
| RESEARCH ARTICLE

Comparative analysis of explainable machine learning models for cancer classification using cytological features

Md Ismail Hossain Siddiqui

Master of Science in Engineering/Industrial Management, Westcliff University, Irvine, California, USA

Md. Soebur Rahman

Master of Business Administration in Management Information Systems, International American University, Los Angeles, California, USA

Abdul Aziz Kabir

Master of Business Administration in Data Analytics, Westcliff University, Irvine, California, USA

Farhad Uddin Mahmud

Master of Business Administration in Management Information Systems, International American University, Los Angeles, California, USA

Saeed Ur Rashid

Master of Business Administration in Data Analytics, Westcliff University, Irvine, California, USA.

Ramisa Samin Shammah

College of Technology and Engineering, Westcliff University, Irvine, USA

Corresponding Author: Md Ismail Hossain Siddiqui, **E-mail:** ismailhossainsiddiqui.ce@gmail.com

| ABSTRACT

Breast cancer is among the causes of cancer related deaths globally with the greatest impact being in the low resource and high volume health care facilities where timely and accurate screening is paramount. This research report is a explainable machine learning model of breast cancer diagnosis based on quantitative features of fine needle aspirate images of breast masses. The data set contains 569 samples and 30 real-valued predictors of cell nuclei morphology, and it does not contain missing values, and the class distribution is moderate. It uses a structured preprocessing pipeline, such as the division of data into training and held-out test sets, feature normalization, and the careful management of class imbalance. Several classification models are compared, among them, Random Forest and Gaussian Naive Bayes, to compare the predictive accuracy and reliability of the model. Experimental outcomes have shown that the Random Forest model obtains the best performance with an accuracy of 0.96 on the held out test set, and balanced precision and recall on the benign and malignant classes. The confusion matrix shows that there is low misclassification rate and only false positive and false negative are three and three respectively. Contrastingly, Gaussian Naive Bayes has a higher accuracy of 0.93, and is less sensitive to malignant cases because of its independence assumptions which are not completely met in the dataset as verified by correlation analysis. These results are also supported by receiver operating characteristic analysis, whose area under the curve value is 1.00 in random forest and 0.99 in Gaussian naive bayes. The findings emphasize the role of model selection in clinical decision support systems, especially in cases where false negatives have to be minimized. The suggested structure focuses on interpretability and high practical use thus appropriate in deployment-oriented screening processes in resource-constrained settings. This paper shows that explainable machine learning models are trained on structured cytological features and can give effective and reliable support in early detection of breast cancer.

KEYWORDS

Breast cancer classification, explainable machine learning, fine needle aspirate cytology, cytological feature analysis, Random Forest, Gaussian Naive Bayes

ARTICLE INFORMATION**ACCEPTED:** 02 October 2023**PUBLISHED:** 29 October 2023**DOI:** 10.32996/jmhs.2023.4.5.14

Introduction

Breast cancer is a common malignancy and the number one cause of cancer-related deaths among women in the world. Early diagnosis is a very important factor in enhancing the survival rate and alleviating the burden of treatment. Nevertheless, in low-resource and high-volume healthcare facilities, it is common to have limited access to advanced imaging systems and professional radiological interpretation. In these circumstances, the dependable, cost efficient, and scalable diagnostic support systems are necessary to aid in making of clinical decisions as well as to direct the high-risk cases towards further examination. Fine needle aspirate (FNA) cytology is among those diagnostic tests that can be done on a mass in the breast with minimal invasiveness. The digitalization of FNA samples allows the isolation of quantitative attributes that characterize morphological attributes of cell nuclei, such as size, shape, texture, and structural anomalies. These characteristics can be used to give important information on differentiating between benign and malignant tumors. As more structured biomedical data becomes available, machine learning techniques have become a viable way of processing such characteristics and assisting in automated diagnosis.

Although machine learning is increasingly being used in medical diagnostics, there are still various challenges. To begin with, most of the classical models suppose independence of features, which is usually broken in biomedical data which have strong correlations between morphological properties. Second, the performance of a model, in its own right, is not enough to be adopted in clinical practice; interpretability and cross-class consistency are also crucial, especially in the context of screening where false negatives may be fatal. Third, small to medium data sizes involve the need to design an experiment with great caution to prevent overfitting and to provide a dependable assessment of data not seen. In the research, a machine learning interpretable framework is constructed to analyze breast cancer based on structured features of FNA images. Analysis is performed on an established diagnostic dataset of 569 samples having 30 real-valued features and without missing values. Data integrity (such as proper data partitioning and normalization) is maintained with the help of a controlled preprocessing pipeline. Two model types, Random Forest and Gaussian Naive Bayes are compared to analyze how the assumptions of the model influence the performance of the classification. The comparative analysis is not only centered on accuracy but also on class-wise precision, recall and distribution of errors with specific attention being on the detection of malignant cases. There are three times the contributions of this work. Firstly, it offers a tight assessment of machine learning models on structured cytological data in an experimental environment that can be replicated. Second, it emphasizes the deficiencies of independence-based models when there are correlated biomedical features, which is justified by correlation analysis. Third, it illustrates the appropriateness of ensemble-based techniques in providing balanced and reliable performance and hence are suitable in deployment-oriented screening support in resource-constrained healthcare.

