ISSN: 2538-5887; Pathobiology Research. 2022;25(4): 7-15

Classifying Breast Tumors as Malignant or Benign Using Digitized Images of Fine Needle Aspiration Samples of Breast Mass Tissue: An Application of Classification Tree Algorithms

ARTICLE INFO

Article Type
Original Research

Authors

Mina Jahangiri[1]

Anoshirvan Kazemnejad[*]

1- Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran.

***Corresponding author:**

Anoshirvan Kazemnejad

Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran, P.O.Box: 14115-111, Tel: +982182883875, Fax: +982182884555, E-mail: kazem_an@modares.ac.ir.

ABSTRACT

Introduction: Breast cancer represents a major public health issue worldwide, highlighting the critical role of early detection in facilitating effective treatment. Fine needle aspiration (FNA) serves as a minimally invasive method for obtaining cellular material from breast masses for subsequent analysis. Nonetheless, pathologists' assessment of FNA samples may be characterized by subjectivity and protracted evaluation times, leading to variability in diagnostic results. Integrating machine learning algorithms, including classification tree models, can potentially improve the consistency and precision of breast tumor classification. Using computational capabilities and sophisticated machine learning methodologies, these models can proficiently categorize digitized images of FNA samples as malignant or benign.

Methods: We used classification tree algorithms such as CART, Ctree, Evtree, QUEST, CRUISE, and GUIDE to distinguish between malignant and benign tumors in the Wisconsin Breast Cancer Dataset (WBCD). The models' performance was evaluated using accuracy metrics, such as sensitivity, specificity, false positive and negative rates, positive and negative predictive values, Youden's Index, accuracy, positive and negative likelihood ratios, diagnostic odds ratios, and AUC (area under the ROC curve).

Results: The results showed that the CRUISE algorithm showed excellent diagnostic performance in distinguishing between malignant and benign tumors.

Conclusion: The results emphasize the critical role of integrating machine learning models into clinical practice to assist pathologists, improve diagnostic outcomes, and reduce subjectivity in cancer classification.

Keywords: Breast cancer, benign tumor, malignant tumor, prediction, diagnostic scheme, classification trees.

INTRODUCTION

Breast cancer is a significant and widespread life-threatening disease that is a primary concern in women's health globally. Early detection is vital for effective treatment and better survival rates. Fine Needle Aspiration (FNA) is a minimally invasive procedure that involves collecting cellular material from breast masses for cytological analysis (1-6).

FNA is essential for assessing the nature of breast tumors by extracting a small number of cells from breast tissue. However, evaluating FNA samples by pathologists can be time-consuming and subjective, leading to considerable variability in diagnostic outcomes among professionals (7). Incorporating machine

learning algorithms into the diagnostic process could improve the consistency and accuracy of breast tumor classification (8).

Recent advancements in computational power and machine learning techniques have enabled the creation of advanced models for analyzing medical images. These models can effectively classify digitized images of FNA samples as either malignant or benign. One commonly used machine learning method in this area is the classification tree algorithm, a well-established supervised learning technique suitable for classification and regression tasks. This method is particularly advantageous in medical contexts due to its interpretability, ability to handle nonlinearity, and robustness against noisy data (9-15).

Breast tumors are generally categorized as either malignant or benign, and distinguishing between these types is crucial for determining the appropriate treatment. Nonneoplastic cases typically do not require aggressive treatment, while malignant tumors demand prompt and targeted therapies. A key challenge is accurately identifying features from FNA images that differentiate malignant tumors from benign ones. Traditionally, pathologists have focused on various characteristics, such as cell size, shape, uniformity, and the presence of mitoses. However, automating the feature extraction process and feeding this data into classification models like decision trees could significantly improve diagnostic accuracy (16, 17).

Decision trees work by recursively partitioning the feature space, assessing the most informative features at each stage to create a tree structure that links input features to output labels—benign or malignant. These models provide valuable insights into decision-making, essential for their acceptance in clinical settings (9-15).

This study investigates the application of classification tree algorithms to identify and categorize breast tumors as malignant or benign using digitized images of FNA samples from breast tissue. These algorithms aim to enhance the accuracy and effectiveness of breast cancer diagnoses, providing valuable support to pathologists in their evaluations. Additionally, the paper will empirically assess the performance of various classification tree algorithms using a well-known breast cancer dataset (the Wisconsin Breast Cancer Dataset).

