Original Article

DOI: 10.4274/haseki.galenos.2022.8440
Med Bull Haseki 2022;60:196-203

Classification of Breast Cancer on the Strength of Potential Risk Factors with Boosting Models: A Public Health Informatics Application

Sami Akbulut*,**, Ipek Balikci Cicek**, Cemil Colak**

*Inonu University Faculty of Medicine, Department of General Surgery, Malatya, Turkey

**Inonu University Faculty of Medicine, Department of Biostatistics and Medical Informatics, Malatya, Turkey

Abstract

Aim: The diagnosis of breast cancer can be accomplished using an algorithm or an early detection model of breast cancer risk via determining factors. In the present study, gradient boosting machines (GBM), extreme gradient boosting (XGBoost) and light gradient boosting (LightGBM) models were applied and their performances were compared.

Methods: The open-access Breast Cancer Wisconsin Dataset, which includes 10 features of breast tumors and results from 569 patients, was used for this study. The GBM, XGBoost, and LightGBM models for classifying breast cancer were established by a repeated stratified K-fold cross validation method. The performance of the model was evaluated with accuracy, recall, precision, and area under the curve (AUC).

Results: Accuracy, recall, AUC, and precision values obtained from the GBM, XGBoost, and LightGBM models were as follows: (93.9%, 93.5%, 0.984, 93.8%), (94.6%, 94%, 0.985, 94.6%), and (95.3%, 94.8%, 0.987, 95.5%), respectively. According to these results, the best performance metrics were obtained from the LightGBM model. When the effects of the variables in the dataset on breast cancer were assessed in this study, the five most significant factors for the LightGBM model were the mean of concave points, texture mean, concavity mean, radius mean, and perimeter mean, respectively.

Conclusion: According to the findings obtained from the study, the LightGBM model gave more successful predictions for breast cancer classification compared with other models. Unlike similar studies examining the same dataset, this study presented variable significance for breast cancer-related variables. Applying the LightGBM approach in the medical field can help doctors make a quick and precise diagnosis.

Keywords: Breast cancer, boosting algorithm, gradient boosting algorithm, XGBoost algorithm, LightGBM algorithm

Introduction

Breast cancer, one of the most frequently diagnosed tumors in women worldwide, has become the second-largest cause of cancer-related deaths. Breast cancer is a serious global health problem: it is the most commonly diagnosed cancer worldwide, with an estimated 2.26 million cases in 2020, and the leading cause of cancer death among women. With the developments in medical treatment, the 5-year survival rate has reached 91%, and the 10-year survival rate has reached 86% (1). According to the World Health Organization, the global incidence of breast cancer is increasing rapidly due to advances in lifestyle, reproductive factors, and life expectancy. 58% of all breast cancer fatalities occur in middle-and low-income nations. While breast cancer survival rates are typically 80% in rich countries, they decrease to 60% in the middle- and 40% in low-income countries due to a lack of early screening programs, which results in incurable diagnoses in 80% of late-stage tumors (2).

Breast cancer is a leading cause of morbidity and death worldwide, and its prevalence is increasing daily. According to Global Cancer Statistics 2020 data, the incidence of breast cancer in Turkey was calculated as

Address for Correspondence: Sami Akbulut,
Inonu University Faculty of Medicine, Department of General Surgery; Department of Biostatistics and Medical Informatics, Malatya, Turkey
Phone: +90 422 341 06 60 E-mail: akbulutsami@gmail.com ORCID: orcid.org/0000-0002-6864-7711
Received: 30.04.2022 **Accepted:** 18.05.2022

10.6% (22345 individuals, both genders, all ages). In the same report, it is reported that approximately one out of every four women (24.4%) in Turkey is diagnosed with breast cancer, and 4.7% (5452 individuals) die of breast cancer in the second rank in terms of mortality rates (3). From a public health perspective, breast cancer incidence is most commonly associated with age, among other risk factors. The incidence of breast cancer increases rapidly at the age of 40-50 toward the end of the active reproductive age and decreases slightly after menopause around the age of 50. The relatively low incidence and mortality of breast cancer causes it to be the most prevalent type of cancer (4).

