International Journal of Advanced Science Computing and Engineering
Vol. 5, No. 1, April 2023, pp. 15-24

Comparative Study of Machine Learning Models on Multiple Breast Cancer Datasets

Md. Arman Hussain Sujon [a], Hossen Mustafa [a]

[a] Institute of Information and Communication Technology, Bangladesh University of Engineering and Technology Dhaka 1000, Bangladesh
[1] sujon1990@gmail.com, hossen_mustafa@iict.buet.ac.bd
* corresponding author

ARTICLE INFO

Article history
Received December 10, 2022
Revised January 7, 2023
Accepted February 17, 2023

Keywords
Classification
Breast cancer prediction
Data Science

ABSTRACT

Carcinoma is one of the scariest and most frequently occurring cancers nowadays among females. It affects nearly around 10% of females all over the world at some point in their lives. Although the cure for this cancer is currently obtainable, the treatment is not effective enough if the disease is not identified at the early stages. Generally, some contemporary medical tests: roentgenogram, breast ultrasound, biopsy, etc., are used for identifying breast cancer. As an alternative, researchers are exploring machine learning techniques for classifying tumors at different stages, e.g., benign and malignant. Classification and data processing strategies can be effective mechanisms for the prediction of cancer. In this paper, we analyze six classification models: Decision Tree, K Nearest Neighbors, Random Forest, Logistic Regression, Extra Trees, and Support Vector Machine on three different datasets. We applied simple principle component analysis (PCA) to reduce dimensions of the datasets. Experimental results show that Random Forest obtained the best accuracy, recall, and F1 score among the six classification techniques for all three datasets. We also find that data attributes and values are important for accurate classification.

1. Introduction

Accurate identification of important information from medical data is challenging in bio-science. The diagnosis of the sickness could be a crucial task in bio-science. There is an enormous quantity of medical diagnosing information accessible that can be used for quick and correct diagnosis of various types of health issues. Manual identification of diseases is vulnerable to human errors, unwanted biases, and time waste. Such delays and errors can be fatal for cancer patients. Data suggests that females are diagnosed more with breast cancer compared to all carcinoma [9].

Breast cancer is the abnormal growth of some cells within any part of a breast. Several diagnostic processes are available for the identification of carcinoma. Mammogram has been proposed to diagnose carcinoma [18]. Ultrasound [19] is also a very efficient technique for the identification of carcinoma. In this process, the wave of sound is distributed within a specific area of the body to observe the condition inside. Positron emission tomography (PET [14] imaging illustrates F-fuorodeoxyglucose which allows doctors to get knowledge of the tumor's position within the human body.

It is created specifically for the recognition of traces of radio-labelled cancer. Flexman et al. [11] used dynamic tomography with the spread of cancer cells. Elastography [20] is a recent technique which supports imaging technology that can be used when carcinoma tissue supports more substantial than the adjacent regular functional tissue. In recent years, neural network [25], different types of computational intelligence techniques [1], predictive data mining [12] and support vector machine and ensemble classification [13] technologies are designed in many medical predictions. Current machine learning methods to detect breast cancer uses different types of Naive Bayes, Support Vector Machine (SVM), KNN, etc., and Xu et al. [27] reported the highest 98.53% accuracy on the University of Wisconsin Hospital dataset [26]. However, there is still room for improvement in carcinoma detection performance.

In this paper, we have analyzed six machine learning classifier models: (i) Decision Tree (DT), (ii) K Nearest Neighbors (KNN), (iii) Random Forest (RF), (iv) Logistic Regression (LR), (v) Extra Trees (ET) and (vi) Support Vector Machine (SVM). We applied these models to three datasets to compare the performance of the classifier that is best suited to predict breast carcinoma at the very initial stage. We also compare our experimental results with alternative schemes that used a similar dataset. Our performance comparison shows that Random Forest outperformed the other five machine learning techniques with the best accuracy of 98.57%, 97.82% precision, and 100% recall. Additionally, our research work also makes an effort to evaluate the performance of three different datasets applying ensemble machine learning algorithms in terms of accuracy, precision, recall, and F1 score. We find the data quality is critical for accurate classification and applying similar techniques to different types of datasets may not be effective.

The rest of the paper is organized as follows. Background is discussed in Section II. In Section III, we discuss data pre- processing and proposed mechanism. Then, the performance analysis is given in Section IV. Finally, we conclude the paper in Section V.