Related Work

Breast cancer diagnosis has been extensively studied using machine learning techniques applied to structured clinical and cytological data [1,2]. Early approaches primarily relied on classical statistical models and conventional machine learning algorithms, leveraging handcrafted features derived from fine needle aspirate and imaging data. Logistic regression, support vector machines, and k-nearest neighbors have been widely explored due to their simplicity and interpretability [3,4]. These models demonstrated reasonable performance on benchmark datasets; however, their effectiveness is often limited by linear assumptions and sensitivity to feature scaling and multicollinearity [5,6]. Ensemble learning methods have shown significant improvements in classification performance for breast cancer detection tasks. Random Forest, in particular, has been widely adopted due to its ability to model nonlinear relationships and handle high-dimensional feature spaces without requiring strict assumptions about data distribution [7,8]. By aggregating multiple decision trees, Random Forest reduces variance and improves generalization, making it suitable for biomedical datasets with correlated features. Several studies have reported strong performance using tree-based models on cytological datasets, highlighting their robustness and reliability in binary classification tasks [9,10].

Probabilistic models such as Gaussian Naive Bayes have also been applied to breast cancer classification. These methods are computationally efficient and perform well under certain conditions, particularly when features are approximately independent [11,12]. However, in real-world biomedical datasets, feature independence is rarely satisfied. Morphological descriptors such as radius, perimeter, and area are inherently correlated, which can negatively impact the performance of Naive Bayes models [13,14]. As a result, although Gaussian Naive Bayes provides a useful baseline, its applicability is limited in scenarios with strong feature dependencies. More recently, deep learning approaches have been explored for breast cancer diagnosis, particularly in image-based settings such as mammography, histopathology, and ultrasound imaging [15,16]. Convolutional neural networks have demonstrated strong performance in extracting hierarchical features directly from raw images, reducing the need for manual feature engineering [17,18]. However, these approaches typically require large-scale annotated datasets and substantial computational resources, which may not be feasible in low-resource healthcare environments. In contrast, structured datasets derived from cytological analysis offer a more practical alternative for developing lightweight and interpretable models. Another important direction in recent research is the incorporation of explainability into machine learning models for clinical applications [19,20]. Explainable artificial intelligence techniques aim to provide transparency into model predictions, enabling clinicians to understand the contribution of individual features to diagnostic decisions. This is particularly important in healthcare, where trust and accountability are essential for adoption. Tree-based models, such as Random Forest, naturally support feature importance analysis, making them more suitable for explainable diagnostic frameworks compared to black-box models. Despite these advancements, several gaps remain in the literature. Many studies focus primarily on predictive performance without addressing the impact of feature correlation, class imbalance, and experimental design on model reliability. Additionally, limited attention has been given to deployment-oriented considerations, such as maintaining consistent performance across classes and minimizing false negatives in screening applications. This study addresses these gaps by providing a structured and reproducible evaluation of machine learning models on cytological features, with a focus on interpretability, robustness, and practical applicability in resource-constrained clinical settings.

Methodology

The figure 1 shows the entire workflow of the machine learning pipeline applied in this study. It starts with data collection where the data is collected and formatted to be analyzed. This is followed by a preprocessing step which involves cleaning the data, normalizing the values of features and dividing the data into training and testing data. Exploratory data analysis follows preprocessing to learn the distribution of features, as well as to examine variable relationships. The step assists in recognizing patterns, correlations and possible problems including redundancy or imbalance in the data. The second step, which is the model development stage, is the selection of the suitable machine learning algorithms and their subsequent training on the processed training data. At this phase, model performance measures like the ability to classify and discriminate are also taken. After model training, the evaluation phase measures performance on a basis of confusion matrix analysis and receiver operating characteristic curves. The step gives a clue on the accuracy and error distribution of the models. Lastly, the findings are inferred and confirmed to draw out significant information. This involves the determination of key features, model behavior validation and the evaluation of the possible clinical utility of the system. Generally, the figure is organized and sequential pipeline that guarantees stable model development and assessment of breast cancer classification.

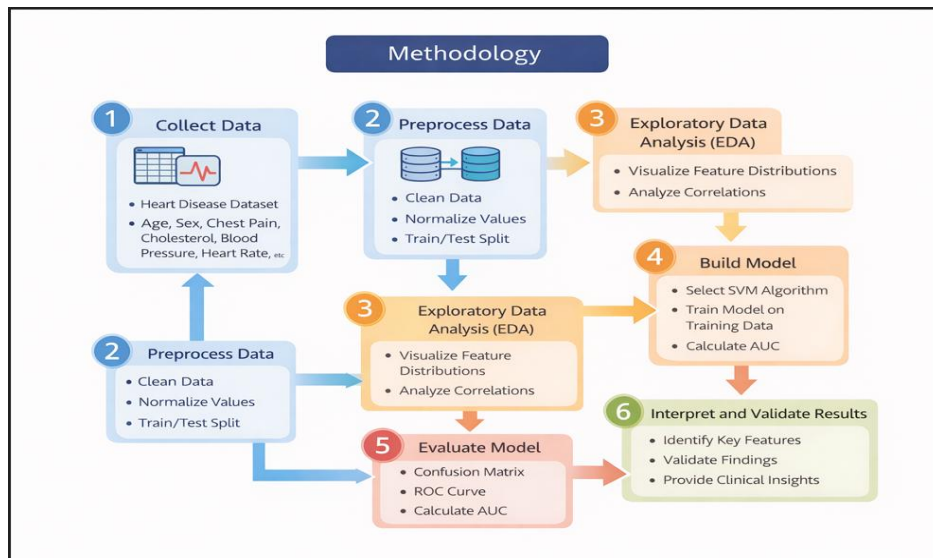


Figure 1 :Machine learning workflow for cancer classification using fine needle aspirate derived features