Breast cancer is thought to be a genetically varied and physiologically diverse illness. Disparities in gene expression are linked to long-known clinical and phenotypic differences. Previous research on breast cancer has revealed five distinct subtypes [luminal A (estrogen receptor (ER +); luminal B (ER +); HER2 overexpression; normal breast-like and basallike] that are associated with varying clinical outcomes. Early detection and classification of breast cancer development allows patients to obtain proper treatment. The basal-like subtype is typically ER and HER2 negative (i.e. not amplified) and resembles breast myoepithelial cells. The basal-like subtype has been associated with BRCA1-associated carcinomas and has the highest proliferation rates and poor clinical outcomes (5,6).

The diagnosis of breast cancer can be accomplished using an algorithm or an early detection model of breast cancer risk via determining factors. This model is used to detect breast cancer risk and is a preventive action that uses machine learning to classify the risk of breast cancer associated with variable predictors, making it easier to classify.

Machine learning is a type of artificial intelligence that allows computers to learn without being explicitly programmed. Machine learning is an area of artificial intelligence technology that employs algorithms to synthesize the underlying relationship between data and information (7). The scientific field of machine learning concerns how computers learn from data; it is also a type of artificial computer intelligence that enables computers to learn automatically and without human involvement or aid (8). Machine learning has been used in cancer detection and diagnosis. Using machine learning algorithms, tumors and other malignancies have been identified, classified, detected, and distinguished. In other words, machine learning has mostly been used to assist in diagnosing and detecting cancer (9,10). Scientific studies on the application of machine learning methods in health care have demonstrated that machine learning significantly affects health quality and safety (11,12).

This study compares the breast cancer classification performance of the gradient boosting machines (GBM), extreme gradient boosting (XGBoost), and light gradient boosting (LightGBM) models, which are boosting algorithms on the open-access breast cancer dataset, and evaluates the associate with breast cancer are determined.

Materials and Methods

Compliance with Ethical Standards

Ethical approval was not applicable and not obtained for this study because the open access dataset was used in this study. This study was conducted in accordance with the Declaration of Helsinki.

Study Design

The data were collected from the Department of General Surgery at the University of Wisconsin-Madison and presented to users as open access. The open-access dataset "Breast Cancer Wisconsin (Diagnostic) Data Set" was collected from the UCI Machine Learning Repository to study the light gradient boosting method's operation and to evaluate the model (13).

Variables in the Study

The following variables were used for this study: diagnosis (malignant, benign), radius mean (mean of distances from the center to points on the perimeter), texture mean (the standard deviation of grayscale values), perimeter mean (mean size of the core tumor), area mean, smoothness mean (mean of local variation in radius lengths), compactness mean (mean of perimeter 2/area - 1.0), concavity mean (mean of the severity of concave portions of the contour), concave points mean (mean for the number of concave portions of the contour), symmetry mean, fractal dimension mean (mean for "coastline approximation" - 1).

Boosting Algorithm

It was developed by Schapire in 1989. Recent work by Freud and Schapire and Friedman has further developed this algorithm. The boosting algorithm is a sequential method based on slow learning that tries to learn from errors. These algorithms combine several low-precision models to create high-precision models. As a general principle, it tries to obtain a strong model by combining the models obtained in each iteration within the framework of specific rules (14). First, random samples are generated from the training data during the boosting algorithm process. A classifier was trained for this sample, and the entire training data was tested. An error was calculated for each sample estimate. If the sample is misclassified, the weight is increased for that sample, and another sample is

created. These processes are repeated until a high degree of accuracy is obtained from the system (15). Within the scope of boosting algorithms, Light Gradient Boosting, one of the tree-based methods, will be used in this study.

Gradient Boosting Algorithm

GBM are learning algorithms that fit new models sequentially to obtain a more accurate estimate of the response variable. This strategy's basic idea is to generate new base learners with the highest correlation to the ensemble's negative gradient of the loss function. Although the loss functions can be chosen arbitrarily, for clarity's sake, if the error function is the conventional squared-error loss, the learning strategy will result in consecutive error fitting. In general, the researcher can choose the loss function, given the breadth of already determined loss functions and the possibility of constructing one's own task-specific loss (16,17).

