Article

A Hybrid Ensemble Method for Accurate Breast Cancer Tumor Classification using State-of-the-Art Classification Learning Algorithms

Adel Saad Assiri[1], Saima Nazir[2], and Sergio A. Velastin[3,4,5,*]

1 College of Business, Management Information Systems, Department King Khalid University, Saudi Arabia
2 Department of Software Engineering, Fatima Jinnah Women University, The Mall Rawalpindi, Pakistan
3 Department of Computer Science, Applied Artificial Intelligence Research Group, Universidad Carlos III de Madrid, Spain
4 School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK
5 Cortexica Vision Systems Ltd., London, UK
* sergio.velastin@ieee.org

Abstract: Breast cancer is the most common cause of death for women worldwide. Thus, the ability of artificial intelligence systems to predict and classify breast cancer is very important. In this paper, a hybrid ensemble method classification mechanism is proposed based on a majority voting mechanism. First, the performance of different state-of-the-art machine learning classification algorithms for the Wisconsin Breast Cancer Dataset (WBCD) were evaluated. The three best classifiers were then selected based on their F3 score. F3 score is used to emphasize the importance of false negatives (recall) in breast cancer classification. Then, these three classifiers, simple logistic Regression learning, stochastic gradient descent learning and multilayer perceptron network, are used for ensemble classification using a voting mechanism. We also evaluated the performance of hard and soft voting mechanism. For hard voting, majority-based voting mechanism was used and for soft voting we used average of probabilities, product of probabilities, maximum of probabilities and minimum of probabilities-based voting methods. The hard voting (majority-based voting) mechanism shows better performance with 99.42% as compared to the state-of-the-art algorithm for WBCD.

Keywords: Breast Cancer Tumor; Classification; Majority-based Voting Mechanism; Multilayer Perceptron Learning Network; Simple Logistic Regression; Stochastic Gradient Descent Learning; Wisconsin Breast Cancer Dataset

1. Introduction

Breast cancer is one of the leading causes of death among women. According to a 2013 World Health Organization report, "It is estimated that over 508,000 women died worldwide in 2011 due to breast cancer" . Breast cancer can be cured and prevented in the primary stages. However, many women are diagnosed with cancer when it is too late. Breast cancer occurs in the breast cells, fatty tissues or fibrous connective tissues within the breast. Breast cancer tumors tend to gradually worsen and grow faster, which causes death. Although it is more common among women, it can also occur among men. Different factors such as age and family history can also increase the risk of breast cancer. Breast cancer tumors are classified into two classes, benign and malignant. Benign tumors are not dangerous to health while malignant tumors can be hazardous [1].

Benign: This tumor type is not dangerous for the human body and rarely causes death in humans. This type of tumor grows in one part of the body and has limited growth.

Malignant: A malignant tumor is more dangerous and can cause death in humans. This type of tumor grows rapidly because of the abnormal growth of cells.

The main types of breast cancer are invasive ductal carcinoma, ductal carcinoma in situ and invasive lobular carcinoma. Ductal carcinoma in situ is the earliest stage of breast cancer and is curable.

Invasive ductal carcinoma originates in the milk duct and is the most common breast cancer. Invasive lobular carcinoma can quickly spread to lymph nodes and other areas of the body. It starts in a lobule of the breast.

Approximately one million women are diagnosed with breast cancer every year worldwide. In the early stage, the rate of survival can be high; the five-year survival rate at this stage is 81%. However, only 35% women with late or advanced-stage breast cancer survive for five years.

This paper proposes a novel ensemble classification method for breast cancer classification using state-of-the-art machine learning classification methods. We evaluated the performance of the following classification algorithms: simple logistic Regression learning, stochastic gradient descent learning and multilayer perceptron network, random decision tree method, random decision forest method, sequential optimization method for support vector machine learning, K-nearest neighbor classifier, and Naïve Bayes classification. The prediction of these three best classification algorithm are then used for ensemble classification. For ensemble classification, we used unweighted voting mechanism including majority-based voting and four minimum probabilities, maximum probabilities, product of probabilities, and the average of probabilities voting mechanisms. The performance of proposed approach is evaluated on the publicly available Wisconsin Breast Cancer Dataset.

The rest of the manuscript is arranged as follows: Section 2 reviews the existing work on breast cancer tumor prediction and discuss the importance of use of appropriate performance measure in medical diagnostics. In Section 3, the proposed methodology is explained in detail, followed by the experimentation and results discussion in Section 4. Section 5 concludes the proposed work and suggests some future direction of research for breast cancer tumor classification.