3.1 Dataset description

The data set in this research is the Breast Cancer Wisconsin Diagnostic dataset which has attributes that were obtained through digitization of images of fine needle aspirate samples of masses in the breast. The samples are the cases of one patient each and characterised by quantitative evaluation of cell nuclei morphology. These characteristics reflect valuable structural and textural features that are significant in differentiating benign and malignant tumors. The total number of samples in the dataset is 569, 357 of which are benign and 212 are malignant, which means that there is an intermediate level of class imbalance. Each of the samples is described as having 30 real-valued features based on ten underlying morphological properties, such as radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension. Each of these properties is calculated with three different measurements, namely: mean, standard error and worst value, which produces an overall representation of each cell nucleus. All the features are continuous and are recorded with uniform numerical accuracy. There are no missing values in the dataset thus it can be directly used to apply machine learning models without imputation. Diagnosis, the target variable, is binary coded in which one category is that of malignant tumors and the other benign cases. Because of the difference in the scale of the features and the existence of correlated attributes, proper preprocessing methods like normalization and proper model selection are needed. The data is also popular as a reference point in medical machine learning studies because of its clear structure, clear feature representation, and good discriminative properties. These characteristics render it appropriate in the assessment of classification models in breast cancer diagnosis and in the creation of decision support that is beneficial to the clinician.

3.1 Data preprocessing

The data comprises 569 samples with 30 real-valued features based on the fine-needle aspirates images. The samples are either benign or malignant. The preprocessing pipeline is designed in such a way that it preserves the integrity of data and gives credible model training. First, the data are separated into training and held-out test sets. There are 134 samples in the test set and it is only utilized in final evaluation Figure 2. The data split is done before all the preprocessing operations are done to prevent information leakage. The standardization method is used to perform feature scaling by ensuring that each feature has mean equal to zero and unit variance. This is necessary because the magnitude of the features is very different among the dataset. Only the training data is used to compute the scaling parameters which are then applied to test set. Class imbalance is handled at the training stage. The initial data set shows a moderate disproportion of benign and malignant classes. To counteract this, resampling is used on the training set only, so as to obtain a more balanced distribution of classes, without resampling the test set. This makes evaluation representative of real-life.

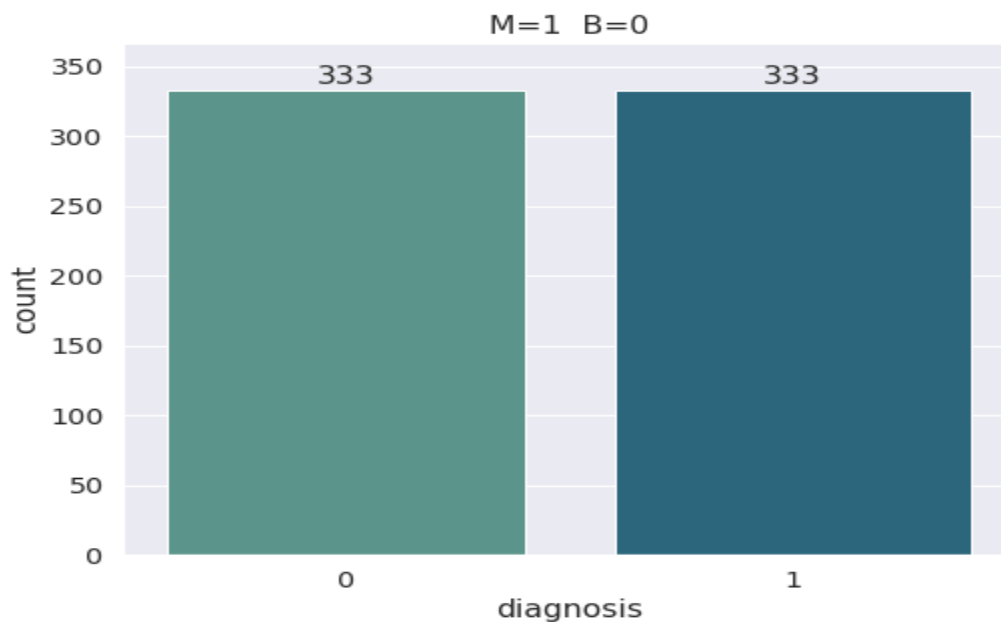


Figure 2: Class distribution of benign and malignant samples after training data balancing

3.2 Model development

Two classification models are employed in this study: Random Forest and Gaussian Naive Bayes. These models are selected to represent two fundamentally different learning paradigms. Random Forest is an ensemble-based, non-parametric method capable of capturing complex nonlinear relationships, while Gaussian Naive Bayes is a probabilistic model based on strong statistical assumptions. The comparison between these models provides insight into how feature dependencies influence classification performance in structured biomedical data.

