# Improving Breast Cancer Detection in BUS Images Using Multimodal Deep Learning and Grad-CAM Fusion

Zeeshan Mubeen[1], Zulfiqar Ali[2], Rahmat Ullah[2](✉), Vishal Krisha Singh[2], and Muhammad Haroon Ahmad[1]

[1] Department of Computer Science, Riphah International University, Lahore, Pakistan
{zeeshan.mubeen,muhammad.haroon}@riphah.edu.pk

[2] School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK
{z.ali,rahmat.ullah,v.k.singh}@essex.ac.uk

**Abstract.** Breast cancer is the second leading cause of death among women, with new cases rising globally each year. Early detection can significantly reduce mortality risk. Mammography, the primary screening method, uses X-ray images to identify breast abnormalities. Recent advancements in deep learning have enhanced medical image processing. This article describes a hybrid convolutional neural network (CNN) method for mammography scan-based breast cancer diagnosis that uses minimum-redundancy maximum-relevance (mRMR). The study combined top-performing pre-trained CNN architectures—AlexNet, EfficientNet-B0, and Inception V3—from eight deep-learning models. Features derived from Gradient-weighted Class Activation Mapping (Grad-CAM) were merged with trained features. The mRMR technique optimized these features, which were then classified using SVM and KNN algorithms. The hybrid model achieved a 99.35% accuracy rate in detecting breast cancer with the SVM classifier on the Breast Ultrasound Images Dataset (BUSI). These results demonstrate the effectiveness of combining feature selection methods with CNN models for breast cancer classification.

**Keywords:** Deep Learning · Fusion · Breast Cancer · Diagnosis and Classification · Transfer Learning · Grad-CAM · Image Processing

## 1 Introduction

Breast cancer is the second most common cause of cancer-related mortality among women, characterized by abnormal cell development and dissemination within the breast [4]. It accounts for 12.5% of new cancer cases globally and continues to be the most common cancer in women, accounting for around 1.96 million new cases and 609,820 deaths annually in the U.S. in 2023 [11]. Tumors are classified as malignant or benign based on cell characteristics. Early detection, crucial for improving survival rates, utilizes imaging techniques like mammograms, ultrasound [4], CT scans, and X-rays [2].

Machine learning has shown efficacy in healthcare for disease diagnosis, although conventional methods require manual feature extraction, which is challenging and expertise-dependent. Deep learning models can autonomously extract features and adapt to input data [13]. This study presents a hybrid deep learning technique for ultrasonic scan-based breast cancer detection as depicted in Fig. 1. Features from eight pre-trained CNN architectures were classified using SVM and KNN algorithms. The best-performing architectures—AlexNet, EfficientNet-B0, and Inception V3—combined with Grad-CAM and mRMR optimization achieved a 99.35% accuracy rate on the BUSI dataset, simplifying diagnosis and reducing the need for specialized expertise.

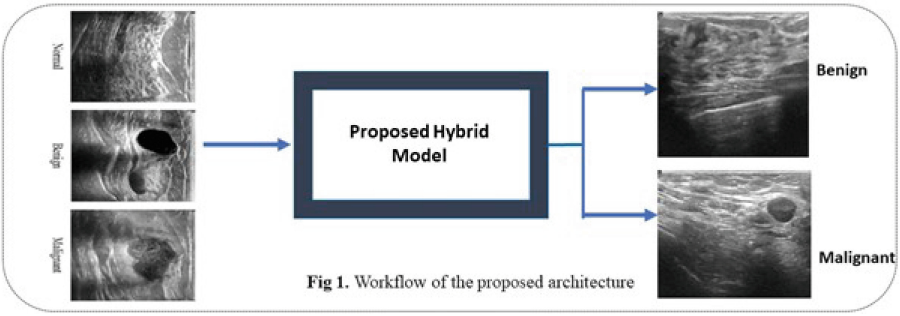



**Fig. 1.** Workflow of the proposed architecture.

### 1.1 Key Contributions

Key contributions of this study include:

- Proposing a deep CNN for classifying two types of breast cancer diagnoses.
- Demonstrating superior performance of the hybrid model using selected features.
- Eliminating manual feature extraction for radiologists with the hybrid CNN.
- Enhancing classification performance for breast cancer detection in MR images.
- Classifying 2-class breast tumor MR images with 8 CNN architectures from a widely used dataset.
- Combining feature maps from the top three pre-trained CNNs with Grad-CAM features to distinguish breast tumor characteristics.
- Using AlexNet, EfficientNet-B0, and Inception V3 to consolidate distinct features of the same image.
- Applying mRMR for dimension reduction to optimize model performance.
- Using two distinct ML classifiers to achieve accurate 2-class breast tumor diagnoses.