2. Background

In this section, we discuss the supervised machine learning algorithms which are analyzed in this paper. We also discuss Principal Component Analysis (PCA) which is used for data processing.

A. Supervised Learning Algorithms

In supervised learning, known information is used to predict future unknown classes. Regression and classification are common ways in the supervised learning category [5]. In this paper, we evaluate the following six supervised machine learning algorithms.

1) Support Vector Machine (SVM) is classification model that builds hyper-planes as decision surfaces to map training data to a high dimensional feature space. Classification of a test data is performed based on the feature space it is mapped to.
2) K Nearest Neighbors (KNN) is a neighbors-based classification technique. KNN stores instances of the training data in terms of distance between data. Classification is computed using majority voting of the k nearest neighbors. Euclidean distance is generally used for distance calculation.
3) A Decision Tree (DT) model produces a sequence of rules that can be accustomed to classify the data. It works with categorical data, so continuous data are converted to discrete values for DT. There are several algorithms to build decision tree, e.g., ID3, C4.5, C5.0, CART, etc.
4) Random Forest (RF) classifier is a meta-estimator. Using varied sub-samples of datasets, random forest matches a variety of decision trees and uses an average to enhance the predictive accuracy of the model. Random forest works well with high dimensional data.

5) Logistic Regression (LR) uses a logistic function that generates probabilities describing the possible outcomes of test data. It works on multiple iterations during training phase to find optimal co-efficient values for best prediction.
6) The Extra Trees (ET) is another technique that generates a large number of unpruned decision trees from the training dataset and then uses these trees for classifications. The sampling of each tree is random without replacement. It computationally performs better compared to random forest

B. Principal Component Analysis

Principal component analysis (PCA) is a feature extraction technique where we can drop least important variables. The main setup of PCA is to cut back the dimensionality of a dataset consisting of the various variables connected with each other [10]. Covariance matrix is calculated to identify correlations of variables. Using the covariance matrix, eigenvectors and eigenvalues are calculated and then, eigenvalues are sorted. PCA is completed by transforming a variable set to a replacement set of variables that is remarked as the principal part (PCs). A number of PCs are selected that gives the best performance. Thus, PCA is used to reduce data dimension whereas keeping the maximum data quality.

3. Proposed Mechanism

In this section, we discuss the details of our proposed data pre-processing technique and experimental setup for the different machine learning models.

A. Data Pre-processing

Data pre-processing is necessary to handle missing values, process outliers, and solve self-contradiction. We use the mean of attributes to process absent data for a category. In addition, a random choice of data set is utilized to verify the correct circulation of data. The number of variances of the first data is calculated as the ratio between the variance of the residual data for the parts from one to nine; and therefore the variance of the initial data.

Principal Component Analysis (PCA) is applied to reduce the dimensionality of the feature columns. We got nine features within the data for dataset-1 [26] (Section IV); therefore we needed to reduce the number of feature columns whereas maintaining the variance in data. We applied the variance as 0.95. By applying PCA, we transform the present set of features into a new set of reduced features for all three datasets.

B. Building Classification Models

We built 6 different classification models for breast cancer prediction: (i) SVM, (ii) KNN, (iii) LR, (iv) DT, (v) RF, and (vi) ET. We discuss details in the following.

1) SVM: Parameter selection for kernel functions is important to the robust classification performance of SVM. It must be one of ‘linear’, ‘poly’, ‘rbf’, or ‘sigmoid’. If a callable is given, it is used to pre-compute the kernel matrix from data matrices. In our mechanism, we used 'poly' as kernel function.
2) KNN: Here, K is the number of nearest neighbors which is the core deciding factor. In our mechanism, we found that K=20 is the best choice.
3) LR: LR does not really have any critical hyper-parameters to tune. In our mechanism, we used L-BFGS as a solver. Regularization (penalty) can be helpful for better performance; we have used the penalty of l2.

4) DT: DT complexity has a crucial effect on its accuracy. In our mechanism, the tree complexity is measured by one of the following metrics: the total number of nodes, the total number of leaves, tree depth, and the number of attributes used.
5) RF: For RF, the most important parameter is the number of random features to sample at each split point (maximum features). We have used bootstrap = true, criterion = gini, maximum depth = 10, and maximum features = sqrt. Another important parameter for the RF is the number of trees (n-estimators). In addition, this should be increased until no further improvement is seen in the model. In our mechanism, the n-estimators value is 10.
6) ET: ET implements a meta estimator that fits a number of randomized decision trees, i.e., extra-trees, on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. The number of trees can be set via the n-stimators and we used 100. In our mechanism, parameter selection for ET is bootstrap =true, criterion = entropy, maximum depth = 40, and maximum features = sqrt.

