International Journal of Basic and Applied Sciences, 14 (SI-2) (2025) 169-174

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/mg38r469
Research paper

A Soft Computing-Based Intelligent Framework for Enhanced Breast Cancer Detection

Charu Sharma , Dr. Kavita Gupta

University Institute of Computing, Chandigarh University, Mohali, India
**Corresponding author E-mail:Charusharma2009@rediffmail.com*

Received: July 4, 2025, Accepted: August 1, 2025, Published: August 12, 2025

Abstract

Breast cancer continues to be a critical health concern globally, and patient prognoses are markedly improved with early identification. Deep learning techniques have emerged as excellent instruments for the analysis of medical pictures, and their application in breast cancer detection has the potential to enhance analytical accuracy. This research introduces a unique methodology called mutual information filter force-driven NeuroConvolveX (MIFF-NCX) for feature selection aimed at improving early-stage breast cancer identification. Proposed methodologies integrate the advantages of deep learning in extracting intricate features from medical datasets with feature selection techniques that emphasize the most valuable attributes. We utilize the original Breast Cancer dataset to assess proposed methodologies and data analysis.
In terms of measures like F-score, sensitivity, ROC curve, specificity, and accuracy curves, the presentation of the suggested models is assessed and contrasted with existing techniques. The findings of this study show that recommended techniques significantly increase the efficacy and efficiency of deep-learning models for identifying breast cancer in its early stages.

***Keywords**: Mutual information filter force-driven NeuroConvolveX (MIFF-NCX), BC Dataset, Early detection dataset, deep learning algorithm*

1. Introduction

The accurate and early breast cancer detection continues to be essential challenge in medical diagnostics. Leveraging advanced computational methods has shown promising potential in enhancing both the efficiency and reliability of diagnosis. In this study, we introduce new soft computing framework developed for improving detection and classification of BC (breast cancer) through the integration of intelligent algorithms and feature optimization techniques [1]. This section discusses performance outcomes of our proposed system [2], compares it to existing approaches, and provides insights into the strengths and limitations based on empirical evaluation. The results are analyzed using standard performance metrics, like sensitivity, accuracy, precision, specificity, F1-score, for ensuring comprehensive evaluation.
Literature Review
By effectively evaluating medical pictures and identifying problematic locations, DL (deep learning) models and feature selection algorithms could enhance early BC diagnosis [3]. Improving treatment results and survival rates requires early identification. Better research, screening programs, treatment planning, and increased diagnostic accuracy are just a few benefits of precision medicine. Overall, the review's conclusions demonstrate that advanced learning algorithms and feature selection may change the way BC is detected and treated, resulting in better patient outcomes, better efficient resource utilization, and advancements in clinical practice and research. By applying emerging technologies like PCA and logistic regression to BC diagnosis, medical professionals could enhance their diagnostic skills [4]. Technology integration could lead to more advanced and precise diagnostic techniques.
The use of AI (artificial intelligence) and ML (machine learning) algorithms for BC diagnosis is covered in the paper [5]. By overcoming the drawbacks of conventional diagnosis techniques, such as their low identification efficiency and expensive procedures, these algorithms are essential for the quick and accurate detection of breast cancerous tumors.
The study specifically suggests a method that combines principal component analysis (PCA) with random forest (RF) to diagnose BC quickly and accurately [6]. To reduce diagnostic time while preserving high identification accuracy, the approach comprises feature collection and characteristic selection. Categorical data from malignant breasts is handled by filtering and prediction algorithms, which also choose pertinent attributes, derive features, and build classification models [7]. By employing these algorithms, the study hopes to increase efficacy as well as accuracy of BC diagnosis while also establishing foundation for intelligent BC detection.
To sum up, AI, ML algorithms are used in improving speed alongside accuracy of BC diagnosis, potentially resolving the problems with conventional diagnostic techniques. In addition to using PCA for feature reduction along with an uncomplicated MLP, the work illustrated the usefulness of transfer learning model as per ViT [8]. Comparable to advances in BC imaging classification, the results demonstrated average precision, accuracy, recall, and F1-score of 98% for DSMM dataset alongside 95% for most datasets. To diagnose cancer, the

authors experimented with a number of data mining and DL techniques [9]. They observed that the bulk of the investigations exclusively employed genetics.