2. Literature Review

Haifeng et al. [2] proposed a novel method for breast cancer prediction using data mining techniques. They formulated an effective way to predict breast cancer based on patients' clinical records. They evaluated the results of two publicly available datasets, the WBCD and the Wisconsin Breast Cancer(Diagnostic) dataset. They evaluated the performance of support vector machines (SVMs), eight hybrid learning models, artificial neural networks (ANNs), Naive Bayes classification and AdaBoost Tree. They proposed a hybrid model based on principal component analysis and other data mining models for feature reduction and suggested that other models such as k-mean could be used for feature space reduction.

In contrast, Quang et al. [3] used both supervised and unsupervised classification models for the breast cancer classification. They proposed the combination of scaling and principal component analysis for feature selection. They proved that an ensemble voting approach is the best breast cancer prediction model. After feature selection, various classification models were tested and trained on the data. Among all the models used for the prediction, only four models: ensemble-voting classifier, logistic regression (LR), SVM and AdaBoost shows better performance, with accuracy around 90% based on the models' results for precision and recall, Area Under the Receiver Operating Characteristics (AUC-ROC), F1 measure and computational time.

Ahmed et al. [4] used decision trees (DTs), ANNs and SVMs [17] to predict breast cancer classification. The implementation of different algorithms indicated that SVMs [5] outperform other classifiers on the WBCD. DT, ANN and SVMs showed 93.6%, 94.7% and 95.7% accuracy, respectively. SVMs show accurate results with the lowest error rate and highest accuracy and DTs show the lowest accuracy. Ten-fold cross validation is used for evaluation.

Mandal et al. [6] used LR, Naive Bayes (NB) and DTs for breast cancer classification. They also analyzed the time complexity of each classifier. LR outperforms other classifiers with the highest accuracy. However, Borges et al. [7] evaluated the performance of Bayesian networks and DT for breast cancer classification and found that Bayesian networks perform better, with 97.80% accuracy.

Chaurasia et al. [8] used Bayesian theorem, DT, radial basis function network and J48 for breast cancer prediction. They evaluated the results on the dataset acquired from the UCI repository and applied preprocessing, data selection and data transformation to form a prediction model. They found that Naive Bayes outperforms other models with classification accuracy of 97.36%. RBF Network and J48 showed classification accuracies of 96.77% and 93.41%, respectively.

Kumar et al. [9] predicted breast cancer using twelve classification algorithms: AdaBoost, J-Rip, LR, lazy learner, decision table, IBK, J48, lazy K-star, multiclass classifier, multilayer perceptron, random forest, Naive Bayes and random tree. They stated that other than Naive Bayes classification, the algorithms performed with greater than 94% accuracy and that Lazy and Tree classifications outperformed other classification algorithms, with 99% accuracy.

Most of the reported literature evaluates the performance of the proposed method based on 'accuracy'. Accuracy is higher when the occurrence of true positive (tp) and true negative (tn) is higher as compared to the false positive (fp) and false negative (fn). Other than accuracy, precision and recall are also important for performance reporting. However, for medical diagnostic the performance of artificial intelligence systems should consider the importance of false negative (recall) more than the false positive (precision). Therefore, it is important to evaluate the performance based on f-measures that weigh recall more than precision.

3. Proposed Methodology

In this paper, a novel algorithm is proposed for breast cancer tumor prediction using a voting mechanism. The publicly-available WBCD is used for evaluation and comparison with existing state-of-the-art methods. As shown in fig. 1, the input dataset is supplied to the different classification learning models to detect malignant and benign tumors.

The performance of eight different state-of-the-art machine learning classification algorithms were tested for tumor classification: simple logistic regression model, stochastic gradient descent learning, multilayer perceptron network, random decision tree method, random decision forest method, sequential minimal optimization method for support vector machine learning, k-nearest neighbor classification and Naive Bayes classification algorithm.

These classifiers are evaluated and three classification algorithms are selected based on the best performance calculated using f-measure. These selected classifiers are used to proposed a new hybrid ensemble classification method based on a voting mechanism. In this paper, two different voting mechanism were evaluated: hard voting (majority-based voting) mechanism and soft voting mechanism. Soft voting mechanism further includes average of probabilities voting, product of probabilities, minimum of probabilities and maximum of probabilities voting methods.

