International Journal for Research in Biology and Pharmacy

GREEN PUBLICATION

ISSN (Online): 2208-2093
Volume 11 Issue 01 March 2025
Available on-line at: https://ijrbp.com/index.php/bp/issue/view/539

DOI:*https://doi.org/10.69980/bp.v11i1.2504* *Received: 27-Dec-2024 Accepted: 27-Feb-2025*

CYTOLOGICAL FEATURE-BASED MACHINE LEARNING CLASSIFICATION OF BENIGN AND MALIGNANT BREAST TUMORS USING THE WISCONSIN DIAGNOSTIC BREAST CANCER

Dr. Hannah E. Collins[1], Dr. Victor Ramirez Ortega[2], Dr. Mei-Ling Zhou[3], Dr. Stefan Müller[4], Dr. Rania El-Haddad[5]

[1] Department of Biomedical Data Science, University of Edinburgh, Edinburgh, United Kingdom
[2] Department of Artificial Intelligence in Medicine, University of Barcelona, Barcelona, Spain
[3] Department of Computational Oncology, National Taiwan University, Taipei, Taiwan
[4] Institute of Medical Informatics and Biostatistics, University of Zurich, Zurich, Switzerland
[5] Department of Pathology and Cancer Research, American University of Beirut, Beirut, Lebanon

Corresponding Author: Dr. Hannah E. Collins
Email: hannah.collins.mlresearch@researchmail.org

Abstract
Breast cancer is still a serious worldwide health issue, and prompt diagnosis and treatment planning depend on the ability to distinguish between benign and malignant tumors. This study aimed to classify benign and malignant breast tumors using cytological through supervised machine learning techniques. The dataset contained 569 tumor records and 30 numerical cytological features derived from breast tumor cell nuclei. After removing non-informative columns, the diagnosis variable was encoded into benign and malignant classes, and the numerical features were standardized. Six ML models were applied: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbors, and Naive Bayes. Model performance was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC. The descriptive results showed that malignant tumors had higher mean values for radius, perimeter, area, concavity, and concave points, indicating greater nuclear size and structural irregularity. The classification results showed strong performance across all models. Random Forest and Support Vector Machine achieved the highest accuracy of 0.9737 and F1-score of 0.9630, while LR achieved the highest ROC-AUC value of 0.9960. These findings indicate that cytological features provide strong diagnostic separation between benign and malignant tumors. The study concludes that ML models can serve as effective decision-support tools for breast tumor classification, although external validation is required before clinical application.

Keywords: *Breast Cancer; Wisconsin Diagnostic Breast Cancer; Cytological Features; Machine Learning; Benign Tumor; Malignant Tumor*

1. Introduction

One of the most serious social health problems is breast cancer globally, and it continues to burden the health care systems with high costs of diagnosis and treatment. According to the cancer statistics of the world, breast cancer is one of the most common types of cancer that is diagnosed in women and a significant cause of cancer-related deaths (Bray et al., 2018; Ferlay et al., 2019). Epidemiological data more recently suggest that the current and future burden of breast cancer will likely only decrease due to population increase, ageing, lifestyle transitions, and disparities in access to early diagnosis services (Arnold et al., 2022; Giaquinto et al., 2022). Rapid and precise diagnosis is thus a crucial factor in enhancing survival rates, minimize unnecessary invasive interventions, and aiding in timely treatment choices. The distinction between benign and malignant breast tumors is of great importance since benign tumors can be treated by monitoring or conservative measures, whereas malignant tumors need to be treated immediately by the clinical team.

When diagnosing breast cancers, cytological analysis is crucial since cellular and nuclear features tend to be an indicator of the nature of tumors biologically. Malignant breast tumors are often characterized by increased nuclear irregularity, enlargement, increased concavity, distorted perimeter and distorted structural pattern of the tumor incursion as compared to benign tumors. These cytological variations give quantifiable diagnostic data that can be transformed into numerical variables to be analyzed numerically. Features such as radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry. These features are computed from digital images of fine needle aspirate samples. Due to their close relationship to the tumor's form, these characteristics can be used to differentiate between benign and malignant tumors. Breast cancer biology studies demonstrate the significance of cellular abnormalities, biomarkers and tumor behavior in the diagnosis and progression of breast cancer, which supports the relevance of such diagnostic characteristics (Barzaman et al., 2020).