Due to this great degree of adaptability, GBM may be tailored to any data-driven job. It introduces a great deal of flexibility into the model design, making the selection of the optimal loss function a question of trial and error. However, Boosting algorithms are reasonably simple to implement, allowing for experimentation with various model designs. Additionally, GBM has demonstrated tremendous effectiveness in various machine-learning and data-mining difficulties (18).

XGBoost Algorithm

XGBoost, short for extreme gradient boosting, is a machine learning method based on gradient boosting and decision tree algorithms. Friedman developed the original version of the XGBoost algorithm in 2002 (15). XGBoost is a viral algorithm, and it is used for health, energy, finance, etc. It has found applications in the fields. Compared to other algorithms, it is in a very advantageous position in terms of speed and performance. XGBoost has high accuracy for both classification and regression models. It is also 10 times faster than other algorithms. XGBoost includes a set of tweaks that improve performance and reduce overfitting or overlearning, thus achieving better performance. In addition, it ensures that the accuracy of the model is maximized by cross-validating itself without considering any parameters (19).

LightGBM Algorithm

The LightGBM algorithm is a different gradient boosting model that uses decision trees. Regression was used for classification and ranking analysis. Two strategies can be used when training each decision tree and separating data, focusing on the level of the tree (level-wise) and focusing on the leaves of the tree (leaf-wise). In the level-wise condition, the tree grows while maintaining the balance of the tree, while the leaf-wise strategy continues to split the leaf that reduces the loss the most. LightGBM's leaf-wise growing tree structure selects the losses in a particular branch and splits them based on their contribution to the overall loss. In most cases, trees with lower error rates learn faster than other depth-focused growing tree-based models (20). While the leaf-wise growth strategy can create any tree by level-wise training, the reverse is not true. Because of these features, over-learning can be prevented since the LightGBM model grows mainly horizontally and the tree depth does not increase much. This gives better results, especially for large datasets (21). Another advantage of the LightGBM model is that it does not require processes such as one-hot encoding to numerically analyze data with categorical variables. It shortens the training time of the model and reduces resource usage by converting the variables with continuous values into categorical values. In the studies conducted on different data sets, it has been concluded that the data learning process of the LightGBM model is 20 times faster than that of other models (22).

Repeated Stratified K-fold Cross-validation

This approach repeats the stratified K-fold cross validator n-times, with each repetition including distinct randomization. Stratified K-fold is similar to stratified K-fold in that the entire data set is partitioned into k subgroups. It approximately maintains the same percentage of samples from each target class during each cycle (23).

Statistical Analysis and Modeling

Quantitative variables are summarized using the median (minimum-maximum) method, while qualitative variables are expressed in terms of numbers and percentages. The Kolmogorov-Smirnov test was used to determine whether the distribution was normal. The Mann-Whitney U test was performed to determine whether there is a statistically significant difference between the categories of the dependent variable in terms of the input variables. Statistical significance was defined as $p<0.05$ values. All analyses were conducted using the IBM SPSS Statistics 26.0 for Windows package application and the Python 3.9.7 programming language (24). According to the studies, these models are superior over other machine learning methods in terms of performance, and these models are included in the study. In the study, during the modeling phase, it was divided into training (80%) and test (20%) data sets. Analysis was carried out using the repeated stratified K-fold cross-validation method.

Results

The data set in this study included 569 patients, 357 (62.7%) benign and 212 (37.3%) malignant breast

lesions. The correlation of the variables with each other is given in Figure 1.

Descriptive statistics for the variables included in the study are given in Table 1. When Table 1 is examined; there was a statistically significant difference between the dependent variable classes (benign/malignant) in terms of texture mean (p<0.001), radius mean (p<0.001), area mean (p<0.001), perimeter mean (p<0.001), smoothness mean (p <0.001), concavity mean (p<0.001), compactness mean (p<0.001), concave points mean (p<0.001) and symmetry mean (p<0.001). However, there was no statistically significant difference between the dependent variable classes (benign/malignant) in terms of fractal dimension mean variables (p=0.537).