Description of the data set

The dataset utilized in this study is the Wisconsin Breast Cancer Dataset (WBCD), which contains digitized images of FNA samples of breast mass tissue. This dataset includes features extracted from FNA samples, such as cell size, shape, and texture, essential for distinguishing between benign and malignant tumors. The WBCD is a widely used data set for a wide range of breast cancer research. Most analyses have focused on the development of machine learning and statistical models applied for diagnosis and prognosis. It is a clinical dataset gathered through fine-needle aspirates of breast masses to determine whether the breast mass of concern is benign or malignant. It is publicly available from the UCI Machine Learning Repository and finds frequent applications in various cancer classification and diagnosis research studies (18-20). The WBCD is a detailed dataset that effectively captures cell nuclei morphology from breast masses, whose contents are boundlessly valuable in deriving diagnosis tools for the condition under discussion. It has continued to lead, through research in machine learning, to further advances in medical diagnostic capabilities, especially in cancer detection. Completeness and thoroughness make the WBCD a reliable source for researchers and data scientists in cancer research. The WBCD includes 569 examples, each of which refers to a single observation of a breast mass. Of these, 357 (62.74%) are labeled benign, and 212 (37.26%) are labeled malignant. The data describes the measurements taken from digitized images of the FNAs and includes 30 features describing the properties of cell nuclei in breast mass tissues. These features are mean radius, mean texture, mean perimeter, mean area, mean smoothness, mean compactness, mean concavity, mean concave points, mean symmetry, mean fractal dimension, radius error, texture error, perimeter error, area error, smoothness error, compactness error, concavity error, concave points error, symmetry error, fractal dimension error, worst radius, worst texture, worst perimeter, worst area, worst smoothness, worst compactness, worst

concavity, worst concave points, worst symmetry, and worst fractal dimension.

Statistical analysis

Several classification tree algorithms were employed to classify the breast tumors as malignant or benign. The algorithms used in this study include CART (Classification and Regression Trees) (21), QUEST (Quick, Unbiased, and Efficient Statistical Tree) (22), GUIDE (Generalized, Unbiased, Interaction Detection and Estimation) (23), CRUISE (Classification Rule with Unbiased Interaction Selection and Estimation) (24, 25), Ctree (Conditional Inference Trees) (26), and Evtree (Evolutionary Learning of Globally Optimal Trees) (27). All algorithms were fitted based on the 10-fold cross-validation.

The dataset was split into training and testing sets using a stratified sampling technique to maintain a balanced distribution of classes. So, 80% of the observations (456 examples) were utilized as the training dataset, while the last 20% (113 examples) were used as the testing dataset. Each classification tree algorithm was trained on the training set and then assessed on the testing set. The models' performance was evaluated using accuracy metrics, such as sensitivity, specificity, false positive and negative rates, positive and negative predictive values, Youden's Index, accuracy, positive and negative likelihood ratios, diagnostic odds ratios, and AUC (area under the ROC curve). The study structure for the statistical analysis plan is shown in Figure 1.