The classification results of the three classifiers are supplied to the ensemble classification method, which provides a final classification result. Details of the classification models and voting mechanism used are given in the following sub-sections.

3.1. Classification Methods

3.1.1. Simple Logistic Regression Model

The LR classification model [10] is a popular choice for modeling binary classifications. For this model, the conditional probability of one of the two output classes is assumed to be equal to a linear combination of the input features [11]. The logistic equation used for modeling this classification model is:

$$Z_i = ln(\frac{P_i}{1 - P_i}) \tag{1}$$

where P is the probability of the occurrence of event i.

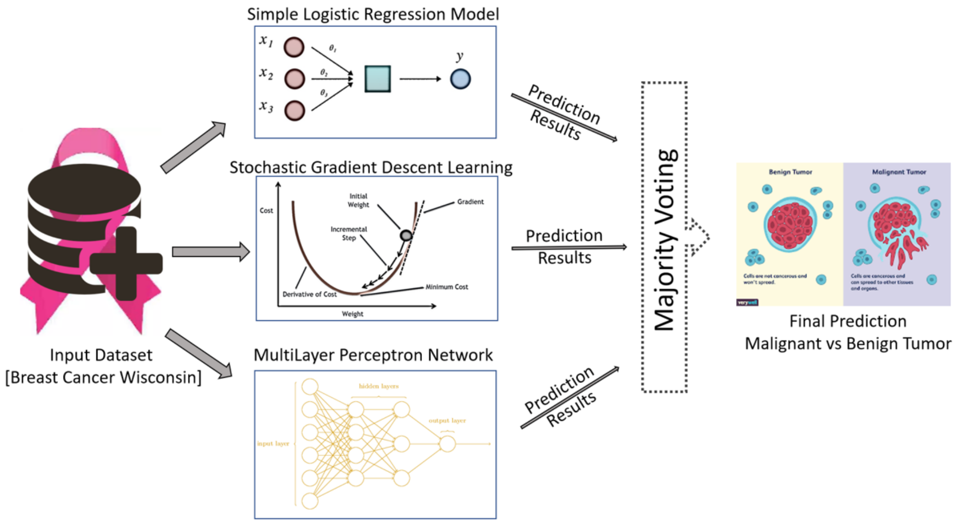

Figure 1. A hybrid ensemble method based on majority-based voting mechanism for breast cancer tumor classification using different machine learning models

3.1.2. Stochastic Gradient Descent Learning

Gradient descent is one of the most popular optimization algorithms used in deep learning and machine learning algorithms. Stochastic gradient descent [12] selects random samples from a dataset instead of selecting the batch data as in batch gradient descent. Stochastic gradient descent optimization performs iteration using the single sample to perform each iteration. The samples are randomly shuffled and selected for the respective iteration. The cost function at each iteration is calculated as:

$$for\ i\ in\ range\ (m) : \theta_j \; - \; \alpha \; (\widehat{y^i} \; - \; y^i) \; x_j^i \tag{2}$$

3.1.3. Multilayer Perceptron Network

Multilayer perceptron (MLP) [13] is a deep artificial neural network that contains many perceptrons. It is structured like a neural network and divided into three groups of layers, the input layer, output layer and the (hidden) layers between these two. These hidden layers hold the key to multilayer perceptron computation. MLPs are often applied on supervised learning algorithms and learn the model based on the correlation between input and output variables.

3.1.4. Random Decision Tree

Random decision tree is one of the popular supervised machine learning algorithm used for the graphical representation of all the possible solutions [14]. The decisions are based on some conditions and are easy to interpret. It identifies and chooses the significant attributes that are helpful in classification. It selects only those attribute that returns the highest information gain (IG). IG is defined as:

$$IG \; = \; E(ParentNode) \; - \; AverageE(ChildNodes) \tag{3}$$

where Entropy (E) is defined as:

$$E \; = \; \sum_i -Prob_i \; (\log_2 Prob_i) \tag{4}$$

and $Prob_i$ is the probability of class i.

3.1.5. Random Decision Forest

Random decision forest is similar to the bootstrapping algorithm with decision tree (CART) model. Random decision forest tries to build k different decision trees by picking a random subset S of training sample [15]. It generates the full Iterative Dichotomiser 3 (ID3) [16] trees with no pruning. It makes a final prediction based on the mean of each prediction. Random decision trees can interpret and handle irrelevant attributes in a simple manner. It is very compact and can handle missing data.

