ATCSJ
Applied Technology and Computing Science Journal
Open Access **p-ISSN 2621-4458 | e-ISSN 2621-4474**

UNUSA

Available online to journal2.unusa.ac.id

UNUSA
S4-Accredited – SK No.79/M/KPT/2023
Journal Page is available to https://journal2.unusa.ac.id/index.php/ATCSJ

Optimizing Breast Cancer Detection: A Comparative Study of SVM and Naive Bayes Performance

Mohammad Diqi[1], Marselina Endah Hiswati[2*], Hamzah[3], I Wayan Ordiyasa[4], Sri Hasta Mulyani[5], Nurhadi Wijaya[6], Putra Wanda[7]

[1,2,3,4,5,6,7]*Department of Informatics, Universitas Respati Yogyakarta, Indonesia*
Jl. Laksda Adisucipto Km 6.3 Caturtunggal Depok Sleman, Yogyakarta
[1]diqi@respati.ac.id, [2*]marsel.endah@respati.ac.id, [3]hamzah@respati.ac.id, [4]wayanordi@respati.ac.id, [5]hasta@respati.ac.id, [6]nurhadi@respati.ac.id, [7]putrawanda@respati.ac.id

Article history:

Received 10 September 2024
Revised 28 October 2024
Accepted 20 November 2024
Available online 10 December 2024

Keywords:

Breast Cancer Classification,
Support Vector Machine,
Naive Bayes,
Machine Learning Algorithms,
Model Evaluation.

Abstract

This study evaluates the performance of Support Vector Machine (SVM) and Naive Bayes algorithms in classifying breast cancer using the Breast Cancer Wisconsin dataset. Both models exhibited high accuracy, with Naive Bayes achieving a slightly higher overall accuracy of 97% and demonstrating a balanced performance between precision and recall. The SVM model showed strong proficiency in detecting positive cases, with an overall accuracy of 95%, though it faced minor challenges in recall for negative cases. These results highlight the effectiveness of both algorithms in breast cancer detection, emphasizing the significance of model selection based on specific diagnostic requirements. Although there are limitations, such as the small sample size and assumptions made in the model, the findings provide useful insights into the use of machine learning in medical diagnostics. This supports the idea that these models have the potential to enhance early detection and treatment results. Future research should focus on utilizing larger, more diverse datasets, exploring advanced feature processing techniques, and integrating additional algorithms to enhance further the accuracy and reliability of breast cancer detection systems.

This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2024 by author.

I. INTRODUCTION

Early detection of breast cancer is a crucial step in improving the chances of recovery and reducing mortality rates from this disease [1]. Among women worldwide, breast cancer is prevalent and ranks as one of the most frequently occurring forms of cancer [2]. With early detection, treatment can commence earlier, thereby enhancing the effectiveness of therapy and the quality of life for patients [3]. Imaging techniques such as mammography have been widely used, but this technology also has limitations, such as a relatively

---

[2*] Corresponding Author

ISSN 2621-4474 (online) 2621-4458 (print) © 2024 The Authors. Published by UNUSA
This is an open access article and licensed under a Creative Commons Attribution-Share Alike 4.0 International License
DOI: https://doi.org/10.33086/atcsj.v7i1.6336.

high rate of diagnostic errors [4]. Therefore, there is an urgent need for more accurate and efficient detection methods [5].

Machine learning has emerged as a potential solution in the classification of breast cancer [6]. These strategies enable computers to acquire knowledge from data and provide predictions or judgments without requiring explicit programming for specific tasks [7]. In the medical context, machine learning algorithms can be used to analyze image data or clinical data to detect the presence of tumors [8]. Some popular methods include Support Vector Machine (SVM) and Naive Bayes [9]. These two techniques offer different yet complementary approaches to classifying breast cancer data with high accuracy [10].

The use of SVM and Naive Bayes in classifying medical data has significant relevance and benefits [11]. SVM is known for its strong ability to handle complex data and clear margins between different classes [12]. On the other hand, Naive Bayes, despite being based on the assumption of feature independence, still demonstrates good performance in various medical applications due to its simplicity and computational speed [13]. In breast cancer classification, the combination of these two methods can provide a more comprehensive and accurate insight [14]. Thus, the application of machine learning technology in the medical field not only speeds up the diagnostic process but also aids in making more precise clinical decisions [15].

