, 1998, pp. 2396–2401.
[11] D. Nauck, R. Kruse, Obtaining interpretable fuzzy classification rules from medical data, Artif. Intell. Med. 16 (2) (1999) 149–169.
[12] C.A. Pena-Reyes, M. Sipper, A fuzzy-genetic approach to breast cancer diagnosis, Artif. Intell. Med. 17 (2) (1999) 131–155.
[13] R. Setiono, Generating concise and accurate classification rules for breast cancer diagnosis, Artif. Intell. Med. 18 (3) (2000) 205–219.
[14] D. Goodman, L. Boggess, A. Watkins, Artificial immune system classification of multiple-class problems, Proc. the rtif. Neural Networks Eng. 2002 (2002) 179–183.
[15] A.A. Albrecht, G. Lappas, S.A. Vinterbo, C. Wong, L. Ohno-Machado, Two applications of the LSA machine, in: Neural Information Processing, 2002. ICONIP'02. Proceedings of the 9th International Conference on, IEEE, 2002, pp. 184–189.
[16] J. Abonyi, F. Szeifert, Supervised fuzzy clustering for the identification of fuzzy classifiers, Pattern Recognit. Lett. 24 (14) (2003) 2195–2207.
[17] E.D. Übeyli, A mixture of experts network structure for breast cancer diagnosis, J. Med. Syst. 29 (5) (2005) 569–579.
[18] K. Polat, S. Günes, Breast cancer diagnosis using least square support vector machine, Digital Signal Process. 17 (4) (2007) 694–701.
[19] E.D. Übeyli, Implementing automated diagnostic systems for breast cancer detection, Expert Syst. Appl. 33 (4) (2007) 1054–1062.
[20] M.F. Akay, Support vector machines combined with feature selection for breast cancer diagnosis, Expert Syst. Appl. 36 (2) (2009) 3240–3247.
[21] M. Karabatak, M.C. Ince, An expert system for detection of breast cancer based on association rules and neural network, Expert Syst. Appl. 36 (2) (2009) 3465–3469.
[22] F. Paulin, A. Santhakumaran, Classification of breast cancer by comparing back propagation training algorithms, Int. J. Comput. Sci. Eng. 3 (1) (2011) 327–332.
[23] J. Dheeba, N.A. Singh, S.T. Selvi, Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach, J. Biomed. Inform. 49 (2014) 45–52.
[24] M. Karabatak, A new classifier for breast cancer detection based on Naïve Bayesian, Measurement 72 (2015) 32–36.
[25] K.B. Nahato, K.N. Harichandran, K. Arputharaj, Knowledge mining from clinical datasets using rough sets and backpropagation neural network, Comput. Math. Methods Med. 2015 (2015) 1–13, https://doi.org/10.1155/2015/460189.
[26] A. Mert, N. Kılıç, E. Bilgili, A. Akan, Breast cancer detection with reduced feature set, Comput. Math. Methods Med. 2015 (2015) 1–11, https://doi.org/ 10.1155/2015/265138.
[27] H. Turabieh, M. Muhanna, GA-based feature selection with ANFIS approach to breast cancer recurrence, Int. Journal of Comput. Sci. Issues (IJCSI) 13 (1) (2016) 36.
[28] M. Kumari, V. Singh, Breast cancer prediction system, Procedia Comput. Sci. 132 (2018) 371–376.
[29] H. Asri, H. Mousannif, H. Al Moatassime, T. Noel, Using machine learning algorithms for breast cancer risk prediction and diagnosis, Procedia Comput. Sci. 83 (2016) 1064–1069.
[30] E. Alicˇkovic´, A. Subasi, Breast cancer diagnosis using GA feature selection and Rotation Forest, Neural Comput. Appl. 28 (4) (2017) 753–763.
[31] A. Dhull, G. Gupta, An Intelligent Two Phase Fuzzy Decision Tree Based Clustering Model for Design of Computer Aided Detection/Diagnosis (CADe/ CADx) System, MAPAN (2017) 1–13.