According to [10], studies advance the field by simultaneously learning genetic sequencing and imaging to predict BC and get additional information to enable early identification and treatment. Researchers who wish to conduct studies in this area can also get advice from us. The authors of [11] looked into recent studies that treated BC by combining DL with many imaging modalities. In order to provide a thorough analysis of the research on BC imaging using DLR-based CAD, the author concentrated on three DL frameworks for breast imaging modalities. The findings showed that NIC algorithms were useful for identifying many types of cancer, with directed ABC and neural network technologies being used in the identification of BC. [12]

From the start of 2012 until the conclusion of 2018, the King Abdul [13] Aziz University Hospital was examined in retrospect. 740 cases of BCs were divided into four major molecular subgroups using immunohistochemistry: triple negative, HER2-positive, lumen A, and lumen B. The chi-square test was employed to assess the relationship between these groupings and clinical and pathological characteristics. The most prevalent subtypes were triple-negative (16%), HER2-positive (11.5%), luminal A (58.5%), and luminal B (14%).

The tumor's average size at diagnosis was 3.2 cm, and the patient's normal age was 49 years. The majority of the time, 85% of cases were pancreatic, and 11.4% were lobular. The lymphatic nodes under the arms were affected in 66.6% of cases. Although the majority of globulin tumors (77%) were of the triple-negative and luminal A tumor subtypes, unilateral resection was used to modify 69.5% of the patients. Carcinoma-in-situ was less common in HER2-positive tumors. Furthermore, a strong association between the axillary lymph node condition and the molecular subtypes was found.

2. Mutual information filter orce-driven NeuroConvolveX (MIFF-NCX)

In this research, a novel framework for the early detection and classification of BC, namely MIFF-NCX, is proposed. To improve image quality and draw attention to significant characteristics, the method combines sophisticated data preparation techniques like Histogram Equalization with feature extraction techniques like PCA and LDA. To effectively identify the most useful features, it uses a specific feature selection technique called MIFF-NCX [14], which combines neural networks and statistical relationship preservation. When compared to current methods, performance evaluations of the methodology's testing on real-world datasets revealed notable gains in accuracy, sensitivity, specificity, and other parameters. All things considered, the study shows that the suggested strategy can successfully improve early BC detection, which could improve patient outcomes.

To improve early-stage BC detection, the study presents a unique method dubbed Mutual Information Filter Force-driven NeuroConvolveX (MIFF-NCX). By optimizing feature selection, this system significantly increases the efficiency and accuracy of diagnosis. MIFF-NCX improves existing BC diagnosis techniques by fusing DL with efficient statistical feature preservation. The study opens the door for the use of this framework in clinical settings by showing how it can enhance early detection results. To further improve diagnostic performance, future research may investigate hybrid models that incorporate DL approaches. According to the report, even though histopathology image classification has advanced significantly, model validation can still be strengthened to increase dependability and clinical application. Here are some specific strategies to improve validation: 1. Using Larger and Diverse Datasets: Validation should use large, diverse, and well-annotated datasets to guarantee that models generalize well across various populations and imaging settings. [3]2. Utilizing Cross-Validation Techniques: Using techniques like k-fold cross-validation can lessen overfitting and yield more reliable model performance estimates. Validation from the outside, which involves testing models on separate external datasets that weren't utilized for training or preliminary validation. With this method, the model's performance is evaluated in various clinical settings and real-world situations. [4]. 4.Selection of Multiple Achievement Measurements such as In imbalanced data sets, relying only on accuracy might be deceptive. The performance of the model is more thoroughly assessed when metrics like precision, recall, F1-score, ROC-AUC, and specificity are included. 5. Handling variations and Quality of Data: Model robustness can be increased by making sure that data preprocessing takes staining, imaging, and tissue preparation variations into account. 6.Model Readability and explication: Applying explainability strategies can help with clinical trust, validation by medical professionals, and comprehension of model decisions. [5]

To sum up, adding these methods to validation processes can make models more robust, reliable, and useful for diagnosing breast cancer in the clinic.