## 2 Related Work

Recent advancements in deep learning have profoundly impacted the analysis of medical imaging, particularly in detecting and classifying breast cancer. Notably, Chen et al. [3] implemented a sophisticated multi-feature extraction method to analyze breast ultrasound (BUS) images, achieving an impressive diagnostic accuracy of 96.91% with a substantial dataset of 1615 images. This approach underscores the capability of intricate feature extraction methods to enhance diagnostic accuracy in medical imaging.

Expanding on this foundation, Tanaka et al. [16] applied state-of-the-art convolutional neural networks (CNNs) such as VGG19 and ResNet152 to BUS images, which resulted in an Area Under the Curve (AUC) of 0.951. Their study exemplifies the potent applicability of CNNs in interpreting intricate imaging data, setting a benchmark for subsequent research endeavors.

Furthermore, a study integrating ResNet-101, ShuffleNet, and MobileNet-V2 showcased a collective enhancement in accuracy, achieving 96.54% on B-mode kidney ultrasound images [16]. This multi-architectural approach highlights the synergistic potential of employing diverse neural networks to capitalize on their unique strengths, thereby enhancing diagnostic precision.

In mammography, hybrid models have shown significant promise. R. Sathesh Raaj et al. [8] devised a hybrid CNN capable of classifying mammogram images with a notable accuracy of 97.91%. This demonstrates the efficacy of hybrid models in handling the varied nuances present in different imaging types and conditions. Additionally, Melekoodappattu et al. [6] combined the robustness of Extreme Learning Machines (ELM) with the optimization capabilities of the Fruitfly Optimization Algorithm, achieving a precision of 99.04%. This innovative approach points to the advantageous integration of bio-inspired algorithms to optimize the training processes of neural networks for intricate diagnostic tasks.

Further enhancing the field, Sharmin et al. [10] utilized ResNet50V2 alongside ensemble-based techniques to achieve a 95% accuracy on the Invasive Ductal Carcinoma (IDC) dataset, thus affirming the robustness of ensemble methods in enhancing predictive performance. Similarly, Srikantamurthy et al. [14] amalgamated CNNs with Long Short-Term Memory (LSTM) networks, achieving a remarkable accuracy of 99% in binary classification tasks. This integration of spatial and temporal feature extraction models enhances classification outcomes and sets a new standard in medical image analysis.

In this study, we have developed a novel hybrid model that integrates features from multiple architectures using Grad-CAM visualization and Minimum Redundancy Maximum Relevance (mRMR) optimization techniques. This model not only enhanced classification performance but also improved operational efficiency, achieving results that surpass those of previous studies. Our findings indicate that integrating multiple diagnostic approaches can substantially refine the accuracy and efficiency of breast cancer detection systems.

## 3 Materials and Methods

This section discusses the dataset, the Grad-CAM heatmap display, the proposed hybrid model, and the deep learning architectures utilized to extract features. It also describes how to select features using the mRMR approach.

### 3.1 Proposed Hybrid Model

The best-performing three deep learning models—AlexNet, EfficientNet-B0, and Inception V3—are used in this hybrid approach to extract features from ultrasound scans of breast cancer patients. These architectures generate different numbers of features: $780 \times 4097$ from AlexNet, $780 \times 1280$ from EfficientNet-B0, and $780 \times 2048$ from Inception V3. These features covering normal, benign, and malignant classes are concatenated to form a combined feature matrix of $780 \times 7224$.

Additionally, Grad-CAM visualization assists in extracting detailed tissue features from each breast cancer ultrasound image within these architectures. The mRMR feature selection method finds the most pertinent characteristics and hence minimizes the size of the feature matrix, condensing the total to 1000 while keeping the optimized features consistent. The selected feature matrix of $780 \times 1000$ is then classified using SVM and KNN classifiers. The proposed hybrid model's architecture is detailed in Fig. 2. This model classifies breast cancer ultrasound images into two levels using SVM and KNN classifiers, and the resulting images are presented for expert evaluation.