The performance metrics [accuracy, recall, area under the curve (AUC), and precision (positive predictive value)] computed from the models developed to classify breast cancer are listed in Table 2. For the GBM model; accuracy, recall, AUC and precision (positive predictive value) values obtained from the model were 93.9%, 93.5%, 0.984, and 93.8% respectively. For the XGBosst model; accuracy, recall, AUC and precision (positive predictive value) values obtained from the model were 94.6%, 94%, 0.985, 94.6% respectively. For the LightGBM model; accuracy, recall, AUC and precision (positive predictive value) values obtained from the model were 95.3%, 94.8%, 0.987, 95.5% respectively. The best performance metrics were obtained from the LightGBM model. The number of times a feature is used in a model determines its importance in LightGBM. The data set is partitioned into several folds based on the feature importance score. We calculate the importance of each feature in each fold and average the importance of each feature across all folds. The feature importance score will be the average of the results. The reasoning behind this is that randomization is used in each run of LightGBM fitting. As a result, the ensemble mean can provide significant evidence for the significance of traits. For the LightGBM model with the best performance metrics, the importance values of breast cancer-related factors are given in Figure 2. The five most important factors are the mean of concave points, texture, concavity, radius, and perimeter, respectively.

Discussion

Breast cancer has the greatest fatality rate among women and is the second most common cancer form worldwide. Breast cancer is one of the most serious health issues due to its poor prognosis, high mortality rate, and new cases. A large number of deaths in women are recorded each year, indicating that there is still an urgent need for more effective and timely diagnosis for appropriate therapy in breast cancer. Despite advancements in cancer diagnosis and treatment, this disease is still a major problem in health (25).

The effectiveness of machine learning approaches in classification and definition has given computer technology the power to make judgments. These benefits of machine learning approaches have resulted in enhanced decision