3.1.6. Sequential Minimal Optimization methods

SMO [17] is the optimization method for Support Vector Machine (SVM) [18] training problem. SMO is more efficient than the quadratic programming solvers. SMO uses heuristics to divide the whole training problem into minor problems so that these smaller problems can be solved analytically. Usually, it reduce the training time by a wide margin. SMO uses "John Platt's sequential minimal optimization algorithm" [19] for training an SVM [20].

3.1.7. K-Nearest Neighbor Classification

KNN [21] stores all the training data and classifies the query data based on a similarity measure. In KNNs, k is used to refer to the number of nearest neighbors that are to be included in the voting processes. KNN use feature similarity. To get better performance, KNN parameter tuning is done by choosing the right value of k. The similarity between two points is calculated using the Euclidean distance. The distance between *(P1(a,b) and P2(c,d))* is defined as:

$$dist(d) = \sqrt{(a-c)^2 + (b-d)^2} \tag{5}$$

3.1.8. Naive Bayes Classification

Naive Bayes classification algorithm [22] is normally famous for its simplicity as well as effectiveness. The Naive Bayes classification model is fast to build and it makes quick predictions. Naive Bayes is the probabilistic classifier and it learns the probabilities of features based on the target class. It assumes that the occurrence of a particular attribute is independent of the occurrence of the other attributes. Even if it depends on the other attributes, Naïve Bayes can show better performance as it does not require accurate probabilities estimates provided that the highest probability is allocated to the correct class. It is based on the Bayes theorem which states that:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \tag{6}$$

where $P(A \mid B)$ and $P(B \mid A)$ are the conditional probabilities of occurrence of event A given that event B is true and vice versa. *A, P(A), P(A | B) and P(B | A)* are called proposition, prior probability, posterior probability and likelihood.

3.2. Voting Mechanism

3.2.1. Majority-Based Voting Mechanism (Hard Voting)

Majority-based voting [23] is most widely used in ensemble classification. It is also known as plurality voting. In the proposed approach, after applying the three above-mentioned classification algorithms, a majority-based voting mechanism is used to improve the classification results [24]. Each of these model classification results is computed for each test instance and the final output is predicted based on the majority results. In majority voting, the class label y is predicted via majority (plurality) voting of each classifier C:

$$y = mode\ \{C_1(x),\ C_2(x),\ ..,\ C_n(x)\} \tag{7}$$

3.2.2. Soft Voting

In soft voting, the final output is predicted based on the predicted probabilities p for the classifiers [25]. In each case below, the probability of class labels assigned by the classifier C to input x is defined as:

- Average of Probabilities Voting Mechanism:

$$y = AVERAGE \ \{ C_1(x), C_2(x), ..., C_L(x) \} \tag{8}$$

- Product of Probabilities Voting Mechanism:

$$y = PROD \ \{ C_1(x), C_2(x), ..., C_L(x) \} \tag{9}$$

- Minimum of Probabilities Voting Mechanism:

$$y = MIN \ \{ C_1(x), C_2(x), ..., C_L(x) \} \tag{10}$$

- Maximum of Probabilities Voting Mechanism:

$$y = MAX \ \{ C_1(x), C_2(x), ..., C_L(x) \} \tag{11}$$

4. Experimentation and Results

4.1. Wisconsin Breast Cancer Dataset

The publicly available WBCD was analyzed (it can be downloaded from the UCI repository). This dataset contains the information taken from the microscopic examination of breast masses. A digital scan of fine-needle aspirates is used to compute features. Fine-needle aspirate is one of the best methods to evaluate the presence of malignant tumors.

This dataset comprises of 569 instances. Each instance contains 32 features that are extracted from images of nuclei. These include the nucleus texture, radius, perimeter, smoothness, area, compactness, concave points, concavity,fractal dimension and symmetry. The remaining features are computed by taking the mean, standard error and worst or largest mean of the above-mentioned features. Figure 2 shows the visualization of the above-mentioned features. The first and second features shown in Figure 2 are ID and diagnosis class (malignant/benign). Results shown in red and blue represent benign and malignant classes, respectively. A 70%-30% training-testing split is used for evaluation, as commonly reported in the literature.