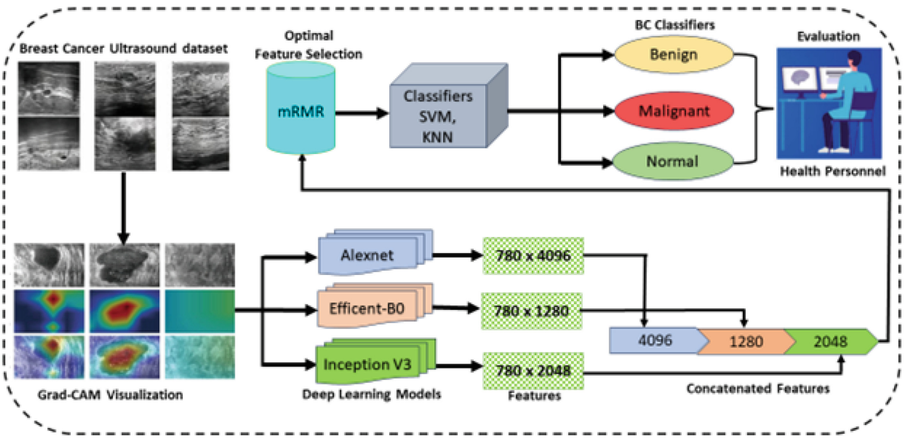


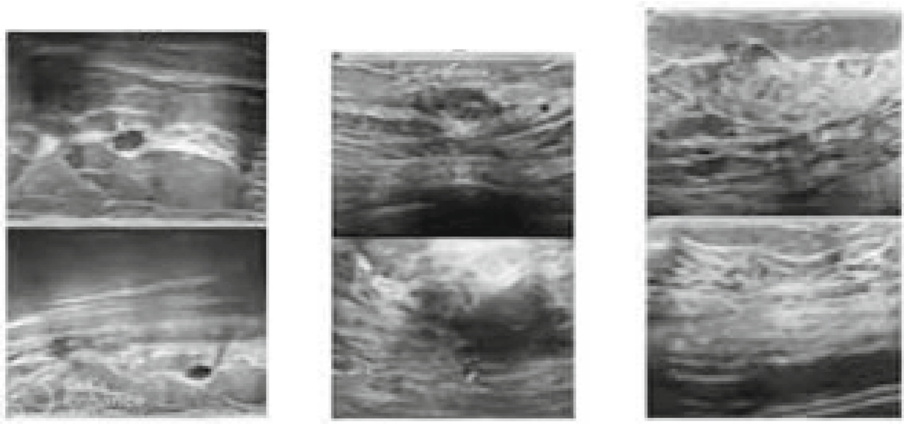

**Fig. 2.** Architecture of the proposed hybrid model.

### 3.2 Breast Ultrasound Dataset Description

The dataset includes breast ultrasound images from 600 women aged 25 to 75, which were collected in 2018. It comprises 780 images, each 500 by 500 pixels in PNG format paired with ground truth images. Three categories—normal, benign, and malignant—are applied to the photographs. In particular, there are 133 normal, 210 malignant, and 437 benign images. Table 1 shows the data distribution, and Fig. 3 displays examples from each category. The dataset is publicly available at [1].

**Table 1.** Distribution of Images with 80:20 Training and Testing Ratio

| Dataset | Benign | Malignant | Normal | Total |
|---|---|---|---|---|
| Train | 350 | 168 | 106 | 624 |
| Test | 87 | 42 | 27 | 156 |
| Total | 437 | 210 | 133 | 780 |



(a) Benign (b) Malignant (c) Normal

**Fig. 3.** Samples of ultrasound images from the dataset: (a) Benign, (b) Malignant, and (c) Normal.

### 3.3 Grad-CAM Visualization

Grad-CAM visualizes a CNN's exploration space, generating class-specific visualizations without explicit position annotations, aiding in weakly supervised segmentation. It utilizes gradients of the classification score to highlight regions impacting the network's decision for a specific class. Our study benefits from Grad-CAM's ability to:

- Generate class-specific heatmaps using input images, a trained CNN, and a chosen class.
- Be applicable across various CNN architectures.
- Provide weakly supervised segmentation and localization.

