A Novel Ensemble Model for Breast Cancer Diagnosis

Aalif Hamid

Abstract Breast cancer is one of the world's second-leading causes of death. One out of every eight women experiences this illness at some point in their lives. Patients with early detection of the disease have a higher likelihood of survival and have a better chance of recovery. There is a critical requirement to classify tumors for malignancy. This study investigates automatic breast cancer prediction using a variety of machine learning techniques. The most prevalent type of classification is binary (benign cancer versus malignant cancer). Machine learning (ML) techniques are now being employed to detect breast cancer. They have a high degree of classification accuracy and a wide range of diagnostic capabilities. This research work presents a novel ensemble model for binary classification of breast mass tumors on the WDBC dataset. The outcomes of seven distinct machine learning individual models (logistic regression, KNN, SVC, etc.) are compared to the hybrid ensemble model of the above seven techniques. The evaluation of the model is done in terms of numerous performance indicators, such as accuracy, precision and recall. Compared to stand-alone models, the results demonstrate that the ensemble model performs remarkably well, with an accuracy of 0.98%.

Keywords Machine learning · Ensemble · Breast cancer · Support vector machine · Stacking classifier

1   Introduction

Cancer is a chronic condition caused by excessive growth and cell division that affects the cellular structure. It causes accelerated cell growth in some cases, depending on cancer type, and in others, cell growth is significantly decreased. This division abnormality affects the body in numerous states, resulting in cancers, undermining the immune system, and other significant complications. A record 15.5 million persons with a history of cancer were residents of the United States alone, according to the

A. Hamid (✉)
Department of Information Technology, Central University of Kashmir, Ganderbal, Kashmir, India

2018 report. According to the World Health Organization (WHO) [19], the illness will have a massive effect on the planet. About one in five individuals will face cancer diagnosis once in their lifetime by 2020. 18.1 million persons worldwide had cancer in 2018, and 9.6 million died of the disease. The figures will likely double by the year 2040. Accelerating global cancer containment by early diagnosis, detection, care and observation is the need of the hour. Pulmonary cancer is the most frequently diagnosed cancer (11.6%), followed by female breast cancer (11.6%) and bowel cancer (10.2%). The study's primary objective is to apply machine learning to diagnose the disease early, potentially saving millions of lives worldwide. The use of several machine learning techniques such as decision trees, artificial neural networks (ANN) and support vector machines (SVM) in cancer diagnosis yields positive outcomes. The adoption of machine learning in the medical field will transform disease diagnosis and help doctors get better insights into the disease.

Breast cancer is the most common type of cancer in women [2], affecting about 2.1 million women each year and resulting in the highest number of cancer-related fatalities. Breast cancer claimed the lives of around 15% of cancer-related deaths in 2018 or 627,000 women. Research conducted by [1] shows that one woman dies every 13 min due to breast cancer, while one case of breast cancer gets diagnosed every two minutes worldwide. Since the shift in clinical practice from cure-based evidence medicine to care-based evidence medicine, disease detection and identification have been given top importance while still in the early stages of development. Breast cancer death rates can be reduced if early identification of the disease is possible. An essential and often used diagnostic method is the triple-test; it consists of three medical examinations, i.e., self-examination, ultrasonography and fine needle biopsy (FNB) of the breast used to produce high accuracy in the disease diagnosis. FNB is an invasive technique that consists of obtaining material from the tumor directly. The examination of the content is performed under the microscope for the occurrence of cancer cells.

Data mining is a discipline in which computer science and analytics converge to detect correlations in the database. The data mining process's fundamental goal is to retrieve and shape valuable information from the data file into a coherent framework for potential use. It employs a combination of strong analytical abilities, domain knowledge and an explicit knowledge base to uncover hidden trends and patterns in heterogeneous information [7]. Many forensic organizations, pharmacies, hospitals and research institutions have many medical diagnosis data available. To make the system computerized and rapid to identify diseases, it is barely necessary to classify them [9]. Forecasting trends can aid in the development of new treatments by assisting in the identification of high-risk individuals based on known pathological and experimental risk factors. Accurate breast cancer risk assessments can help with lifestyle changes, immunotherapy, personalized screening and risk-stratified follow-up therapy [2]. The suggested research project is a study to see if a breast tumor is benign or malignant. In this regard, machine learning has been applied to the dataset. Machine learning is a series of instruments that are used to build and test algorithms that enable inference, identification of patterns and classification. Machine learning focuses on four steps: data collection, a compilation of models,