4.2. Performance Evaluation Measures

Given the true positive (TP), false positive (FP), true negative (TN) and false negative (FN) counts, the following performance evaluation measures were calculated:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$F_\beta = (1+\beta^2) \frac{Precision * Recall}{(\beta^2 * Precision) + Recall} \tag{15}$$

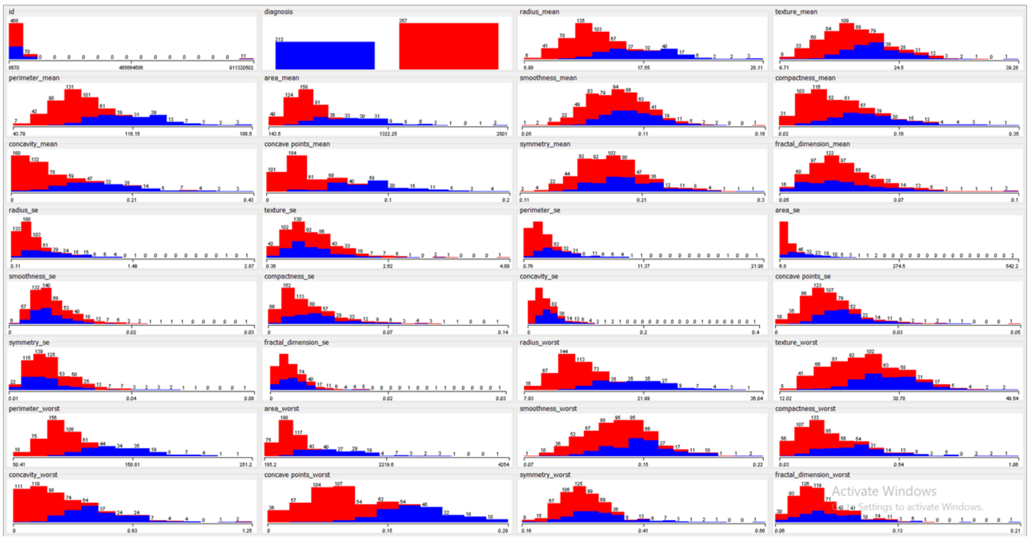

Figure 2. Feature visualization result for WBCD

Table 1. Performance of state-of-the-art machine learning algorithms on WBCD.

Classification Algorithms	Accuracy	Precision	Recall	F1 Score	F2 Score	F3 Score
Simple Logistic Regression Learning	98.246%	0.9830	0.9820	0.9825	0.9822	0.9821
Stochastic Gradient Descent Learning	97.876%	0.9791	0.9789	0.9710	0.9710	0.9710
Multilayer Perceptron Network	97.661%	0.9770	0.9770	0.9770	0.9770	0.9770
Random Decision Tree Method	91.813%	0.9200	0.9180	0.9190	0.9184	0.9182
Random Decision Forest Method	96.491%	0.9650	0.9650	0.9650	0.9650	0.9650
Sequential Minimal Optimization Method	97.076%	0.9710	0.9710	0.9710	0.9710	0.9710
K-Nearest Neighbor Classification	97.076%	0.9710	0.9710	0.9710	0.9710	0.9710
Naïve Bayes Classification	91.813%	0.9190	0.9180	0.9185	0.9182	0.9181

where β^2 is a positive real number. For F1, F2, F3 score calculation, β is set to 1, 2 and 3 respectively.

In medical diagnostics research, precision and recall are the main metrics used for reporting performance. F-measure is also important in reporting performance in medical diagnostics, because it combines precision and recall into a single figure which is simpler to use for comparisons. In the medical community, a false negative is normally more devastating than a false positive therefore, recall is considered to be a more important measurement. An FP (precision) might not be as important as an FN (recall) because FPs can be discarded by clinicians but a missed condition could be very serious for a patient. In the light of that a modified form of F1 that weighs recall more heavily than precision is a better metric. Therefore in this work F2 and F3 have been used.

4.3. Results and Discussion

The performances of eight different classification algorithms are evaluated on the WBCD dataset. Table 1 shows the results. The SLR model chooses the attributes that results at lowest possible squared error. SGD learning removes all the missing values and converts nominal attributes into binary ones. All attributes are also normalized. MLP learning uses a back propagation algorithm for learning. All the nodes in MLP network are sigmoid. At each node, random decision tree considers randomly chosen k attributes without performing pruning. The SMO method makes use of John Platt's optimization algorithm for training SVM. KNN selects an appropriate k based on cross validation and performs distance weighting for learning. Fig. 3 compares the performance of the all eight classifiers in terms of F3 score. The simple logistic regression (SLR) learning model, stochastic gradient descent (SGD) learning and multilayer perceptron network (MLP) show better performance in terms of F3 score than the other five classification algorithms.