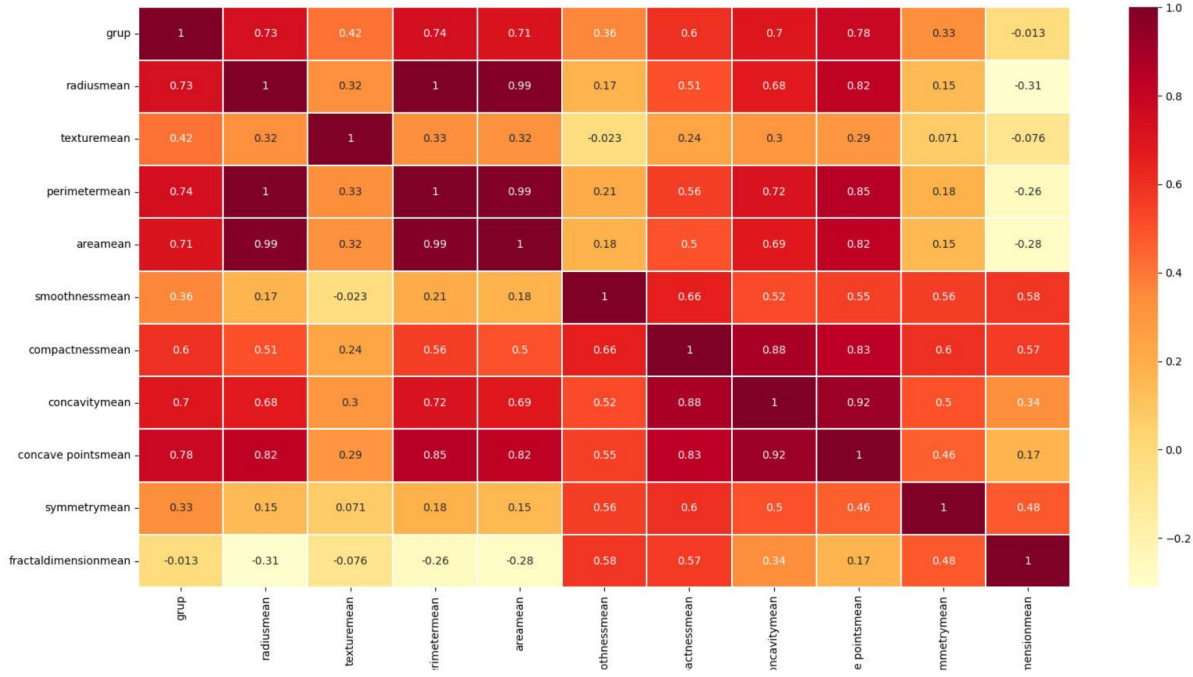

Figure 1. Correlations of the variables under question

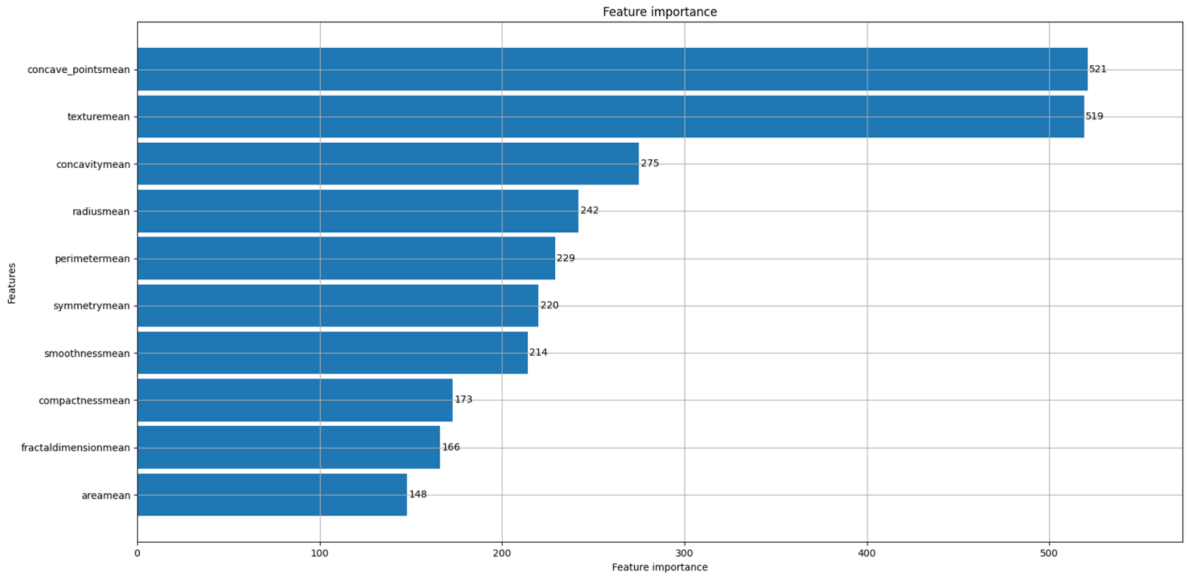

Figure 2. Importance values of variables according to LightGBM model

Table 1. Descriptive statistics for independent variables			
Variables	**Diagnosis**		**p-value**
	Benign (n=357)	**Malignant (n=212)**	
	Median (IQR)	**Median (IQR)**	
Texture mean	17.39 (4.665)	21.46 (4.4725)	<0.001
Radius mean	12.2 (2.305)	17.325 (4.525)	<0.001
Perimeter mean	78.18 (15.34)	114.2 (31.3)	<0.001
Smoothness mean	0.091 (0.018)	0.102 (0.017)	<0.001
Area mean	458.4 (175.35)	932 (500.55)	<0.001
Concavity mean	0.037 (0.0399)	0.151 (0.0939)	<0.001
Compactness mean	0.075 (0.0423)	0.132 (0.0642)	<0.001
Symmetry mean	0.171(0.0312)	0.19 (0.0364)	<0.001
Concave points mean	0.023 (0.0176)	0.086 (0.0394)	<0.001
Fractal dimension mean	0.062 (0.0073)	0.062 (0.01072)	0.537
*Mann-Whitney U test			

Table 2. The values of performance metrics			
Performance metrics / **Model**	**GBM value**	**XGBoost value**	**LightGBM value**
Accuracy (%)	93.9	94.6	95.3
Recall (%)	93.5	94	94.8
Precision (Positive predictive value) (%)	93.8	94.6	95.5
AUC	0.984	0.985	0.987
AUC: Area under the curve, GBM: Gradient boosting machines, XGBoost: Extreme gradient boosting, LightGBM: Light gradient boosting			

support systems that can assist specialists in diagnosis and treatment processes in the field of health. It is possible to achieve high success in the diagnosis of diseases using decision-support systems. Machine learning approaches, which are frequently used in cancer diagnosis processes, play an essential role in inference (26).