model training and model verification [20]. The association between breast cancer and machine learning is not new. It has been used for decades to diagnose cancers and other autoimmune disorders, predict gene mutations that trigger cancer and evaluate the prognostic. Data mining and machine learning techniques are widely used in the medical industry since they are discovered to be quite useful in medical practitioners' decision-making processes. The implementation of multiple machine learning algorithms such as support vector machine (SVM), artificial neural network (ANN) and decision trees has shown promising results in the field of cancer detection. In addition to transforming illness identification, the adoption of artificial learning in medical imaging can enable doctors to obtain more in-depth insights into the disease.

The triple-test is an essential and frequently used diagnostic method; it consists of three medical tests, i.e., self-examination, ultrasonography and breast fine needle biopsy (FNB) used to diagnose the disease to achieve high precision. FNB is an invasive technique that consists of directly obtaining tumor material. The contents are checked for the presence of cancer cells under the microscope. The study work uses the WDBC dataset. The dataset consists of attributes that measure a digitized description of a fine needle aspirate from the image of mammary mass (FNA). The features were recorded by analyzing the characteristics of the cell nuclei under the microscope.

The following is a description of the paper's structure. The past work in breast cancer diagnosis using machine learning approaches is briefly described in Sect. 2. The materials and procedures employed in the research study are described in Sect. 3. The work's experimentation results are discussed in Sect. 4, while the paper's conclusion is discussed in Sect. 5.

2 Related Work

Large number of researches has been done in the past decade on the diagnosis of breast cancer. Some researchers have used mammograms, histopathological images, while others have worked on calculative data. The ability of the machine learning algorithms to find new hidden patterns in a large amount of data has shown quite great results in the healthcare sector. Cancer diagnosis at an early stage increases the chances of recovery in patients. For the diagnosis of breast cancer, researchers use a variety of machine learning methods such as artificial neural networks (ANN), decision trees, support vector machines (SVM), convolutional neural network and random forest. On the WDBC dataset, SVM and ANN were applied, and the results of the applied machine learning (ML) techniques were compared using performance measures.

In comparison with ANN, SVM demonstrated the best accuracy of 96% based on the performance criteria of the employed ML approaches [8]. The dataset WDBC was used to test a feed forward backpropagation neural network. The ANN in this study comprises nine input and one output neuron, with 1, 2, 3, 4 and 5 hidden layers. The results demonstrate that using the TANSIG transfer function, a 20-neuron feed

forward backpropagation single hidden layer neural network using the TANSIG transfer function achieves the highest classification accuracy (98.9% on training set and 99% on test set) [10]. The methodology of adaptive feature selection was used, and classification was enabled using KNN, random forest and multilayer perceptron [22]. Results proved that before feature selection KNN, SVM and PNN obtained an accuracy of 100%, 100%, 96%, respectively, and after selection of best attributes, the accuracy got jumped to 97.19% for PNN. Breast Cancer Coimbra dataset was used on ANN, ELM, KNN and SVM algorithms by [23]; ELM is same as that of artificial neural network with one difference, and ELM consists of only one hidden layer instead of many hidden layers in case of ANN. Results showed that ELM outperformed the other two algorithms in terms of accuracy and training time and is preferable for more samples. The research [11] used six different algorithms, such as AdaBoost, random forest, extra tree classifier and the stacking classifier, on basic learners on the WDBC dataset and obtained an accuracy of 92.9% on voting classifier, with extra trees classifier receiving the highest score of 95.1739%, followed by XGBoost 95.1691% and AdaBoost 94.7343%. AlexNet architecture was used by [24] on the CBIS-DDSM dataset in which the CLAUDE algorithm for image enhancement was done with image segmentation to increase the sample size of the dataset. In the proposed work, linear SVM obtained the highest accuracy of 80.5%. Table 2 shows the brief summary of various methodologies adapted by researchers for breast cancer.