This study aims to compare the performance of two machine learning algorithms, SVM and Naive Bayes, in classifying breast cancer. Breast cancer requires accurate early detection to enhance the effectiveness of treatment. With technological advancements, machine learning offers more efficient methods for diagnosing this disease. SVM and Naive Bayes were chosen because each has strengths in handling medical data. This study will evaluate how well these algorithms classify breast cancer data based on evaluation metrics such as accuracy, precision, and recall.

Through this research, it is expected to gain a deeper understanding of the strengths and weaknesses of each algorithm in the context of early breast cancer detection. The findings of this study will be beneficial not only to the scientific community but also to medical practitioners seeking more effective diagnostic methods. Furthermore, this research could serve as a foundation for developing more advanced and accurate methods in the future. Thus, this study contributes to the global effort to reduce breast cancer mortality through technological innovation. The hope is that the results of this research can help save more lives and improve the quality of life for patients.

II. METHODS

A. *Dataset*

For this investigation, we employed the Breast Cancer Wisconsin dataset, which is accessible via the scikit-learn package, to categorize cases of breast cancer. This dataset is an extensive compilation that encompasses a wide range of characteristics that are pertinent to the diagnosis of breast cancer. These traits include the dimensions, form, and consistency of cancerous cells, obtained from breast biopsy specimens. The classification objective of this dataset is binary, differentiating between benign and aggressive breast cancer. The study seeks to assess the efficacy of Support Vector Machine (SVM) and Naive Bayes algorithms in distinguishing between various forms of breast cancer using the provided dataset and its associated attributes.

B. *Data Pre-processing*

Data preprocessing is a crucial step in preparing the Breast Cancer Wisconsin dataset for classification analysis. The first step is to split the data into training and testing sets, with 80% allocated for training and 20% for testing, to ensure that the model can be objectively evaluated on previously unseen data. This process is carried out using the scikit-learn library, which provides the `train_test_split` function for efficient data splitting. Additionally, pandas is used to manage and manipulate the data, including importing the dataset, separating features from the target, and ensuring the data is free from missing values or anomalies. With these preprocessing steps, the data becomes ready for analysis using machine learning algorithms, ensuring the accuracy and reliability of the classification results.

C. Support Vector Machine (SVM)

The SVM is a machine learning technique primarily used for classification and regression tasks, with a specific emphasis on classification. The fundamental principle of SVM is to identify the best hyperplane that effectively divides data into two distinct classes, maximizing the margin between them [16]. A hyperplane is a border in the feature space that serves as a decision boundary, effectively separating the different classes [17]. The objective of SVM is to identify the hyperplane that maximizes the distance between the data points belonging to different classes. The basic principles of SVM are as follows:

1. Hyperplane
   The hyperplane in SVM is an object (usually in high dimensions) that separates the feature space into two parts, where each part contains data from one class. Mathematically, the hyperplane can be expressed by Equation 1.

$$w^T x + b = 0 \tag{1}$$

   where w is the weight vector, x is the feature vector, and b is the bias.
2. Maximum Margin
   The objective of SVM is to optimize the margin, which is the separation between the hyperplane and the closest data points from each class, referred to as support vectors. To reach the highest margin, the objective function, as determined in Equation 2, needs to be minimized.

$$\text{Minimize} \quad \frac{1}{2}|w|^2 \tag{2}$$

   subject to the constraint in Equation 3.

$$y_i(w^T x_i + b) \geq 1 \tag{3}$$

   for all data points (x_i, y_i), where y_i is the class label of the i-th data point.