In this study, the breast cancer classification performances of the GBM, XGBoost, and LightGBM models, which are boosting algorithms on the open-access breast cancer dataset, are compared and the factors associated with breast cancer are determined.

Asri et al. (27) applied Support Vector Machines, decision tree (C4.5), Naive Bayes, and k-Nearest Neighbor machine learning algorithms to the Wisconsin breast cancer dataset in the UCI Machine Learning Repository in the WEKA environment. In their study, accuracy, sensitivity, sensitivity and specificity parameters were used while evaluating the classification models.

Abdel-Zaher and Eldeib (28) developed a clinical support system for detecting breast cancer. In the model used in this study, the weights were obtained from the deep belief network and the learning function of Liebenberg Marquardt, and the back propagation neural network was used. Promising accuracy was achieved compared to the previously published studies.

When the studies with the same data set are examined, the Wisconsin Original Data Set, which contains 569 records and 31 features/variables (30 predictors, 1 target), was used to improve the accuracy of breast cancer diagnosis using several machine learning algorithms. The accuracy of the suggested support vector machine model was determined to be 0.9766, and the study's findings indicated that the proposed model has a reasonable performance rate and will help increase breast cancer accuracy, a critical issue in modern times. In this study, the accuracy value for breast cancer classification was 0.9491 when only 11 features or variables (10 predictors, 1 dependent) were used on the same dataset (29). Breast cancer classification was successfully accomplished in this investigation by relying on fewer variables/features, and identical performance metrics were attained in the stated study.

Bayrak et al. (30) applied Support Vector Machines and Artificial Neural Network machine learning methods in the WEKA environment in their study on the Wisconsin (original) breast cancer dataset. The Support Vector Machines (SMO algorithm) performed best when the results of the algorithms were compared according to performance metrics such as accuracy, precision, sensitivity, and ROC area.

In the years 2020 and beyond, numerous studies have been conducted to study the classification of breast cancer using machine learning algorithms using the same data set. Rawal et al. (31) present a comparative analysis of the Wisconsin Diagnostic Data Set using several machine learning methods such as Support Vector Machine, Nave Bayes, Decision Tree, K-Nearest Neighbor, k-means clustering, and Artificial Neural Networks to detect early breast cancer.

Guldogan et al. (32) created a deep learning model for the classification of breast cancer with a 10-fold cross-validation method. Breast cancer-related factors were predicted from the deep learning model with accuracy, specificity, sensitivity, F1-score, positive and negative predictive values, and AUC. In this study, breast cancer classification was successfully performed, and similar performance success was achieved. When the effects of the variables on breast cancer were evaluated, the concave point mean and perimeter mean, two of the five most important variables, were similarly obtained during the process.

Harinishree et al. (33) have recently proposed several computer-aided frameworks to minimize multiple unnecessary breast biopsies. The mentioned article explores accessible directories of information for preparing machine learning models and presents a wide-ranging correlation between the various models to predict breast cancer.

Assegie et al. (34) analyzed the decision tree and adaptive boosting models' prediction performance. The adaptive boosting model is 92.53 percent accurate, whereas the decision tree is 88.80 percent accurate. Overall, the adaboost algorithm outperformed the decision tree approach.

Magesh and Swarnalatha (35) used decision tree, support vector machine, and SVM algorithms to predict breast cancer. The authors choose the best algorithm according to the accuracy and error rate. In the related study, data visualization and descriptive statistics were presented, and precision, recall, and F1-score measures were nearly 95% in SVM. After adjusting the SVM hiper-parameters, the accuracy increased to 97%.

Sakib et al. (36) made a comparison between machine learning and deep learning methods for breast cancer detection and diagnosis. Classification was carried out using five supervised machine learning techniques (i.e., SVM, decision tree, logistic regression, random forest, K-nearest neighbor) and a deep learning technique. The Breast Cancer Wisconsin (diagnosis) dataset was used as a training set to evaluate and compare the effectiveness and efficiency of each algorithm, including classification accuracy, recall, specificity, precision, false-negative rate, false-positive rate, F1-score, and Matthews correlation coefficient.

