Cihan University-Erbil Scientific Journal (CUESJ)

Research Article

Application of Support Vector Machines Kernel Functions for Breast Cancer Detection and Classification: A Comparative Study

Raz M. Hama Salih, Saman H. Mahmood

Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University-Erbil, Iraq

ABSTRACT

This study deals with the linear, polynomial, and radial basis function (RBF) kernel-based support vector machines (SVM) applied to the classification of breast cancer. It checks the efficiency of different kernel-based techniques for distinguishing benign and malignant tumors against medical datasets. This analysis is conducted on the public Wisconsin Breast Cancer Dataset borrowed from the UCI Machine Learning Repository, which consists of 569 cases, 357 benign and 212 malignant. It aims to develop the assessment necessary for each Kernel based on its accuracy, precision, recall, computational efficiency, F1 score, and area under the curve (AUC). The outcome reveals that, in general, the polynomial Kernel will provide top-ranking results across the various features, thereby enhancing accuracy (96.84%), specificity (97.76%), and AUC (0.9966) compared to the linear and RBF kernels for use in disease classifications of cancers. These findings indicate that the polynomial kernel adeptly captures complex, non-linear data relationships, making it a strong candidate for developing more accurate SVM-based breast cancer detection models.

Keywords: Machine learning, support vector machines, kernel functions, breast cancer, classification

INTRODUCTION

Breast cancer is one of the leading causes of cancer-related mortality among women worldwide, presenting a significant global health challenge. Early and accurate diagnosis is important in the improvement of patient outcomes and survival rates.[1] Traditional diagnostic approaches include mammography, ultrasound, and biopsy – all with their disadvantages. Mammography has wide usage, though one may also expect false positives and negatives, especially in dense breast tissue. However, while highly sensitive, biopsies are invasive and require a great deal of time. Moreover, they may not be suitable for all patients.[2] The inherent variability in the interpretation of medical images adds to diagnostic inconsistencies and can further delay diagnosis.[3] Such limitations call for newer methods of diagnosis. While these traditional methods are good in many cases, they can sometimes have difficulty with data with complex relationships or many dimensions.

Machine learning can be a very promising avenue for improving medical diagnosis by enabling the analysis of complex datasets with high accuracy and speed.[4] Supervised learning algorithms, especially those developed for classification, are quite effective at distinguishing.

While the field has dealt mainly with segregating between cancerous and non-cancerous tissues using different features of medical images or clinical data, a variety of machine learning algorithms are used to achieve the end goal. Support vector machines (SVM) are among them that have acquired great attention for being robust and capable of dealing with high-dimensional data on the availability of small-sized samples. Basically, kernel functions map input data into a higher-dimension feature space to create non-linear separating hyperplanes; therefore, increasing classification efficiency on complicated datasets can be developed correspondingly.

In light of existing related studies on machine learning for the detection of breast cancer, comparative studies, which would focus comprehensively on the performance of various SVM kernel functions using some specific breast cancer datasets, are few and far between. While various studies have been conducted on the application of different kernel functions

Corresponding Author:
Raz M. Hama Salih, Department of Statistics and Informatics, College of Administration and Economics, Salahaddin University-Erbil, Iraq.
E-mail: raz.hamasalih@su.edu.krd

Received: February 4, 2025
Accepted: March 14, 2025
Published: April 20, 2025

DOI: 10.24086/cuesj.v9n1y2025.pp41-46

for the detection of breast cancer, there is no comparative study on the performances of Linear, Polynomial, and radial basis function (RBF) kernels on the different publicly available datasets of breast cancer.[5] This will indeed be very helpful to point out which of these kernel functions would best suit the optimum detection and classification of breast cancer in various scenarios.

The main purpose of this research is to gauge the performances for different kernel functions of SVM on breast cancer datasets, such as a Linear, Polynomial, and RBF kernel for the most appropriate kernel regarding their performance on breast cancer data effective detection and classification. This study will carry out the assessment of the various functions of kernels and their efficiency in performance and develop insight into choosing the appropriate kernel function among the options given for the choice of a kernel function in certain datasets.