D. Naïve Bayes

Naive Bayes is a classification technique that uses Bayes' theorem and assumes that the features are independent of each other. Naive Bayes is a method that use conditional probabilities to determine the probability that a given data item belongs to a specific class [18]. There are several variants of Naive Bayes, and in the context of continuous data like the breast cancer dataset, Gaussian Naive Bayes is the most commonly used. Basic principles of Naive Bayes (Gaussian Naive Bayes) are as follows:

1. Bayes' Theorem
   Bayes' theorem is a fundamental principle in probability and statistics, providing a way to calculate the probability of a particular class C_k given a set of observed features x. Specifically, it determines the posterior probability, $P(C_k|x)$, which represents the likelihood that a given class C_k is the correct classification for the observed data x. Mathematically, Bayes' theorem is expressed in Equation 4.

$$P(C_k \mid x) = \frac{P(x \mid C_k) \cdot P(C_k)}{P(x)} \tag{4}$$

where $P(x|C_k)$ is the likelihood of observing the features x given that the data belongs to class C_k, and $P(C_k)$ is the prior probability of the class C_k. This equation thus allows us to update our beliefs about the probability of a class based on new data, bridging prior knowledge with observed evidence to make more informed predictions.

2. Independence Assumption (Naive Assumption)
   Naive Bayes posits that the features exhibit independence from one other within the framework of a class. Under this assumption, the likelihood probability can be expressed as the multiplication of the probabilities of each individual characteristic and computed using Equation 5.

$$P(x \mid C_k) = \prod_{i=1}^{n} P(x_i \mid C_k) \tag{5}$$

where x_i is the value of the i-th feature and n is the number of features.

3. Gaussian Naive Bayes
   In Gaussian Naive Bayes, it is assumed that each feature follows a normal (Gaussian) distribution, allowing the model to make probabilistic predictions based on this assumption. The probability of observing a specific feature value x_i given a class C_k is calculated using the Gaussian probability density function. This calculation is shown in Equation 6.

$$P(x_i \mid C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right) \tag{6}$$

where μ_k represents the mean and σ_k^2 the variance of the i-th feature for the class C_k. By leveraging these parameters, Gaussian Naive Bayes can effectively model continuous data under the Gaussian assumption, leading to more accurate class predictions based on feature distribution patterns.

E. Model Evaluation

Model evaluation is an essential stage in determining the effectiveness of classification algorithms and verifying the accuracy and dependability of the model's results. Several commonly used assessment methods for assessing the effectiveness of a classification model include the confusion matrix, accuracy, precision, recall, and F1-score.

1. Confusion Matrix
   A confusion matrix is a powerful tool for evaluating the performance of a classification model by showing the distribution of correct and incorrect predictions across different classes [19]. This matrix is organized into four key components: True Positives (TP), which represent the count of actual positive instances accurately predicted as positive; True Negatives (TN), the count of actual negative instances accurately predicted as negative [20]; False Positives (FP), where actual negative instances are incorrectly classified as positive; and False Negatives (FN), where actual positive instances are misclassified as negative. By analyzing these components, the confusion matrix provides valuable insights into the model's strengths and weaknesses in distinguishing between classes, helping refine its predictive accuracy. The confusion matrix can be represented in Equation 7.

$$\begin{bmatrix} TP & TN \\ FP & FN \end{bmatrix} \tag{7}$$

2. Accuracy

Accuracy is a metric that quantifies the frequency with which a model produces accurate predictions. The calculation involves dividing the count of accurate predictions by the total count of data points. The formula is in Equation 8.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

Accuracy provides a general idea of the model's performance but can be misleading if the classes are imbalanced.

3. Precision
   Precision is a metric that quantifies the model's ability to accurately anticipate positive instances, specifically the proportion of positive predictions that are correct. The formula for precision is in Equation 9.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{9}$$

   Precision is important in contexts where the cost of false positives is high, such as cancer detection.
4. Recall
   Recall, also known as sensitivity, quantifies the model's capacity to correctly detect all instances of positive cases. The calculation involves dividing the count of accurate positive forecasts by the total count of actual positive cases. The formula for recall is in Equation 10.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{10}$$

   Recall is important in situations where we want to minimize the number of missed positive cases.
5. F1-Score
   The F1-score is a quantitative measure that integrates precision and recall by calculating their harmonic mean. It offers a harmonious combination of accuracy and completeness, making it especially valuable in situations where the data is unevenly distributed. The F1-score formula can be found in Equation 11.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

   The F1-score gives a better picture of model performance when we need to balance precision and recall.