3.2.1 Random Forest

Random Forest is an ensemble learning method which involves a combination of multiple decision trees to make a strong and generalized forecast. Every tree is trained with a random sampled subset of the training data and at each decision node, a random subset of features are taken into consideration to split Figure 3. This two-way randomness minimizes the variance of models and prevents overfitting, especially in small-sample datasets. Random Forest suits well in this study because of a number of features of the data. First, features obtained on the images of fine needle aspirates have high correlations, which are supported by the correlation analysis. By nature, tree-based methods can handle such multicollinearity without necessarily involving explicitly eliminating features. Second, the data have nonlinear associations among morphological features and diagnostic results, which can be well represented by decision tree structures. One more benefit of the Random Forest is that it gives the measurements of feature importance. These scores are used to determine the contribution of every feature to the predictive power of the model, and therefore can be interpreted in clinical settings. In breast cancer diagnosis, it is critical to have the knowledge of the morphological features that add the most to the classification decisions in order to instill trust and aid in the clinical reasoning. Moreover, the Random Forest can withstand noise and outliers that tend to be common in biomedical data. It has the benefit of averaging irregularities over individual trees and hence making consistent predictions. These features render it a good contender to deployment-oriented diagnostic systems, especially in the context where data quality and consistency might fluctuate.

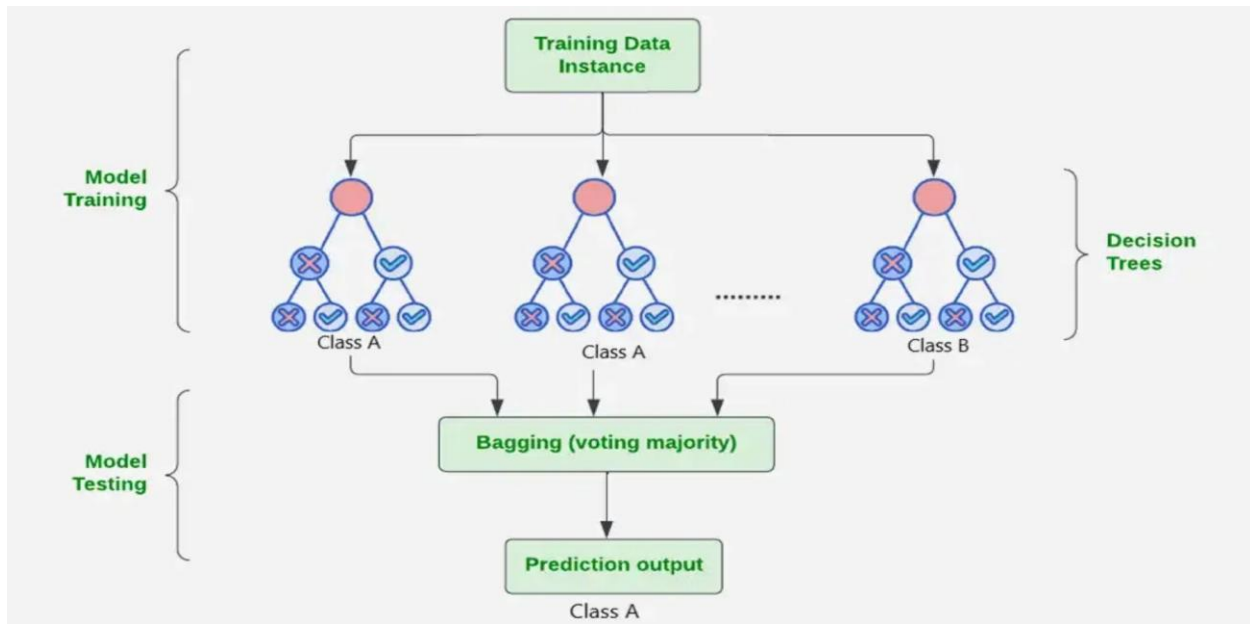


Figure 3: Architecture of the Random Forest classifier with ensemble decision trees and majority voting mechanism

3.2.2 Gaussian Naive Bayes

Gaussian Naive Bayes is a probabilistic classification algorithm that uses the Bayes theorem with the assumption that features are conditionally independent with the class label. All the features are assumed to be normally distributed in each category. Although it is simple, the Gaussian Naive Bayes has been widely applied in medical and diagnostic applications because it is efficient and has a relatively low cost of computation. In this experiment, Gaussian Naive Bayes is used as a control experiment to test how effective its assumptions are to the performance. The independence assumption is especially relevant in this regard, where the dataset has several features based on similar morphological features, including radius, perimeter, and area. These characteristics

are correlated, by definition, and can impact model accuracy in violation of the independence assumption. The strength of Gaussian Naive Bayes is that it can work even with small data sets. It has less parameters than more complicated models and is less likely to overfitting when using limited data. The probabilistic structure of it also offers class probabilities estimates, which could be of use in threshold-based decision-making in clinical screening contexts. The weaknesses of the Gaussian Naive Bayes are however seen when the features are highly dependent. In these instances, the model might under or overestimate the likelihood of some classes and therefore give suboptimal predictions. This is especially essential in medical diagnosis where erroneous diagnosis of a malignant case may have grave repercussions.

3.3 Evaluation protocol

A strictly held-out test set is used to evaluate model performance to be certain that it is an objective measure of the generalization capability. The test set does not play any role in the training, preprocessing parameter estimation, or model selection, and therefore, provides a realistic estimate of how the models would do on unknown clinical data. This method is especially significant in the medical field, where the false estimation of the performance can result in the unreliable diagnostic conclusions. An all-encompassing set of evaluation measures is used to reflect various ways of classification performance. The measurement of overall correctness of predictions is done in terms of accuracy. Precision measures the fraction of correctly identified positive cases of all predicted positives whereas recall is the measure of how well the model identifies the real cases of malignancy. The F1-score offers a balanced score, as it incorporates both precision and recall hence suitable in cases of imbalanced classification. A combination of all these metrics gives a clear picture of how the models behave in both classes. Besides aggregate measures, confusion matrices are evaluated to investigate the results of the prediction by classes and the distribution of the errors. This enables the direct observation of false positive and false negative. The special focus is given to false negatives because the inability to detect malignant cases may lead to severe clinical outcomes. A screening application model has to be sensitive enough and at the same time, it should not be so specific as to make it unacceptable. The receiver operating characteristic curves are also used to determine the performance of a model based on the different classification thresholds. The threshold-independent measure of discriminative ability is the area under the curve. The higher the value, the more separation there is between benign and malignant classes, and this means that the model is very strong in differentiating the two conditions.