According to the findings obtained, GBM, XGBoost, and LightGBM, one of the boosting models showed that the classification performance on the open-access "Breast Cancer Wisconsin Dataset" gave successful predictions in classifying breast cancer according to the metric values. Among the three models, the LightGBM model gave the most successful results. In addition, the variable importance score values of cancer-related factors were estimated from the model created, unlike similar studies examining the same data set. The data set used in the study was open access and the other clinical and demographic data of the patients could not be reached. For this reason, there is no information about the subtypes of breast cancer in the study, and the findings cannot be interpreted for the subtypes. In future studies, the classification performances of many machine learning models and ensemble learning approaches can be examined to gain insights into the prediction of diseases.

Study Limitations

The main limitation of this study is that the data set was collected from a single center and shared. For this reason, the results obtained cannot be generalized according to multicenter studies and provide inferences for a certain region. The superiority of this study compared to other studies is that the importance of the variables obtained as a result of the modeling is given. Thus, risk factors that may be important related to breast cancer have been revealed.

Conclusion

Any advancements in the early detection and prediction of cancer and the implementation of alternative treatment procedures are essential for treatment. We compared the performance of three algorithms for breast cancer prediction. Applying the LightGBM approach in the medical field can help doctors make a quick and precise diagnosis. So we can help patients and doctors save time, and we can reduce medical testing and time limits.

Ethics

Ethics Committee Approval: Ethical approval was not applicable and not obtained for this study because the open access dataset was used in this study. This study was conducted in accordance with the Declaration of Helsinki.

Informed Consent: The open access dataset study.

Peer-reviewed: Externally peer-reviewed.

Authorship Contributions

Concept: S.A., I.B.C., C.C., Design: S.A., I.B.C., Data Collection, or Processing: S.A., I.B.C., C.C., Analysis, or Interpretation: I.B.C., C.C., Literature Research: S.A., I.B.C., Writing: S.A., I.B.C., C.C.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declare that this study received no financial support.

References

1. Lee J, Lee MG. Effects of Exercise Interventions on Breast Cancer Patients During Adjuvant Therapy: A Systematic Review and Meta-analysis of Randomized Controlled Trials. Cancer Nurs 2020;43:115-25.
2. Ping J, Guo X, Ye F, et al. Differences in gene-expression profiles in breast cancer between African and European-ancestry women. Carcinogenesis 2020;41:887-93.
3. Gul A, Aygin D. Lymphedema and Air Travel After Breast Cancer Surgery. IGUSABDER 2021;15:669-80.
4. Haydaroglu A, Cakar B, Gokmen E, et al. Epidemiological and overall survival characteristics of breast cancer patients in Ege University Hospital database. Ege Journal of Medicine 2019;58:50-7.
5. Peterson AC, Uppal H. Method for predicting response to breast cancer therapeutic agents and method of treatment of breast cancer. Google Patents; 2019.
6. Arslan AK, Tunc Z, Cicek IB, Colak C. A novel interpretable web-based tool on the associative classification methods: an application on breast cancer dataset. The Journal of Cognitive Systems 2020;5:33-40.
7. Yilmaz R, Yagin FH. Early Detection of Coronary Heart Disease Based on Machine Learning Methods. Medical Records 2022;4:1-6.
8. Awad M, Khanna R. Efficient learning machines: theories, concepts, and applications for engineers and system designers. Springer nature; 2015.
9. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009;338:b606.
10. Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques Third Edition [M]. The Morgan Kaufmann Series in Data Management Systems 2011;5:83-124.
11. Buchlak QD, Esmaili N, Leveque JC, et al. Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. Neurosurg Rev 2020;43:1235-53.
12. Yagin FH, Yagin B, Arslan AK, Colak C. Comparison of Performances of Associative Classification Methods for Cervical Cancer Prediction: Observational Study. Turkiye Klinikleri J Biostat 2021;13:266-72.
13. Dua D, Graff C. UCI Machine Learning Repository [http://archive. ics. uci. edu/ml]. Irvine, CA: University of California. School of Information and Computer Science. 2019;25:27.
14. Telli S. [Emotion detection and recognition on twitter using ensemble learning] (Thesis). Izmir (Turkey): Ege Univ; 2019.
15. Yangin G. [Application of XGboost and decision tree based algorithms on Diabetes Data] (Thesis). Istanbul (Turkey): Mimar Sinan Fine Arts Univ; 2019.