Deeper CNN layers capture intricate visual structures, balancing high-level semantics with precise spatial information. Grad-CAM attributes suitable values to neurons using gradient data, elucidating decisions made at the output layer. The computed weights represent the significance of the feature map for a target class. The resulting heatmap highlights traits impacting the class of interest, aiding in disease differentiation.

In breast cancer dataset samples, Grad-CAM highlights regions of interest, aiding disease diagnosis as shown in Fig. 4. The coloration distinguishes diseases, showcasing distinct features associated with each.

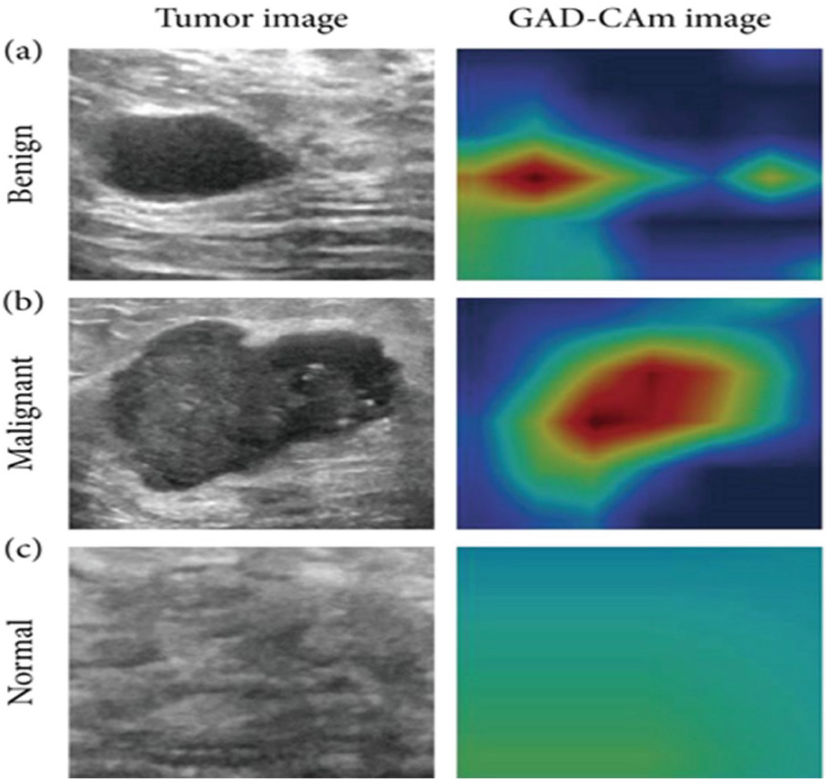



**Fig. 4.** Samples from Grad-CAM results of breast cancer ultrasound images.

### 3.4 Pre-trained CNN Architectures

In this study's proposed hybrid CNN model, the features extracted from eight distinct CNN architectures within the dataset underwent classification using SVM and KNN algorithms. This process aimed to identify which feature extractions from various architectures would be combined. The architectures were then compared based on their accuracy performance, arranging them in order of greatest to lowest accuracy. The outcomes of this ranking are outlined in Table 2. The top three performing architectures in this ranking, namely AlexNet, EfficientNet-B0, and Inception V3, were selected to amalgamate feature extractions.

AlexNet, with 8 layers, processes $227 \times 227$ input data. EfficientNet-B0, with 237 layers, processes $224 \times 224$ data, while Inception V3, with 48 layers, operates on $299 \times 299$ data using ReLU activation.

These models share common layers, with convolution layers being pivotal. They generate feature maps by applying filters to input images. Equation (1) determines the output image dimensions after convolution layer filtering.

$$Y = \frac{x - k + 2p}{s} + 1 \tag{1}$$

where $Y$ represents output size, $x$ indicates input size, $k$ is filter size, $s$ stands for stride, and $p$ represents padding.

The convolution operation is performed as in Eq. (2).

$$O = I * K_{mn} = \sum_{pq} I_{pq} K_{(m-p)(n-q)} \tag{2}$$

where $O$ represents output, $I$ represents input, and $K$ indicates kernel.