In recent years, machine learning (ML) has become increasingly important in medical diagnosis due to its capability to detect intricate patterns in both structured and image health information. ML algorithms are able to work on multiple diagnostic variables at the same time and categorise the cases according to the acquired relationships of features and outcomes. ML has been utilized in study of breast cancer on cytological datasets, mammographic screening, histopathology and other diagnostic data. When ML algorithms are applied it has demonstrated that computational models can be useful in detecting breast cancer (Agarap, 2018). Likewise, deep learning studies have shown good diagnostic capability in breast-related cancer images such as detection of metastasis to lymph nodes and invasive breast cancer in whole-slide images (Cruz-Roa et al., 2017; Ehteshami Bejnordi et al., 2017). Despite the power of deep learning methods, traditional ML models can still help in structured cytological data due to their computational efficiency, interpretability and their ability to accommodate tabular diagnostic features.

Artificial intelligence and ML are now being considered as an assistant in the medical diagnostics process rather than substitutes to clinicians. Most recent sources affirm that AIs have the potential to enhance the accuracy, efficiency, and clinical decision support of diagnoses but also need to be thoroughly validated, transparent, and responsibly interpreted (Al-Antari, 2023; Alzubaidi et al., 2021). Explainable AI is especially significant in high-stakes medical environments since clinicians need to know the foundations of algorithmic decision-making to implement them in practice. This is crucial for the classification of breast cancer since a mistaken benign or malignant diagnosis could have serious clinical ramifications. The necessity of explainable artificial intelligence has been highly highlighted in the diagnostic and decision-support settings (Gunning & Aha, 2019).

Despite the extensive utilization of ML studies, there is still a need to conduct further comparative analysis of the classification models. The various algorithms could be different in the aspects of accuracy, precision, recall, F1-score and ROC-AUC and these differences have a practical impact on the reliability of diagnosis. Specifically, recall will be significant since the inability to detect any malignant tumors can postpone the diagnosis and treatment. Thus, the reliable and interpretable approaches to classification can be determined by comparative analysis of various models based on the same cytological features.

This study seeks to categorize benign and malignant breast tumors based on the cytological characteristics with the help of ML. The purpose of the study is to look at the cytological characteristics of both benign and malignant breast cancers, preprocess data to classify benign and malignant tumors with ML, apply the chosen ML models, assess performance of the models in terms of accuracy, precision, recall, F1-score, and ROC-AUC, and determine the best model to classify benign and malignant tumors.

2. Materials and Methods

2.1 Research Design

This research analysis employed a quantitative research study and ML based research design in which the researcher attempted to classify breast tumors by using cytological diagnostic characteristics as benign or malignant tumors. The supervised classification methodology was adopted since the outcome variable, tumor diagnosis, was known in data. The methodology involved cleaning the dataset, preparing features, developing a model and evaluating the performance.

2.2 Dataset Description

The research utilized dataset that was uploaded had 569 tumor observations and 33 columns. In this, the target variable was diagnosis, the case identification column was id, and there were 30 columns of numerical cytological features and one column without a name. The ID column and the blank unnamed column were eliminated since they were not relevant in classifying the tumors. The final dataset, after cleaning, had 569 observations and 30 predictor variables of diagnosis (UCI Machine Learning, 2016).

The target variable, diagnosis, categorized tumors into two groups: benign and malignant. B and M were used to denote benign and malignant tumors, respectively, in the dataset. There were 357 benign cases and 212 malignant cases (62.74 and 37.26 percent of the total cases, respectively) in the dataset. This rendered the dataset to be analyzed using binary classification.

2.3 Study Variables

Tumor diagnosis was dependent variable of study. To analyze ML, the labels of the diagnosis categories have been transformed into numeric values, where benign tumors corresponded to 0 and malignant tumors to 1. The positive class was considered to be malignant tumors since their accurate diagnosis is more crucial in clinical practice.

The independent variables comprised 30 numerical cytological features which were based on breast tumor cell nuclei. Radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension were among the nuclear features that were explained. Each characteristic was expressed in three different ways: mean value, standard error, and worst value.

These characteristics were deemed proper predictors since malignant tumors tend to possess more nuclear irregularities, increased size measurements, and increased concavity compared to the benign tumors.

