**Journal of Artificial Intelligence and Data Science (JAIDA)** **Vol. 1, No. 1, pp. 11-21, 2021**
**Research Article**

**https://jaida.ikcu.edu.tr/**

# Identification of Breast Cancer Metastasis Using Boosting Algorithms on Cytopathologic Data

Safak KAYIKCI *

Bolu Abant Izzet Baysal University, Faculty of Engineering, Department of Computer Engineering, Turkey

**Abstract**
Breast cancer is the second most common cancer among women after lung cancer. Early diagnosis of cancer can positively affect the recovery process from disease. Several machine learning-based approaches have been studied for cancer detection on histopathological images. In this study, identification of cancer type has been made using Gradient Boosting Machine (GBM), eXtreme Gradient Boost (XGBoost), and Light Gradient Boosting Machine (LightGBM) algorithms. The performances of these techniques have been measured on the Breast Cancer Wisconsin (Diagnostic) dataset. According to the results obtained, Gradient Boosting Machine (GBM) got the highest accuracy rate with 97.02% success. Although there is no pathological prior knowledge about the disease, high success has been achieved in diagnosing with the deep learning architectures used.
***Keywords: Breast cancer; eXtreme gradient boost; gradient boosting machine; light gradient boosting machine.***

## 1. Introduction

Breast cancer is a widespread kind of cancer that is a dangerous type that affects women the most. The World Health Organization (WHO) reported that approximately two million women die from this disease cancer every year. Approximately 627,000 women died of breast cancer in 2018 [1]. Therefore, early detection studies are critical.

Histopathological images of breast cancer are automatically categorized as benign or malignant cancer with the help of computer-aided diagnosis systems, and early treatment must be initiated as soon as possible. While detection of breast cancer, mammography can be applied by techniques like Magnetic Resonance Imaging (MRI), ultrasound, tomography, and a biopsy of the breast tissue are required for the definitive diagnosis [2][3]. By examining histological samples of cells or tissues taken from the body using the biopsy method under a microscope, pathologists try to make a definite diagnosis about the change in the breast [4]. They try to classify as benign, malignant, or normal according to diagnosis results. These samples are analyzed with different microscopic magnification rates. Since the examination of histopathological images is a troublesome duration, evaluation of these images with computer-aided methods will make a serious contribution to the correct diagnosis. The experience and attention of pathologists during the analysis of these images are essential for the correct diagnosis and diagnosis. Computer-based systems can minimize the wrong diagnoses decisions by being affected by negative factors such as fatigue and distraction that may occur in the daily lives of pathologists. Thus, it can enable experts to focalize on crucial cases to diagnose [5].

Many important studies have been carried out in this field using digital image processing techniques. Automatic classification of benign and malignant breast cancer types has become especially important. Various shallow machine learning algorithms such as Artificial Neural Networks (ANN), Random Forest (RF), Support Vector Machines SVM and Principal Component Analysis (PCA) have been used primarily for the diagnosis of breast cancer.

Spanhol et al. published an open-source BreakHis data set containing 7,909 images obtained from 82 patients [6]. The authors proposed LeNet and AlexNet models for the classification of breast cancer images. In this study, the success of the AlexNet model in classifying histopathological images was reported to be better than the LeNet model. The dimensions of the images submitted to AlexNet are defined as 32 × 32 or 64 × 64 pixels. Simple fusion rules such as maximum, product, and some are used in the proposed models for comparison. The authors achieved the highest classification success of %85.6 ± 4.8 with the AlexNet model they proposed for dual breast cancer image classification. In 2017, Han et al. used the Class Structure-based Deep Convolutional Neural Network (CSDCNN) model [7]. Dimensions of the images given as an introduction to the CSDCNN model have been resized to be 256 × 256. Filters of 3 × 3, 5 × 5, or 7 × 7 are applied in the convolution layers. Two different ways have been tried in the training process. First, training of model on the BreakHis data set was made from scratch, but when it was seen that the poor result was obtained, the weights of the model trained on ImageNet with the transfer learning method were used as the second way. The proposed model has reached the highest performance of %96.9 ± 1.9 in the binary classification of images. In 2018, Wang et al. used an ensemble

*Corresponding author e-mail address: safak.kayikci@ibu.edu.tr

Received: 29/06/2021; Accepted: 17/08/2021

algorithm based on a support vector machine (SVM) on images at different magnifications [8]. The recommended Weighted Area Under the Receiver Operating Characteristic Curve Ensemble (WAUCE) model decreased the variance by %97.89, whereas accuracy is increased by %33.34. Wang et al. won first place in the competition in Symposium on Biomedical Imaging (ISBI) challenge [9]. They achieved 0.925 area under the curve (AUC) for the goal of image categorization and a 0.7051 score for evaluating biopsy images. Bejnordi et al. applied machine learning solutions to get exceptional achievement than 11 participant pathologists in a general pathology simulation demo [10]. Yala et al. proposed several models like risk factor-based logistic regression model (RF-LR), deep learning model (image only DL), and a hybrid DL model that used both traditional risk factors and mammograms [11]. In another study, Khan et al. proposed a fully connected dense architecture using pre-trained convolutional neural network frameworks for cancer type categorization with average pooling [12]. Filipczuk et al. achieved a performance of %98 by applying four distinct classifiers trained with a 25-dimensional feature space to distinguish 737 breast cancer images as benign or malignant [13]. In 2016, Albarqouni et al. present an innovative conception as learning from crowds which uses data aggregation straightly in training operation of the convolutional neural networks by added crowdsourcing layer (AggNet) [14]. In a study done in 2019, three different convolutional neural networks (CNNs) of Inception V3, Inception-ResNet V2, and ResNet-101 architectures were successfully achieved by Zhou et al. [15]. Terasa et al. made their categorization at the patient level using a model similar to AlexNet for classifying breast cancer images [16]. In this study, the maximum fusion method using different fusion techniques with an average recognition accuracy of %90 and %85.6 has been reported. In another study, %83.25 classification success was achieved at the patient level using CNN and multitasking CNN (MultiTask Cascaded Convolutional Networks, MTCNN) models [17]. In 2018, Alom et al. performed dual and multiple classifications at the patient and image level by the Inception Recurrent Residual Convolutional Neural Network (IRRCNN) model [18]. The proposed model has an architecture consisting of the combination of Inception-v4, ResNet, and Recurrent Convolutional Neural Network models. 128, 256, 512, and 1024 feature maps were used in the blocks used for the construction of the model, respectively. The model includes approximately 9.3 million parameters. Images submitted to the model were considered as randomly cropped or non-overlapping patches. With the model used in this study, an average performance of %97 was achieved [19].