Table 1 Various breast cancer datasets	Database	Database size (GB)	Total patients
	MIAS	2.3	161
	DDSM	6	2620
	CBIS-DDSm	70.5	237
	ISPY1	76.2	237
	Breast-MRI-NACT-Pilot	19.5	64
	QIN-Breast	11.286	67
	Mouse-Mammary	8.6	32
	QIN Breast DCE-MRI	15.8	10
	BREAST-DIAGNOSIS	60.8	88
	RIDER Breast MRI	0.401	5
	BCDR	1.4 GB	1734
	TCGA-BRCA	53.92 (TB)	1098
	BreakHis	4	82
	Inbreast	40 MB	115
	Coimbra Breast Cancer	7 Kb	116
	WDBC	122 Kb	

There is a growing abundance of researchers drawing their interest attentions to ensembles. Moreover, it has solid proof that they can substantially change the classification performance. Six machine learning techniques were employed to analyze 8942 patients with breast cancer using the local dataset [12]. All methods, including decision trees, neural networks, extreme boost, logistic regression and support vector machines, produced very close accuracy of 79.80%, 82.7%, 81.7%, 81.7% and 81.7%, respectively, with random forest outperforming the others with an accuracy of 82.7%. Kadam et al. [25] suggested a feature ensemble model for categorizing breast cancer tumors into benign and malignant tumors based on sparse autoencoders and soft-max regression. In the analysis, ten-fold cross-validation was used that gave an impressive 98.60% result and showed that the proposed model performs other state-of-the-art techniques. The research presented by [26] is focused on the diagnosis of breast cancer using an ensemble learning algorithm based on SVM to decrease the diagnosis variance and increase the diagnosis accuracy. Twelve separate SVMs are hybridized using the weighted field under the receiver operating characteristic curve ensemble (WAUCE) approach. Wisconsin breast cancer, Wisconsin diagnostic breast cancer and the control, epidemiology and end outcomes (SEER) program breast cancer datasets have been studied to determine the efficacy of the suggested model. The results show that the WAUCE model achieves higher accuracy with a little reduced variance for breast cancer detection when compared to five additional ensemble mechanisms and two conventional ensemble models, namely adaptive boosting and bagging classification tree. The results show that the WAUCE model achieves higher accuracy with a little reduced variance for breast cancer detection when compared to five additional ensemble mechanisms and two conventional ensemble models, namely adaptive boosting and bagging classification tree. The datasets for breast cancer research are listed in Table 1.

3 Materials and Methods

The Wisconsin diagnostic dataset (WDBC) of breast cancer is used to measure the effectiveness of the proposed ensemble model for breast cancer diagnosis. This dataset has been collected by the University of Wisconsin Hospitals, Madison [27]. The dataset is available for download from the UCI machine learning repository. It is composed up of data derived from a digitized description of a fine needle aspirate of breast mass (FNA). The properties of the nuclei of the cells seen in Table 3 are identified by the characteristics. The dataset comprises a total of 32 attributes, of which 30 were used in the experiment as independent variables. There are a total of 569 instances of the patients recorded in the dataset.

Preprocessing is the method to remove the unwanted components from the dataset to fit the machine learning algorithms perfectly. WDBC dataset contains one independent categorical variable under the attribute name, "diagnosis". This column contains values in the form of benign or malignant (B/M). But the machine learning algorithm takes the input in the form of numbers (0/1); so, this column is preprocessed before

Table 2 Brief overview of diverse techniques for cancer diagnosis

Author	Technique	Dataset	Performance
[13]	Extracted characteristics using deep CNN architectures were proposed as a novel framework for the categorization of breast cancers	BreakHis	Accuracy = 97.5%
[14]	VGG-16 as a feature extractor combined with NN classifier for detection	MIAS-DDSM	Accuracy = 90.5%
[15]	Implemented CNN along with other machine learning algorithms	WDBC	Accuracy = 99.67%
[16]	The (SVM) model with radial basis function (RBF) kernel appeared to be the most successful classifier, according to the performance metrics of multicollinearity assessment on machine learning models	Coimbra Breast Cancer	Accuracy = 93.9%
[17]	Using a deep neural network with support value (DNNS), a novel technique to identify breast cancer was presented	Local Dataset	Accuracy = 97.21
[18]	For the diagnosis of breast cancer, an extreme learning machine model with a radial basis function (RBF) kernel was used	WDBC	AccuracyELM = 95.6 SVM = 88.9
[3]	Used convolutional neural networks, as well as several preprocessing strategies such as contrast scaling, dilation, cropping and decision fusion using an ensemble of networks, to tackle the challenge of breast microcalcification classification	DDSM	Recall = 97.3%