By using mutual information to rank the most informative features pertinent to breast cancer detection, the MIFF-NCX (Mutual Information Filter Force-driven NeuroConvolveX) approach improves feature selection. By measuring the shared information between features and the target class, MIFF-NCX effectively reduces uncertainty and focuses on the features that offer the highest discriminative power, in contrast to traditional techniques that might choose features only based on statistical significance or less dynamic criteria. By ensuring that only the most pertinent characteristics are kept, this method raises the precision and effectiveness of early-stage breast cancer detection models. [5]

For clinical application, the MIFF-NCX method's processing time and computing complexity are crucial factors. Even though the method uses sophisticated methods like convolutional operations, mutual information calculation, and feature extraction techniques like PCA and LDA, these procedures can be computationally demanding, particularly when dealing with big or high-dimensional datasets. In clinical situations where prompt decision-making is crucial, efficient algorithm implementation and optimization are required to minimize processing times and guarantee fast results. Although thorough testing of processing times in real-world circumstances would be necessary to prove its acceptability for normal clinical use, the study implies that MIFF-NCX can be made practical for real-world applications by utilizing enhanced training processes and possibly hardware acceleration. [7] [6]

In combination, the NeuroConvolveX (NCX) and the Mutual Information Filter (MIFF)-based feature selection component improve the precision and effectiveness of breast cancer diagnosis. By assessing statistical correlations between features, the MIFF method reduces dimensionality while maintaining pertinent information to identify the most useful features. NeuroConvolveX, a neural network with shared weights intended to effectively learn complicated representations from this optimized feature set, is then fed these chosen features. The distilled characteristics are used by NeuroConvolveX, a multi-layer architecture, to model complex patterns and increase classification accuracy. When combined, the mutual information-driven feature selection lessens processing burden and enhances the neural network's ability to concentrate on important features, producing more accurate and dependable detection results. [10]

2.1 Result

Early detection of BC remains pivotal in saving lives and improving outcomes for affected individuals. Through regular screening programs, advancements in imaging technologies, and increased awareness of risk factors and symptoms, healthcare systems can identify BC at its earliest stages, enabling prompt intervention and treatment. We proposed a new method called mutual information filter force-driven NeuroConvolveX (MIFF-NCX) to improve the early detection of BC. An Intel Core i7 processor with 40 GB of RAM powers MIFF-NCX. The proposed approach has been implemented using Python and trained on an NVIDIA GTX 1070 GPU.

The effectiveness of the proposed model in detecting early breast cancer is assessed by comparing it to existing methods using measures like accuracy, sensitivity, specificity, F1-score, Receiver Operating Characteristic (ROC) curves, and accuracy loss graphs.

A binary classification model's performance is displayed in Figure 1. It shows that the model correctly identified 3579 cases of class 1 (True Positives), accurately recognized 3527 cases of class 0 (True Negatives), mistakenly predicted class 0 for 362 cases that were class 1 (False Negatives), and incorrectly predicted class 1 for 532 cases that were actually class 0 (False Positives). The matrix offers a thorough examination of the model's accuracy while helping in determining the extent to which it can categorize both classes.