LITERATURE REVIEW

Traditional methods for diagnosis of breast cancer include several diagnostic modalities such as mammography, ultrasound, and biopsies. However, these methods are not without their limitations.[6-9]

Machine learning, or more specifically supervised learning, has developed to effectively enhance the detection and classification of breast cancer. Various studies have been carried out that apply several machine learning algorithms, such as artificial neural networks, SVMs, and ensemble methods, on the imaging and clinical datasets for the diagnosis of breast cancer.[10] Of all these, the most promising performance is offered by SVMs, which classify the images and clinical data of breast cancer due to their capability to handle high-dimensional datasets with fewer numbers of total observations effectively.[11]

The SVM works by constructing a hyperplane that maximizes the margin between the data points of different classes and separates them in high-dimensional space. Data points closest to the hyperplane are called support vectors, and they uniquely define the position of the hyperplane. Major performance bases of SVM are kernel functions, as they allow non-linear classification without explicit computation of transformation in higher dimensional space.[12]

The selection of the kernel function in SVM plays a crucial role in choosing the model's implementation. The linear kernel, the simplest model, is sufficient for linearly detachable data but cannot be used when dealing with complex and non-linear datasets. A study has revealed that SVMs employing a linear kernel can achieve high accuracy on smaller datasets.[13] Conversely, the RBF kernel is one of the most adequate functions for breast cancer detection. Studies demonstrated that RBF-based SVMs outperform other kernel functions in accuracy and robustness, especially in non-linear datasets. For instance, one study revealed that the RBF kernel performed an accuracy of 89.1%, exceeding the polynomial and linear functions.[14] Similarly, another investigation revealed that RBF models optimized with genetic algorithms achieved accuracies of 97.71%.[15]

This research paper investigates the effectiveness of kernel function selection, compares the performance of different kernel functions, and discusses optimization techniques to improve SVM classification accuracy in breast cancer data.

MATERIALS AND METHODS

Kernel Functions

Linear kernel

The Linear kernel performs a dot product of input vectors and is suitable for linearly separable data.[16] The linear kernel tries to find a line that maximizes margins and minimizes misclassification. It is intended to find a hyperplane that maximizes the margins between classes and, at the same time, minimizes misclassifications.

$$K(x,y) = x.z + C \quad (1)$$

C is a parameter that controls the trade-off between the margin of the decision boundary and the accuracy of classifying the training data.

As the value of C increases, the tolerance for misclassification decreases, resulting in a narrower margin.[17]

Polynomial kernel

The Polynomial kernel maps input data to a higher-dimensional space using a polynomial function, with different degrees influencing the model's complexity.[18]

$$K(x,y) = (\gamma\, x.z + C)^d \quad (2)$$

γ is a scaling parameter that determines the influence of the inner product *x.z*. *d* is the degree of polynomial

RBF

It is one of the most popular kernels. RBF kernel measures similarity based on the radial distance between data points, using a gamma parameter that controls the influence of each data point.[19,20]

$$K(x,y) = EXP(-\gamma \|x-z\|^2) \quad (3)$$

Where $\gamma > 0$

Each kernel has advantages and disadvantages, and the best choice often depends on the nature of the input data.[9]

Whereas much research has been directed toward machine learning in the detection of breast cancer, the literature indicates a lack of comparative studies that have analyzed and compared the performances of different SVM kernel functions on different datasets regarding classification. Therefore, the literature demonstrates the need to conduct a systematic study of different SVM kernels by using publicly available datasets related to breast cancer. In this respect, therefore, the present work meets this gap in the literature by comparing the different types of SVM kernels in-depth, as much as possible, and intends to determine which kernel best fits the purpose of detecting and classifying breast cancer, including the discussion of advantages and disadvantages associated with each kernel.

Dataset Description

The performance of the various kernel functions in SVM was compared using the open Wisconsin Breast

Cancer (WBC) dataset from the UCI Machine Learning Repository.[21] This dataset contains 569 instances, and each instance is characterized by 30 features (variables) computed from digitized images of the fine needle aspiration. These features describe the cell nuclei's properties in the image, including radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension, with measurements for mean, standard error, and worst values. The dataset is binary, with 357 instances labeled benign and 212 instances labeled malignant.