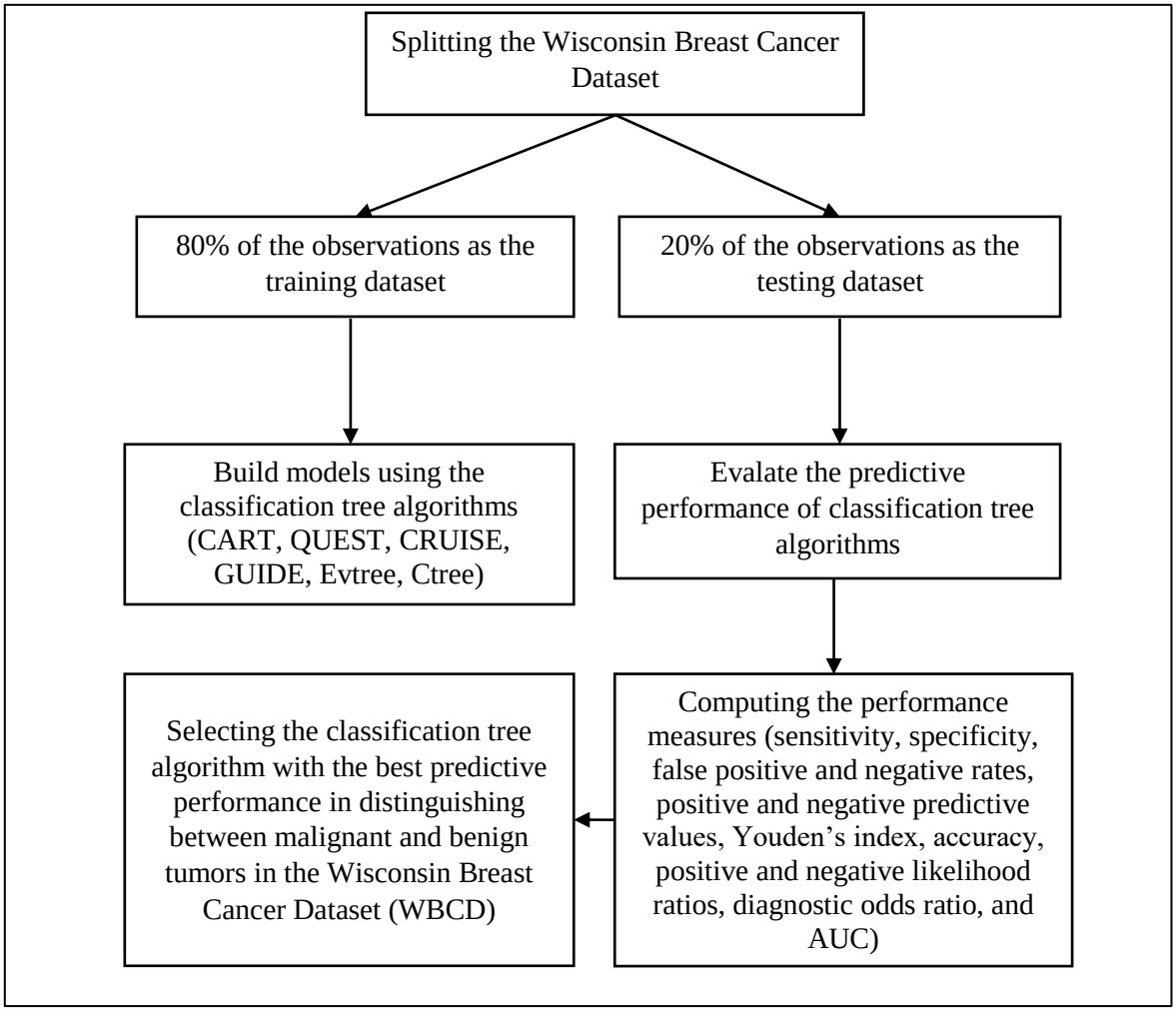


Figure 1. The study structure for the statistical analysis plan

Table 1. The accuracy measures of classification tree algorithms for distinguishing between malignant and benign tumors in the Wisconsin Breast Cancer Dataset with their 95% confidence interval