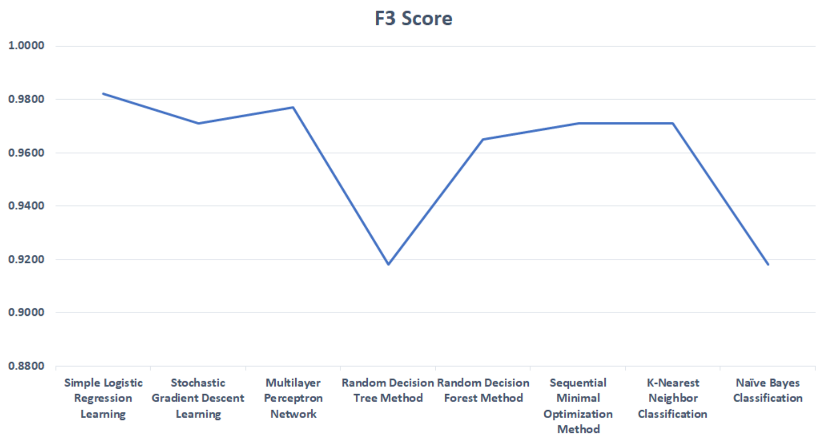

Figure 3. F3 score of state-of-the-art classification algorithms for WBCD

Table 2. Evaluation results of the proposed voting mechanisms for the WBCD

Proposed voting mechanism	Accuracy	Precision	Recall	F1 Score	F2 Score	F3 Score
Majority-based voting mechanism	**99.42 %**	**0.9940**	**0.9940**	**0.994**	**0.9940**	**0.9940**
Average of probabilities voting mechanism	98.83 %	0.989	0.988	0.9885	0.9882	0.9881
Product of probabilities voting mechanism	98.12 %	0.9850	0.9850	0.9850	0.9850	0.9850
Minimum of probabilities voting mechanism	98.46 %	0.986	0.981	0.9835	0.9820	0.9815
Maximum of probabilities voting mechanism	99.41 %	0.9840	0.9840	0.9840	0.9840	0.9840

After analyzing these results, the three best classifiers were selected based on F3 score i.e. simple logistic learning, stochastic gradient descent learning and multilayer perceptron network and passed their classification results to a different voting mechanism. This proposed mechanism predicts the final output based on the voting of the three participating classification algorithms. For ensemble method classification, different unweighted voting schemes are used i.e. minimum probabilities, maximum probabilities, majority voting, product of probabilities, and the average of probabilities,.

Table 2 shows the results of the voting mechanisms for the WBCD. The results show that majority-based voting mechanism performs better than the other voting mechanisms. Usually it is assumed that soft voting performs better than hard voting as it considers more information by using individual classifier's uncertainty in the final prediction. However, these experiments show that for WBCD, majority-based voting (hard voting) performs better than the soft voting mechanisms i.e. average of probabilities voting, product of probabilities, maximum of probabilities and minimum of probabilities.

4.4. Comparison with Existing Work

Table 3 shows a comparison between the proposed majority-based voting mechanism with existing work on breast cancer prediction using the WBCD. Nahato et al. [26] used a back propagation neural network. Chen et al. [27] Kumari et al. [28] and Dumitru et al. [29] used conventional prediction algorithms like support vector machines, K-nearest neighbor and Naive Bayes, respectively. Liu et al. [30] proposed a novel evolutionary neural network approach and Nguyen et al. [3] used feature scaling and principle component analysis prior to model training. As far as we can ascertain, the proposed approach performs better than any in the literature to date.

5. Conclusion and Future Research Directions

In this research, we first discussed and reviewed the existing state-of-the-art method for breast cancer classification. By analyzing the results of the existing work, we realized that the existing literature mostly relies on the reporting the performance in term of accuracy and ignore the importance

Table 3. Comparison with the existing work for breast cancer prediction using WBCD

Work	Proposed Method	Accuracy
Ours	Majority-based voting mechanism	**99.42**%
Nahato et al., [26]	Backpropagation neural network	98.60%
Liu et al., [30]	An evolutionary artificial neural network	97.38%
Chen et al., [27]	A support vector machine classifier	89.20%
Kumari et al., [28]	K-Nearest neighbor classification algorithm	99.28%
Dumitru et al., [29]	Naive bayesian classification	74.24%
Shaikh et al.,[31]	Dimensionality reduction and support vector machine	97.91%
Nguyen et al., [3]	Feature selection and ensemble voting	98.00%
Alickovic et al.,[32]	Normalized multi layer perceptron neural network	99.27%

of the false negative results. Therefore, we proposed to use the modified form of F-measure (F3 score) that weigh recall more than precision.