Data Pre-processing

Before training the SVM classifiers, several pre-processing steps were performed on the dataset. First, an inspection for missing values revealed that none were present in this specific dataset. This guarantees that all features contribute equally to the classification process and helps avoid potential numerical instability issues. In addition, the processed data was divided into a training set (70%) and a testing set (30%) aligning with common practices to balance model training and evaluation.

Evaluation Metrics

The confusion matrix, Accuracy, Sensitivity, and Specificity are key metrics used for evaluating classification models. Each one exhibits distinct strengths and shortcomings depending on the specific problem, particularly with respect to class imbalance.

Confusion matrix

It is a metric that illustrates the performance of the classification algorithm based on the data contained within it. For a binary classification, it can be expressed in table 1.

Accuracy

Accuracy measures how many predictions were made correctly regarding all cases predicted. Accuracy means all positive and negative results were true. It is simple to understand and compute and work well when the classes are balanced.[22]

$$Accuracy = \frac{Nmber\ of\ correct\ predictions}{Total\ Number\ of\ Predictions} = \frac{TP + TN}{TP + TN + FN + FP} \quad (4)$$

Sensitivity

The sensitivity (Recall) is the degree of effectiveness in a classification algorithm that classifies data points into a positive class.[23]

$$\text{Sensitivity(Recall)} = \frac{TP}{TP + FN} \quad (5)$$

Specificity

This is the measure of a degree of effectiveness in a classification approach in classifying data into the negative class.[24]

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (6)$$

$$F_\beta = \frac{\left(1 + \beta^2\right) * Precision * Recall}{\left(\beta^2 * Precision\right) + Recall} \ , \beta = 1,2,3 \quad (7)$$

F-scores

The F1, F2, and F3 scores are obtained by setting β to 1, 2, and 3, and taking a positive real number. F1-Score matric combines two different metrics (precision and recall). These metrics are the most used performance metrics in medical research, which involve diagnosis and classification.[25]

Area under the curve (AUC)

The AUC is a vital metric of performance for any classification models. The higher the AUC, the better the model.[25,26]

SVM Implementation

The SVM classifiers were implemented in MATLAB R2022a, using the (fitcsvm) function for appropriate corresponding types, namely Linear Kernel, Polynomial Kernel, and RBF Kernel, and 5-folds cross-validation for tuning the hyperparameters so that the generalizability of the model could be checked, and overfitting was reduced as far as possible.

RESULTS AND DISCUSSION

This section is followed by comparing the performance of three SVMs: Linear, polynomial, and RBFs. A confusion matrix was then created to evaluate the performance of the models, which allowed several parameters to be calculated, such as accuracy, precision, recall, specificity, and F-scores, such as the F1, F2, and F3 scores. Furthermore, AUC can also be used for the measure in the comparison of classifiers since high AUC will indicate better performance in all thresholds of possible decisions. These metrics provide a comprehensive evaluation of the SVM models, with a focus on recall and AUC being especially significant in the medical domain.

Linear Kernel

Table 2 presents a summary of the test results using the classification model. Out of the (212) positives, (203) were, in fact, positive, while (9) were false negatives; likewise, out of (357) negatives, (345) were true negatives, and the remaining (12) were false positives. Further computation on the accuracy, sensitivity, and specificity evaluation measures could be done as per the following:

$$Accuracy = \frac{203 + 345}{203 + 345 + 9 + 12} = 0.9630$$

Table 1: Confusion matrix for binary classes (negative and positive)

Predicted\Actual	Positive	Negative
Positive	True positive (TP)	False positive (FP)
Negative	False negative (FN)	True negative (TN)

Table 2: Confusion matrix for linear kernel model

Predicted\Actual	Positive	Negative
Positive	203	9
Negative	12	345

$$\text{Sensitivity}\left(\text{Recall}\right)=\frac{203}{203+9}=0.9575$$

$$\text{Specificity}\left(\text{True nagtive rate}\right)=\frac{345}{345+12}=0.9663$$