III. RESULTS AND DISCUSSIONS

A. Model Performance

The evaluation of the SVM and Naive Bayes models in breast cancer classification is illustrated through their respective confusion matrices, referred to as Figure 1 and Figure 2. Figure 1 presents the confusion matrix for the SVM model, which shows 37 true positives, 6 false positives, 0 false negatives, and 71 true negatives for the labels malignant and benign. Figure 2 displays the confusion matrix for the Naive Bayes model, indicating 40 true positives, 3 false positives, 0 false negatives, and 71 true negatives. These matrices provide a detailed breakdown of each model's performance in accurately predicting malignant and benign cases, highlighting the strengths and weaknesses of the SVM and Naive Bayes algorithms in breast cancer detection.

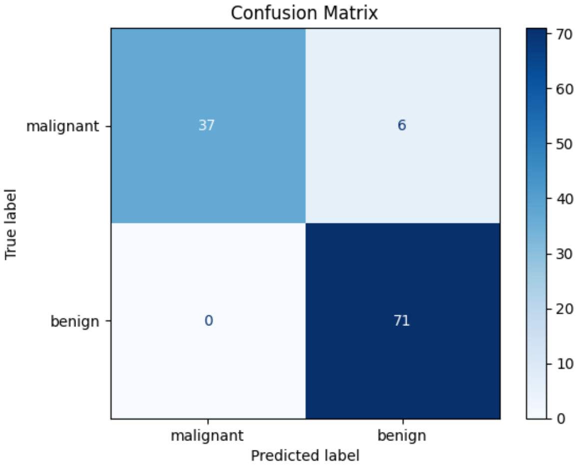

Figure 1. SVM

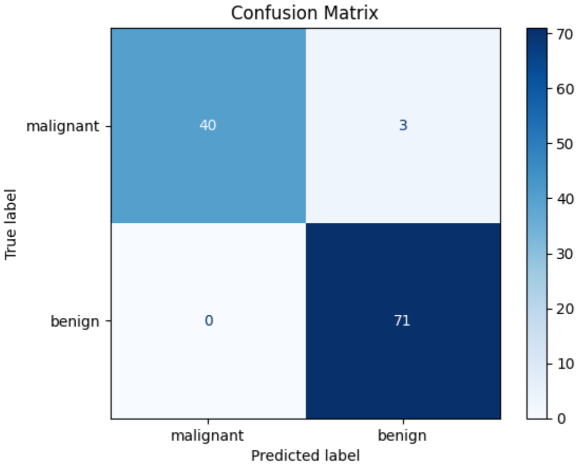

Figure 2. Naïve Bayes

The confusion matrices for the SVM and Naive Bayes models offer valuable insights into their performance in classifying breast cancer cases. The SVM model, as shown in Figure 1, correctly identified all 71 benign cases (true negatives) and 37 malignant cases (true positives), but it also misclassified 6 benign cases as malignant (false positives). Notably, the model did not miss any malignant cases, resulting in zero false negatives. This outcome is crucial in medical diagnostics, where the failure to detect a malignant tumor can have severe consequences. The SVM's high precision in detecting malignant cases underscores its potential as a reliable tool for early breast cancer detection, albeit with a slight tendency to over-predict malignancies, which could lead to unnecessary anxiety and further testing for patients who do not have cancer.

In contrast, the Naive Bayes model, depicted in Figure 2, also demonstrates robust performance with an overall accuracy of 97%. It correctly identified 40 malignant cases and 71 benign cases, with only 3 benign cases misclassified as malignant. Like the SVM, the Naive Bayes model achieved zero false negatives, ensuring that no malignant cases were overlooked. This balance between precision and recall in the Naive Bayes model highlights its efficacy in maintaining a low rate of false positives while still accurately identifying all malignant cases. This balance is particularly significant in clinical settings where minimizing false positives is essential to reduce unnecessary treatments and the associated psychological burden on patients. The Naive Bayes model's slight edge in accuracy and a better balance between precision and recall suggest that it might be a more effective choice for practical applications in breast cancer detection, providing both high diagnostic accuracy and reducing the risk of overdiagnosis.