## 2. Dataset

The Breast Cancer (Wisconsin) Diagnosis dataset [20] contains the diagnosis and a set of 30 features defining the specifics of the cell nucleus that exist in the digital appearance of a fine needle aspirate (FNA) of a breast mass. FNA is a distinguishing process utilize to examine lumps or tumors. In this approach, a slim (23–25 scale) perforated needle is injected into the subject for exemplification under a microscope (biopsy). FNA operations are harmless and riskless light surgical protocols.

There are 569 samples, of which 212 of them are malignant, and 357 of them are benign. The percent of the malignant tumor is %37.3. The percent is curiously big. The dataset does not specify a general medical anatomization distribution. Instead, ten substantial features are measured for every cell: radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension.

Pearson Correlation or Pearson Moment Correlation (PPMC), or bivariate correlation, is the standard measure of correlation in statistics. It represents the linear conjunction among two samples. As the correlation coefficient approaches 1, it implies that for each positive increment in a parameter, there will be a positive increment in the other related parameter at a fixed rate. Zero means there is no positive or negative increase with any increase. These two are not related. The correlation coefficient -1 implies that for each positive increment in a parameter, there will be a positive decrement at a specified rate in the other variable. The correlation coefficient in absolute value gives us the strength of the relationship. The larger the number, the stronger the partnership. With Pearson correlation, only a linear relationship between two continuous variables can be tested (A relationship is linear only when a change in one variable is associated with a proportional change in the other variable). Pearson correlation The highest correlations are between perimeter_mean and radius_worst, area_worst and radius_worst, perimeter_worst and radius_worst, perimeter_mean, area_worst, area_mean, radius_mean, texture_mean, and texture_worst. Pearson Correlation of features is shown in Figure 1.

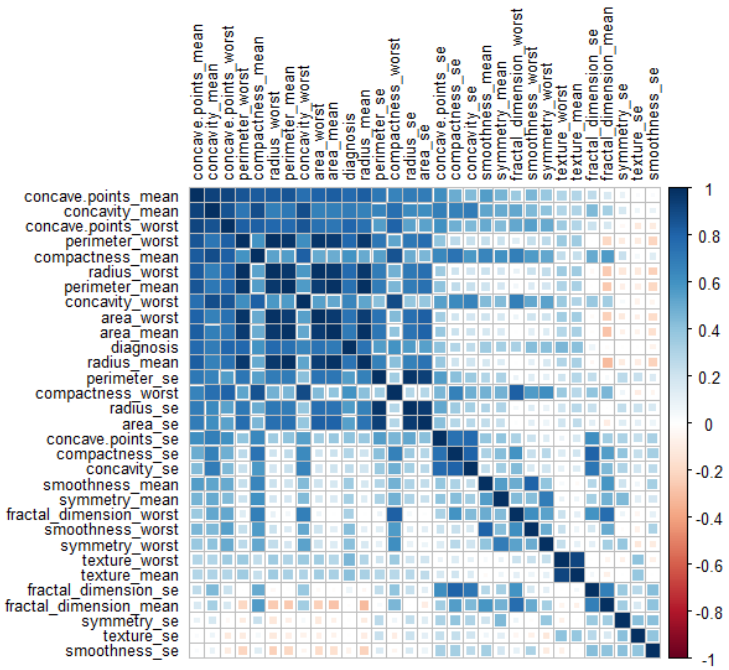


**Figure 1.** *Pearson correlation of features.*

These highly correlated features are represented in Figure 2 with the usage of boxplot2g, showing the scatter plot (in the two dimensions given by the selected features) for the clustered data (grouped by diagnosis), over which are superposed the elliptical-shaped boxes in an equivalent (but still enhanced) way a boxplot will visualize the same information for a single dimension. It is observed that some of the correlated pairs are showing a good separation as well between data with diagnosis B and data with diagnosis M.