Polynomial Kernel

Table 3 shows that out of (312) positive class data points, (202) were correctly identified as true positives and (10) were false negatives. Of the (357) negative class data, (349) were true negatives, while (8) were false positives. The calculated values for metrics are as follows:

$$Accuracy=\frac{202+349}{202+349+10+8}=0.9683$$

$$\text{Sensitivity}\left(\text{Recall}\right)=\frac{202}{202+10}=0.9528$$

$$\text{Specificity}\left(\text{True nagtive rate}\right)=\frac{349}{349+8}=0.9775$$

RBF Kernel

Table 4 shows that out of (212) positive classes, (192) of them were predicted in the positive class, which is called true positive, and the left of them (20) of the data got negative that called False Negative. Furthermore, from (357) negative class test data, True Negative is (346), and False Positive is (11). The following are the performance metrics calculated in this regard, which include accuracy, sensitivity, and specificity.

$$Accuracy=\frac{192+346}{192+20+11+346}=0.9455$$

$$\text{Sensitivity}\left(\text{Recall}\right)=\frac{192}{192+20}=0.9056$$

$$\text{Specificity}\left(\text{True nagtive rate}\right)=\frac{346}{346+11}=0.9691$$

Table 5 presents a direct comparison of the three SVM kernel models using key performance metrics. The Polynomial kernel model demonstrates superior overall performance with the highest accuracy at 0.9684, precision at 0.9619, and specificity at 0.9776. While the Linear kernel achieves a slightly higher recall of 0.9575 compared to the Polynomial's 0.9528, the latter demonstrates a more balanced performance. The polynomial model also has the highest F1-score of 0.9573, a better way of measuring the balance between precision and recall. The RBF kernel lags in most metrics, including the lowest precision at 0.9458, recall at 0.9057, and significantly lower f-scores, where the highest score it achieves is the F1 score of 0.9253. Crucially, the Polynomial kernel model achieves the highest AUC at 0.9966, followed by linear with 0.9955 and RBF with 0.9818, indicating its superior capability in distinguishing between malignant and benign cases overall, solidifying its position as the best-performing model among the three in the table.

Table 3: Confusion matrix for polynomial kernel model

Predicted\Actual	Positive	Negative
Positive	202	10
Negative	8	349

Table 4: Confusion matrix for the matrix of RBF model

Predicted\Actual	Positive	Negative
Positive	192	20
Negative	11	346

As shown in the graphs from Figure 1, the Polynomial Kernel model displays a receiver operating characteristic curve with an AUC of 0.9966, the highest among the three models, indicating superior discriminatory ability. The Polynomial Kernel's curve is closer to the ideal top left corner compared to the curves of the other two models. Next, the Linear Kernel also demonstrates strong performance with an AUC of 0.9859. Its curve is slightly below the Polynomial model's, showing excellent discriminatory power, though not quite as good as the Polynomial Kernel. Finally, the RBF Kernel has the lowest AUC at 0.9818 (rounded from 0.98175), and its curve is visibly further from the top left corner, indicating comparatively lower discriminatory capabilities.

In all the graphs, the dotted diagonal line represents the performance of a random classifier with an AUC of 0.5. Since all models have curves significantly above this line, it suggests that all three models are notably better than a random classifier. Overall, Figure 1 visually confirms the numerical results in Table 4, reinforcing the conclusion that the Polynomial Kernel is the best-performing model for breast cancer classification.

DISCUSSION

This study offers a comparative analysis of three widely used SVM kernel functions: Linear, polynomial, and RBF, specifically for the classification of breast cancer using the WBC dataset. We assessed each kernel's performance using various evaluation metrics, including accuracy, precision, recall, specificity, F-scores, and AUC.