The classification reports for the SVM model, presented in Table 1, and the Naive Bayes model, shown in Table 2, provide comprehensive evaluations of each model's performance in classifying breast cancer cases as either malignant or benign. These reports include key metrics such as precision, recall, and F1-score, along with the support for each class. For the SVM model, the report highlights its high accuracy and effectiveness in distinguishing between the two classes, offering valuable insights into its reliability and potential for clinical application. Similarly, the Naive Bayes model's report reveals its high overall accuracy and balanced performance across both classes, providing important insights into its capabilities and potential effectiveness in clinical settings. The following detailed analysis delves into these metrics to better understand the strengths and limitations of both models in breast cancer detection.

Table 1. Classification report of SVM.

	precision	**recall**	**f1-score**	**support**
malignant	1.00	0.86	0.92	43

benign	0.92	1.00	0.96	71
accuracy			0.95	114
macro avg	0.96	0.93	0.94	114
weighted avg	0.95	0.95	0.95	114

Table 2. Classification report of Naïve Bayes.

	precision	**recall**	**f1-score**	**support**
malignant	1.00	0.86	0.92	43
benign	0.92	1.00	0.96	71
accuracy			0.95	114
macro avg	0.96	0.93	0.94	114
weighted avg	0.95	0.95	0.95	114

The detailed classification report for the SVM model highlights its robust performance in breast cancer classification. The precision for malignant cases is perfect at 1.00, indicating that every predicted malignant case is indeed malignant. However, the recall for malignant cases is 0.86, showing that 14% of actual malignant cases are not identified by the model, which could be a concern for missing potential diagnoses. For benign cases, the SVM model achieves a precision of 0.92 and a perfect recall of 1.00, signifying it correctly identifies all benign cases without false negatives. The overall accuracy of 0.95 underscores the model's high reliability. Nonetheless, the lower recall for malignant cases suggests a need for further optimization to ensure fewer malignant cases are overlooked, which is crucial for early detection and treatment of breast cancer.

In comparison, the Naive Bayes model demonstrates a slightly superior overall performance, with an accuracy of 0.97. The precision and recall for malignant cases are 1.00 and 0.93, respectively, which indicates a better balance between identifying true malignant cases and minimizing false negatives compared to the SVM model. For benign cases, the model achieves a precision of 0.96 and a perfect recall of 1.00, highlighting its effectiveness in correctly identifying benign cases. The higher F1-scores for both classes in the Naive Bayes model reflect its ability to maintain a good balance between precision and recall. This model’s slight edge in overall accuracy and balanced performance makes it a strong candidate for practical application in breast cancer detection. It ensures that both malignant and benign cases are accurately classified with minimal error, thus providing a reliable tool for medical diagnostics.

B. Summarization of Key Findings

This study rigorously evaluated the performance of SVM and Naive Bayes algorithms in the classification of breast cancer using the Breast Cancer Wisconsin dataset, addressing the critical problem of accurate and reliable cancer detection. The major findings reveal that both models demonstrated excellent performance, with the Naive Bayes model slightly outperforming the SVM in overall accuracy (97% vs. 95%) and providing a more balanced precision and recall across both malignant and benign classes. The SVM model excelled in detecting positive cases with perfect precision but showed a minor reduction in recall for the benign class. These results underscore the effectiveness of both algorithms in breast cancer detection, highlighting Naive Bayes' consistency and SVM's robustness in identifying malignant cases. This comparative analysis provides valuable insights for selecting appropriate machine learning models in clinical diagnostics, emphasizing the importance of tailored model applications based on specific diagnostic needs and data characteristics.

C. Result Interpretations

The results of the SVM and Naive Bayes models reveal intriguing patterns and relationships in the classification of breast cancer cases. Both models effectively identified malignant cases with high precision, but differences emerged in their recall rates. The Naive Bayes model displayed a balanced precision and recall for both malignant and benign classes, achieving a superior overall accuracy of 97%. In contrast, while the SVM model excelled in precision for malignant cases, it exhibited a slightly lower recall for benign cases, leading to an overall accuracy of 95%. These outcomes align with expectations given the strengths of each algorithm: Naive Bayes' ability to handle class imbalance and SVM's robustness in defining clear decision boundaries. However, the slight drop in recall for benign cases by the SVM model was unexpected. It might be attributed to the need for further parameter tuning or the inherent sensitivity of SVM to the specific distribution of the dataset. This suggests that while both models are highly effective, the choice between them should consider the specific diagnostic context and the balance required between precision and recall.