3.4 Experimental design considerations

In order to establish the reliability and reproducibility of the experimental results, structured and controlled experimental design is adopted. Other preprocessing operations such as feature scaling and class balancing are done only after the dataset is split into training and test subsets. This will also avoid any influence of information of the test set to the training process, which will remove data leakage. The test set is rigidly and optimally not utilized, but only after the final evaluation. It does not do hyperparameter tuning or intermediate validation, so that performance metrics indicate actual generalization and not optimization bias. The selected design plays an important role in preserving the integrity of the reported findings, particularly in a clinical research setting. The issue of class imbalance is considered only in the training set to enhance the learning of the model without changing the natural distribution of the test set. This makes the evaluation reflective of the reality in the field where the class distributions are not balanced. By keeping the original distribution in the test set, a more realistic evaluation of model performance in real application contexts is possible. Lastly, both the Random Forest and Gaussian Naive Bayes models are subjected to the same preprocessing, training and evaluation conditions. This will provide a fair comparison of the two methods and any differences in performance can only be due to the nature of the model but not experimental anomalies. Collectively, these design decisions form a solid structure of how to test machine learning models in a clinical and reproducible way.

4. Results

This section presents the experimental results obtained from the evaluation of Random Forest and Gaussian Naive Bayes models on the held-out test set. The analysis focuses on overall performance, class-wise behavior, and error distribution to assess the suitability of each model for clinical screening applications.

4.1 Random Forest

Random Forest model shows good and stable results on the test set that was not held, which means that it is efficient in separating benign and malignant breast tumors. The model has a total accuracy of 0.96, and this represents the high degree of accuracy in predictions between the two classes (table 1). The performance in terms of classes also confirms the balanced performance. In the case of the benign category, the model has a precision of 0.95 and a recall of 0.96 with an F1-score of 0.95. In the case of the malignant class, its precision is 0.96 and the recall is 0.95, and F1-score of 0.96. These findings suggest that the model is as effective

in the identification of both types of cases, and is not biased to any of the two classes of benign and malignant cases. The confusion matrix analysis indicates that among the 134 test samples only six are misclassified. Particularly, there is an error of three benign samples being predicted as malignant and three malignant samples being predicted as benign. The fact that the false negative numbers are minimal is especially relevant to a clinical setting since it implies that the model is highly useful in identifying malignant cases and reducing the number of missed diagnoses figure 4. The analyses of the receiver operating characteristic also justify the strength of the Random Forest model. The space below the curve is 1.00 which suggests that there is almost perfect separability of the two classes. This implies that the model at various classification thresholds is highly sensitive and specific (figure 5).

Table 1: Classification performance of Random Forest model on the test set

	precision	recall	f1-score	support
0	0.95	0.96	0.95	60
1	0.96	0.95	0.96	74
accuracy			0.96	134
macro avg	0.95	0.95	0.95	134
weighted avg	0.96	0.96	0.96	134