The results are quite consistent: the polynomial kernel surpasses both the linear and RBF kernels in all respects, particularly regarding overall accuracy, specificity, and most impressively-AUC. Such results complement other works in which the efficiency of polynomial kernels on complex classification problems was underlined. In fact, in one of

Table 5: Performance metric comparison

Metric	Polynomial kernel	RBF kernel	Linear kernel
Accuracy	0.9684	0.9455	0.9630
Precision	0.9619	0.9458	0.9442
Recall	0.9528	0.9057	0.9575
Specificity	0.9776	0.9692	0.9663
F1 score	0.9573	0.9253	0.9508
F2 score	0.9546	0.9134	0.9548
F3 score	0.9537	0.9095	0.9562
AUC	0.9966	0.9818	0.9859

AUC: Area under the curve, RBF: Radial basis function

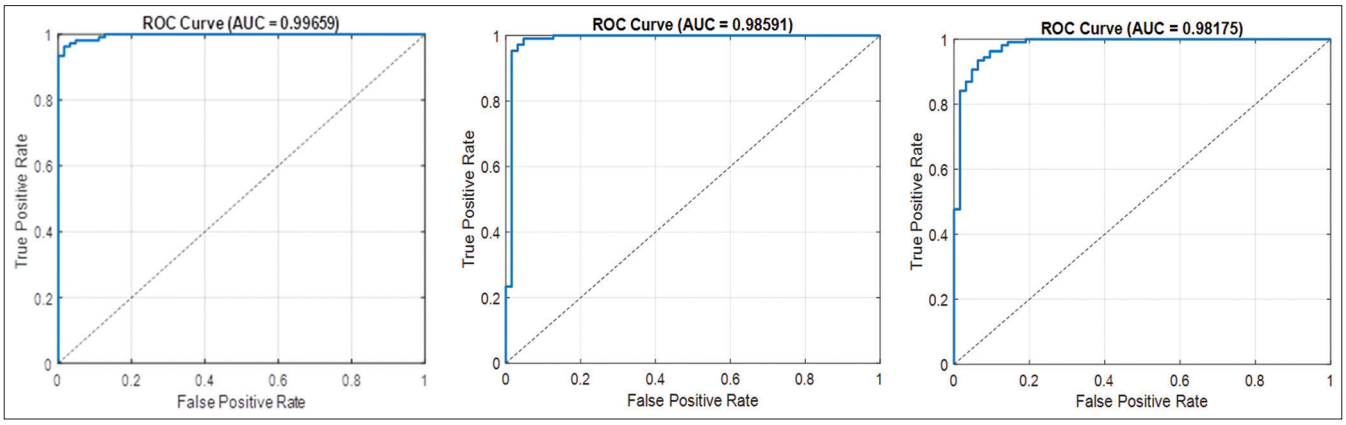

Figure 1: Area under the curve for each Kernel function of (polynomial, linear, and radial basis function)

the studies, similar results were obtained, representing the polynomial kernel as one of the best for classification tasks, mainly due to its non-linear connections' detectability within data.[5] Furthermore, many researchers have proved that even though the RBF kernel may be a suitable choice for most of the intricate data, it always requires appropriate parameter tuning.[20,27] Our case reflects that the RBF kernel was not able to achieve similar performances attained by the polynomial kernel.

It is justified that this kernel is performing better because the polynomial kernel maps the input data into higher-dimensional space using a polynomial function and hence grasps more complex relations within the data.[23] Increased flexibility becomes important in modeling the intricacy of medical datasets for which classes cannot always be separated by a linear boundary.

The linear kernel also performed very well with an AUC of 0.9859 and Sensitivity, proving that there was a certain degree of linearity inside the dataset. This is important because linear kernels are computationally simpler and sometimes can give decent results with far lower computational costs. However, our analysis suggests that due to its capability of capturing more intricate patterns, the polynomial kernel achieves higher accuracy and specificity. Although effective for many problems, the RBF kernel exhibited considerably lower performance on this classification problem.