Algorithm / Metric	CART	Ctree	Evtree	QUEST	CRUISE	GUIDE
Sensitivity	0.93 (0.81, 0.99)	0.93 (0.81, 0.99)	0.90 (0.77, 0.97)	0.93 (0.81, 0.99)	0.93 (0.81, 0.99)	0.90 (0.77, 0.97)
Specificity	0.93 (0.84, 0.98)	0.96 (0.88, 0.99)	0.99 (0.92, 1)	0.96 (0.88, 0.99)	0.97 (0.90, 1)	0.93 (0.84, 0.98)
False negative rate	0.07 (0.01, 0.19)	0.07 (0.01, 0.19)	0.1 (0.03, 0.23)	0.07 (0.01, 0.19)	0.07 (0.01, 0.19)	0.1 (0.03, 0.23)
False positive rate	0.07 (0.02, 0.16)	0.04 (0.01, 0.12)	0.01 (0, 0.08)	0.04 (0.01, 0.12)	0.03 (0, 0.10)	0.07 (0.02, 0.16)
Positive predictive value	0.89 (0.75, 0.96)	0.93 (0.81, 0.99)	0.97 (0.87, 1)	0.93 (0.81, 0.99)	0.95 (0.83, 0.99)	0.88 (0.75, 0.96)
Negative predictive value	0.96 (0.88, 0.99)	0.96 (0.88, 0.99)	0.95 (0.87, 0.99)	0.96 (0.88, 0.99)	0.96 (0.88, 0.99)	0.94 (0.86, 0.98)
Youden's Index	0.86 (0.65, 0.96)	0.89 (0.69, 0.98)	0.89 (0.70, 0.97)	0.89 (0.69, 0.98)	0.90 (0.71, 0.98)	0.83 (0.62, 0.95)
Accuracy	0.93 (0.87, 0.97)	0.95 (0.89, 0.98)	0.96 (0.90, 0.99)	0.95 (0.89, 0.98)	0.96 (0.90, 0.99)	0.92 (0.85, 0.96)
Positive likelihood ratio	13.19 (5.64, 30.83)	21.98 (7.24, 66.72)	64.24 (9.15, 450.87)	21.98 (7.24, 66.72)	32.96 (8.39, 129.57)	12.85 (5.49, 30.08)
Negative likelihood ratio	0.08 (0.03, 0.23)	0.07 (0.03, 0.22)	0.10 (0.04, 0.25)	0.07 (0.03, 0.22)	0.07 (0.02, 0.22)	0.10 (0.04, 0.26)
Diagnostic odds ratio	171.60 (39, 758)	294.67 (57, 1531)	665 (72, 616)	294.67 (57, 1531)	448.50 (72, 2800)	125.40 (32, 496)
AUC	0.929 (0.880-0.980) P < 0.001	0.943 (0.897, 0.989) P < 0.001	0.945 (0.898, 0.992) P < 0.001	0.943 (0.897, 0.989) P < 0.001	0.950 (0.906, 0.994) P < 0.001	0.917 (0.863, 0.971) P < 0.001

A classification tree algorithm that achieves sensitivity, specificity, positive and negative predictive values, Youden's Index, and accuracy close to 1 demonstrates superior differential performance. An algorithm with a likelihood ratio exceeding 10, a negative likelihood ratio below 0.1, and a high diagnostic odds ratio indicates strong diagnostic capability in distinguishing between malignant and benign tumors. Additionally, ROC (receiver operating characteristic) curve analysis was employed to compute the AUC and compare the AUC values of different classification tree algorithms, with a higher AUC signifying better overall performance for each algorithm (28). Data balance was checked using Shannon entropy (SE) and SE = 0 or SE =1 shows that the data set is unbalanced or balanced, respectively (29).

Software

The data analysis was conducted with R software. The data splitting and fitting of the classification tree algorithms such as CART, Ctree, and Evtree and data splitting were done using the caret package (30). Software for the classification tree algorithms like QUEST, CRUISE, and GUIDE

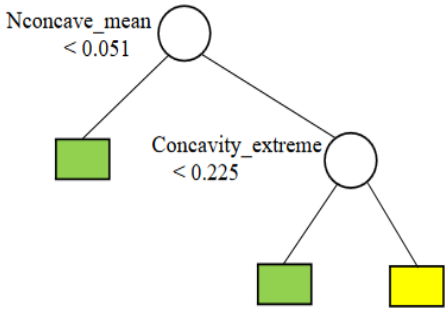
was obtained from http://pages.stat.wisc.edu/~loh/research.html. Package epiR (31) and package pROC (32) were used to compute the accuracy metrics and ROC curve analysis.

RESULTS

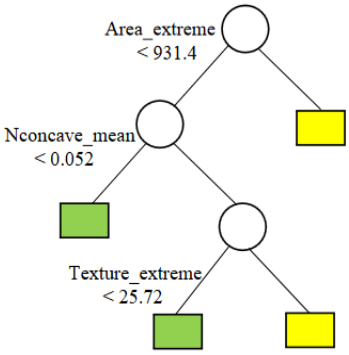
The Shannon entropy value was determined to be 0.953, which is close to 1. As a result, the data is well-balanced, and no remedial methods were needed. All classification tree algorithms demonstrated highly acceptable diagnostic performance in sensitivity, specificity, positive and negative predictive values, Youden's index, accuracy, positive and negative likelihood ratios, diagnostic odds ratio, and AUC in distinguishing between malignant and benign tumors in the WBCD (Table 1). The AUCs for all algorithms were statistically significant (P < 0.001). Among the tree algorithms, CRUISE showed the highest AUC of 0.950, or excellent diagnostic performance, while GUIDE had the lowest AUC of 0.917.