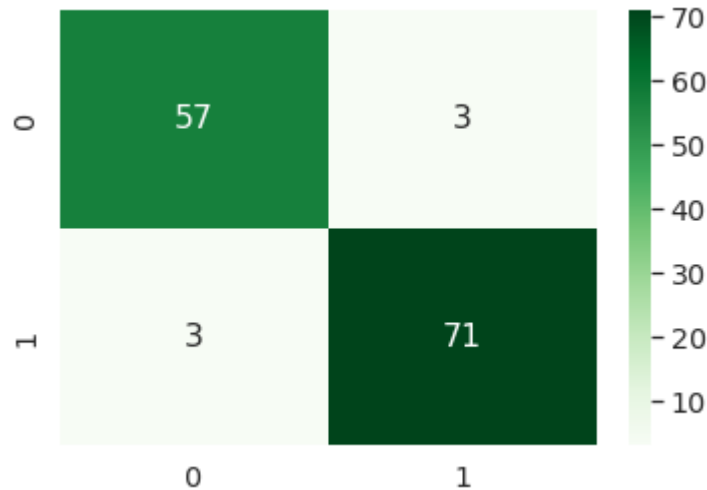


Figure 4: Confusion matrix of Random Forest model on the held out test set

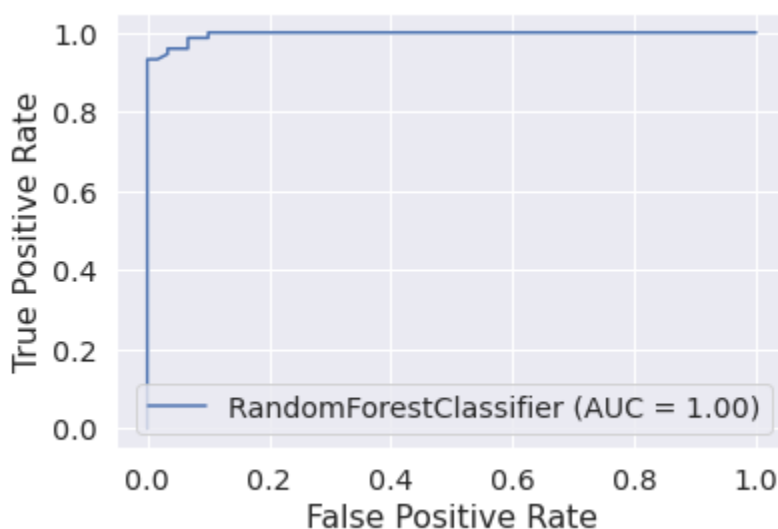


Figure 5: Receiver operating characteristic curve of Random Forest model

4.2 Gaussian Naive Bayes

The Gaussian Naive Bayes model achieves strong overall performance on the held-out test set, with an accuracy of 0.93. While slightly lower than the Random Forest model, the results indicate that Gaussian Naive Bayes remains a competitive baseline for breast cancer classification using structured features table 2. The class-wise performance reveals some variation between benign and malignant detection. For the benign class, the model achieves a precision of 0.89 and a recall of 0.95, resulting in an F1-score of 0.92. For the malignant class, the precision is higher at 0.96, but the recall decreases to 0.91, with an F1-score of 0.93. This indicates that while the model is effective in correctly identifying predicted malignant cases, it is less sensitive in capturing all actual malignant instances. The confusion matrix provides further insight into the error distribution. Out of 134 test samples, a total of ten instances are misclassified figure 6. Specifically, three benign samples are incorrectly predicted as malignant, and seven malignant samples are incorrectly classified as benign. The higher number of false negatives compared to the Random Forest model is particularly important, as it reflects missed cancer cases, which are critical in clinical screening scenarios. The receiver operating characteristic analysis shows an area under the curve of 0.99, indicating excellent discriminative ability figure 7. Although slightly lower than the Random Forest model, this value still demonstrates that Gaussian Naive Bayes can effectively separate benign and malignant classes across different thresholds.

Table 1: Classification performance of Naive Bayes model on the test set

	precision	recall	f1-score	support
0	0.89	0.95	0.92	60
1	0.96	0.91	0.93	74
accuracy			0.93	134
macro avg	0.92	0.93	0.92	134
weighted avg	0.93	0.93	0.93	134

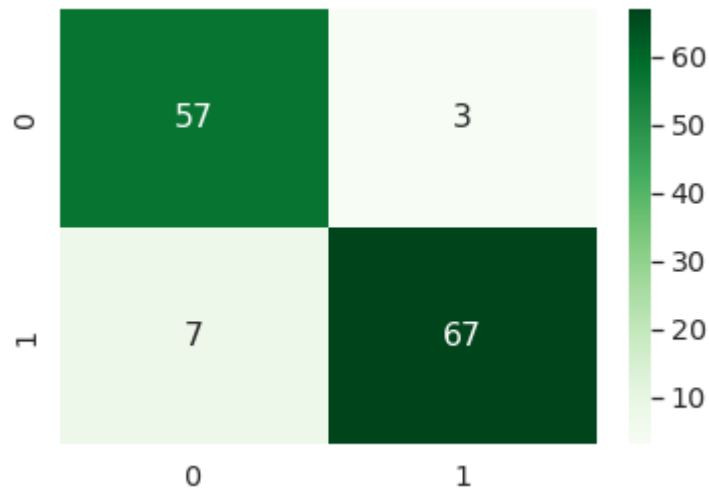


Figure 6: Confusion matrix of Gaussian Naive Bayes model on the held out test set

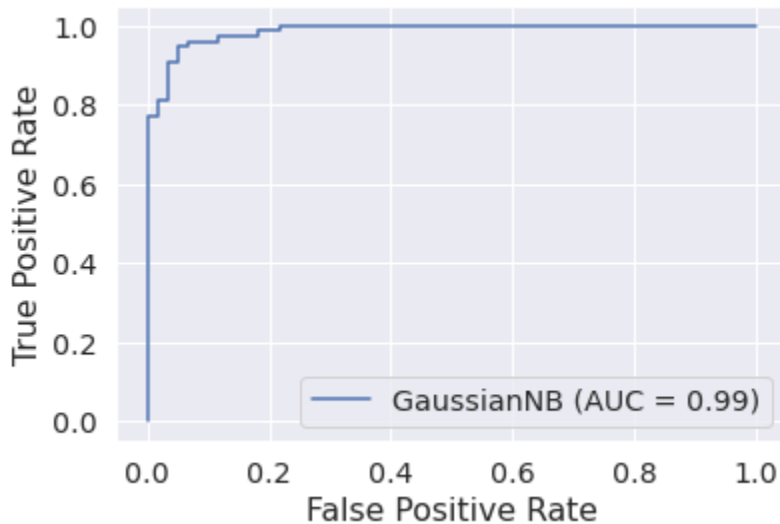


Figure 7: Receiver operating characteristic curve of Gaussian Naive Bayes

4.3 Receiver operating characteristic analysis

Receiver operating characteristic analysis further supports the superiority of the Random Forest model. The area under the curve for Random Forest reaches 1.00, indicating near-perfect separability between benign and malignant classes. Gaussian Naive Bayes also demonstrates strong performance with an area under the curve of 0.99, but slightly underperforms compared to Random Forest. While both models exhibit high discriminative ability, the marginal difference becomes important in clinical contexts where even small improvements in classification performance can lead to better patient outcomes.