Lower AUC and lower precision, recall, and F-scores of the RBF kernel indicate poor generalization of the Radial Distance-Based similarity measure for the chosen dataset. This may be due to either the nature of the dataset or some problems in RBF hyperparameters, indicating that kernel selection needs to be specific to the nature of a dataset. This is most significant in medical diagnosis, in which early detection is critical. A higher AUC for the polynomial kernel means increased chances of successfully differentiating between malignant and benign cases and fewer false negatives. Sensitivity is of utmost importance for better patient outcomes, reducing the chances of late intervention, and is very significant for kernel selection in medical contexts.[22]

Although this work presents a critical study of the SVM kernel functions using the given dataset, one should also ensure that it is not without its set of limitations. The model does not fully generalize to other datasets on breast cancer, each of which may have variable features and dimensions. Further research work must exercise this comparative analysis with other datasets and variations in parameters to assess the generalizability of the polynomial kernel overall in different scenarios of cancer detection and robustly develop our conclusion.

Further work could be extended to other types of kernels in SVM or other machine-learning approaches to see whether these produce superior results. The conclusion of this research really proved that kernel function selection is very important in any SVM-based medical diagnosis system. So far, the polynomial kernel has resulted in the best performance in classifying breast cancer using the WBC dataset and thus serves as a very strong tool toward more effective and reliable model development with an emphasis on higher AUC and increased Sensitivity.

CONCLUSION

This study certainly establishes the supremacy of the polynomial kernel function of SVM for the classification of breast cancer using a WBCD dataset. Upon rigorous evaluation, a polynomial kernel function demonstrated the highest accuracy with specificity and AUC, with a significant lead over linear and RBF kernel functions.

These results point out the importance of a proper kernel function in the development of the SVM-based model for medical diagnosis. This can be interpreted that the polynomial kernel captured non-linear patterns with complexity, therefore finer discrimination of benign and malignant tumors was allowed. The higher AUC and specificity in the medical context mean the reduction of the number of false negatives for early and correct diagnosis.

These study findings also provide the basis for the optimum performance of SVM models in breast cancer datasets. It is important to contribute to the formation of a basis upon which future studies can be conducted based on these results with other datasets of breast cancer using different machine-learning methods. In general, the study postulates that a polynomial kernel represents an important step forward in machine learning toward developing effective models for the detection of breast cancer with the ultimate goal of improving both the care as well as the outcomes of the patients.