Table 3 Features of WDBC dataset [5]

Feature	Feature name
Feature 1	Patient ID
Feature 2	Diagnosis (B/M)
Feature 3	Nuclei radius
Feature 4	Nuclei texture
Feature 5	Nuclei perimeter
Feature 6	Nuclei area
Feature 7	Nuclei smoothness
Feature 8	Nuclei compactness
Feature 9	Nuclei concavity
Feature 10	Nuclei concave Points
Feature 11	Nuclei symmetry
Feature 12	Nuclei fractal dimension

fetching it into the algorithms. Label encoder which is available under Scikit-Learn package is used for the processing of the dependent variable. The dataset is divided using an 80:20 ratio, i.e., 80% of the data is used for algorithm training, while 20% of the data is the unseen data used to assess the performance assessment algorithms as suggested by [11]. Seven different machine learning algorithms like logistic regression, SVM extra trees classifier, Gaussian NB classifier, KNN, SGD classifier, etc., have been implemented in the research work. Combination of the said algorithms proved very beneficial and gave outstanding results. Implementation of work is done on the WDBC dataset which is preprocessed first and split into testing and training data.

3.1 Implementation of Work

The capability of computers to train from experience without being instructed personally is machine learning [21]. Both rules are coded in the standard programming style, and the computer can produce output depending on the logical argument, i.e., the rules. However, more rules need to be written as the structure gets complicated, so it becomes unfit to manage. However, it is assumed that machine learning can fix this dilemma. The computers here learn how the input and output data are associated and then write a law accordingly. Any time a new data or condition occurs, the programmer does not need to write new rules. The main focus of machine learning is learning and inference. Machine learning is graded into supervised, unsupervised and reinforcement learning [6]. In this research work, supervised machine learning algorithms such as logistic regression, random forest and K-nearest neighbor (KNN) are implemented.

Logistic Regression

By incorporating data onto a straight line, linear regression algorithms predict the values. It is said that these algorithms are unbounded since the value is not simply between 0 and 1. So, they are not ideal for problems with grouping, giving rise to "Logistic Regression". Logistic regression is a statistical model that is better adapted for problems of classification. "These algorithms use a mathematical function called "Sigmoid" to squeeze a number between 0 and 1.Two logistic regression algorithms were implemented in this research work, one with the solver equal to "saga" and the other with the solver parameter equal to "lbfgs".

Random Forest Classifier

The classifier averages the number of classifiers (decision-tree classifiers) on various data sub-samples. To boost prediction accuracy, it uses averaging of all classifiers. A parameter called "Criterion" is used in the random forest classifier. It is a feature that tests the efficiency of the division. This split parameter performs the task of splitting the data into clusters and should be so that a single class (0 or 1) outweighs

each category. By applying variability, random forest models minimize the risk of overfitting by:

Constructing many trees (n-estimators)
Drawing substitute findings (i.e., a bootstrapped sample)

The best split nodes are split between a random subset of features chosen for each node.

A random forest classifier is implemented in the ensemble model with n_estimators set to 10 and criterion parameter set to "gini".

Extra Trees Classifier

This puts together a variety of randomized decision trees on various data sub-samples. The average estimate of individual classifiers is the approximation of this ensemble. We used two extra tree classifier models with distinct parameters in our ensemble model, as described in the previous classifier. Extra trees is like random forest. It creates several trees and breaks nodes using random subsets of characteristics, but with two key differences: It does not bootstrap observations (meaning it tests without replacement), and nodes are broken on random splits rather than best splits. An extra tree classifier is implemented in the ensemble model with the same n_estimators as the random forest classifier and criterion parameter set to "gini".