4.4 Impact of feature dependencies

The observed performance differences can be attributed to the underlying assumptions of each model. The dataset contains highly correlated features, as demonstrated in the correlation analysis. Random Forest effectively handles such dependencies by constructing decision trees that capture nonlinear interactions among features. In contrast, Gaussian Naive Bayes assumes conditional independence between features, which is not satisfied in this dataset. As a result, its probabilistic estimates become less accurate, leading to increased misclassification, particularly for malignant cases. This finding highlights the importance of selecting models that align with the structural properties of biomedical data.

5. Discussion

The findings of the study prove that machine learning models are capable of differentiating between benign and malignant breast tumors with the help of the structured features based on fine needle aspirate data. Random Forest and Gaussian Naive Bayes have high classification performance but their behaviour is very different because of their assumptions and their ability to represent the relationships between features. This part will discuss these results with regard to clinical relevance, data nature, and practical deployment factors. An essential point is that the Random Forest model is better in all evaluation measures. The fact that it is highly accurate and recalls both classes means that it makes consistent and dependable predictions. Most importantly, the few false negatives indicate its usefulness in the detection of malignancies. False negative reduction is of the essence in a clinical screening environment because missed cancer diagnosis will lead to postponement of treatment and poor patient outcomes. The equal error distribution also points out to the fact that the model gives no preference to any of the classes, which is necessary in terms of providing diagnostic support in a fair manner. Conversely, Gaussian Naive Bayes, despite its competitive overall accuracy, is less sensitive to malignant cases. This can be directly attributed to the fact that, higher number of false negatives was found in this model because it assumed that the features are independent of each other. In this paper, the correlation analysis shows that there are strong dependencies between a number of morphological features, e.g., radius, perimeter and area. These dependencies are a contravention of the assumptions of Gaussian Naive Bayes resulting in poorer estimates of probabilities and misclassification of critical cases. This observation supports the role of matching the model selection to the statistical characteristics of the data. The findings also indicate the importance of the feature characteristics in the model performance determination. The dataset has clear and highly differentiated features, especially the ones that are associated with tumor size and structural abnormalities. These properties allow high separability of classes as indicated by the almost perfect receiver operating characteristic curves. Nevertheless, these high performance results should be viewed with some reservations, particularly due to relatively small size of the dataset. Unless the experimental design is done carefully, it runs a risk of overestimating the performance of a model. In this work, this risk is reduced by means of strict separation of training and testing data, as well as controlled preprocessing. Practically speaking, the results justify the deployment of ensemble-based tools to the deployment-oriented diagnostic systems, especially in healthcare settings with a limited number of resources. Random Forest is a good trade off of performance, strength, and interpretability. The fact that it can give scores on the importance of features enables clinicians to interpret the foundations of the predictions that is necessary to trust and adopt. Also, it is robust to noise and correlated characteristics, which are appropriate in real-world data, where the ideal conditions are seldom achieved. Regardless of these benefits, it is important to note a number of weaknesses. The data employed in this research is comparatively small and is in the form of pre-extracted features as opposed to uncropped medical images. Consequently, the results might not be generalizable to more complicated clinical conditions with heterogeneous data sources. Moreover, the research considers a small number of models, and other algorithms like gradient boosting or deep learning techniques might be useful in offering more information about performance trade-offs. Lastly, though the research focuses on deployment-related aspects, real-world implementation must be validated on external datasets and real workflows.

6. Conclusion

This paper proposes a systematic and repeatable machine learning model to classify breast cancer based on the features obtained through data on the fine needle aspirates. This analysis shows that both random forest and Gaussian Naive Bayes can perform very well when it comes to predictive performance on this dataset, although each is very different in terms of their effectiveness, based on their modeling assumptions. Random Forest model is more accurate and balanced in the class-wise evaluation than Gaussian Naive Bayes in all metrics of evaluation. More importantly it greatly minimises the false negatives, which is an essential fact in clinical screening systems where a missed malignant case can be very dangerous. Gaussian Naive Bayes, on the other hand, is a computationally efficient, and competitive baseline but is restricted by the assumption of independence of features, which is not true in this dataset as the morphological features are strongly correlated. The results emphasize that the choice of the model should be based not only on the general accuracy but also the characteristics of the data and the clinical consequences of various types of errors. Random Forest and other ensemble-based approaches are more appropriate to structured biomedical data with correlated features to provide robustness, interpretability, and reliable performance. Practically, the proposed framework shows how explainable machine learning models can be used to aid breast cancer screening, especially in resource-limited and high-volume healthcare settings. The analysis of structured cytological features is efficient and scalable because it does not require complex imaging pipelines, which is why the method is deployable-friendly. Nonetheless, there are a number of shortcomings. The research is founded on one dataset that has a rather small sample size, and more comprehensive clinical datasets should be validated to ascertain the generalizability. Future research ought to examine sophisticated models, introduce external validation, and look into their integration into clinical practice.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