CRUISE

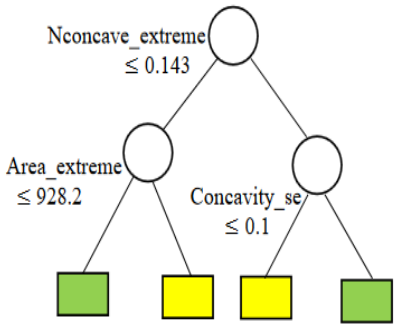
CART



Evtree



Ctree



QUEST

GUIDE

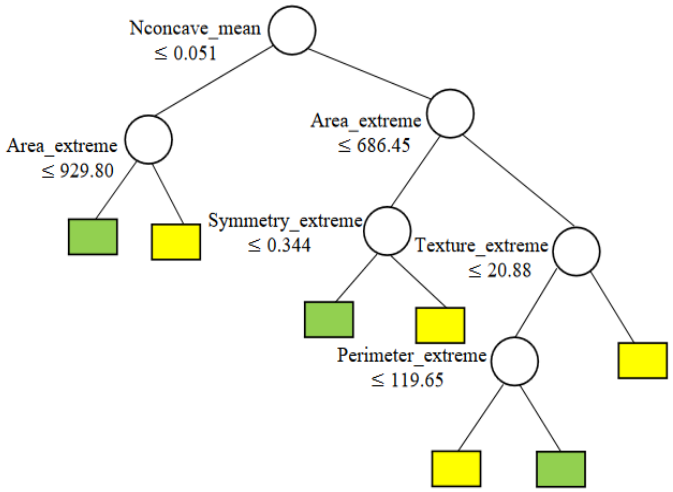
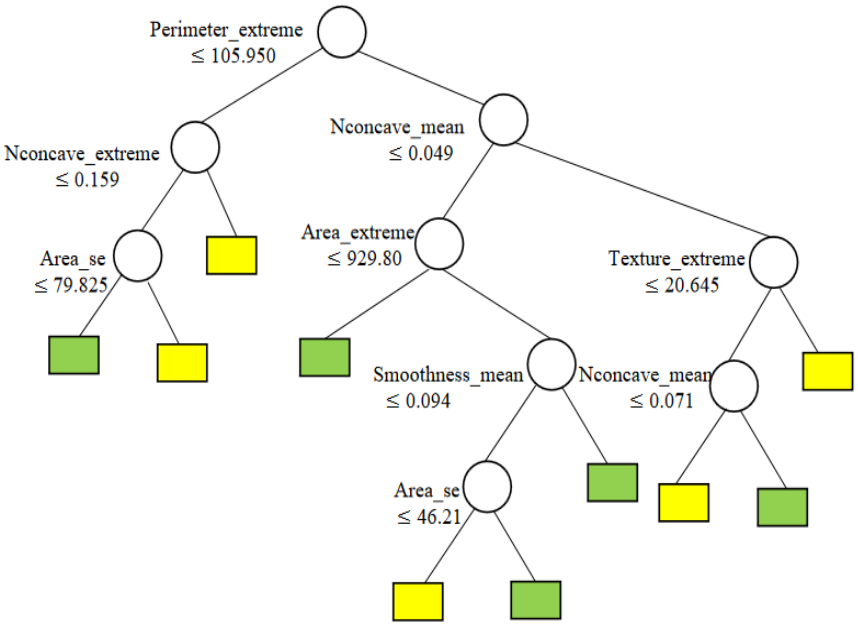


Figure 2. The tree structure of classification tree algorithms (CART, Evtree, Ctree, and GUIDE) for distinguishing between malignant and benign tumors in the Wisconsin Breast Cancer Dataset (Green: benign and yellow: malignant)