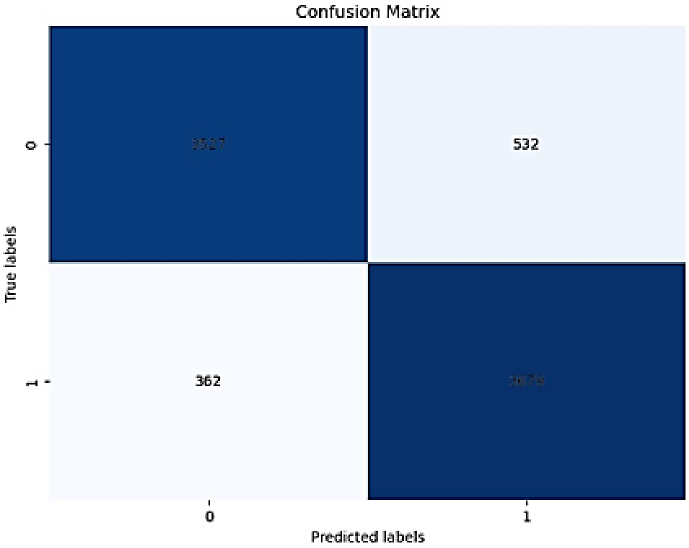


Fig. 1: Confusion Matrix

To assess performance, I generated ROC curve and assessed AUC. It illustrates the sensitivity and 1-specificity interaction with various decision thresholds. Better discriminative capacity was demonstrated by the ROC curve produced by the proposed model. Model's ability to differentiate between positive and negative cases of BC has been evidenced by its area under ROC curve. The proposed model can operate as advantageous diagnostic tool for BC screening, predicated on improved performance. Figure 2 depicts the output result of AUC-ROC.

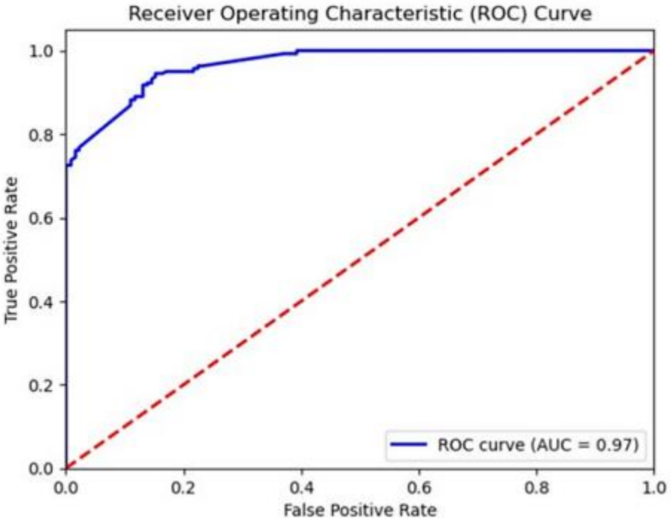
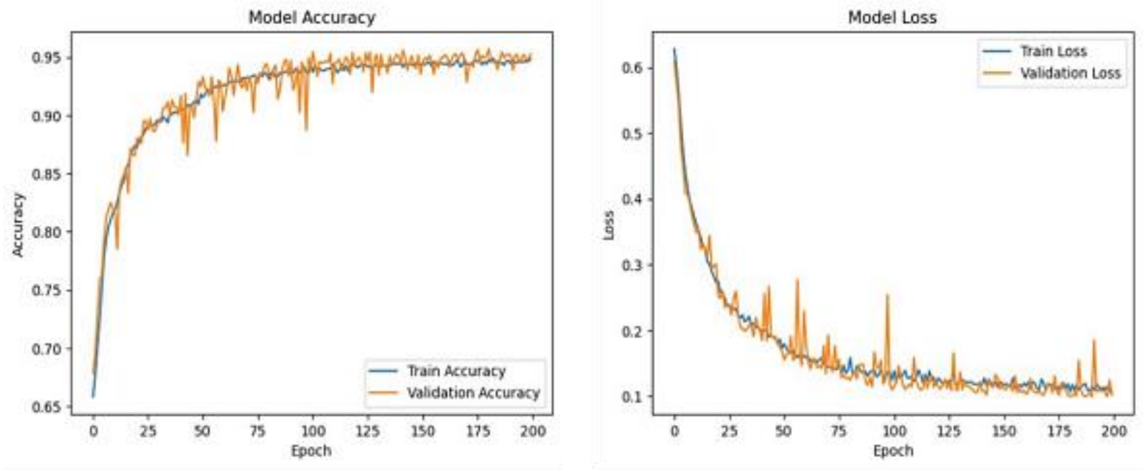