The ReLU activation function is depicted in Eq. (3).

$$f(x) = \max(0, x) \tag{3}$$

Normalization techniques are elucidated in Eqs. (4), (5), and (6).

$$T_i = \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \epsilon}} \tag{4}$$

$$\sigma_\beta = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_\beta)^2} \tag{5}$$

$$\mu_\beta = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{6}$$

where $T_i$ stands for the normalized value, $\sigma_\beta$ is the standard deviation, $\mu_\beta$ is the mean, and $N$ is the number of inputs.

### 3.5 mRMR Method

The mRMR method aids in selecting significant features, minimizing computational costs and interdependence among attributes. It selects features that have maximum relevance to the target variable and minimum redundancy among themselves.

## 4 Experimental Setup and Results

Table 2 presents accuracy results from pre-trained CNN architectures with Grad-CAM and two classifiers. This section discusses experimental findings of the hybrid model, including confusion matrices and various metrics. Deep learning models must be evaluated using assessment metrics such as accuracy, F1-Score, precision, recall/sensitivity, specificity, and the confusion matrix.

Accuracy measures overall correctness, accounting for true positive, true negative, false positive, and false negative outcomes. Precision evaluates correct positive predictions, while recall/sensitivity assesses accurately identified true positive cases. The F1-Score combines precision and recall. Confusion matrices offer a comprehensive view of predicted results, aiding in metric evaluation.

**Table 2.** Classification Results of 8 CNN Architectures with Two Classifiers: SVM and KNN

| Models | Grad-CAM + SVM (%) | Grad-CAM + KNN (%) |
|---|---|---|
| AlexNet | 97.4 | 96.7 |
| EfficientNet-B0 | 96.1 | 96.1 |
| Inception V3 | 95.5 | 94.8 |
| MobileNetV2 | 95.5 | 93.2 |
| ResNet101 | 94.8 | 93.1 |
| DarkNet53 | 94.2 | 94.2 |
| VGG19 | 92.2 | 91.7 |
| GoogleNet | 92.2 | 90.2 |

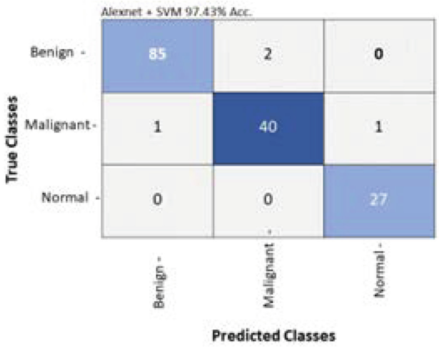
### 4.1 Experimental Setup

The hybrid model operates within the Google Colab Pro framework, expediting training and model assessment. The experimental setup involves Python programming, Keras library, and TensorFlow backend on a Tesla T4 GPU with 25 GB RAM.

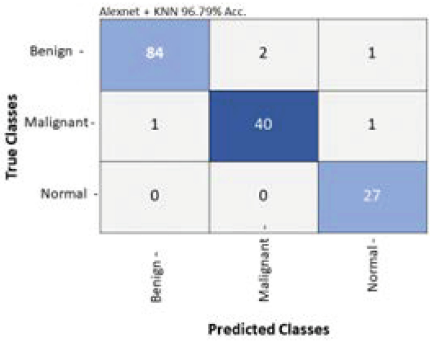
Performance analysis revealed AlexNet with SVM achieved the highest accuracy at 97.4% (Table 2). Our hybrid model, utilizing features from all three architectures classified via SVM and KNN, achieved 99.35% accuracy with mRMR feature selection. This outperforms recent methods, setting a new benchmark.

### 4.2 Performance Metrics

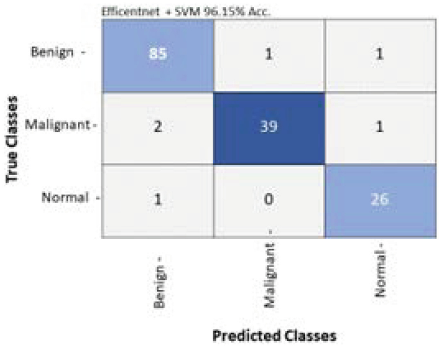
This research employs five performance evaluation metrics: accuracy, precision, sensitivity (recall), specificity, and F1-score to assess the proposed methodology. Figure 5 illustrates confusion matrices generated by SVM and KNN classifiers for AlexNet, EfficientNet-B0, Inception V3, and the hybrid model.