2.4 Data Preprocessing

Preprocessing of data was done before the model development. The dataset was initially examined in terms of structure, type of variables, missing, duplicate records and class distribution. The id column was eliminated as it was not informative and was a unique identifier only. The blank column with no name was also eliminated as it had no analytical data.

The variable diagnosis was numerically codedMalignant patients were categorized as 1, whereas benign ones were coded as 0. The dataset has been analyzed concerning missing values of the diagnostic feature columns and the target variable. The retained variables did not have any missing values; hence, no imputation was necessary. Repetitive records were also verified and were not found.

The 30 numerical predictor variables were standardized on features since the cytological measurements were on various scales. To illustrate, the numerical range of variables in the area category was bigger than smoothness or fractal-dimension variables. Standardization puts all features on a similar scale and it is especially significant to algorithms like SVM and K-Nearest Neighbors which are sensitive to size of features.

The data was preprocessed and then separated into predictor variables and target variable. Independent variables were the 30 standardized cytological features, and the dependent variable was a coded diagnosis variable. An 80:20 split was applied to the dataset to create a training and a testing set. Stratified sampling was used to preserve the percentage of benign and malignant cases in both categories.

2.5 Machine Learning (ML) Models

Six ML models LR, Decision Tree, Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors, and Naive Bayes were selected for classification. These models were selected as they are typically applied in binary classification and they permit comparisons of linear, trees, distance-based, margin-based, ensemble, and probabilistic methods. LR was used as a baseline model due to its interpretability and ease of usage.

A decision tree was also added as it offers rule-based classification and also explains the decision paths. The use of RF as an ensemble model was used to enhance predictive stability and minimize overfitting. The reason behind the use of SVM was based on its capability to produce the best separating boundary amongst diagnostic classes. K-Nearest neighbors was added as a distance classifier, and Naive Bayes was a probabilistic comparative model. To ensure a fair comparison, same training and testing datasets were used for each model's training and testing.

2.6 Model Evaluation

Accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix were used to assess the model's performance. Accuracy was total percentage of tumors that were correctly classified. Precision was used to measure the percentage of predicted malignant that were actually malignant. Recall, or sensitivity, was a measure of the percentage of correct detection of actual malignant cases by model. The aspect of recall was particularly significant in this research due to the possibility of diagnosing and treating a malignant tumor as benign and postponing the diagnosis and treatment.

Precision and recall were balanced using the F1-score, and ROC-AUC was used to evaluate the performance of each model to differentiate between benign and malignant tumors at various thresholds. True benign, true malignant, fake benign, and false malignant classifications were considered using the confusion matrix. These measures gave an all-encompassing framework to compare the diagnostic effectiveness of chosen ML models.

3. Results

3.1 Dataset Characteristics and Class Distribution

The resulting analytical data set comprised of 569 breast tumor records and 30 cytological features following the elimination of the non-informative identification column and the empty unnamed column. The target variable was diagnosis which categorized the tumors into benign and malignant. There were 357 benign cases out of all the observations, which constituted 62.74 percent of the data and 212 cases were malignant, which constituted 37.26 percent of the data. This distribution shows that benign tumors were prevalent over malignant tumors in the dataset; but both classes were represented adequately to be used in supervised binary classification. Table 1 demonstrates the general structure and distribution of classes. The data set had 569 tumor records and 30 cytological features, benign cases (62.74) and malignant cases (37.26) constitute 62.74 and 37.26 of the total sample respectively. Figure 1 gives the distribution of classes of the dataset. The figure indicates that benign cases were prevalent compared to malignant cases and there were 357 and 212 benign tumors respectively. Though the dataset was not balanced, malignant cases were sufficient to train and test binary classification models.