Fig. 2: The output result of AUC-ROC

Figure 3 compares performance on training and validation sets to show the model accuracy throughout 200 epochs. The blue line represents training accuracy, whereas the orange line denotes validation accuracy. As the model learns from the data, accuracy numbers initially increase rapidly, eventually stabilizing at 95%. Two curves' approximate alignment suggests that there is minimal overfitting and that the model is generalizing well. Model loss, which calculates the model's error, is presented over 200 epochs in Figure 3. An orange line indicates validation loss, and a blue line indicates training loss. With occasional spikes in validation loss, which indicate variations in model performance, both loss values gradually decline as the model gets better. The relatively low and stable loss at the end suggests that the model has learned well with little under fitting or overfitting.



(a) Accuracy (b) Loss

Fig. 3: The results of accuracy and loss

The existing methods for the metrics sensitivity and specificity are SVM [3], ANN [4], and DT [5]. ML approach effectiveness is evaluated based on several performance measure factors. The actual and anticipated data are transformed into TP, FP, TN, and FN to evaluate the parameters. The ramifications of the term are explained below.
TP represents true positivity.
TN represents true negativity.
FP represents a false positive.
FN represents false negatives.
These attributes are often utilized to evaluate different phrases by their corresponding formula, and they are used to evaluate the effectiveness of the research. These metrics can be utilized to evaluate a system's performance. <Sensitivity is the percentage of genuine positive instances that the test identified as true positives.> Sensitivity is represented by

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP + FN}} \tag{1}$$

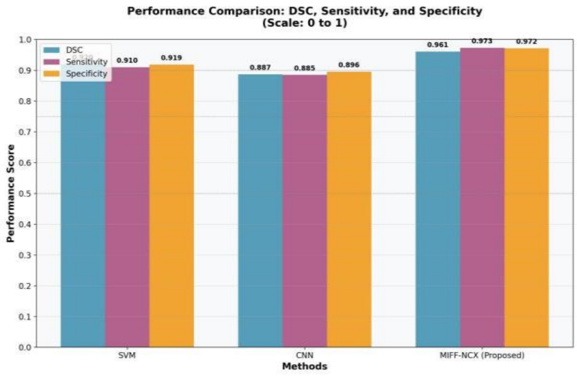


Fig. 4. Comparative Analysis of Sensitivity

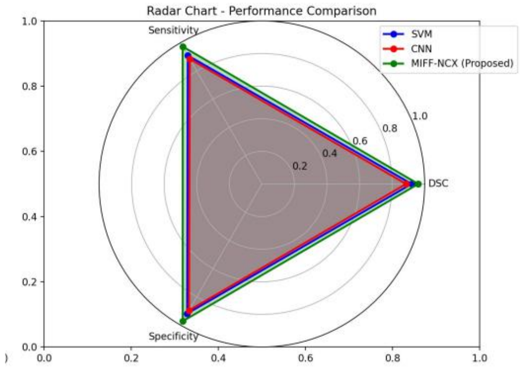


Fig. 4: Comparative Analysis of Sensitivity (Radar chart)

Sensitivity refers to the capacity to recognize affirmative situations with accuracy. With a sensitivity of 97%, our suggested model outperformed SVM, DSC, CNN which had corresponding sensitivity values of 91%, 88%, and 86% (Figure 4). The result validates the enhanced ability of our approach to detect a higher proportion of real positive cases.
Specificity, defined as the correlation between observed negative instances and all negative examples, is the rate at which BC's existence is projected to occur, including all cases. Figure 4 and figure 4(a) represent the specificity of the model and radar chart of the same.