REFERENCES

1. A. Roheel, A. Khan, F. Anwar, Z. Akbar, M. F. Akhtar, M. I. Khan, M. F. Sohail and R. Ahmad. Global epidemiology of breast cancer based on risk factors: A systematic review. *Frontiers in Oncology*, vol. 13, p. 1240098, 2023.
2. K. Hashimoto, S. Nishimura, T. Ito, N. Oka and M. Akagi. Limitations and usefulness of biopsy techniques for the diagnosis of metastatic bone and soft tissue tumors. *Annals of Medicine and Surgery*, vol. 68, p. 102581, 2021.
3. A. Schmidt, P. Morales-Alvarez and R. Molina. Probabilistic Modeling of Inter-and Intra-observer Variability in Medical Image Segmentation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 21097-21106, 2023.
4. M. M. Ahsan, S. A. Luna and Z. Siddique. Machine-learning-based disease diagnosis: A comprehensive review. *Healthcare* (*Basel*), vol. 10, p. 541, 2022.
5. D. Virmani and H. Pande. Comparative Analysis on Effect of Different SVM Kernel Functions for Classification. In: *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2022*. Vol. 3. Springer, Cham, pp. 657-670, 2022.
6. L. J. Grimm and M. A. Mazurowski. Breast cancer radiogenomics: Current status and future directions. *Academic Radiology,* vol. 27, no. 1, pp. 39-46, 2020.
7. A. S. Rana, J. Rafique and H. Riffat. *Advances in Breast Ultrasound Imaging: Enhancing Diagnostic Precision and Clinical Utility*. IntechOpen, London, 2024.
8. J. H. Catani, R. Matsumoto, F. Horigome, T. Tucunduva, M. Costenaro and N. de Barros. A Pictorial Review of Breast Biopsy Complications. In: *2017: European Congress of Radiology-ECR*, 2017.
9. J. Chen, Z. Gandomkar and W. M. Reed. Investigating the impact of cognitive biases in radiologists' image interpretation: A scoping review. *European Journal of Radiology*, vol. 166, p. 111013, 2023.
10. R. M. Al-Tam and S. M. Narangale. Breast cancer detection and diagnosis using machine learning: A survey. *Journal of Scientific Research,* vol. 65, no. 5, pp. 265-285, 2021.
11. A. B. Yusuf, R. M. Dima and S. K. Aina. Optimized breast cancer classification using feature selection and outliers detection. *Journal of the Nigerian Society of Physical Sciences*, vol. 3, pp. 298-307, 2021.
12. D. Valkenborg, A. J. Rousseau, M. Geubbelmans and T. Burzykowski. Support vector machines. *American Journal of Orthodontics and Dentofacial Orthopedics*, vol. 164, no. 5, pp. 754-757, 2023.
13. M. W. Huang, C. W. Chen, W. C. Lin, S. W. Ke and C. F. Tsai. SVM and SVM ensembles in breast cancer prediction. *PLoS One*, vol. 12, no. 1, p. e0161501, 2017.
14. S. H. Hasanah. Classification support vector machine in breast cancer patients. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 16, no. 1, pp. 129-136, 2022.
15. U. A. Contardi, P. R. Scalassara and D. V. Thomaz. Benign and malign breast cancer classification using support vector machines optimised with particle swarm and genetic algorithms. *Learning and Non-Linear Models*, vol. 20, no. 2, pp. 21-23, 2022.
16. A. Tharwat. Parameter investigation of support vector machine classifier with kernel functions. *Knowledge and Information Systems,* vol. 61, pp. 1269-1302, 2019.
17. G. Elefante, W. Erb, F. Marchetti, E. Perracchione, D. Poggiali and G. Santin. Interpolation with the polynomial kernels. *arXiv preprint arXiv:2212.07658,* 2022.
18. L. Muflikhah and D. J. Haryanto. High performance of polynomial kernel at SVM Algorithm for sentiment analysis. *Journal of Information Technology and Computer Science,* vol. 3, no. 2, pp. 194-201, 2018.
19. S. P. Rana, M. Dey, R. Loretoni, M. Duranti, L. Sani, A. Vispa, M. Ghavami, S. Dudley and G. Tiberi. Radial basis function for breast lesion detection from MammoWave clinical data. *Diagnostics,* vol. 11, no. 10, p. 1930, 2021.
20. X. Ding, J. Liu, F. Yang and J. Cao. Random radial basis function kernel-based support vector machine. *Journal of the Franklin Institute*, vol. 358, no. 18, pp. 10121-10140, 2021.
21. W. Wolberg, O. Mangasarian, N. Street and W. Street. *Breast Cancer Wisconsin* (*Diagnostic*). UCI Machine Learning Repository, California, 1993.
22. D. Valero-Carreras, J. Alcaraz and M. Landete. Comparing two SVM models through different metrics based on the confusion matrix. *Computers & Operations Research,* vol. 152, p. 106131, 2023.
23. J. Miao and W. Zhu. Precision-recall curve (PRC) classification trees. *Evolutionary Intelligence*, vol. 15, no. 3, pp. 1545-1569, 2022.
24. A. Kovács, P. Palásti, D. Veréb, B. Bozsik, A. Palkó and Z. T. Kincses. The sensitivity and specificity of chest CT in the diagnosis of COVID-19. *European Radiology*, vol. 31, pp. 2819-2824, 2021.
25. A. S. Assiri, S. Nazir and S. A. Velastin. Breast tumor classification using an ensemble machine learning method. *Journal of Imaging*, vol. 6, no. 6, p. 39, 2020.
26. N. H. Mahmood and D. H. Kadir. An elastic net approach to logistic regression for genetic selection in high-dimensional brain cancer data. *Cihan University-Erbil Scientific Journal,* vol. 9, no. 1, pp. 14-23, 2025.
27. M. S. Reza, U. Hafsha, R. Amin, R. Yasmin and S. Ruhi. Improving SVM performance for type II diabetes prediction with an improved non-linear kernel: Insights from the PIMA dataset. *Computer Methods and Programs in Biomedicine Update,* vol. 4, p. 100118, 2023.