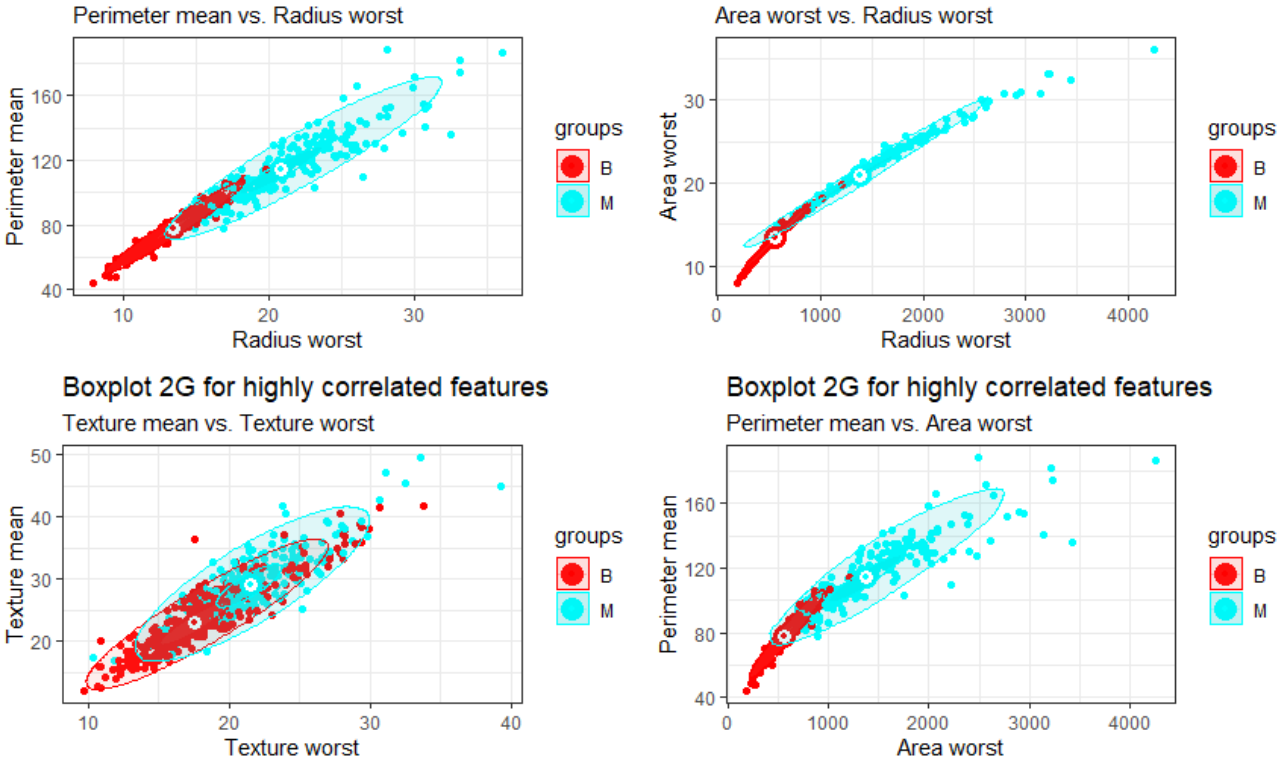


**Figure 2.** *Highly correlated pairs.*

Boxplots for some inverse correlated and low correlated pairs of features are shown in Figure 3 and Figure 4. It is observed that low correlated features that have in the same time a considerable overlap for the two 'M' and 'B' groups (ex: 'fractal_dimension_worst' and 'area_se') as well as low correlated features that have in the same time a good selectivity for 'M' and 'B' groups (ex: 'perimeter_worst'and 'fractal_dimension_se').