With specificity of 97%, the suggested model demonstrated a significant capacity to accurately detect negative situations. Specificity is a measure of predictive accuracy. It depicts specificities of 91%, 88%, and 97% for SVM, CNN, and MIFF-NCX respectively.

Table 1: Depicts the outcomes of the study.Outcomes of Sensitivity and Specificity

Metrics	SVM	CNN	DSC	MIFF-NCX(Proposed)
Sensitivity (%)	92	94	92	97
Specificity (%)	91	88	86	97

The increased specificity of the model indicates a lower false positive rate, demonstrating its effectiveness in accurately categorizing negative cases.

For assessing accuracy, precision, recall, and F1-score, SVM [5], , and DSC [7] are assessed. A common metric in the classification or early detection of BC is accuracy. It is computed as the ratio of successfully predicted occurrences to all instances and assesses the overall accuracy of a model's predictions. The accuracy formula is

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP + FN + TN + FP}} \tag{3}$$

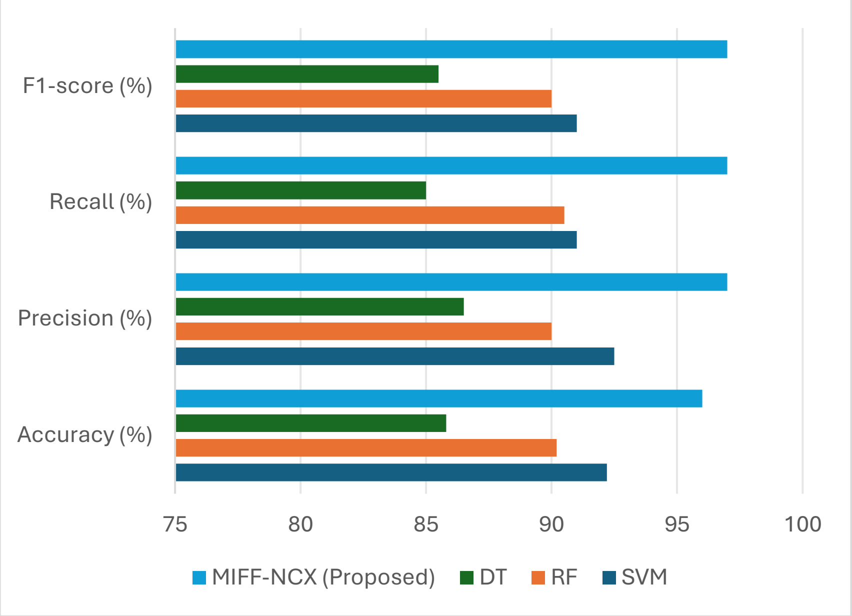


Fig. 5: Comparative Analysis of Accuracy, Precision, Recall, F1-Score

The accuracy of the proposed BC detection model has been compared to other models in the study. Results demonstrated that the proposed model's accuracy was higher at 97% compared with the accuracy of other models that include SVM, DSC, CNN at 92%, 89%, and 88%, respectively (Figure 5). This implies that the proposed model works better than existing methods in accurately categorizing cases of BC. The division of the cases that are positive among all the examples that are expected to be positive is called precision.

$$\text{Precision} = \frac{\text{TP}}{\text{TP + FP}} \tag{4}$$

The proposed BC detection approach outperformed previous models with a precision of 97%. Figure 5 shows the precision values of the models. The precision values for the existing methods, which comprise SVM,DSC and CNN, were 92.5%, 90%, and 86.5%, respectively, in comparison to the proposed method. This figure suggests that our model has a greater percentage of real positives among the cases it correctly predicted to be positive, highlighting its capacity to reduce false positives when compared to other models.

The ability of a model to accurately identify each positive event from all actual positive instances is crucial and is termed as recall. It is described as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}} \tag{5}$$

The proposed model's recall of 97% was discovered to be superior to that of SVM, DSC, CNN which had recall values of 91%, 90.5%, and 85%, respectively (Figure 6). This implies that, in comparison to other models, the proposed method is more sensitive to instances of BC because it is better at detecting a larger percentage of true positive cases.