CRUISE



QUEST

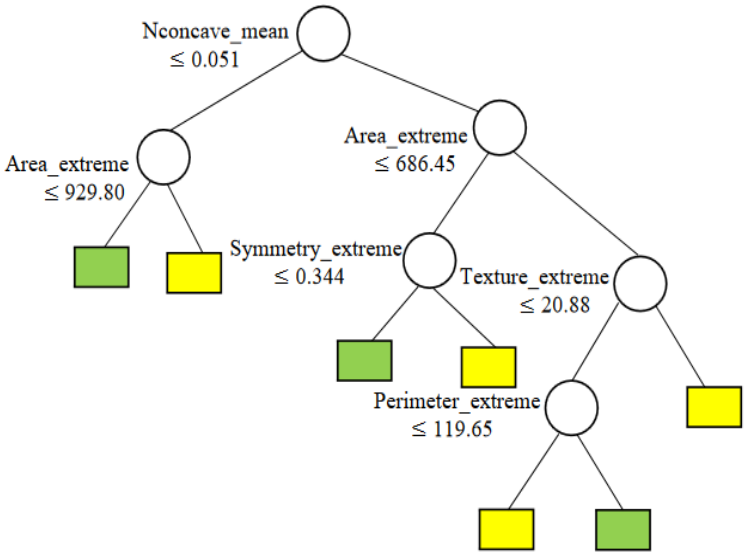


Figure 3. The tree structure of classification algorithms (CRUISE and QUEST) for distinguishing between malignant and benign tumors in the Wisconsin Breast Cancer Dataset (Green: benign and yellow: malignant)

Figures 2 and 3 illustrate the tree structure of these methods. According to the tree structure of the CRUISE algorithm (tree with the best diagnostic performance), we can conclude that subjects with Perimeter_extreme ≤ 105.950 & Nconcave_extreme ≤ 0.159 & Area_se ≤ 79.825 or Perimeter_extreme > 105.950 & Nconcave_mean ≤ 0.049 & Area_extreme ≤ 929.80 or Perimeter_extreme > 105.950 & Nconcave_mean ≤ 0.049 & Area_extreme ≤ 929.80 & Smoothness_mean ≤ 0.094 & Area_se > 46.21 or Perimeter_extreme > 105.950 & Nconcave_mean ≤ 0.049 & Area_extreme ≤ 929.80 & Smoothness_mean > 0.094 or Perimeter_extreme > 105.950 & Nconcave_mean > 0.049 & Texture_extreme ≤ 20.645 &

Nconcave_mean > 0.071 were classified as malignant tumor.

Figure 4 also shows the plot of variable importance based on the CRUISE tree algorithm (algorithm with the best diagnostic performance) for distinguishing between malignant and benign tumors in the WBCD. According to this plot, predictor variables such as Nconcave_extreme, Perimeter_extreme, Nconcave_mean, Radius_extreme, and Area_extreme are the five most important variables for distinguishing between malignant and benign tumors in the WBCD.

DISCUSSION

This study evaluates the performance of several decision tree algorithms, including CART, Ctree, Evtree, QUEST, GUIDE, and CRUISE, in differentiating between malignant and benign tumors in the Wisconsin Breast Cancer Dataset. The high diagnostic performance exhibited by CRUISE, with AUC values of 0.950, highlights their effectiveness in distinguishing between benign and malignant tumors. The key findings highlight the significance of interpretability in clinical decision-making, where decision tree algorithms stand out by providing ease of understanding, which is essential for usability in medical settings.

The diagnostic reliability of the CRUISE algorithm highlights its potential to enhance breast cancer diagnostics. Their performance, characterized by high sensitivity, specificity, and AUC, indicates that these models could substantially decrease the rates of misdiagnoses and false positives, which are prevalent issues in breast cancer detection. This is crucial, as over-diagnosis can result in unnecessary treatments, leading to both physical and psychological harm to patients.

The findings are consistent with previous research, which also demonstrated the effectiveness of machine learning techniques in breast cancer diagnostics. However, decision tree algorithms provide a significant advantage in interpretability compared to models such as neural networks and support vector machines.