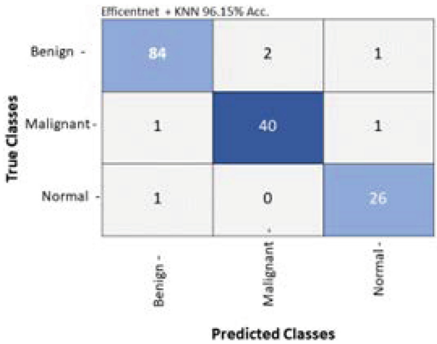

(a) AlexNet + SVM




(b) EfficientNet-B0 + SVM




(c) Inception V3 + SVM




(d) Hybrid Model + SVM

**Fig. 5.** Confusion matrices obtained from AlexNet, EfficientNet-B0, Inception V3, and Hybrid model using SVM classifier.

The results are summarized in Table 3.

## 5 Discussion

Breast cancer, the second leading cause of female mortality globally, necessitates early detection for mortality reduction. Mammography serves as the primary screening tool for detecting abnormalities like tumors. Recent advancements

**Table 3.** Performance Metrics (in %) of the Proposed Hybrid Model

| Model | Class | Accuracy | Sensitivity | Specificity | Precision | F1 Score |
|---|---|---|---|---|---|---|
| Hybrid Model | Malignant | 98.8 | 99.4 | 99.5 | 98.8 | 99.1 |
| | Benign | 100 | 99.2 | 99.4 | 99.2 | 99.25 |
| | Normal | 100 | 99.4 | 100 | 100 | 99.81 |
| | **Overall** | **99.35** | **99.3** | **99.6** | **99.3** | **99.3** |

in deep learning, especially CNNs, have revolutionized medical imaging. This study introduces a hybrid approach combining CNNs and mRMR technique for breast cancer detection from mammography scans. Among eight models, AlexNet, EfficientNet-B0, and Inception V3 were identified as top performers for feature extraction. These pre-trained CNN architectures are integrated into a hybrid setup, combining features from Grad-CAM with trained ones. Optimized via mRMR, these features are classified using SVM algorithms. Achieving a remarkable 99.35% accuracy rate on the BUSI dataset, this approach showcases the potential of hybrid models and advanced feature selection methods in bolstering breast cancer detection accuracy and reliability, marking a significant stride in medical image analysis. Table 4 showcases the outstanding outcomes of the proposed hybrid model compared to existing studies in the literature.

**Table 4.** Comparison Analysis of the Proposed Architecture with Existing Studies

| Study | Approach | Year | Accuracy (%) |
|---|---|---|---|
| Sureshkumar, V. et al. [15] | Hybrid CNN and Extreme Learning | 2024 | 86.00 |
| Sameh Zarif et al. [17] | CNN+EfficientNetV2B3 | 2024 | 96.30 |
| G. Sajiv et al. [9] | Hybrid Deep Learning Classifier | 2024 | 98.20 |
| Karthiga, R. et al. [5] | Hybrid deep neural network | 2024 | 92.00 |
| R. Sathesh Raaj et al. [8] | Hybrid CNN | 2023 | 98.44 |
| Othman, N.A. et al. [7] | LSTM and GRU | 2023 | 98.00 |
| Selina Sharmin et al. [10] | ResNet50V2 | 2023 | 95.00 |
| Singh, L. et al. [12] | Faster R-CNN | 2022 | 95.20 |
| **Proposed Hybrid Model** | AlexNet, EfficientNet-B0, Inception V3 + mRMR | 2024 | **99.35** |

## 6 Conclusion and Future Work

This research demonstrates the potential of integrating CNN with the mRMR feature selection technique to improve breast cancer detection from ultrasound images. By combining CNN architectures like AlexNet, EfficientNet-B0, and Inception V3 with feature optimization methods such as Grad-CAM and mRMR, our hybrid model achieved a 99.35% accuracy rate using SVM classifiers on the BUSI dataset. The findings highlight the value of advanced deep learning models and feature selection in medical imaging, emphasizing the role of hybrid approaches in enhancing diagnostic tools. However, the limited size and scope of the BUSI dataset may affect the generalizability of these results. Future work will validate the model on larger datasets, such as INBreast and CBIS-DDSM, to assess its robustness in diverse clinical settings.