KNN

It is among the straightforward and non-parametric techniques for classification that stores and groups all data based on some similarity function (e.g., distance). The plurality vote of neighbors controls the classification/grouping of any data point. The number of neighbors eligible for voting is determined by the K. In this research work, K is set to 5.

Support Vector Classifier

It is a supervised learning technique which is most commonly used to solve classification problems. Data points reflect the dataset in the SVM. SVM constructs hyperplanes that have the highest margin in multi-dimensional space to categorize the results. A margin represents the longest interval between the closest data points. The linear kernel is proven to show promising results. SVC with the linear kernel is used in the model.

Gaussian NB Classifier

This model utilizes the Gaussian Naïve Bayes algorithm to predict/forecast the result. The constant values connected to each attribute are considered to have a Gaussian distribution here, i.e., Gaussian is said to be the likelihood of attributes. All default values are used for this algorithm in the model.

SGD Classifier

Stochastic gradient descent (SGD) is a simple but powerful optimization algorithm used to find the parameter/feature coefficient values that minimize a cost function.

In other words, it is used under convex loss functions such as SVM and logistic regression for discriminatory learning of linear classifiers. Since the update to the coefficients is done for each training instance, it has been successfully extended to large-scale datasets rather than at the end of cases. Each parameter is set to its default values for this algorithm in the proposed work.

Ensembling Approach

The research work combines the eight different sets of algorithms to boost the performance of the model. All the algorithms are merged and passed to the stacking classifier (voting classifier). A voting classifier is a machine learning model that trains on a range of diverse models and selects an output or class based on its most excellent chance of determining the class as the output. It essentially aggregates the outcomes of each classifier passed into the stacking classifier and, based on the most significant majority of votes, forecasts the output class. Rather than creating individual-specific models and finding each of them' performance, we build a specific single model that trains these models and forecasts output based on their cumulative plurality of votes for each type of output. It supports two types of voting: hard voting and soft voting. In hard voting, the predicted result class is a class with the most significant majority of votes in hard voting, i.e., the class that was most likely to be predicted by one of the classifiers. In soft voting, the output class is the forecast for soft voting, based on the average likelihood given to that class. Hard voting is implemented in the research work. The ensemble model consisting of eight algorithms surpasses the individual model performance and achieves an accuracy of 98.2%. The category of stand-alone models surpasses the individual models in cancer diagnosis, classifiers and gain competitive outcomes (Fig. 1).
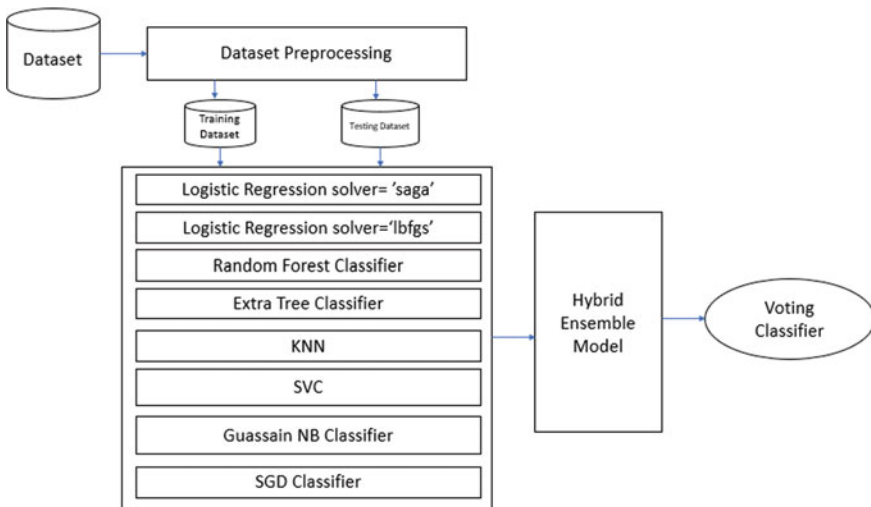

Fig. 1 Proposed ensemble model

4   Results and Discussion

4.1   Results Using Individual Classifiers

Each algorithm's confusion matrix is calculated. Following data preprocessing, the performance of the classifiers is shown using several performance parameters like accuracy, F1 score, recall, precision, etc.