Table 1. Dataset Characteristics and Class Distribution

Dataset Characteristic	Value
Total records	569
Total cytological features	30
Target variable	Diagnosis
Benign cases	357
Malignant cases	212
Benign percentage	62.74%
Malignant percentage	37.26%

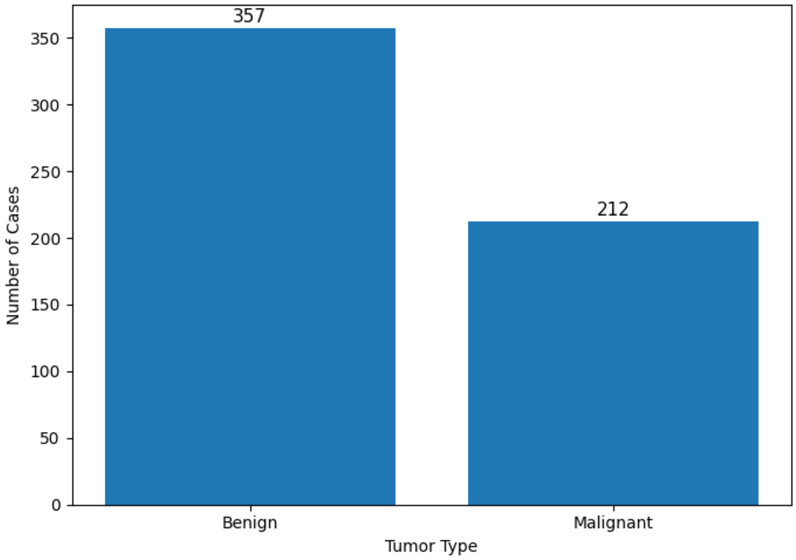

Figure 1. Distribution of Benign and Malignant Breast Tumor Cases

3.2 Descriptive Analysis of Cytological Features

Descriptive analysis was used to compare the chosen cytological characteristics between the benign and malignant tumor groups. Radius mean, texture mean, perimeter mean, area mean, concavity mean, and concave points mean were the features selected. These variables have been chosen since they are important nuclear size, texture, boundary, and shape-irregularity factors.

The findings reveal definite contrasts between benign and malignant tumors. All the chosen cytological features had higher mean values in malignant tumors. The average radius was 17.463 on the malignant tumors and 12.147 on the benign tumors. In the same way, the mean perimeter of malignant tumors was significantly greater (115.365), whereas, the mean perimeter of benign tumors was 78.075. The average size of the malignant tumors was 978.376, and was over twice that of the benign tumors, at 462.790.

The descriptive statistics of selected cytological features according to tumor type are shown in Table 2. Findings show that the mean values of malignant tumors were higher than those of benign tumors in radius, texture, perimeter, area, concavity, and concave points.

Table 2. Descriptive Statistics of Selected Cytological Features by Tumor Type

Tumor Type	**Radius Mean**	**Radius SD**	**Texture Mean**	**Texture SD**	**Perimeter Mean**	**Perimeter SD**	**Area Mean**	**Area SD**	**Concavity Mean**	**Concavity SD**	**Concave Points Mean**	**Concave Points SD**
Benign	12.147	1.781	17.915	3.995	78.075	11.807	462.790	134.287	0.046	0.043	0.026	0.016
Malignant	17.463	3.204	21.605	3.779	115.365	21.855	978.376	367.938	0.161	0.075	0.088	0.034

There were also significant differences in features belonging to shape irregularity. The average value of concavity of malignant tumors was 0.161 and 0.046 for benign tumors. Similarly, the average concave points were 0.088 with malignant tumors and 0.026 with benign tumors. These observations indicate that the nuclear size and structural abnormality of malignant tumors are more apparent compared to benign tumors. Figure 2 presents the comparison of the mean values of the chosen cytological characteristics of benign and malignant tumors. The figure illustrates that malignant tumors have always had higher radius, texture, area, perimeter, concavity, and concave point values. The greatest differences in visuals were seen in area, perimeter, radius, concavity and concave points and it was established that malignant tumors tend to have larger and more irregular cell nuclei.