The F1 score aggregates recall and accuracy into a single number. It is especially helpful when attempting to strike a balance between recall and accuracy. The F1 score is calculated as follows:

$$\text{F1 score} = \frac{\text{2TP}}{\text{2TP + +FP+ FN}} \tag{6}$$

The F1 score, a measure of recall and accuracy balanced at 97%, was maximized by the proposed strategy. On the other hand, Figure 5 shows that SVM, DSC, and CNN, respectively, achieved F1 scores of 91%, 90%, and 85.5%.

Table 2: Depicts research results. Outcome of the Accuracy, Precision, Recall, and F1-Score

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SVM	92.2	92.5	91	91
RF	90.2	90	90.5	90
DT	85.8	86.5	85	85.5
MIFF-NCX (Proposed)	96	97	97	97

This demonstrates that the proposed method exceeds previous methods in striking a better balance between accuracy and recall.

3. Conclusion

MIFF-NCX is intended to tackle the difficulties associated with detecting BC at an early stage by meticulously choosing the most relevant aspects from the intricate and multi-dimensional medical imaging data. It is also notable that MIFF has been incorporated into the model, as it makes it possible to identify the informative patterns that are essential for precise cancer diagnosis by capturing the interdependence between its features. The force-driven Neuro ConvolveX method further enhances this process by utilizing the concepts of force-directed graph architecture to data-driven demonstrate the most crucial features. When compared to existing techniques, MIFF-NCX shows several benefits. Conventional techniques could have trouble navigating the complex interactions seen in medical imaging data, which could result in a feature selection that is not ideal. MIFF-NCX provides a more advanced and subtle method. MIFF-NCX modifies the weight of variables based on their ability to contribute to overall prediction value of the model to account for complexities involved in early-stage BC diagnosis. In case of early-stage BC, MIFF-NCX approach is expected to significantly enhance capacity to differentiate between benign and malignant instances. It reduces the possibility of overfitting and improves the model's capacity to generalize to new data by concentrating on the most relevant elements. This is especially important for medical diagnostics, as precise and dependable forecasts are essential for providing patients with high-quality treatment.

In conclusion, future research should prioritize validation, interpretability, multi-modal data integration, and the creation of easily available, reliable algorithms to further enhance breast cancer diagnosis and treatment, even though present models show encouraging performance.

References

[1] Militello, C., L. Rundo, M. Dimarco, A. Orlando, V. Conti, R. Woitek, I. D'Angelo, T. V. Bartolotta, and G. Russo. (2022). Semi automated and interactive segmentation of contrast enhancing masses on breast DCE MRI using spatial fuzzy clustering. Biomedical Signal Processing and Control 71: 103113.

[2] Sharma, R., J. B. Sharma, R. Maheshwari, and D. Baleanu. (2021). Early anomaly prediction in breast thermogram by hybrid model consisting of superpixel segmentation, sparse feature descriptors and extreme learning machine classifier. Biomedical Signal Processing and Control 70: 103011.

[3] Charu, and K. Gupta. (2024). Systematic review of contemporary breast cancer detection techniques using machine learning. In Proceedings of the 7th International Conference on Contemporary Computing and Informatics (IC3I 2024), Greater Noida, India, 338–344. https://doi.org/10.1109/IC3I61595.2024.10829220.

[4] Gupta, K. K., Rituvijay, P. Pahadiya, and S. Saxena. (2022). Detection of cancer in breast thermograms using mathematical threshold based segmentation and morphology technique. International Journal of System Assurance Engineering and Management: 1–8.

[5] Chaudhury, S., A. N. Krishna, S. Gupta, K. S. Sankaran, S. Khan, K. Sau, A. Raghuvanshi, and F. Sammy. (2022). Effective image processing and segmentation-based machine learning techniques for diagnosis of breast cancer. Computational and Mathematical Methods in Medicine 2022.