- [1]. Shinde, Swati, Madhura Kalbhor, and Pankaj Wajire. "DeepCyto: a hybrid framework for cervical cancer classification by using deep feature fusion of cytology images." *Math Biosci Eng* 19.7 (2022): 6415-34.
- [2]. Yu, Bo, Peng Yin, Hechang Chen, Yifei Wang, Yu Zhao, Xianling Cong, Jouke Dijkstra, and Lele Cong. "Pyramid multi-loss vision transformer for thyroid cancer classification using cytological smear." *Knowledge-Based Systems* 275 (2023): 110721.
- [3]. Fang, Ming, Xiujuan Lei, Bo Liao, and Fang-Xiang Wu. "A deep neural network for cervical cell classification based on cytology images." *IEEE Access* 10 (2022): 130968-130980.
- [4]. Fei, Manman, Xin Zhang, Dongdong Chen, Zhiyun Song, Qian Wang, and Lichi Zhang. "Whole slide cervical cancer classification via graph attention networks and contrastive learning." *Neurocomputing* 613 (2025): 128787.
- [5]. Alias, Nur Ain, Wan Azani Mustafa, Mohd Aminudin Jamlos, Hiam Alquran, Hafizul Fahri Hanafi, Shahrina Ismail, and Khairul Shakir Ab Rahman. "Pap smear images classification using machine learning: a literature matrix." *Diagnostics* 12, no. 12 (2022): 2900.
- [6]. Kalbhor, Madhura, Swati Vijay Shinde, and Hemant Jude. "Cervical cancer diagnosis based on cytology pap smear image classification using fractional coefficient and machine learning classifiers." *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 20, no. 5 (2022): 1091-1102.
- [7]. Tsukamoto, Tetsuya, Atsushi Teramoto, Ayumi Yamada, Yuka Kiriya, Eiko Sakurai, Ayano Michiba, Kazuyoshi Imaizumi, and Hiroshi Fujita. "Comparison of fine-tuned deep convolutional neural networks for the automated classification of lung cancer cytology images with integration of additional classifiers." *Asian Pacific journal of cancer prevention: APJCP* 23, no. 4 (2022): 1315.
- [8]. Alquran, Hiam, Mohammed Alsalatie, Wan Azani Mustafa, Rabah Al Abdi, and Ahmad Rasdan Ismail. "Cervical net: A novel cervical cancer classification using feature fusion." *Bioengineering* 9, no. 10 (2022): 578.
- [9]. Xie, Xiaofeng, Chi-Cheng Fu, Lei Lv, Qiuyi Ye, Yue Yu, Qu Fang, Liping Zhang, Likun Hou, and Chunyan Wu. "Deep convolutional neural network-based classification of cancer cells on cytological pleural effusion images." *Modern pathology* 35, no. 5 (2022): 609-614.
- [10]. Zhao, Chen, Renjun Shuai, Li Ma, Wenjia Liu, and Menglin Wu. "Improving cervical cancer classification with imbalanced datasets combining taming transformers with T2T-ViT." *Multimedia tools and applications* 81, no. 17 (2022): 24265-24300.
- [11]. Dey, Pranab. "Artificial neural network in diagnostic cytology." *Cytojournal* 19 (2022): 27.
- [12]. Kalbhor, Madhura, Swati Shinde, Sagar Lahade, and Tanupriya Choudhury. "DeepCerviCancer-deep learning-based cervical image classification using colposcopy and cytology images." *EAI Endorsed Transactions on Pervasive Health and Technology* 9 (2022).
- [13]. Yaman, Orhan, and Turker Tuncer. "Exemplar pyramid deep feature extraction based cervical cancer image classification model using pap-smear images." *Biomedical Signal Processing and Control* 73 (2022): 103428.
- [14]. Ishii, Shuhei, Manabu Takamatsu, Hironori Ninomiya, Kentaro Inamura, Takeshi Horai, Akira Iyoda, Naoko Honma et al. "Machine learning-based gene alteration prediction model for primary lung cancer using cytologic images." *Cancer cytopathology* 130, no. 10 (2022): 812-823.
- [15]. Dai, Jiao, Heng Zheng, Ju Jin, Ye Cheng, and Haimiao Xu. "Claudin18. 2 expression and clinicopathological features in cytology effusion specimens from gastric adenocarcinoma: a comparative study with tissue specimens." *Cancer Cytopathology* 131, no. 6 (2022): 365-372.
- [16]. Yu, Wenjin, Yangyang Liu, Yunsong Zhao, Haofan Huang, Jiahao Liu, Xiaofeng Yao, Jingwen Li et al. "Deep learning-based classification of cancer cell in leptomeningeal metastasis on cytomorphologic features of cerebrospinal fluid." *Frontiers in Oncology* 12 (2022): 821594.
- [17]. Duc, Nguyen Thanh, Yong-Moon Lee, Jae Hyun Park, and Boreom Lee. "An ensemble deep learning for automatic prediction of papillary thyroid carcinoma using fine needle aspiration cytology." *Expert Systems with Applications* 188 (2022): 115927.
- [18]. Mohammed, Badiea Abdulkarem, Ebrahim Mohammed Senan, Zeyad Ghaleb Al-Mekhlafi, Meshari Alazmi, Abdulaziz M. Alayba, Adwan Alownie Alanazi, Abdulrahman Alreshidi, and Mona Alshahrani. "Hybrid techniques for diagnosis with WSIs for early detection of cervical cancer based on fusion features." *Applied Sciences* 12, no. 17 (2022): 8836.
- [19]. Sood, Anubhuti, Deepika Mishra, Rahul Yadav, Krushna Bhatt, and Harsh Priya. "The application of the Bethesda system for reporting cervical cytology to oral cytology: an institutional study." *Clinical Cancer Investigation Journal* 11, no. 4-2022 (2022): 25-32.
- [20]. Aljuaid, Hanan, Nazik Alturki, Najah Alsubaie, Lucia Cavallaro, and Antonio Liotta. "Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning." *Computer Methods and Programs in Biomedicine* 223 (2022): 106951.