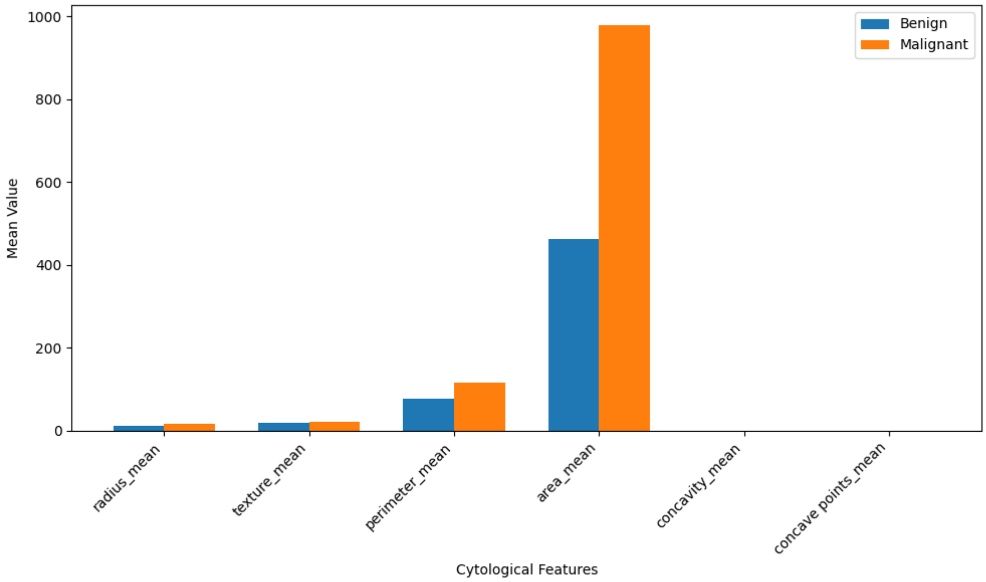

Figure 2. Comparison of Key Cytological Features Between Benign and Malignant Tumors

3.3 ML Model Performance

Six supervised ML models were assessed for their ability to classify benign and malignant breast cancers. K-Nearest Neighbors, Naive Bayes, RF, Decision Tree, LR, and SVM were the models. Model performance was assessed using ROC-AUC, F1-score, accuracy, precision, and recall.

The findings indicate that models all had high classification performance with accuracy levels of 0.9298 to 0.9737. Random Forest and SVM were best in terms of accuracy with both having an accuracy of 0.9737. The two models also achieved the best F1-score of 0.9630. This demonstrates that accuracy and recall have a good trade-off.

Perfect precision ratings were highest for Random Forest and S of 1.0000, which implies that all malignant tumors identified by the two models were indeed malignant. Nevertheless, they had a recall value of 0.9286 which implies that a low percentage of malignant cases were missed. LR was also very strong with the accuracy of 0.9649, precision of 0.9750, recall of 0.9286, F1-score of 0.9512, and the highest ROC-AUC of 0.9960. Table 3 summarizes the comparative performance of the chosen ML models. The scores indicate that the accuracy and F1-score of Random Forest and SVM are the highest whereas the ROC-AUC of the LR is the highest.

Table 3. Performance Comparison of ML Classification Models

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.9649	0.9750	0.9286	0.9512	0.9960
Decision Tree	0.9298	0.9048	0.9048	0.9048	0.9246
Random Forest	0.9737	1.0000	0.9286	0.9630	0.9929
Support Vector Machine	0.9737	1.0000	0.9286	0.9630	0.9947
K-Nearest Neighbors	0.9561	0.9744	0.9048	0.9383	0.9823
Naive Bayes	0.9386	1.0000	0.8333	0.9091	0.9934

Naive Bayes had a high ROC-AUC of 0.9934 and an ideal precision of 1.0000, which implies that it overlooked more cases of malignancy than the other models. Decision Tree generated the smallest value of ROC-AUC; 0.9246, but the overall performance was not very high.The comparison of the ROC curves of the ML models is given in Figure 3. The ROC curves indicate that the majority of the models possessed outstanding discriminatory power with regard to benign and malignant tumors. LR had the highest ROC-AUC value of 0.9960, then SVM, 0.9947, Naive Bayes, and finally Random Forest was 0.9929. Such findings suggest that the chosen cytological characteristics were powerful diagnostic predictors of benign and malignant tumors of the breast.