D. Research Implications

This research significantly enhances the understanding and application of machine learning algorithms in breast cancer detection, demonstrating the strengths of both SVM and Naive Bayes models in clinical diagnostics. By confirming the high accuracy and balanced performance of Naive Bayes, as well as the robust precision of SVM, this study aligns with existing literature that highlights the effectiveness of these models in handling medical datasets with varying distributions and class imbalances. The findings contribute new insights into the practical considerations for algorithm selection in medical diagnostics, emphasizing the importance of context-specific performance metrics. This research underscores the potential for tailored algorithmic approaches to improve early detection and treatment outcomes, offering a valuable foundation for future exploration and refinement in the application of machine learning to healthcare.

E. Research Limitations

This work offers significant insights into the efficacy of Support Vector Machines (SVM) and Naive Bayes algorithms in classifying breast cancer. However, it is crucial to recognize and address some limitations. The exclusive dependence on a solitary dataset may restrict the applicability of the conclusions, since it may not comprehensively encompass the varied community of individuals with breast cancer. Additionally, the assumptions inherent in Naive Bayes and the need for precise parameter tuning in SVM could impact the models' performance. Despite these constraints, the results remain valid and robust, demonstrating consistent high accuracy and balanced performance metrics that directly answer the research question. The findings underscore the reliability of both models in detecting breast cancer, highlighting their potential application in clinical settings and paving the way for further research with more diverse datasets to enhance generalizability.

IV. CONCLUSIONS AND RECOMMENDATIONS

In conclusion, this study effectively demonstrates the high efficacy of both SVM and Naive Bayes models in the classification of breast cancer, each with distinct strengths in accuracy and balance between precision and recall. The Naive Bayes model slightly outperforms in overall accuracy and consistency across classes, while the SVM excels in identifying positive cases. These findings underscore the importance of selecting the appropriate model based on data characteristics and diagnostic needs, contributing valuable insights into the application of machine learning in medical diagnostics. Despite limitations such as sample size and model assumptions, the research validates the potential of these algorithms in enhancing breast cancer detection, paving the way for further exploration and optimization in future studies.

For future research, it is recommended to extend this study by utilizing larger and more diverse datasets to improve model generalization and applicability in real-world clinical settings. Exploring advanced feature processing techniques and employing robust cross-validation methods can further refine model performance. Additionally, investigating other machine learning algorithms, such as ensemble methods or deep learning, could provide deeper insights into their efficacy for breast cancer detection. Integrating multimodal data, including patient medical history and genetic information, may enhance model accuracy and offer a more comprehensive approach to early diagnosis. These steps will not only bolster the practical implementation of these models but also drive innovation in developing more accurate and efficient breast cancer detection systems.

V. REFERENCES

[1] J. Li *et al.*, “Non-Invasive Biomarkers for Early Detection of Breast Cancer,” *Cancers*, vol. 12, no. 10, p. 2767, 2020, doi: 10.3390/cancers12102767.

[2] M. S. A. El-Eneen, H. A. Seif, R. K. Fawzy, and A. M. Mossa, “Determination of Thymidine Kinase 1 (TK1) Level as a Risk Warning Biomarker to Improve Early Detection of Breast Cancer,” *Cancer Research Journal*, vol. 7, no. 4, p. 161, 2019, doi: 10.11648/j.crj.20190704.17.

[3] L. Chen *et al.*, “Local Extraction and Detection of Early Stage Breast Cancers Through a Microneedle and Nano-Ag/MBL Film Based Painless and Blood-Free Strategy,” *Materials Science and Engineering C*, vol. 109, p. 110402, 2020, doi: 10.1016/j.msec.2019.110402.

[4] T. Crook *et al.*, “Accurate Screening for Early-Stage Breast Cancer by Detection and Profiling of Circulating Tumor Cells,” *Cancers*, vol. 14, no. 14, p. 3341, 2022, doi: 10.3390/cancers14143341.