Integrating the CRUISE algorithm into clinical practice could enhance the precision and efficiency of breast cancer diagnostics. By offering objective support during diagnostic evaluations, these models complement human expertise, improve the consistency of diagnoses, and reduce the need for invasive procedures in benign cases. This integration could streamline workflows, reduce patient stress, and conserve healthcare resources.

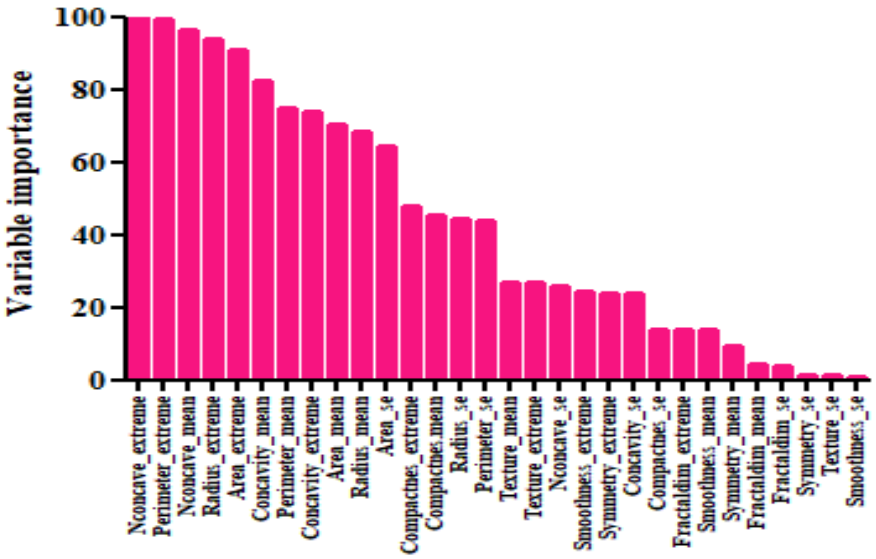


Figure 4. The plot of variable importance based on the CRUISE tree algorithm for distinguishing between malignant and benign tumors in the Wisconsin Breast Cancer Dataset

The study acknowledges several limitations, including using the WBCD, which may only partially represent the diversity of real-world cases. Additionally, the dataset's size and demographic limitations could hinder the generalizability of the findings. Future research should also investigate the potential of ensemble methods, such as random forests, bagging, and boosting, as these techniques may enhance predictive power.

CONCLUSION

This research highlights the efficacy of classification tree algorithms in enhancing the accuracy of breast cancer diagnoses, utilizing the Wisconsin Breast Cancer Dataset. A comparative analysis of various classification trees demonstrated the significant ability of the CRUISE algorithm to differentiate between benign and malignant tumors based on features derived from Fine Needle Aspiration samples. The results emphasize the critical role of integrating machine learning models into clinical practice to assist pathologists, improve diagnostic outcomes, and reduce subjectivity in cancer classification. Future investigations may aim to incorporate additional datasets and further refine existing algorithms to optimize strategies for breast cancer detection and treatment.

COMPETING INTERESTS

The authors declare that they have no competing interests.

FUNDING

No funding was received for this study.

ACKNOWLEDGMENT

This breast cancer database was obtained from the University of Wisconsin Hospitals, Madison by Dr. William H. Wolberg.

REFERENCES

[1] Siegel RL, Miller KD, Jemal AJCacjfc. Cancer statistics, 2018. 2018;68(1):7-30.

[2] Wang J, Wu S-GJBCT, Therapy. Breast cancer: an overview of current therapeutic strategies, challenge, and perspectives. 2023:721-30.

[3] Bukhari MH, Arshad M, Jamal S, Niazi S, Bashir S, Bakhshi IM, et al. Use of Fine-Needle Aspiration in the Evaluation of Breast Lumps. 2011;2011(1):689521.

[4] Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman DJCacjfc. Global cancer statistics. 2011;61(2):69-90.

[5] Modjtabai A, Khatib OM. Guidelines for the early detection and screening of breast cancer: World Health Organization, Regional Office for the Eastern Mediterranean; 2006.

[6] Mort E, Esserman L, Tripathy D, Hillner B, Houghton J, Bunker J, et al. Diagnosis and management of early-stage breast cancer. 1995;2:25-42.