In this paper, we proposed a hybrid ensemble classification by evaluating the performance of simple logistic regression learning, stochastic gradient descent learning and multilayer perceptron network, random decision tree method, random decision forest method, sequential optimization method for support vector machine learning, K-nearest neighbor classifier, and Naive Bayes classification algorithms. We reported the performance of these eight classifiers using different performance measures i.e. accuracy, precision, recall, F1 score, F2 score, and F3 score. Later we selected the three classifiers with the best F3 score and proposed a hybrid ensemble method using voting mechanism.

We used different unweighted voting mechanisms including majority-based voting, average of probabilities, product of probabilities, minimum of probabilities and maximum of probabilities. The majority-based voting mechanism shows better f3 score than the other voting mechanism.

For future research, we plan to evaluate different feature selection algorithms that can help us determine the smallest subset of features that can assist in accurate classification of breast cancer as either benign or malignant. Instead of unweighted voting mechanism, researchers can evaluate the performance of different weighted voting mechanisms including simple weighted voting, rescaled weighted voting, best-worst weighted voting, and quadratic best-worst weighted voting. The finding of this study can be a good starting point for breast cancer classification using the image dataset too.

Acknowledgment

Adel Saad Assiri would like to express his gratitude to King Khalid University, Saudi Arabia for providing administrative and technical support.

References

1. Sharma, G.N.; Dave, R.; Sanadya, J.; Sharma, P.; Sharma, K. Various types and management of breast cancer: an overview. *Journal of advanced pharmaceutical technology & research* **2010**, *1*, 109.
2. Wang, H.; Yoon, S.W. Breast cancer prediction using data mining method. IIE Annu. Conf. Expo 2015, 2015, pp. 818–828.
3. Nguyen, Q.H.; Do, T.T.; Wang, Y.; Heng, S.S.; Chen, K.; Ang, W.H.M.; Philip, C.E.; Singh, M.; Pham, H.N.; Nguyen, B.P.; others. Breast Cancer Prediction using Feature Selection and Ensemble Voting. 2019 International Conference on System Science and Engineering (ICSSE). IEEE, 2019, pp. 250–254.
4. Ahmad, L.G.; Eshlaghy, A.; Poorebrahimi, A.; Ebrahimi, M.; Razavi, A.; others. Using three machine learning techniques for predicting breast cancer recurrence. *J Health Med Inform* **2013**, *4*, 3.
5. Nazir, S.; Ghazanfar, M.A.; Aljohani, N.R.; Azam, M.A.; Alowibdi, J.S. Data analysis to uncover intruder attacks using data mining techniques. 2017 5th International Conference on Information and Communication Technology (ICoIC7). IEEE, 2017, pp. 1–6.