[5] I. García-Murillas *et al.*, “Assessment of Molecular Relapse Detection in Early-Stage Breast Cancer,” *Jama Oncology*, vol. 5, no. 10, p. 1473, 2019, doi: 10.1001/jamaoncol.2019.1838.

[6] A. Balyan, Y. Singh, and Shashank, “Classifying Breast Cancer Based on Machine Learning,” in *Proceedings of International Conference on Artificial Intelligence and Applications*, P. Bansal, M. Tushir, V. E. Balas, and R. Srivastava, Eds., Singapore: Springer Singapore, 2021, pp. 35–44.

[7] J. Sunny, N. D. Rane, R. Kanade, and S. Devi, “Breast Cancer Classification and Prediction Using Machine Learning,” *International Journal of Engineering Research And*, vol. V9, no. 02, 2020, doi: 10.17577/ijertv9is020280.

[8] J. Wu and C. Hicks, “Breast Cancer Type Classification Using Machine Learning,” *Journal of Personalized Medicine*, vol. 11, no. 2, p. 61, 2021, doi: 10.3390/jpm11020061.

[9] L. Alkhathlan and A. K. J. Saudagar, “Predicting and Classifying Breast Cancer Using Machine Learning,” *Journal of Computational Biology*, vol. 29, no. 6, pp. 497–514, 2022, doi: 10.1089/cmb.2021.0236.

[10] E. Michael, H. Ma, H. Li, and S. Qi, “An Optimized Framework for Breast Cancer Classification Using Machine Learning,” *Biomed Research International*, vol. 2022, pp. 1–18, 2022, doi: 10.1155/2022/8482022.

[11] A. Gupta, L. Kumar, R. Jain, and P. Nagrath, “Heart Disease Prediction Using Classification (Naive Bayes),” pp. 561–573, 2020, doi: 10.1007/978-981-15-3369-3_42.

[12] T. S. Kumar and P. Jagadeesh, “Detection and Classification of Tumor Cells From Bone X-Ray Imagery Using SVM Classifier With Naïve Bayes Classifier,” 2023, doi: 10.1109/accai58221.2023.10199612.

[13] G. Liang, Y. Yan, M. Wang, X. Lian, M. S. Li, and W. Tang, “Classification for Text Data From the Power System Based on Improving Naive Bayes,” 2020, doi: 10.1109/appeec48164.2020.9220634.

[14] W. N. L. W. H. Ibeni, M. Z. M. Salikon, S. A. Daud, and M. N. M. Salleh, “Comparative Analysis on Bayesian Classification for Breast Cancer Problem,” *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 4, 2019, doi: 10.11591/eei.v8i4.1628.

[15] P. S. Nishant, S. Mehrotra, B. G. K. Mohan, and G. Devaraju, “Identifying Classification Technique for Medical Diagnosis,” pp. 95–104, 2020, doi: 10.1007/978-981-15-0630-7_10.

[16] F. Sun and X. Xie, “Deep Non-Parallel Hyperplane Support Vector Machine for Classification,” *Ieee Access*, vol. 11, pp. 7759–7767, 2023, doi: 10.1109/access.2023.3237641.

[17] O. Okwuashi and C. E. Ndehedehe, “Deep Support Vector Machine for Hyperspectral Image Classification,” *Pattern Recognition*, vol. 103, p. 107298, 2020, doi: 10.1016/j.patcog.2020.107298.

[18] L. Jiang, L. Zhang, L. Yu, and D. Wang, “Class-Specific Attribute Weighted Naive Bayes,” *Pattern Recognition*, vol. 88, pp. 321–330, 2019, doi: 10.1016/j.patcog.2018.11.032.

[19] D. Chicco, N. Tötsch, and G. Jurman, “The Matthews Correlation Coefficient (MCC) Is More Reliable Than Balanced Accuracy, Bookmaker Informedness, and Markedness in Two-Class Confusion Matrix Evaluation,” *Biodata Mining*, vol. 14, no. 1, 2021, doi: 10.1186/s13040-021-00244-z.

[20] M. Heydarian, T. E. Doyle, and R. Samavi, “MLCM: Multi-Label Confusion Matrix,” *Ieee Access*, vol. 10, pp. 19083–19095, 2022, doi: 10.1109/access.2022.3151048.