C. Training and Testing Phases

We applied our model to 3 different datasets. For each, the whole dataset is divided into training and test data. The model is built based on training data. Test data is used to analyze the trained model. We used k-fold cross-validation for our analysis. Cross-validation is a technique used to minimize overfitting. In our study, we have used a k=10 to partition.

Table 1. Experimental Setup

Model	Param Name	Value
SVM	Kernel function	poly
KNN	K	20
LR	Solver	L-BFGS
	Penalty	12
RF	Max depth	10
	Criteria	Gini
	n-estimator	10
ET	Max depth	40
	Criteria	entropy
	n-estimator	100

D. Performance Metrics

We used several performance metrics to determine the performance of our machine-learning models. During this process, a confusion matrix is formed for actual data and predicted data. The confusion matrix is formed using TP, TN, FP, and FN as described below:

1) TP: True Positive is the number of positive data that are predicted correctly.
2) TN: True Negative is the number of negative data that are predicted correctly.
3) FP: False Positive is the number of the negative data that are predicted positive incorrectly.
4) FN: False Negative is the number of positive data that are predicted negative incorrectly.

In this paper, we used accuracy, precision, recall, and F1 score [18] to measure the performance of our proposed models.

1) Accuracy (Acc) is the quantitative relation of correctly predicted data to total data.

$$Acc = \frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

2) Precision (P) is the proportion of the data that are literally positive among all the predicted positives by the model.

$$P = \frac{TP}{TP+FP} \tag{2}$$

3) Recall (R) is the proportion of data that are literally positive among all positive data.

$$R = \frac{TP}{TP+FN} \quad (3)$$

4) F1 Scores is the weighted average of precision and recall. Therefore false positives and false negatives are taken by this score into consideration. F1 is helpful for evaluation compared to accuracy.

$$F1_{SCORE} = \frac{P \, x \, R}{P+R} x \, 2 \quad (4)$$

Table 2. Attributes of Dataset-1

No.	Attributes	Domain
1	Clump thickness	1-10
2	Uniformity of cell size	1-10
3	Uniformity of cell shape	1-10
4	Marginal adhesion	1-10
5	Single epithelial cell size	1-10
6	Bare nuclei	1-10
7	Bland chromatin	1-10
8	Normal nuclei	1-10
9	Mitoses	1-10

Table 3. Attributes of dataset-2 where each attribute has 3 values: mean, largest and standard error

No.	Attributes	Domain
1	Radius	Real
2	Texture	Real
3	Perimeter	Real
4	Area	Real
5	Smoothness	Real
6	Compactness	Real
7	Concavity	Real
8	Concave points	Real
9	Symmetry	Real
10	Fractal dimension	Real

Table 4. Attributes of Dataset-3

No.	Attributes	Domain
1	Age	1-10
2	Menopause	1-10
3	Tumor size	1-10
4	Inv nodes	1-10
5	Node caps	1-10
6	Deg malig	1-10
7	Breast	1-10
8	Breast quad	1-10
9	Irradiat	1-10

4. Performance Analysis

A. *Dataset Description*

In this work, we used three datasets for performance analysis of the model. The datasets are: (i) Dataset-1: Wisconsin Breast Cancer Dataset [26], (ii) Dataset-2: Breast Cancer Wisconsin

(Diagnostic) Data Set [17], (iii) Dataset-3: Breast Cancer Dataset of University Medical Centre, Institute of Oncology, Ljubljana, Yugoslavia [6]. All these datasets are publicly available.

Dataset-1 [26] was created by Dr. William H. Wolberg, a physician at the University of Wisconsin Hospital, USA. The samples were analyzed based on a digital scan. This dataset contains 699 instances where the cases are either non- cancerous/benign or infectious. Among all instances, 65.50% are from benign, and 34.50% are from malignant class. The benign cases are set as a positive category and the malignant cases are set as a negative category in our analysis. This dataset has 10 attributes with domain values between 1 and 10. For instance, there are 3 attributes for radius: radius_mean, radius_se, radius_worst. All these 30 attributes have real values. The attributes except the Identifier are shown in Table 2.