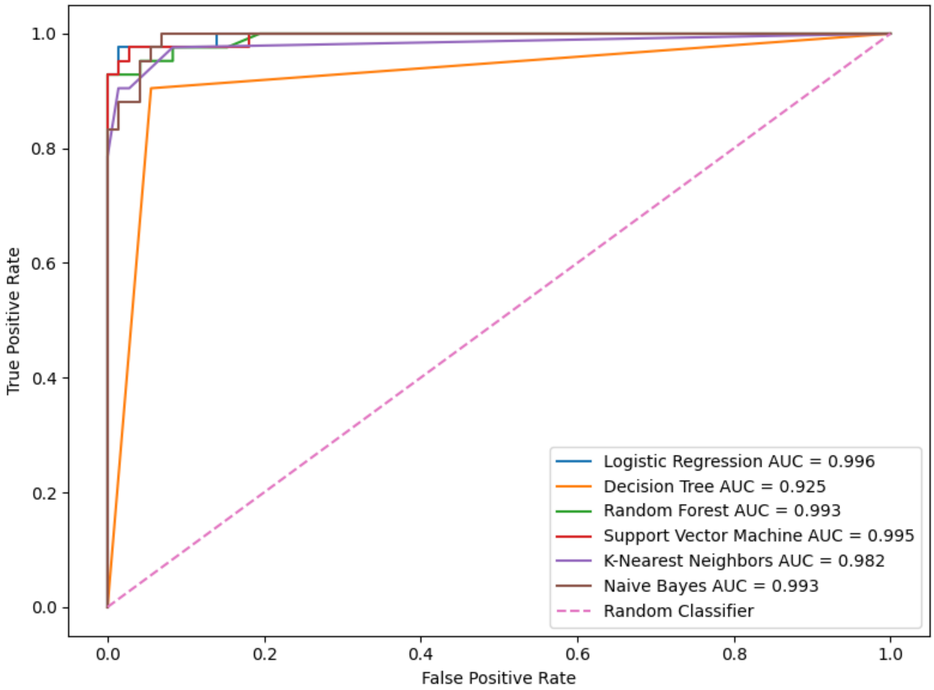

Figure 3. ROC Curve Comparison of ML Models

3.4 Summary of Results

All in all, the findings affirm descriptive analysis revealed that values of nuclear size and other features used to determine irregularity were always higher in malignant tumors, especially in area, perimeter, radius, concavity, and concave points. These differences are a good biological and statistical basis of classification.

Additionally, the ML results demonstrated the utility of cytologic characteristics in distinguishing between benign and malignant tumors. LR got the best ROC-AUC value, while Random Forest and SVM had the highest accuracy and F1-score. The models that performed the best in this investigation were SVM, Random Forest, and LR based on diagnostic reliability. The results suggest that decision support for the classification of breast tumors may be provided by ML models trained on cytological markers.

4. Discussion

The results of this study indicate classify benign and malignant breast tumors in a binary way. The sample comprised 569 cases, 357 of which were benign and 212 malignant tumors, which means that both diagnostic groups were well represented. Even though the distribution was not that skewed, the size of the malignant cases was adequate to conduct supervised ML. This is significant as breast cancer is a key issue in global health, and the incidence and mortality rates differ significantly between the population and age groups (Heer et al., 2020; Lei et al., 2021). The persistent nature of breast cancer demonstrates the importance of high-quality diagnostic support systems that could be used to classify the tumor early and correctly (Loibl et al., 2021; Momenimovahed & Salehiniya, 2019).

Descriptive results indicated the existence of clear cytological differences in benign and malignant tumors. Radius, texture, perimeter, area, concavity, and concave points were significantly higher in mean values in malignant tumors. These variations imply that the nuclear dimensions and structural irregularity of malignant tumors are often larger than those of benign tumors. Specifically, the increased value of area, perimeter, concavity and concave points indicates that the nuclei of malignant cells tend to be larger and irregular in shape. These results align with the biological knowledge on breast cancer, in which malignant transformation is linked with abnormal cellular growth, irregular nuclear boundaries, and more complex architecture (Loibl et al., 2021; Strelcenia & Prakoonwit, 2023). Other related studies have also highlighted that diagnostic feature selection and feature engineering can enhance the classification of breast cancer cases by determining the most discriminative cytological features.

All selected models showed good categorization in ML results. RF and SVM recorded best accuracy of 0.9737 and F1-score of 0.9630 which implies good predictive reliability. LR was also very accurate, with the highest accuracy of 0.9649 and the maximum ROC-AUC of 0.9960. These results indicate that the dataset cytological features are strong separators of benign and malignant tumors. The good results of LR can be attributed to the fact that the data is structured and highly informative during the numerical format whereas the good results of RF can be attributed to the fact that it is capable of capturing nonlinear relationships amongst cytological variables. The effectiveness of SVM must have been high since it works well in high-dimensional classification tasks whose classes are well separated. These findings are in line with other breast cancer classification studies that have reported that the traditional ML models can achieve high diagnostic performance when trained on well-prepared cytological or clinical feature datasets (Naji et al., 2021; Reshan et al., 2023).