[32] L. Dora, S. Agrawal, R. Panda, A. Abraham, Optimal breast cancer classification using Gauss-Newton representation based algorithm, Expert Syst. Appl. 85 (2017) 134–145.
[33] C.J. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Disc. 2 (2) (1998) 121–167.
[34] V. Vapnik, The Nature of Statistical Learning Theory, Springer Science & Business Media, 2013.
[35] D. Tao, X. Tang, X. Li, X. Wu, Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval, IEEE Trans. Pattern Anal. Mach. Intell. 28 (7) (2006) 1088–1099.
[36] C. Zhou, K. Yin, Y. Cao, B. Ahmed, Application of time series analysis and PSO– SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir, China, Eng. Geol. 204 (2016) 108–120.
[37] M.S. Elbisy, Sea wave parameters prediction by support vector machine using a genetic algorithm, J. Coastal Res. 31 (4) (2013) 892–899.
[38] Y.C. Wu, J.C. Yang, Y.S. Lee, An approximate approach for training polynomial kernel SVMs in linear time, Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions (pp. 65-68), Association for Computational Linguistics, 2007.
[39] V.N. Vapnik, V. Vapnik, Statistical Learning Theory, Vol. 1, Wiley, New York, 1998.
[40] S. Walczak, Artificial neural networks, in: Encyclopedia of Information Science and Technology, Fourth Edition, IGI Global, 2018, pp. 120–131.
[41] G.S. Fesghandis, A. Pooya, M. Kazemi, Z.N. Azimi, Comparison of multilayer perceptron and radial basis function neural networks in predicting the success of new product development, Eng., Technol. Appl. Sci. Res. 7 (1) (2016) 1425– 1428.
[42] E. Ebrahimi, K. Mollazade, A. Arefi, An expert system for classification of potato tubers using image processing and artificial neural networks, Int. J. Food Eng. 8(4) (2012).
[43] G. Li, G. Zhao, C. Zhou, M. Ren, Stochastic elastic properties of composite matrix material with random voids based on radial basis function network, Int. J. Comput. Methods 15 (01) (2018) 1750082.
[44] S.H. Yoo, S.K. Oh, W. Pedrycz, Optimized face recognition algorithm using radial basis function neural networks and its practical applications, Neural Networks 69 (2015) 111–125.
[45] W. H. Wolberg, 1992. UCI Machine Learning Repository [https://archive.ics.uci. edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29]. Irvine, CA: University of California, School of Information and Computer Science [accessed July 1, 2016].
[46] V. Cherkassky, F.M. Mulier, Learning From Data: Concepts, Theory, and Methods, John Wiley & Sons, 2007.
[47] B. Schölkopf, J. Burges, A. Smola, Advances in Kernel Methods: Support Vector Machine, MIT Press, Cambridge, MA, 1999.
[48] M. Abdar, M. Zomorodi-Moghadam, R. Das, I.H. Ting, Performance analysis of classification algorithms on early detection of liver disease, Expert Syst. Appl. 67 (2017) 239–251.
[49] R. Das, A. Sengur, Evaluation of ensemble methods for diagnosing of valvular heart disease, Expert Syst. Appl. 37 (7) (2010) 5110–5115.
[50] R. Das, A comparison of multiple classification methods for diagnosis of Parkinson disease, Expert Syst. Appl. 37 (2) (2010) 1568–1572.
[51] T. Nguyen, A. Khosravi, D. Creighton, S. Nahavandi, Classification of healthcare data using genetic fuzzy logic system and wavelets, Expert Syst. Appl. 42 (4) (2015) 2184–2197.
[52] W.K. Mashwani, A. Salhi, Multiobjective memetic algorithm based on decomposition, Appl. Soft Comput. 21 (2014) 221–243.
[53] W.K. Mashwani, A. Salhi, Multiobjective evolutionary algorithm based on multimethod with dynamic resources allocation, Appl. Soft Comput. 39 (2016) 292–309.
[54] Q. Kotimah, W.F. Mahmudy, V.N. Wijayaningrum, Optimization of fuzzy tsukamoto membership function using genetic algorithm to determine the river water, Int. J. Electr. Comput. Eng. (IJECE) 7 (5) (2017) 2838–2846.
[55] V.N. Wijayaningrum, W.F. Mahmudy, Optimization of ship's route scheduling using genetic algorithm, Indonesian J. Electr. Eng. Comput. Sci. 2 (1) (2016) 180–186.
[56] P. Pławiak, Novel genetic ensembles of classifiers applied to myocardium dysfunction recognition based on ECG signals, Swarm Evol. Comput. 39 (2018) 192–208.
[57] P. Pławiak, Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system, Expert Syst. Appl. 92 (2018) 334–349.
[58] W. Ksia¸ zek, M. Abdar, U.R. Acharya, P. Pławiak, A novel machine learning approach for early detection of hepatocellular carcinoma patients, Cognit. Syst. Res. 54 (2019) 116–127.

[59] M. Hassoon, M.S. Kouhi, M. Zomorodi-Moghadam, M. Abdar, Rule Optimization of Boosted C5. 0 Classification Using Genetic Algorithm for Liver disease Prediction, Computer and Applications (ICCA), 2017 International Conference on (pp. 299-305), 2017.

[60] M. Abdar, M. Zomorodi-Moghadam, X. Zhou, R. Gururajan, X. Tao, P.D. Barua, R. Gururajan, A new nested ensemble technique for automated diagnosis of breast cancer, Pattern Recognit. Lett. (2018).