Dataset-2 [17] consists of 32 attributes and the features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. There are 569 instances where 212 (37.19%) are malignant and the other 357 (62.81%) are benign. Except the Identifier and Class attribute, the rest 30 attributes are based on 10 attributes where each attribute has 3 feature values: mean, highest and standard error. The attributes except the Identifier are shown in Table IIIs.

Dataset-3 [6] consists of 286 instances with ten attributes. In this dataset, 29.72% instances are defined as malignant and the other 70.28% are defined as benign. This dataset contains categorical data. To fit this dataset in the proposed model, the One-Hot Encoding method is used. The class attribute of this dataset is defined as recurrence-events and no-recurrence-events. The attributes except the Identifier are shown in Table IV.

B. Experimental Results

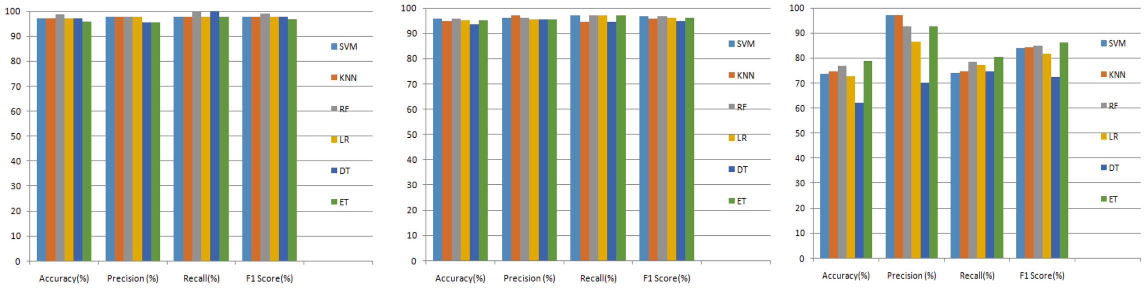
We implemented the proposed models using Python; a machine learning library, Scikit-learn [21], was used in our implementation for building the models. The dataset was divided into 10 partitions in each iteration for cross- validation. A confusion matrix for the six machine learning techniques is generated using TP, FP, TN, and FN. Using the confusion matrix, we calculated the accuracy, precision, recall, and F1 score for all the models. Fig.-1 summarizes the performance metrics for all six models for the three datasets.

For Dataset-1 [26], we used 90% (629 instances) of the overall data to train all six models and the rest 10% (70 instances) for testing. The experimental results depict that random forest performed best among the 6 machine learning techniques with the best accuracy of 98.57%. Among the others, the accuracy of DT, SVM, LR, and KNN was 97.14%. ET got the lowest accuracy of 95.71% in our analysis. The precision for Random Forest is comparatively low (97.83%) but it achieved the highest recall (100%) and F1 score (98.9). Among others, DT has the highest recall of 100% while we see 97.83% recall for SVM and LR. Overall, we can see that random forest performs best among these six models for Dataset-1 [26].

Dataset-2 [17] also gave similar performance like Dataset-1 [26] where RF and SVM performed the best with 95.91% accuracy, 96.3% precision, and 97.2% recall. KNN, LR, ET got 94.74%, 95.32%, and 95.32% accuracy. Decision tree performed worst with 93.57% accuracy, 94.94% F1 score.

For Dataset-3 [6], the model performance is low; this is because this dataset contains categorical data. In this case, ET performed the best with 78.95% accuracy, 92.54% precision, and 86.11% recall. Even though RF performed best for the other two datasets, it achieved accuracy of76.84% for Dataset-3 with 78.48% recall. In this case as well, decision tree performed worst with 62.11% accuracy and 72.31% F1 score.

We show a comparison of performance for all three datasets in Fig.-2. We can see that each model achieved the highest accuracy for Dataset-1 [26] compared to the other two datasets in our analysis.



(A) Performance for Dataset-1 (B) Performance for Dataset-2 (C) Performance for Dataset-3

Figure 1. This figure shows the comparative analysis in terms of accuracy, precision, recall, and f1 score for the six machine learning models for (a) dateset-1, (b) dateset-2, and (c) dateset-3 respectively

C. Comparative Study

A comparative study for breast cancer prediction of existing works which also used Dataset-1 [26] is illustrated in Table V. Among these, the accuracy of Kernel-based orthogonal transform [27] was the best (98.53%).