[7] Gomes Pinto D, Schmitt FCJAC. Overcoming Pitfalls in Breast Fine-Needle Aspiration Cytology: A Practical Review. 2024;68(3):206-18.

[8] Yue W, Wang Z, Chen H, Payne A, Liu XJD. Machine learning with applications in breast cancer diagnosis and prognosis. 2018;2(2):13.

[9] De'ath G, Fabricius KE. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology. 2000;81(11):3178-92.

[10] Lemon SC, Roy J, Clark MA, Friedmann PD, Rakowski W. Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. Annals of behavioral medicine. 2003;26(3):172-81.

[11] Feldesman MR. Classification trees as an alternative to linear discriminant analysis. American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists. 2002;119(3):257-75.

[12] Malehi AS, Jahangiri M. Classic and Bayesian Tree-Based Methods. Enhanced Expert Systems. 2019:27.

[13] Jahangiri M, Khodadi E, Rahim F, Saki N, Saki Malehi A. Decision-tree-based methods for differential diagnosis of β-thalassemia trait from iron deficiency anemia. Expert Systems. 2017;34(3):e12201.

[14] Rahim F, Kazemnejad A, Jahangiri M, Malehi AS, Gohari K. Diagnostic performance of classification

trees and hematological functions in hematologic disorders: an application of multidimensional scaling and cluster analysis. BMC medical informatics and decision making. 2021;21(1):1-13.

[15] Jahangiri M, Rahim F, Saki N, Saki Malehi A. Application of Bayesian Decision Tree in Hematology Research: Differential Diagnosis of β-Thalassemia Trait from Iron Deficiency Anemia. Computational and Mathematical Methods in Medicine. 2021;2021.

[16] Rasool A, Bunterngchit C, Tiejian L, Islam MR, Qu Q, Jiang QJIjoer, et al. Improved machine learning-based predictive models for breast cancer diagnosis. 2022;19(6):3211.

[17] Sachdeva RK, Bathla PJIJoSI. A machine learning-based framework for diagnosis of breast cancer. 2022;10(1):1-11.

[18] Street WN, Wolberg WH, Mangasarian OL, editors. Nuclear feature extraction for breast tumor diagnosis. Biomedical image processing and biomedical visualization; 1993: SPIE.

[19] Wolberg WH, Street WN, Mangasarian OLJCl. Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. 1994;77(2-3):163-71.

[20] Mangasarian OL, Street WN, Wolberg WHJOr. Breast cancer diagnosis and prognosis via linear programming. 1995;43(4):570-7.

[21] Loh WY. Classification and regression tree methods. Encyclopedia of statistics in quality and reliability. 2008.

[22] Loh W-Y, Shih Y-S. Split selection methods for classification trees. Statistica sinica. 1997:815-40.

[23] Loh W-Y. Improving the precision of classification trees. The Annals of Applied Statistics. 2009:1710-37.

[24] Kim H, Loh W-Y. Classification trees with bivariate linear discriminant node models. Journal of Computational and Graphical Statistics. 2003;12(3):512-30.

[25] Kim H, Loh W-Y. Classification trees with unbiased multiway splits. Journal of the American Statistical Association. 2001;96(454):589-604.

[26] Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: A conditional inference framework. Journal of Computational and Graphical Statistics. 2006;15(3):651-74.

[27] Grubinger T, Zeileis A, Pfeiffer K-P. evtree: Evolutionary learning of globally optimal classification and regression trees in R. Working Papers in Economics and Statistics; 2011.

[28] Šimundić A-MJM, sciences b. Measures of diagnostic accuracy: basic definitions. 2008;22(4):61-5.

[29] Wang K, Phillips CA, Saxton AM, Langston MA. EntropyExplorer: an R package for computing and comparing differential Shannon entropy, differential coefficient of variation, and differential expression. BMC research notes. 2015;8(1):1-5.

[30] Kuhn MJJoss. Caret package. 2008;28(5):1-26.

[31] 31. Stevenson M, Stevenson MM, BiasedUrn IJTftaoedRpv. Package 'epiR'. 2015.

[32] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. Package 'pROC'. 2021.