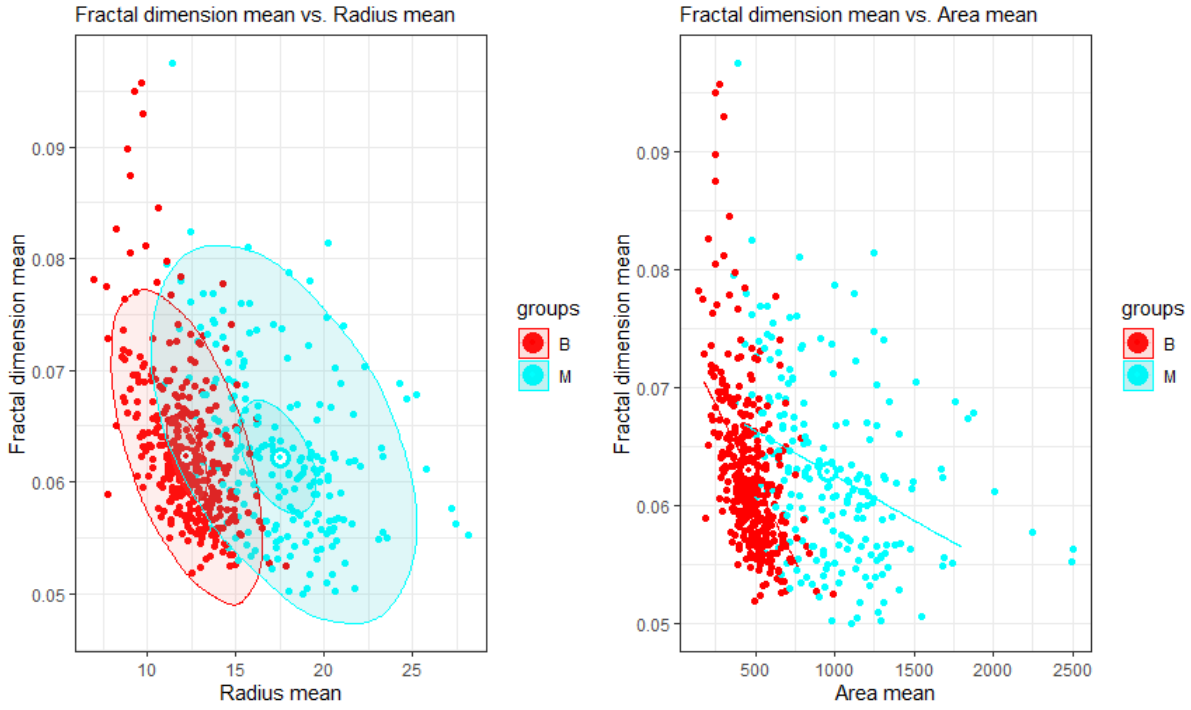


**Figure 3.** *Inverse correlated pairs.*

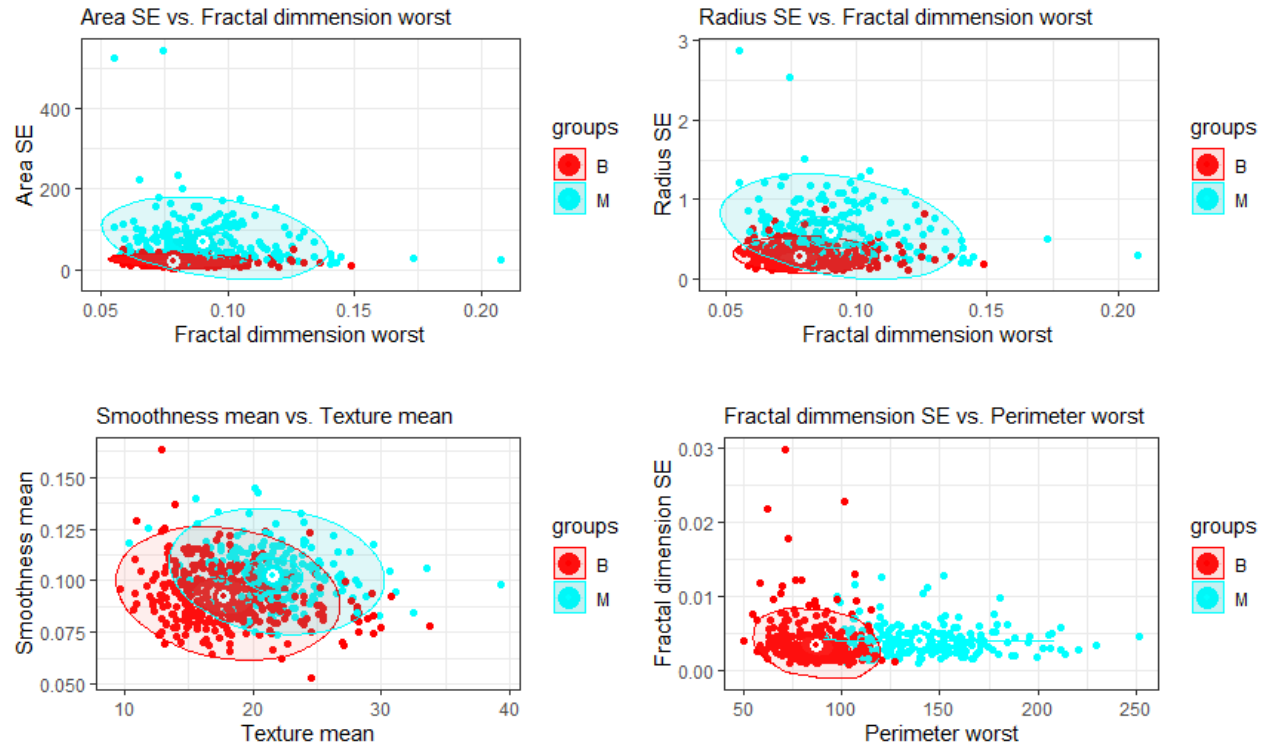


**Figure 4.** *Low correlated pairs.*

## 3. Method

### 3.1. Principal Components Analysis (PCA) transform

Principal component analysis (PCA) is a numerical method that converts mutually related variables group into another group of variables that are not mutually related. These new variables, made up of eigenvectors, represent the original data differently. It is the mere of true eigenvector-based multivariate analysis. Most of the time, this work is seen as disclosing the inner composition of the data by revealing the variances in the data in the best way. If a multivariable dataset is considered as a coordinate system with one variable per axis, the principal component analysis provides the user with a low-dimensional shadow picture that contains the most informative view of the data at hand. PCA moves data to a new coordinate system with a linear orthogonal transformation. Thus, the largest variance obtained from the initial data is the first coordinate and is considered the first fundamental component. It can be defined as the oldest statistical tool used to analyze multivariate databases and reduce large data to lower dimensions.

PCA was first proposed by Pearson in 1901 [21], and the development of the theory was made by Hotelling in 1933 [22]. It can be used to show the relationship, similarity, or differences of information in a multidimensional database. The biggest advantage of PCA is that once the model of the information in the database is defined, the data size is reduced and compressed, and data loss is minimized. This feature of PCA is also used in compressing pictures. In this way, the same picture can be obtained again without losing much information. Out of these specialties, PCA also has other advantages such as low sensitivity to noise, reduced memory and capacity needs, and more efficient operation in small-sized spaces. It is one of the most preferred methods in the field due to its ease of application based on Karhunen-Loeve (K-L) or Hotelling Transformation. In addition, PCA is a linear method that can be used to reduce redundancies based on the least mean square error rate with the linear transformation of the coordinates of the data.

In this study, the data is projected in the plane of the two principal components. The direction of the features is represented in the same plane. Two ellipses are showing the 0.68 probability boundary for the distribution of the two groups of diagnosis, B (benign) and M (malignant). A circle superposed over the scatter plot data helps evaluate the relative ratio between the features in the most important principal components plane. The attributes with the most magnitudes or aligned with the governing principle component have the maximum variance.

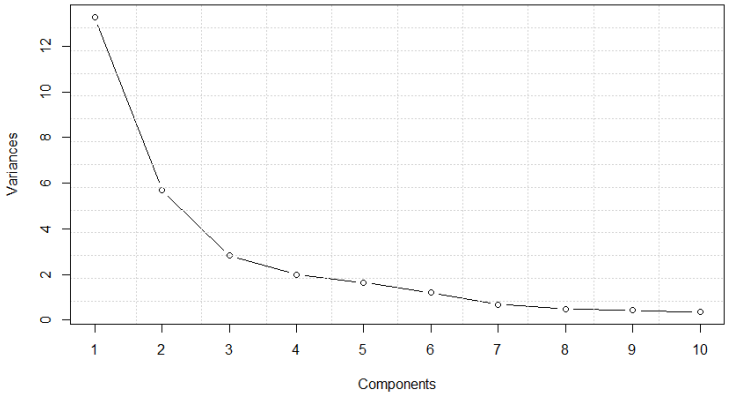


**Figure 5.** *Principle component analysis weights.*

The first two Principal Components explain together a %63.3 from the variance as shown in Figure 6.