   The confusion matrix for the actual and forecast classes is expressed in true positive (TP), which indicates that the original and predicted classes were correctly classified [4]. True negative (TN) is incorrectly classified. False positive (FP) indicates that something has been misclassified. FN stands for false negative, which indicates that something has been misclassified. All classifiers' performance is evaluated using these formulas.

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{1}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{2}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

$$F1\text{-}score = 2 * \left(\frac{Precision * Recall}{Precision + Recall} \right) \tag{4}$$

   Before testing the model on the hybrid ensemble, the dataset is passed to each of the seven machine learning algorithms, i.e., logistic regression, random forest classifier, etc., to test the performance in terms of different evaluation metrics. Hybrid ensemble of the different algorithms is combined and passed to the stacking classifier the results of the experimentation on various algorithms along with the stacking classifier are as shown in Table 4. Ensemble model surpasses the individual models on the WDBC dataset with the accuracy of 0.98% for the prediction of breast cancer on different independent features of the dataset like radius, smoothness, concavity, etc.

5   Conclusion

Breast cancer is one of the most deadly diseases, impacting millions of women worldwide. As a result, any advancement in cancer illness diagnosis and prediction is vital for healthy living. Machine learning-based illness detection has proven to be quite valuable for the early detection of a variety of fatal diseases. This study uses the WDBC dataset for breast cancer diagnosis using a novel ensemble method. In the research work, seven different machine learning models are hybridized and sent

Table 4 Comparison of individual model with the ensemble model

Algorithm	Accuracy	F1-score	Recall	Precision
Logistic regression Solver = "lbfgs"	0.95	0.95	0.95	0.94
Logistic regression Solver = "saga"	0.89	0.88	0.86	0.91
Random forest classifier	0.96	0.95	0.95	0.96
Extra tree classifier	0.95	0.95	0.94	0.95
KNN	0.94	0.94	0.94	0.94
SVC	0.95	0.95	0.95	0.94
Gaussian NB classifier	0.93	0.93	0.93	0.93
SGD classifier	0.90	0.90	0.90	0.90
Stacking classifier	0.98	0.98	0.98	0.96

to a stacking classifier for malignancy diagnosis. The proposed hybrid ensemble was found to more effective than the individual model like KNN, extra tree classifier, SVC, etc.

In future work, feature selection can be utilized to determine the most relevant features to provide to the algorithms and investigate their impact on the outcomes. The findings can also be tested against a range of breast cancer databases. In addition to breast cancer, the suggested ensemble model can be used to diagnose a variety of diseases. Moreover, in a future study, numerous optimization and parallel computing approaches can increase the model's performance.

References

1. The official NVIDIA blog. https://blogs.nvidia.com/blog/category/deep-learning/. Last accessed 2020/08/08
2. Ming C, Viassolo V, Probst-Hensch N, Chappuis PO, Dinov ID, Katapodi MC (2019) Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT and BOADICEA models. Breast Cancer Res 21(1):75
3. Sert E, Ertekin S, Halici U (2017) Ensemble of convolutional neural networks for classification of breast microcalcification from mammograms. In: 39th annual international conference of the IEEE engineering in medicine and biology society 2017, EMBC, IEEE, Korea
4. Mushtaq Z, Yaqub A, Hassan A, Su SF (2019) Performance analysis of supervised classifiers using PCA based techniques on breast cancer. In: International conference on engineering and emerging technologies 2019, ICEET, pp 1–6. IEEE, Pakistan
5. Osman A, Aljahdali H (2020) An effective of ensemble boosting learning method for breast cancer virtual screening using neural network model. IEEE Access 8(39):165–39174
6. Abdullah AJ, Hasan TM, Waleed J (2019) An expanded vision of breast cancer diagnosis approaches based on machine learning techniques. In: International engineering conference 2019, IEC. IEEE, Iraq, pp 177–181

7. Prabadevi B, Deepa N, Krithika LB, Vinod V (2020) Analysis of machine learning algorithms on cancer dataset. In: 2020 international conference on emerging trends in information technology and engineering 2020, IC-ETITE, India, pp 1–10