Azar et al. [2] studied the performance of different decision tree models to predict breast cancer and got the best accuracy of 97.07% for boosted decision tree. Local linear wavelet neural network (LLWNN [23] secured an accuracy level of 97.2%. On the other hand, Azar et al. analyzed different types of support vector machine models and got the best accuracy of 97.1429%, by Linear Programming SVM (LPSVM [3]. The proposed system in [15] included Naive Bayes, SVM, and J48 maltreatment as classifier methodology to realize an accuracy of 97.13%. Latchoumi et al. [16] also used a weighted smooth SVM and got 98.42% accuracy. Sakri et al. [22] reported 81.3%, 80%, and 75%, accuracy for Fast Decision Tree Learner (RepTree), NB, and KNNs using particle swarm optimization feature selection. In [4], with the assistance of gradient boosting, 91.7% accuracy was achieved by BBN, and BAN, and 94.11% gained for TAN. Chaurasia et al. [7] reported an accuracy of 97.36% using Naive Bayes.

In our analysis for Dataset-1 [26], we found that the Random Forest model performs relatively higher than the other techniques with 98.57% accuracy, 97.83% precision, 100% recall, and 98.9 F1-score. If we compare existing techniques, Random Forest technique outperformed all of these in terms of accuracy.

Table 6. Performance Measure by Using Ensemble Classifier

Dataset	Acc.	Precision	Recall	F1 Score
Wisconsin Breast Cancer Dataset [26]	98.91	99.27	97.84	98.55
Breast Cancer Wisconsin (Diagnostic) Dataset [17]	97.66	99.03	97.14	98.08
Breast Cancer Dataset of University Medical Centre [6]	84.21	95.71	84.81	89.93

D. Performance Measure by Ensemble Classifier

Table 5 shows the results of the ensemble model in contrast to the individual ML technique. Dataset-1 has the highest accuracy when compared with other Datasets. For the ensemble classifier, we first train and test the dataset with the individual algorithms where we used 70% of the data for training and the rest 30% of data as test data. Evaluation of this ensemble approach is performed by some performance parameters such as accuracy, precision, recall, and F1-score.

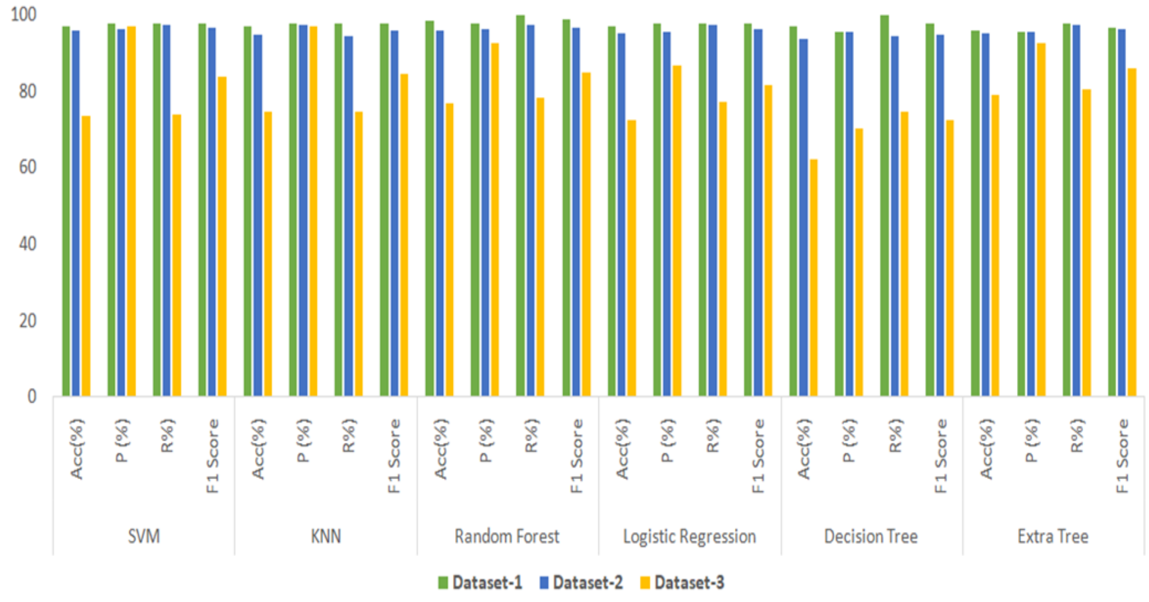


Figure 2. This figure shows the comparative analysis for the three datasets in terms of accuracy, precision, recall, and F1 score