While traditional classifiers like SVM and KNN performed well, they may not fully exploit deep learning's potential. Future research will explore more advanced classifiers and end-to-end deep learning methods, including Vision Transformers, to further enhance performance. This study provides a foundation for future research aimed at refining diagnostic algorithms and expanding their applicability in real-world medical diagnostics.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data Brief **28**, 104863 (2020)
2. Basurto-Hurtado, J.A., Cruz-Albarran, I.A., Toledano-Ayala, M., Ibarra-Manzano, M.A., Morales-Hernandez, L.A., Perez-Ramirez, C.A.: Diagnostic strategies for breast cancer detection: from image generation to classification strategies using artificial intelligence algorithms. Cancers **14**(14), 3442 (2022)
3. Chen, H., Ma, M., Liu, G., Wang, Y., Jin, Z., Liu, C.: Breast tumor classification in ultrasound images by fusion of deep convolutional neural network and shallow LBP feature. J. Digit. Imaging **36**(3), 932–946 (2023)
4. Guo, R., Lu, G., Qin, B., Fei, B.: Ultrasound imaging technologies for breast cancer detection and management: a review. Ultras. Med. Biol. **44**(1), 37–70 (2018)
5. Karthiga, R., et al.: A novel exploratory hybrid deep neural network to predict breast cancer for mammography based on wavelet features. Multimed. Tools Appl. 1–27 (2024)
6. Melekoodappattu, J.G., Subbian, P.S.: Automated breast cancer detection using hybrid extreme learning machine classifier. J. Ambient. Intell. Humaniz. Comput. **14**(5), 5489–5498 (2023)
7. Othman, N.A., Abdel-Fattah, M.A., Ali, A.T.: A hybrid deep learning framework with decision-level fusion for breast cancer survival prediction. Big Data Cogn. Comput. **7**(1), 50 (2023)
8. Raaj, R.S.: Breast cancer detection and diagnosis using hybrid deep learning architecture. Biomed. Signal Process. Control **82**, 104558 (2023)
9. Sajiv, G., Ramkumar, G., Shanthi, S., Chinnathambi, A., Alharbi, S.A.: Predicting breast cancer risk from histopathology images using hybrid deep learning classifier. Med. Eng. Phys. 104149 (2024)
10. Sharmin, S., Ahammad, T., Talukder, M.A., Ghose, P.: A hybrid dependable deep feature extraction and ensemble-based machine learning approach for breast cancer detection. IEEE Access (2023)
11. Siegel, R.L., Miller, K.D., Wagle, N.S., Jemal, A.: Cancer statistics, 2023. CA: Cancer J. Clin. **73**(1) (2023)

12. Singh, L., Alam, A.: An efficient hybrid methodology for an early detection of breast cancer in digital mammograms. J. Ambient. Intell. Humaniz. Comput. **15**(1), 337–360 (2024)
13. Singh, S.P., Urooj, S., Lay-Ekuakille, A.: Breast cancer detection using PCPCET and ADEWNN: a geometric invariant approach to medical X-ray image sensors. IEEE Sens. J. **16**(12), 4847–4855 (2016)
14. Srikantamurthy, M.M., Rallabandi, V.S., Dudekula, D.B., Natarajan, S., Park, J.: Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning. BMC Med. Imaging **23**(1), 19 (2023)
15. Sureshkumar, V., Prasad, R., Balasubramaniam, S., Jagannathan, D., Daniel, J., Dhanasekaran, S.: Breast cancer detection and analytics using hybrid CNN and extreme learning machine. J. Pers. Med. **14**(8), 792 (2024)
16. Tanaka, H., Chiu, S.W., Watanabe, T., Kaoku, S., Yamaguchi, T.: Computer-aided diagnosis system for breast ultrasound images using deep learning. Phys. Med. Biol. **64**(23), 235013 (2019)
17. Zarif, S., Abdulkader, H., Elaraby, I., Alharbi, A., Elkilani, W.S., Pławiak, P.: Using hybrid pre-trained models for breast cancer detection. PLoS One **19**(1), e0296912 (2024)