6. Mandal, S.K. Performance analysis of data mining algorithms for breast cancer cell detection using Naïve Bayes, logistic regression and decision tree. *International Journal Of Engineering And Computer Science* **2017**, *6*, 20388–20391.
7. Borges, L.R. Analysis of the Wisconsin Breast Cancer Dataset and Machine Learning for Breast Cancer Detection. *Group* **1989**, *1*.
8. Chaurasia, V.; Pal, S.; Tiwari, B. Prediction of benign and malignant breast cancer using data mining techniques. *Journal of Algorithms & Computational Technology* **2018**, *12*, 119–126.
9. Kumar, V.; Mishra, B.K.; Mazzara, M.; Verma, A. Prediction of Malignant & Benign Breast Cancer: A Data Mining Approach in Healthcare Applications. *arXiv preprint arXiv:1902.03825* **2019**.
10. Landwehr, N.; Hall, M.; Frank, E. Logistic model trees. *Machine learning* **2005**, *59*, 161–205.
11. Sumner, M.; Frank, E.; Hall, M. Speeding up logistic model tree induction. European conference on principles of data mining and knowledge discovery. Springer, 2005, pp. 675–683.
12. Bottou, L. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*; Springer, 2010; pp. 177–186.
13. Pal, S.K.; Mitra, S. Multilayer perceptron, fuzzy sets, and classification. *IEEE Transactions on neural networks* **1992**, *3*, 683–697.
14. Rokach, L. Decision forest: Twenty years of research. *Information Fusion* **2016**, *27*, 111–125.
15. Daho, M.E.H.; Chikh, M.A. Combining bootstrapping samples, random subspaces and random forests to build classifiers. *Journal of Medical Imaging and Health Informatics* **2015**, *5*, 539–544.
16. Khedr, A.E.; Idrees, A.M.; El Seddawy, A.I. Enhancing Iterative Dichotomiser 3 algorithm for classification decision tree. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2016**, *6*, 70–79.
17. She, J.; Schmidt, M. Linear convergence and support vector identification of sequential minimal optimization. 10th NIPS Workshop on Optimization for Machine Learning, 2017, p. 5.
18. Nazir, S.; Yousaf, M.H.; Velastin, S.A. Inter and intra class correlation analysis (IICCA) for human action recognition in realistic scenarios **2017**.
19. Ibarra, J.B.; Caya, M.V.C.; Bentir, S.A.P.; Paglinawan, A.C.; Monta, J.J.; Penetrante, F.; Mocon, J.; Turingan, J. Development of the Low Cost Classroom Response System Using Test-Driven Development Approach and Analysis of the Adaptive Capability of Students Using Sequential Minimal Optimization Algorithm. 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA). IEEE, 2019, pp. 689–693.
20. Nazir, S.; Yousaf, M.H.; Velastin, S.A. Feature Similarity and Frequency-Based Weighted Visual Words Codebook Learning Scheme for Human Action Recognition. Pacific-Rim Symposium on Image and Video Technology. Springer, 2017, pp. 326–336.
21. Nazir, S.; Yousaf, M.H.; Velastin, S.A. Evaluating a bag-of-visual features approach using spatio-temporal features for action recognition. *Computers & Electrical Engineering* **2018**, *72*, 660–669.
22. Al-Sabbah, S.A.; Mohammad, S.F.; Eanad, M.M. Use of the Naive Bayes Function and the Models of Artificial Neural Networks to Classify Some Cancer Tumors. *Indian Journal of Public Health Research & Development* **2019**, *10*, 1563–1569.
23. Delgado, J.; Ishii, N. Memory-based weighted majority prediction. SIGIR Workshop Recomm. Syst. Citeseer, 1999.
24. Kang, X.b.; Lin, G.f.; Chen, Y.j.; Zhao, F.; Zhang, E.h.; Jing, C.n. Robust and secure zero-watermarking algorithm for color images based on majority voting pattern and hyper-chaotic encryption. *Multimedia Tools and Applications* **2019**, pp. 1–34.
25. Du, K.L.; Swamy, M. Combining Multiple Learners: Data Fusion and Ensemble Learning. In *Neural Networks and Statistical Learning*; Springer, 2019; pp. 737–767.
26. Nahato, K.B.; Harichandran, K.N.; Arputharaj, K. Knowledge mining from clinical datasets using rough sets and backpropagation neural network. *Computational and mathematical methods in medicine* **2015**, *2015*.
27. Chen, H.L.; Yang, B.; Liu, J.; Liu, D.Y. A support vector machine classifier with rough set-based feature selection for breast cancer diagnosis. *Expert Systems with Applications* **2011**, *38*, 9014–9022.
28. Kumari, M.; Singh, V. Breast Cancer Prediction system. *Procedia computer science* **2018**, *132*, 371–376.
29. Dumitru, D. Prediction of recurrent events in breast cancer using the Naive Bayesian classification. *Annals of the University of Craiova-Mathematics and Computer Science Series* **2009**, *36*, 92–96.

30. Liu, L.; Deng, M. An evolutionary artificial neural network approach for breast cancer diagnosis. 2010 Third International Conference on Knowledge Discovery and Data Mining. IEEE, 2010, pp. 593–596.
31. Shaikh, T.A.; Ali, R. Applying Machine Learning Algorithms for Early Diagnosis and Prediction of Breast Cancer Risk. Proceedings of 2nd International Conference on Communication, Computing and Networking. Springer, 2019, pp. 589–598.
32. Alickovic, E.; Subasi, A. Normalized Neural Networks for Breast Cancer Classification. International Conference on Medical and Biological Engineering. Springer, 2019, pp. 519–524.