Table 5. Performance Comparison of Different Method for Breast Cancer Prediction

No.	Method	Accuracy
1	Kernel orthogonal transform [27]	98.53
2	Single decision tree [2]	95.75
3	Boosted decision tree [2]	97.07
4	Decision tree forest [2]	95.51
5	Local Linear Wavelet Neural Network [23]	97.2
6	Linear Programming SVM [3]	97.14
7	Lagrangian SVM [3]	95.42
8	Smooth SVM [3]	96.57
9	Proximal SVM [3]	96
10	Lagrangian SVM [3]	96.57
11	Standard SVM [3]	94.86
12	Weighted-Particle Swarm Optimization Smooth SVM [16]	98.42
13	- a e ayse-J48 [15]	97.13
14	a e ayes [22]	81.3
15	Fast Decision Tree Learner [22]	80
16	K-Nearest Neighbor [22]	75
17	Bayes Belief Network [24]	91.7
18	oosted ugmented a e ayes [4]	91.7
19	ree ugmented a e ayes [4]	94.11
20	a e ayes [7]	97.36
21	Proposed Random Forest	98.57

Finally, we use the *voting ensemble strategy* that combines all predicted output to achieve greater accuracy. The result shows that the ensemble technique is more accurate at each performance parameter as expected. Dataset-1 has the highest Precision 99.27%, Recall 98.55%, and F1 Score 98.91% of all for predicting breast cancer. Using ensemble technique, we got 97.66% and 84.21% accuracy for Dataset-2 and Dataset-3 respectively. Based on our experimental result, we can say that Dataset-1 is the best among the three datasets for predicting the risk of developing breast cancer.

5. Conclusion

Early prediction of breast cancer can be beneficial for the survival of a patient. In this paper, we present a generic mechanism for feature selection and model building for the prediction of breast cancer. The proposed mechanism has been used to generate six different machine learning models, and three different datasets are used for comparative analysis. Among all these techniques, the Random Forest came out with the very best accuracy of 98.57% for the UC Irvine breast cancer dataset (Dataset-1). We also used the ensemble classifier with the voting technique and got the best performance for Dataset-1. The other two datasets are used to compare the accuracy and determine the model consistency. Experimental results show that the model works well for both numeric and categorical data. Our analysis shows that the quality of data is very important for breast cancer prediction, and numerical data is more useful than categorical data. This paper shows that machine learning techniques could be highly effective with a proper dataset for the early detection of breast cancer.

References

[1] D. P. Acharjya and C. L. Chowdhary. *"Breast cancer detection using hybrid computational intelligence techniques"*. In Handbook of Research on Emerging Perspectives on Healthcare Information Systems and Informatics, pages 251–280. IGI Global, 2018.

[2] Ahmad Taher Azar and Shereen M El-Metwally. Decision tree classifiers for automated medical diagnosis. Neural Computing and Applications, 23(7):2387–2403, 2013.

[3] Ahmad Taher Azar and Shaimaa Ahmed El-Said. Performance analysis of support vector machines classifiers in breast cancer mammography recognition. Neural Computing and Applications, 24(5):1163–1177, 2014.

[4] A Bazila Banu and Ponniah Thirumalaikolundusubramanian. Comparison of bayes classifiers for breast cancer classification. Asian Pacific journal of cancer prevention: APJCP, 19(10):2917, 2018.

[5] Giuseppe Bonaccorso. Mastering Machine Learning Algorithms: Expert techniques for implementing popular machine learning algorithms, fine-tuning your models, and understanding how they work. Packt Publishing Ltd, 2020.

[6] Vikas Chaurasia and Saurabh Pal. Data mining techniques: to predict and resolve breast cancer survivability. International Journal of Computer Science and Mobile Computing IJCSMC, 3(1):10–22, 2014.

[7] Vikas Chaurasia, Saurabh Pal, and BB Tiwari. Prediction of benign and malignant breast cancer using data mining techniques. Journal of Algorithms & Computational Technology, 12(2):119–126, 2018.

[8] Hercules Dalianis. Evaluation Metrics and Evaluation. Springer International Publishing, 2018.