8. Bayrak EA, Kırcı P, Ensari T (2019) Comparison of machine learning methods for breast cancer diagnosis. In: 2019 scientific meeting on electrical-electronics and biomedical engineering and computer science 2019, EBBT, Turkey, pp 1–3

9. Islam MM, Haque MR, Iqbal H et al (2020) Breast cancer prediction: a comparative study using machine learning techniques. SN Comput Sci 1:290

10. Osmanović A, Halilović S, Ilah LA, Fojnica A, Gromilić Z (2019) Machine learning techniques for classification of breast cancer. In: Lhotska L, Sukupova L, Lacković I, Ibbott G (eds) World congress on medical physics and biomedical engineering 2018. IFMBE proceedings, vol 68/1. Springer, Singapore

11. Chaurasia V, Pal S (2020) Applications of machine learning techniques to predict diagnostic breast cancer. SN Comput Sci 1:270

12. Ganggayah MD, Taib NA, Har YC, Lio P, Dhillon SK (2019) Predicting factors for survival of breast cancer patients using machine learning techniques. BMC Med Inform Decis Mak 9(1):48

13. Khan S, Islam N, Jan Z, Din IU, Rodrigues JJPC (2019) A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. Pattern Recogn Lett 125:1–6

14. Guan S, Loew M (2017) Breast cancer detection using transfer learning in convolutional neural networks. In: 2017 IEEE applied imagery pattern recognition workshop, AIPR, USA

15. Khuriwal N, Mishra N (2018) Breast cancer diagnosis using deep learning algorithm. In: International conference on advances in computing, communication control and networking, ICACCCN

16. Rahman MM, Ghasemi Y, Suley E, Zhou Y, Wang S, Rogers J (2020) Machine learning based computer aided diagnosis of breast cancer utilizing anthropometric and clinical features, IRBM

17. Vaka AR, Soni B, Sudheer Reddy K (2020) Breast cancer detection by leveraging machine learning. ICT Exp 6(4):320–324

18. Mojrian S, Pinter G, Joloudari JH, Felde I, Szabo-Gali A, Nadai L, Mosavi A (2020) Hybrid machine learning model of extreme learning machine radial basis function for breast cancer detection and diagnosis; a multilayer fuzzy expert system. In: RIVF international conference on computing and communication technologies 2020, RIVF. IEEE, Vietnam, pp 1–7

19. World Health Organization. https://www.who.int/cancer/prevention/diagnosisscreening/breast-cancer/en/#:~:text=In 2018, it is estimated, in nearly every region globally. Last accessed 2020/08/08

20. Amrane M, Oukid S, Gagaoua I et al (2018) Breast cancer classification using machine learning. In: Electric electronics, computer science, biomedical engineerings' meeting 2018, EBBT, pp 1–4

21. Bhardwaj R, Nambiar AR, Dutta D (2017) A study of machine learning in healthcare, vol 2. In: IEEE 41st annual computer software and applications conference 2017, COMPSAC, Italy, pp 236–241

22. Shargabi B, Shami F (2019) An experimental study for breast cancer prediction algorithms, vol 12. In: Proceedings of the Second international conference on data science, e-learning and information systems 2019. Association for Computing Machinery, pp 1–6

23. Aslan M, Celik Y, Sabanci K et al (2018) Breast cancer diagnosis by different machine learning methods using blood analysis data. Int J Intell Syst Appl Eng 6(4):289–293

24. Ragab DA, Sharkas M, Marshall S et al (2019) Breast cancer detection using deep convolutional neural networks and support vector machines. PeerJ 7

25. Kadam V, Jadhav S, Vijayakumar K (2019) Breast cancer diagnosis using feature ensemble learning based on stacked sparse autoencoders and softmax regression. J Med Syst 43(8)

26. Wang H, Zheng B, Yoon S et al (2018) A support vector machine-based ensemble algorithm for breast cancer diagnosis. Eur J Oper Res 267(2):687–699
27. Mangasarian O, Street W, Wolberg W (1995) Breast cancer diagnosis and prognosis via linear programming. Oper Res 43(4):570–577