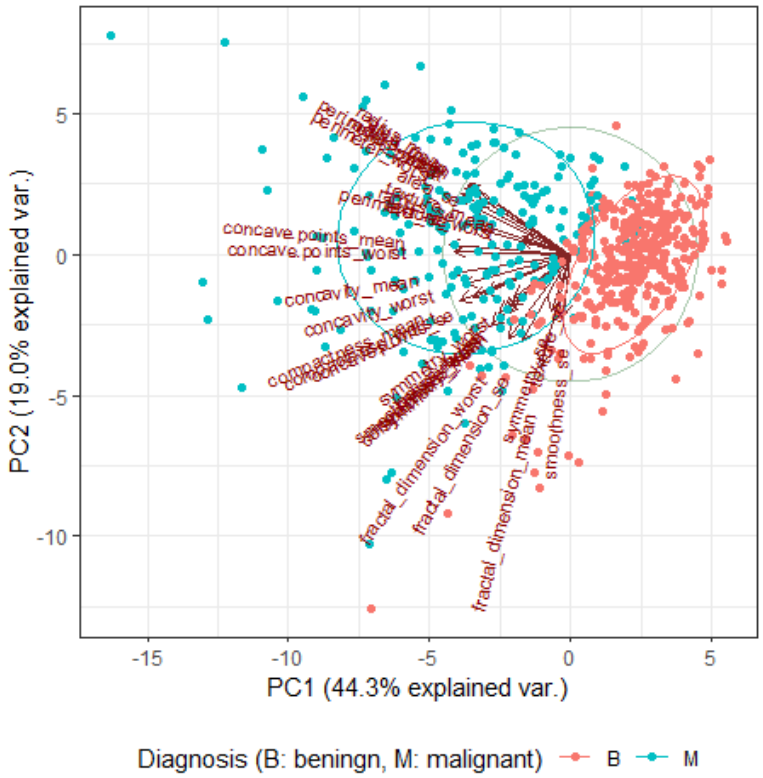


**Figure 6.** *Data distribution in the plan of PC1 and PC2.*

The projection of the data in the {PC3, PC4} and {PC5, PC6} principal components planes is shown in Figure 7. Principal components PC3-PC6 are explaining together %25.5 variations. It is also observed that not only there is no significant alignment of a certain feature with one of the PC3:PC6 principal components but also in the planes {PC3, PC4} and {PC5, PC6} the B and M points are not separated in distinct clusters like it is the case in the {PC1, PC2} plane.

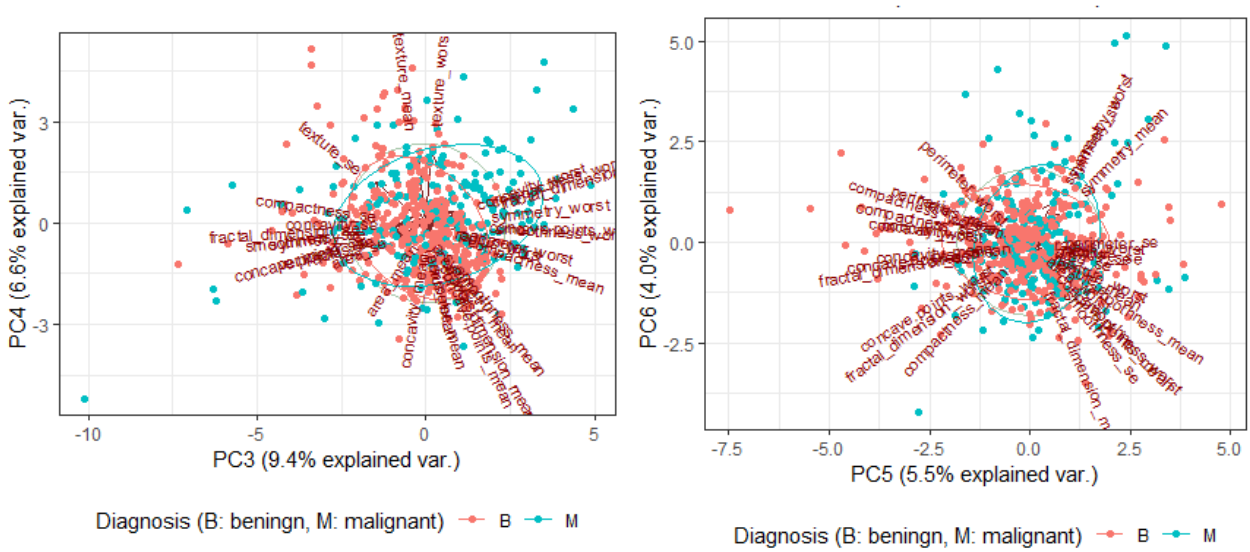


**Figure 7.** *Data distribution in the plan of PC3-PC4-PC5 and PC6.*

### 3.2. Gradient Boosting Machine (GBM)

Gradient Boosting Machine (GBM) was presented by Jerome Friedman in 2001 [23]. Other than this name, GBM is also referred to as MART (Multiple Additive Regression Trees) or GBRT (Gradient Boosting Regression Trees) in the literature. It constructs a forward stage-wise additive model by implementing gradient descent in function space. Cross-validation with five folds is used in the study. To find the best number of trees to use for the prediction for the test data, 'gbm.perf` function is used [24]. Mathematical implementation of boosting in gbm is explained below [25]:

Initialize $\hat{f}(x)$ to be a constant with sampling rate $p$, $\hat{f}(x) = arg\ min_{\rho} \sum_{i=1}^{N} \Psi(y_i, p)$ (1)
For $t$ in $1$,..., $T$ number of iterations do

1. Compute the negative gradient as the working response

$$z_i = -\frac{\delta}{\delta f(x_i)} \Psi\left(y_i, f(x_i)\right) \Big|_{f(x_i) = \hat{f}(x_i)} \quad (2)$$

2. Randomly select $p \times N$ cases from the dataset.
3. Fit a regression tree with the depth of each tree K terminal nodes, $g(x) = E(z|x)$. This tree is fit using only those randomly selected observations.
4. Compute the optimal terminal node predictions, $\rho_1, \ldots, \rho_k$, as

$$p_k = \arg min \rho \sum_{x_i \in S_k} \Psi\left(y_i, \hat{f}(x_i) + p\right) \quad (3)$$

where $S_k$ is the set of $x$s that define terminal node $k$. Again this step uses only the randomly selected observations.

5. Update $\hat{f}(x)$ with learning rate $\lambda$ as

$$\hat{f}(x) \leftarrow \hat{f}(x) + \lambda_{p_{k(x)}} \quad (4)$$

where $k(x)$ indicates the index of the terminal node into which an observation with features $x$ would fall. This function is used because it returns the optimal number of trees for prediction.

GBM parameters used in the study are given in Table 1 and Figure 8.

**Table 1.** *Gradient Boosting Machine parameters.*

| distribution | bernoulli |
|---|---|
| n.trees | 500 |
| shrinkage | 0.1 |
| n.minobsinnode | 15 |
| cv.folds | 5 |
| n.cores | 1 |

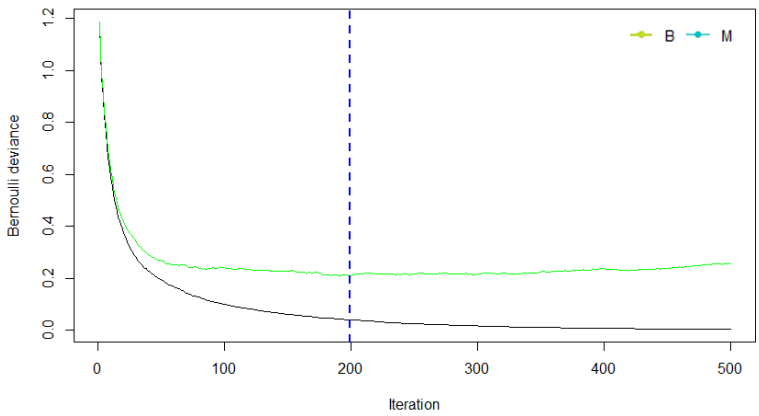


**Figure 8.** *The number of trees in GBM.*

### 3.3. eXtreme Gradient Boost (XGBoost)

Extreme Gradient Boosting (XGBoost) is a machine learning algorithm that has become increasingly popular in both the data science and remote sensing field with its high classification performance and works based on gradient boosted decision trees [26]. XGBoost, one of the new generation community learning algorithms, increases the general accuracy (performance) of the model by preventing the problem of overfitting during the training process of the algorithm. The main reason underlying the success of this method is the purpose function it uses in the learning process. The objective function consists of the loss/loss function and the term regularization. The loss/loss function calculates the difference of each predicted value made by the model from its (predicted class) real value. The term regularization, on the other hand, controls the complexity of the model, and this eliminates the overfitting problem in the model [27][28]. The parameters determined within the scope of the study are given in Table 2.

**Table 2.** *XGBoost parameters.*

| **objective** | **binary:logistic** |
|---|---|
| eval_metric | AUC |
| eta | 0.012 |
| subsample | 0.8 |
| max_depth | 8 |
| colsample_bytree | 0.9 |
| min_child_weight | 5 |
| nfold | 5 |
| nrounds | 5000 |
| nthread | 1 |
| early_stopping_round | 100 |

The most suitable parameter values (for the number of leaves and learning rate parameters) were found by the grid search algorithm. The most suitable parameters were searched between [100-1000] for the number of leaves and [0.1-1] for the learning rate. The grid search algorithm trains all combinations for the parameter pair (number of leaves and learning rate parameters) and calculates the model accuracy with the cross-validation technique. The parameter pair that gives the highest model accuracy is accepted as the most suitable parameter value. The binary logistic objective function is used for the process. The evaluation metric is chosen as AUC (area under the curve). Initial start is given by n = 0.012, subsample=0.8, max_depth=8, colsample_bytree=0.9 and min_child_weight=5. Model is trained using cross-validation with five folds. The number of rounds equal to 5000, with early stopping criteria for 100 steps, are used. Also, the frequency of printing partial is set for results every 100 steps. The area under curve values in Figure 9.

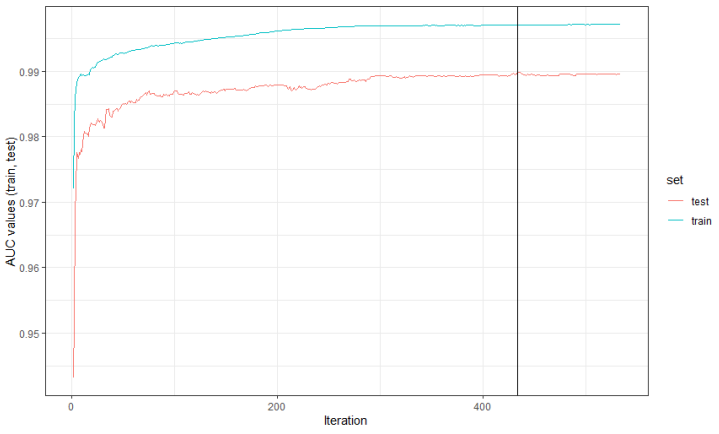


**Figure 9.** *The area under curve values for XGBoost.*

### 3.4. Light Gradient Boosting Machines (LightGBM)

Light gradient enhancement machines (LightGBM) are one of the next-generation community learning algorithms based on decision trees operating under the framework of gradient enhancement [29]. Developed by Microsoft in 2017, LightGBM has attracted the attention of researchers in both data science and remote sensing, especially with its outstanding achievements in machine learning competitions in recent years. As the method's