[9] Carol E DeSantis, Jiemin Ma, Mia M Gaudet, Lisa A Newman, Kimberly D Miller, Ann Goding Sauer, Ahmedin Jemal, and Rebecca L Siegel. Breast cancer statistics, 2019. CA: a cancer journal for clinicians, 69(6):438–451, 2019.

[10] George H Dunteman. Principal components analysis. Number 69. Sage, 1989.

[11] Molly L Flexman, Michael A Khalil, Hyun K Kim, Christopher J Fong, Andreas H Hielscher, Rabah Al Abdi, Randall L Barbour, Elise Desperito, and Dawn L Hershman. Digital optical tomography system for dynamic breast imaging. Journal of biomedical optics, 16(7):076014, 2011.

[12] Ramin Ghorbani and Rouzbeh Ghousi. Predictive data mining approaches in medical diagnosis: A review of some diseases prediction. International Journal of Data and Network Science, 3(2):47–70, 2019.

[13] Min-Wei Huang, Chih-Wen Chen, Wei-Chao Lin, Shih-Wen Ke, and Chih-Fong Tsai. Svm and svm ensembles in breast cancer prediction. PloS one, 12(1):e0161501, 2017.

[14] Rakesh Kumar and Abass Alavi. Fluorodeoxyglucose-pet in the management of breast cancer. Radiologic Clinics, 42(6):1113–1122, 2004.

[15] U Karthik Kumar, MB Sai Nikhil, and K Sumangali. Prediction of breast cancer using voting classifier technique. In 2017 IEEE international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM), pages 108–114. IEEE, 2017.

[16] TP Latchoumi and Latha Parthiban. Abnormality detection using weighed particle swarm optimization and smooth support vector machine. Biomedical Research, 28(11):4749–4751, 2017.

[17] Moshe Lichman. Uc irvine machine learning repository. University of California “http://archive. ics. uci. edu/ml”, Irvine, School of Information and Computer Sciences, 2013.

[18] Tuong L Nguyen, Daniel F Schmidt, Enes Makalic, Gertraud Maskarinec, Shuai Li, Gillian S Dite, Ye K Aung, Christopher F Evans, Ho N Trinh, Laura Baglietto, et al. Novel mammogram-based measures improve breast cancer risk prediction beyond an established mammographic density measure. International Journal of Cancer, 148(9):2193–2202, 2021.

[19] Neslihan Ozmen, Robin Dapp, Michael Zapf, Hartmut Gemmeke, Nicole V Ruiter, and Koen WA van Dongen. Comparing different ultrasound imaging methods for breast cancer detection. IEEE transactions on ultrasonics, ferroelectrics, and frequency control, 62(4):637–646, 2015.

[20] Seong Ho Park and Kyunghwa Han. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Radiology, 286(3):800–809, 2018.

[21] Fabian Pedregosa, a el aro uau , Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. the Journal of machine Learning research, 12:2825–2830, 2011.

[22] Sapiah Binti Sakri, Nuraini Binti Abdul Rashid, and Zuhaira Muhammad Zain. Particle swarm optimization feature selection for breast cancer recurrence prediction. IEEE Access, 6:29637–29647, 2018.

[23] Manas Ranjan Senapati and Pradipta K Dash. Local linear wavelet neural network based breast tumor classification using firefly algorithm. Neural Computing and Applications, 22(7):1591–1598, 2013.

[24] Manas Ranjan Senapati, Ganapati Panda, and Pradipta K Dash. Hybrid approach using kpso and rls for rbfnn design for breast cancer detection. Neural Computing and Applications, 24(3):745–753, 2014.

[25] Fung Fung Ting, Yen Jun Tan, and Kok Swee Sim. Convolutional neural network improvement for breast cancer classification. Expert Systems with Applications, 120:103–115, 2019.

[26] WIlliam Wolberg, OL Mangasarian, and DW Aha. Uci machine learning repository: Breast cancer wisconsin (original) data set. UCI Machine Learning Repository: Breast Cancer Wisconsin (Original) Data Set. University of Wisconsin Hospitals Madison, Wisconsin, USA, nd Web, 2015.

[27] Yong Xu, Qi Zhu, and Jinghua Wang. Breast cancer diagnosis based on a kernel orthogonal transform. Neural Computing and Applications, 21(8):1865–1870, 2012.