[6] Gogoi, U. R., G. Majumdar, M. K. Bhowmik, and A. K. Ghosh. (2019). Evaluating the efficiency of infrared breast thermography for early breast cancer risk prediction in asymptomatic population. Infrared Physics & Technology 99: 201–211.

[7] Selle, J. J., M. Ulaganathan, A. Pranavi, and P. S. Rani. (2021). Automated ROI extraction and pattern classification of breast thermograms. In Proceedings of the 9th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO 2021), 1–5. IEEE.

[8] Hao, Y., L. Zhang, S. Qiao, Y. Bai, R. Cheng, H. Xue, Y. Hou, W. Zhang, and G. Zhang. (2022). Breast cancer histopathological images classification based on deep semantic features and gray level co-occurrence matrix. PLOS ONE 17(5): e0267955.

[9] Zheng, J., D. Lin, Z. Gao, S. Wang, M. He, and J. Fan. (2020). Deep learning assisted with efficient AdaBoost algorithm for breast cancer detection and early diagnosis. IEEE Access 8: 96946–96954.

[10] Zebari, D. A., D. A. Ibrahim, D. Q. Zeebaree, M. A. Mohammed, H. Haron, N. A. Zebari, R. Damaševičius, and R. Maskeliūnas. (2021). Breast cancer detection using mammogram images with improved multi fractal dimension approach and feature fusion. Applied Sciences 11(24): 12122.

[11] Aristokli, N., I. Polycarpou, S. C. Themistocleous, D. Sophocleous, and I. Mamais. (2022). Comparison of the diagnostic performance of MRI, ultrasound, and mammography for detection of breast cancer based on tumor type, breast density, and patient history. Radiography 28(3): 848–856.

[12] Nguyen, S., D. Polat, P. Karbasi, D. Moser, L. Wang, K. Hulsey, M. C. Çobanoğlu, B. Dogan, and A. Montillo. (2020). Preoperative prediction of lymph node metastasis from clinical DCE MRI of the primary breast tumor using a 4D CNN. In Medical Image Computing and Computer Assisted Intervention (MICCAI 2020), Part II, 326–334. Springer.

[13] Piantadosi, G., M. Sansone, R. Fusco, and C. Sansone. (2020). Multi planar 3D breast segmentation in MRI via deep convolutional neural networks. Artificial Intelligence in Medicine 103: 101781.

[14] Vamvakas, A., D. Tsivaka, A. Logothetis, K. Vassiou, and I. Tsougos. (2022). Breast cancer classification on multiparametric MRI: Increased performance of boosting ensemble methods. Technology in Cancer Research & Treatment 21: 15330338221087828.

[15] Inglese, M., M. Ferrante, T. Boccato, A. Conti, C. A. Pistolese, O. C. Buonomo, R. M. D'Angelillo, and N. Toschi. (2023). Dynomics: A novel and promising approach for improved breast cancer prognosis prediction. Journal of Personalized Medicine 13(6): 1004.

[16] Gravina, M., E. Cordelli, D. Santucci, P. Soda, and C. Sansone. (2022). Evaluating tumour bounding options for deep learning based axillary lymph node metastasis prediction in breast cancer. In Proceedings of the 26th International Conference on Pattern Recognition (ICPR 2022), 4335–4342. IEEE.

[17] Jing, X., M. D. Dorrius, M. Wielema, P. E. Sijens, M. Oudkerk, and P. van Ooijen. (2022). Breast tumor identification in ultrafast MRI using temporal and spatial information. Cancers 14(8): 2042.

[18] Gutierrez Cardenas, J. (2024). Breast cancer classification through transfer learning with Vision Transformer, PCA, and machine learning models. International Journal of Advanced Computer Science and Applications 15(4): 1027. http://www.ijacsa.thesai.org.

[19] Gutierrez Cardenas, J. (2023). A comprehensive review on breast cancer detection, classification, and segmentation using deep learning. Archives of Computational Methods in Engineering 30(4): 1145–1168. https://doi.org/10.1007/s11831 023 09968 z.