16. Pittman SJ, Brown KA. Multi-scale approach for predicting fish species distributions across coral reef seascapes. PLoS One 2011;6:e20583.
17. Hutchinson R, Liu L-P, Dietterich T. Incorporating boosted regression trees into ecological latent variable models. Proceedings of the AAAI Conference on Artificial Intelligence 2011;25:1343-8.
18. Johnson R, Tong Zhang. Learning Nonlinear Functions Using Regularized Greedy Forest. IEEE Trans Pattern Anal Mach Intell 2014;36:942-54.
19. Ekiz E. [Prediction of debt collection behaviour with machine learning techniques: A case study on telecommunication company customers] (Thesis). Istanbul (Turkey): Istanbul Technical Univ; 2019.
20. Kesici M. [Wide area measurement based early prediction of power system transient instability and its evolution using deep learning and decision tree based algorithms] (Thesis). Istanbul (Turkey): Istanbul Technical Univ; 2019.
21. Gumustas E. [Classification with ensemble methods on missing and imbalanced data]. (Thesis). Istanbul (Turkey): Mimar Sinan Fine Arts Univ; 2019.
22. Ke G, Meng Q, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems 2017;30:3146-54.
23. Artur M. Review the performance of the Bernoulli Naïve Bayes Classifier in Intrusion Detection Systems using Recursive Feature Elimination with Cross-validated selection of the best number of features. Procedia Computer Science 2021;190:564-70.
24. Python R. Python. Python Releases for Windows. 2019;24.
25. Sadoughi F, Dana PM, Asemi Z, et al. Molecular and cellular mechanisms of melatonin in breast cancer. Biochimie 2022:S0300-9084(22)00067-0.
26. Kavitha T, Mathai PP, Karthikeyan C, et al. Deep Learning Based Capsule Neural Network Model for Breast Cancer Diagnosis Using Mammogram Images. Interdiscip Sci 2022;14:113-29.
27. Asri H, Mousannif H, Al Moatassime H, Noel T. Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Computer Science 2016;83:1064-9.
28. Abdel-Zaher AM, Eldeib AM. Breast cancer classification using deep belief networks. Expert Systems with Applications 2016;46:139-44.
29. Kor H. Classification of Breast Cancer by Machine Learning Methods. SETSCI Conference Proceedings 2019;4:508-11.
30. Bayrak EA, Kırcı P, Ensari T. Comparison of machine learning methods for breast cancer diagnosis. 2019 Scientific meeting on electrical-electronics & biomedical engineering and computer science (EBBT): IEEE; 2019. p. 1-3.
31. Rawal G, Rawal R, Shah H, Patel K. A Comparative Study Between Artificial Neural Networks and Conventional Classifiers for Predicting Diagnosis of Breast Cancer. ICDSMLA 2019. Springer; 2020. p. 261-71.
32. Guldogan E, Tunc Z, Colak C. Classification of Breast Cancer and Determination of Related Factors with Deep Learning Approach. The Journal of Cognitive Systems 2020;5:10-4.
33. Harinishree M, Aditya C, Sachin D. Detection of Breast Cancer using Machine Learning Algorithms–A Survey. 2021 5th International Conference on Computing Methodologies and Communication (ICCMC): IEEE; 2021. p. 1598-601.
34. Assegie TA, Tulasi RL, Kumar NK. Breast cancer prediction model with decision tree and adaptive boosting. IAES International Journal of Artificial Intelligence 2021;10:184.
35. Magesh G, Swarnalatha P. Analysis of breast cancer prediction and visualisation using machine learning models. International Journal of Cloud Computing 2022;11:43-60.
36. Sakib S, Yasmin N, Tanzeem AK, Shorna F, Alam SB. Breast Cancer Detection and Classification: A Comparative Analysis Using Machine Learning Algorithms. Proceedings of Third International Conference on Communication, Computing and Electronics Systems: Springer; 2022. p. 703-17.