name suggests, it takes the prefix "Light" because it is an algorithm with high processing speed. A feature that distinguishes this method from other gradient enhancement algorithms is the growth strategy used in the training of decision trees. While LightGBM uses the vertical growth strategy (leaf growth), other gradient enhancement methods use the level-wise growth strategy. Another important detail that makes LightGBM exclusive and unique is the two new algorithms it contains and enables it to increase processing speed. These algorithms are gradient-based one-sided sampling (Gradient-based One-Side Sampling, GOSS) and exclusive feature bundling (EFB) methods[30]. It uses the subsampled dataset that it generates from the data instead of using the entire data with gradient-based one-sided sampling. It also reduces transaction complexity by converting sparse features to more frequent/dense features with exclusive feature support. The parameters used in the classification process are shown in Table3. The selection of the parameters has been chosen in consideration of the suggestions on the "parameter setting" page on the main page of LightGBM. Light GBM grows trees vertically while another algorithm grows trees horizontally, meaning that Light GBM grows tree leaf-wise while another algorithm grows level-wise. LightGBM parameters are given in Table 3.

**Table 3.** *LightGBM parameters.*

| **params** | **lightGBM.grid** |
|---|---|
| learning_rate | AUC |
| num_leaves | 0.012 |
| num_threads | 0.8 |
| nrounds | 8 |
| early_stopping_rounds | 0.9 |
| eval_freq | 5 |
| eval | |
| nfold | 5 |
| stratified | 5000 |

## 4. Results

The confusion matrix table was used for performance measurements in this study, as is done in most studies. The data set is randomly divided into train (%70 of data = 398) and test parts (%30 of data = 171) at random. The train set is used for finding the optimal model parameters by cross-validation whereas the test set is used only to measure the performance of the model. Accuracy success is measured by the percentage of correctly classified data specimens. Table 4 shows the confusion matrix values for algorithms with indicating true and predicted values.

**Table 4.** *Confuison matrix values.*

| | | |
|---|---|---|
| Malignant | GBM = 1<br>XGBoost = 4<br>LightGBM = 1 | GBM = 62<br>XGBoost = 60<br>LightGBM = 56 |
| Benign | GBM = 105<br>XGBoost = 102<br>LightGBM = 111 | GBM = 3<br>XGBoost = 5<br>LightGBM = 3 |
| | Benign | Malignant |

Receiver Operating Characteristic (ROC) is the ratio of True Positive Rate (TPR) and False Positive Rate (FPR) that is a metric by which we understand whether the models established to solve classification problems are working well. The Roc Curve (Curve) is the curve showing the values that FPR will take if TPR increases, ie 1 convergence. Here, it is desired that the TPR converges to 1, but it is desired that the FPR remain low in addition to the situation where this TPR converges to 1. Area Under the Curve(AUC) is the area under the ROC curves. As this area approaches 1, the performance of the model increases. As shown in Table 5, GBM provided the most successful solution with 0.972. LightGBM had 0.970 and XGBoost had 0.963 AUC values.

**Table 5.** *Area under curve (AUC) values.*

| algorithm | value |
|---|---|
| Gradient Boosting Machine (GBM) | 0.972 |
| eXtreme Gradient Boost (XGBoost) | 0.963 |
| Light Gradient Boosting Machine (LightGBM) | 0.970 |

## 5. Conclusion

In this study, the evaluation of pathology images with deep learning architectures is presented. Gradient Boosting Machine (GBM), eXtreme Gradient Boost (XGBoost), and Light Gradient Boosting Machine (LightGBM) have been used because of their minimum classification error. With these architectures, it is possible to detect cancer cells without manually extracting any feature. A success rate of %97 ± 1 was determined. It has been observed that each deep learning model gives different success results. The highest success is obtained from GBM architecture. The hybrid boosting models can improve the classification performance, but training times and degrees of complexity will increase. Microscopic images of breast tissue must be interpreted in the diagnosis of breast cancer among women. A digital medical photography technique is used by doctors to detect breast cancer. However, for accurate detection, it should be sufficient in the field and spend more time. Computer-aided systems recommended assisting experts are extremely important. Various techniques have been developed for these systems to detect cancer and accurately display cancer cells on the monitor. Advances in digital imaging techniques have automated the diagnostic methods recommended in the pathology workflow. This situation accelerates the diagnosis of the disease. The superior success of deep learning on classification and feature extraction has also shown itself in this area. Comparison of classification methods for diagnosis has been proposed, which can be used in addition to the success of breast cancer diagnosis and the competence of the histopathologist. Continuation of such studies is of great importance and essential in the medical field.

## Declaration of Interest

The authors declare that there is no conflict of interest.

## References

[1] N. Yassin, S. Omran, M. E. Houby, and H. Allam, “Machine learning techniques for breast cancer computer aided diagnosis using different image modalities,” A Systematic Review. Comput Meth Prog Bio, vol. 156, no. 1, pp. 25-45, 2018.

[2] M. Allinen, R. Beroukhim, L. Cai, C. Brennan, J. Lahti-Domenici, H. Huang, and K. Polyak, “Molecular characterization of the tumor microenvironment in breast cancer,” Cancer Cell, vol. 6, no. 1, pp. 17-32, 2004.