Clinically, the findings suggest the possible usefulness of ML as a diagnostic decision-support system on breast tumor classification. ML algorithms can be trained on several diagnostic features at once, and could be useful in aiding clinicians by giving fast, stable, and data-driven classification results. Nevertheless, these models cannot be considered as substitutes for clinical experience. Artificial intelligence systems need to be well-vetted, disclosed, and responsibly incorporated into clinical practices in medical settings (Kelly et al., 2019; Rajkomar et al., 2019). There is also some evidence that AI systems can be used to assist in diagnostic performance, based on breast cancer screening, but they need to be rigorously evaluated and compared with expert practice to be applied in clinical practice (McKinney et al., 2020; Rodriguez-Ruiz et al., 2019).

Another area of concern when using ML to diagnose breast cancer is interpretability. As wrong classification can be deadly, clinicians must know the features that affect model predictions. Explainable AI methods, such as SHAP-based interpretation can be used to determine the value of individual features and enhance trust in model results (Lundberg et al., 2020; Lundberg & Lee, 2017). According to reporting standards of AI-based health research, there is a focus on transparency, methodological transparency, and validation before clinical implementation (Liu et al., 2020; Rivera et al., 2020).

Although this study has promising findings, there are limitations to it. The data was secondary and had a size of 569 observations. It involved cytological variables alone and omitted clinical, genetic, histopathological or imaging variables. The models were tested on an internal train-test split and had not been tested externally using independent, hospital-based data. Thus, the diagnostic performance in the real world can be different from that in this benchmark dataset. Future research involves bigger and more heterogeneous clinical samples, external validation, explainable AI, and combining cytological features with images, genetic, and clinical factors. The extensions can enhance model generalizability and enhance clinical relevance of ML based breast tumor classification.

5. Conclusion

This study explain effectiveness of cytological features for ML based classification of benign and malignant breast tumors. It was discovered that the dataset was suitable for binary classification, with 569 records of tumors, 30 numerical cytological features and reasonably represented benign and malignant diagnostic classes. The descriptive results demonstrated definitive variations between benign and malignant tumors, especially in terms of nuclear size, perimeter, area, concavity and concave points. These features had higher values in malignant tumors, which is an indication of increased cellular irregularity and structural abnormality. ML outcomes revealed that cytological features had the capacity to aid in the classification of breast tumors. RF and SVM are the most accurate and have the highest F1-score, whereas the LR has the highest ROC-AUC. These results indicate that well-organized cytological variables have robust diagnostic differentiation between benign and malignant tumors. Overall, study supports the idea that ML models could provide useful decision support during the diagnosis of breast cancer. Nevertheless, these models are supposed to be viewed as supportive tools as opposed to substitutes to clinical judgment. The study has constraints of using a secondary benchmark dataset and no external clinical validation. The future study ought to utilize larger independent datasets, to include clinical and imaging variables and explainable artificial intelligence methods that enhance transparency, reliability, and clinical applicability.

References

1. Agarap, A. F. M. (2018). On breast cancer detection: An application of machine learning algorithms on the wisconsin diagnostic dataset. *Proceedings of the 2nd International Conference on Machine Learning and Soft Computing*, 5–9. https://doi.org/10.1145/3184066.3184080
2. Al-Antari, M. A. (2023). Artificial intelligence for medical diagnostics—Existing and future AI technology! *Diagnostics*, *13*(4), 688.
3. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. https://doi.org/10.1186/s40537-021-00444-8
4. Arnold, M., Morgan, E., Rumgay, H., Mafra, A., Singh, D., Laversanne, M., Vignat, J., Gralow, J. R., Cardoso, F., Siesling, S., & Soerjomataram, I. (2022). Current and future burden of breast cancer: Global statistics for 2020 and 2040. *The Breast*, *66*, 15–23. https://doi.org/10.1016/j.breast.2022.08.010
5. Barzaman, K., Karami, J., Zarei, Z., Hosseinzadeh, A., Kazemi, M. H., Moradi-Kalbolandi, S., Safari, E., & Farahmand, L. (2020). Breast cancer: Biology, biomarkers, and treatments. *International Immunopharmacology*, *84*, 106535.
6. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424. https://doi.org/10.3322/caac.21492
7. Cruz-Roa, A., Gilmore, H., Basavanhally, A., Feldman, M., Ganesan, S., Shih, N. N., Tomaszewski, J., González, F. A., & Madabhushi, A. (2017). Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. *Scientific Reports*, *7*(1), 46450.
8. Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J. A., consortium, C., Hermsen, M., & Manson, Q. F. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, *318*(22), 2199–2210.

9. Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *International Journal of Cancer*, *144*(8), 1941–1953. https://doi.org/10.1002/ijc.31937
10. Giaquinto, A. N., Sung, H., Miller, K. D., Kramer, J. L., Newman, L. A., Minihan, A., Jemal, A., & Siegel, R. L. (2022). Breast Cancer Statistics, 2022. *CA: A Cancer Journal for Clinicians*, *72*(6), 524–541. https://doi.org/10.3322/caac.21754
11. Gunning, D., & Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI Magazine*, *40*(2), 44–58.
12. Heer, E., Harper, A., Escandor, N., Sung, H., McCormack, V., & Fidler-Benaoudia, M. M. (2020). Global burden and trends in premenopausal and postmenopausal breast cancer: A population-based study. *The Lancet Global Health*, *8*(8), e1027–e1037.
13. Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, *17*(1), 195. https://doi.org/10.1186/s12916-019-1426-2
14. Lei, S., Zheng, R., Zhang, S., Wang, S., Chen, R., Sun, K., Zeng, H., Zhou, J., & Wei, W. (2021). Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020. *Cancer Communications*, *41*(11), 1183–1194. https://doi.org/10.1002/cac2.12207
15. Liu, X., Rivera, S. C., Moher, D., Calvert, M. J., Denniston, A. K., Ashrafian, H., Beam, A. L., Chan, A.-W., Collins, G. S., & Deeks, A. D. J. (2020). Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *The Lancet Digital Health*, *2*(10), e537–e548.
16. Loibl, S., Poortmans, P., Morrow, M., Denkert, C., & Curigliano, G. (2021). Breast cancer. *The Lancet*, *397*(10286), 1750–1769.
17. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67.
18. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
19. McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. S., & Darzi, A. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, *577*(7788), 89–94.
20. Momenimovahed, Z., & Salehiniya, H. (2019). Epidemiological characteristics of and risk factors for breast cancer in the world. *Breast Cancer: Targets and Therapy*, *Volume 11*, 151–164. https://doi.org/10.2147/BCTT.S176070
21. Naji, M. A., El Filali, S., Aarika, K., Benlahmar, E. H., Ait Abdelouhahid, R., & Debauche, O. (2021). Machine learning algorithms for breast cancer prediction and diagnosis. *Procedia Computer Science*, *191*, 487–492.
22. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine. *New England Journal of Medicine*, *380*(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259
23. Reshan, M. S. A., Amin, S., Zeb, M. A., Sulaiman, A., Alshahrani, H., Azar, A. T., & Shaikh, A. (2023). Enhancing breast cancer detection and classification using advanced multi-model features and ensemble machine learning techniques. *Life*, *13*(10), 2093.
24. Rivera, S. C., Liu, X., Chan, A.-W., Denniston, A. K., Calvert, M. J., Ashrafian, H., Beam, A. L., Collins, G. S., Darzi, A., & Deeks, J. J. (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension. *The Lancet Digital Health*, *2*(10), e549–e560.
25. Rodriguez-Ruiz, A., Lång, K., Gubern-Merida, A., Broeders, M., Gennaro, G., Clauser, P., Helbich, T. H., Chevalier, M., Tan, T., & Mertelmeier, T. (2019). Stand-alone artificial intelligence for breast cancer detection in mammography: Comparison with 101 radiologists. *JNCI: Journal of the National Cancer Institute*, *111*(9), 916–922.
26. Strelcenia, E., & Prakoonwit, S. (2023). Effective feature engineering and classification of breast cancer diagnosis: A comparative study. *BioMedInformatics*, *3*(3), 616–631.
27. UCI Machine Learning. (2016). *Breast Cancer Wisconsin (Diagnostic) Data Set*. https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data