[3] S. H. Jafari, Z. Saadatpour, A. Salmaninejad, F. Momeni, M. Mokhtari, J. S. Nahand, and M. Kianmehr, “Breast cancer diagnosis: imaging techniques and biochemical markers,” Journal of Cellular Physiology, vol. 233, no. 7, pp. 5200-5213, 2018.

[4] N. Antropova, B. Q. Huynh, and M. L. Giger, “A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets,” Medical physics, vol. 44, no. 10, pp. 5162-5171, 2017.

[5] W. Yue, Z. Wang, H. Chen, A. Payne, and X. Liu, “Machine learning with applications in breast cancer diagnosis and prognosis” Designs,vol, 2, no. 2, pp. 13, 2018.

[6] F. Spanhol, E. Oliveira, C. Petitjean, and L. Heutte, “Breast cancer histopathological image classification using convolutional neural networks,” International Joint Conference on Neural Networks (IJCNN), vol. 32, no. 4, pp. 2560-2567, 2016.

[7] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, “Breast cancer multi-classification from histopathological images with structured deep learning model,” Scientific Reports, vol. 7, no. 1, pp. 4172-4182, 2017.

[8] H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, “A support vector machine-based ensemble algorithm for breast cancer diagnosis,” European Journal of Operational Research, vol. 267, no. 2, pp. 687-699, 2018.

[9] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck, “Deep learning for identifying metastatic breast cancer,” arXiv preprint arXiv:1606.05718, 2016.

[10] B. E. Bejnordi, M. Veta, P. J. V. Diest, B. V. Ginneken, N. Karssemeijer, G. Litjens, and CAMELYON16 Consortium. “Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer,” Jama, vol. 318, no. 22, pp. 2199-2210, 2017.

[11] A. Yala, C. Lehman, T. Schuster, T. Portnoi, and R. Barzilay, “A deep learning mammography-based model for improved breast cancer risk prediction,” Radiology, vol. 292, no. 1, pp. 60-66, 2019.

[12] S. Khan, N. Islam, Z. Jan, I. U. Din, and J. J. C. Rodrigues, “A novel deep learning based framework for the detection and classification of breast cancer using transfer learning,” Pattern Recognition Letters, vol. 125, pp. 1-6, 2019.

[13] P. Filipczuk, T. Fevens, A. Krzyzak, and R. Monczak, “Computer-Aided Breast Cancer Diagnosis Based on the Analysis of Cytological Images of Fine Needle Biopsies,” IEEE Trans. Med. Imaging, vol. 32, no. 12, pp. 2169-2178, 2013.

[14] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab, “Aggnet: deep learning from crowds for mitosis

detection in breast cancer histology images," IEEE transactions on medical imaging, vol. 35 no. 5, pp. 1313-1321, 2016.

[15] L. Q. Zhou, X. L. Wu, S. Y. Huang, G. G. Wu, H. R. Ye, Q. Wei, and C. F. Dietrich, "Lymph node metastasis prediction from primary breast cancer US images using deep learning," Radiology, vol. 294 no. 1, pp. 19-28, 2020.

[16] T. Araujo, G. Aresta, F. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, and A. Campilho, "Classification of breast cancer histology images using convolutional neural networks," Plos One, vol. 12 no. 6, e0177544, 2017.

[17] N. Bayramoglu, J. Kannala, and J. Heikkila, "Deep learning for magnification independent breast cancer histopathology image classification," 23rd International Conference on Pattern Recognition (ICPR) '2016 pp. 2440-2445.

[18] Z. Alom, C. Yakopcic, M. Taha, and K. Asari, "Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network," J Digit Imaging, vol. 45, no. 3, pp. 1-13, 2019.

[19] M. Veta, P. Pluim, J. Van, and A. Viergever, "Breast cancer histopathology image analysis: A review," IEEE T BIO-MED ENG, vol. 61, no. 5, pp. 1400-1411, 2014.

[20] D. Dua, and C. Graff, "UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]," Irvine, CA: University of California, School of Information and Computer Science, 2019.

[21] Karl Pearson F.R.S. "LIII. On lines and planes of closest fit to systems of points in space" The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, vol. 2, no. 11, pp. 559–572, https://doi.org/10.1080/14786440109462720, 1901.

[22] H. Hotelling, "Analysis of a complex of statistical variables into principal components," Journal of Educational Psychology, vol. 24, no. 6, pp. 417, 1993.

[23] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," Annals of Statistics, pp. 1189–1232, 2001.

[24] B. Greenwell, B. Boehmke, J. Cunningham, G. B. M. Developers, M. B. Greenwell, 2019. Package 'gbm'.

[25] G. Ridgeway, "Generalized Boosted Models: A guide to the gbm package," Update, vol. 1, no. 1, 2007.

[26] T. Chen, and C. Guestrin, "Xgboost: a scalable tree boosting system," In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining '07,2016 pp. 785-794.

[27] A. M. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data," GIScience and Remote Sensing, vol. 57, no. 1, pp. 1-20, 2020. doi: 10.1080/15481603.2019.1650447

[28] L. Rumora, M. Miler, and D. Medak, "Impact of various atmospheric corrections on sentinel-2 land cover classification accuracy using machine learning classifiers," ISPRS International Journal of Geo-Information, vol. 9, no. 4, pp. 277, doi: 10.3390/ijgi9040277.

[29] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, … T.-Y. Liu, "Lightgbm: a highly efficient gradient boosting decision tree," Advances in Neural Information Processing Systems, vol. 30, pp. 3146–3154, 2017.

[30] R. Wang, Y. Liu, X. Ye, Q. Tang, J. Gou, M. Huang, and Y. Wen, "Power system transient stability assessment based on bayesian optimized LightGBM," In 2019 IEEE 3rd Conference on Energy Internet and Energy System Integration (EI2) '11,2019, pp. 263-268.