Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 7, No. 3, July 2025, pp: 611-623 e-ISSN: 2656-8632
RESEARCH ARTICLE OPEN ACCESS

Breast Cancer Classification on Ultrasound Images Using DenseNet Framework with Attention Mechanism

Hanina Nafisa Azka, Wiharto, and Esti Suryani

Department of Informatics, Faculty of Information Technology and Data Science, University of Sebelas Maret, Indonesia

Corresponding author: Wiharto (e-mail: wiharto@staff.uns.ac.id)

Abstract **Breast cancer is one of the most prevalent and life-threatening diseases among women worldwide. Early detection of breast cancer being critical for increasing survival rates. Ultrasound image is commonly used for breast cancer screening due to its non-invasive, safe, and cost-effective. However, ultrasound images are often of low quality and have significant noise, which can hinder the effectiveness of classification models. This study proposes an enhanced breast cancer classification model that leverages transfer learning in combination with attention mechanisms to improve diagnostic performance. The main contribution of this research is the introduction of Dense-SASE, a novel architecture that combines DenseNet-121 with two powerful attention modules: Scaled-Dot Product Attention and Squeeze-and-Excitation (SE) Block. These mechanisms are integrated to improve feature representation and allow the model to focus on the most relevant regions of the ultrasound images. The proposed method was evaluated on a publicly available breast ultrasound image dataset, with classification performed across three categories: normal, benign, and malignant. Experimental results demonstrate that the Dense-SASE model achieves an accuracy of 98.29%, a precision of 97.97%, a recall of 98.98%, and an F1-score of 98.44%. Additionally, Grad-CAM visualizations demonstrated the model's capability to localize lesion areas effectively, avoiding non-informative regions, and confirming the model's interpretability. In conclusion, the Dense-SASE model significantly improves the accuracy and reliability of breast cancer classification in ultrasound images. By effectively learning and focusing on clinically relevant features, this approach offers a promising solution for computer-aided diagnosis (CAD) systems and has the potential to assist radiologists in early and accurate breast cancer detection.**

Keywords Breast Cancer Classification; Ultrasound Image; Transfer Learning; Attention Mechanism.

I. Introduction

Breast cancer is the second most common cancer globally. According to data from the International Agency for Research on Cancer (IARC), breast cancer accounts for 2.3 million cases or 23.8% of all cancer cases among women, with a mortality rate of 15.4%, resulting in approximately 666,000 deaths [1]. In Indonesia, breast cancer is the most prevalent type of cancer, affecting around 66,000 people. Early detection of breast cancer can significantly improve recovery rates and maximize patient survival [2], [3].

Although mammography is considered the gold standard for breast cancer screening, this method has limitations when dealing with dense breast parenchyma [4]. For dense breast tissue, ultrasonography (US) is a powerful diagnostic tool, as it can detect breast tumors that may be missed by mammography [5]. US uses high-frequency sound waves to provide comprehensive information about the dimensions, morphology, and characteristics of breast lesions in the form of ultrasound images [6] [7]. This method is non-invasive and relatively affordable for people from all socioeconomic backgrounds [8]. However, manual interpretation relies heavily on radiologist expertise, which can lead to variability in diagnosis. Therefore, an advanced system is necessary to enhance the accuracy and objectivity of breast cancer diagnosis [9].

In recent years, the implementation of Computer Aided Diagnosis (CAD) systems has shown significant potential in improving breast cancer diagnosis accuracy [10]. Deep learning approaches, particularly Convolutional Neural Networks (CNNs), have proven effective for medical image processing [11], [12] . This method can automatically extract features from images, reducing the need for complex and time-consuming manual feature extraction [13]. The

Manuscript received March 8, 2024; Accepted May 5, 2025; date of publication May 9, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.779

implementation of transfer learning for feature extraction can also improve model performance across various tasks, including breast cancer classification [14], [15].

In a study conducted by [16], DenseNet-121 demonstrated superior performance compared to other architectures, achieving an accuracy of 87.5%. DenseNet incorporates direct connections between all layers within a block, addressing the vanishing gradient problem and promoting feature propagation. However, this model still has limitations, as all features are treated with equal weight, potentially reducing efficiency due to less relevant features.

Incorporating attention mechanisms into dense units can improve model interpretability by assigning attention weights to each input [17]. Self-attention, one such mechanism, computes relationships between pixels globally, assigning higher weights to relevant image features and capturing complex patterns [18]. A study by [19] employed Self-Attention Random Forest (SARF) for breast cancer classification on histopathology images, achieving an accuracy of 92.96% and an AUC of 0.9588. Another study by [20] combined Self-Attention (SA) with Multi-Instance Learning (MIL), resulting an accuracy of 91% and an AUC of 0.912. Heatmap analysis from the study indicated that the color distribution in the attention matrix displayed stripe-like patterns, suggesting that self-attention not only captures relationships between individual elements but also identifies global data patterns.

A study by Deng [21] conducted breast cancer classification on mammography images by integrating a Squeeze-and-Excitation Block attention mechanism. The study demonstrated that adding this mechanism improved classification accuracy across several architectures, including Inception-V4, ResNeXt, and DenseNet, with results increasing from 89.97% to 92.17%, 89.64% to 91.57%, and 89.20% to 91.79%, respectively. The SE Block enhances the model's ability to dynamically learn inter-channel relationships, allowing greater focus on relevant features [22]. This module can be applied to object detection [23], segmentation [24], and image classification [25].

To further improve model performance in medical image analysis, Moon [26] conducted research on ultrasound images using four different types of images: (1) original ROI images, (2) tumor images, (3) Tumor Shape Images (TSI), and (4) fused images. DenseNet-121 architecture was applied to each image type, yielding accuracies of 86.35%, 86.35%, 82.19%, and 89.32%, respectively. These results indicate that image type selection influences the model's performance in breast cancer detection.

This study aims to proposes the Dense-SASE model DenseNet with Scaled-Dot Product Attention and SE Block combining the strengths of transfer learning with additional attention mechanisms. DenseNet is selected for its ability to maintain gradients and extract complex features. The Scaled-Dot Product Attention mechanism enables the model to focus on important regions of the image, while Squeeze-and-Excitation enhances relevant features, improving classification performance. The contribution of this study are: 1) implement transfer learning using DenseNet-121 architecture integrated with attention mechanisms as a novel approach for breast cancer classification on ultrasound images and 2) assess the effectiveness of attention mechanisms in focusing on relevant features essential for accurate breast cancer classification.

This study structured as follows: section II discusses about the dataset used, proposed methods,

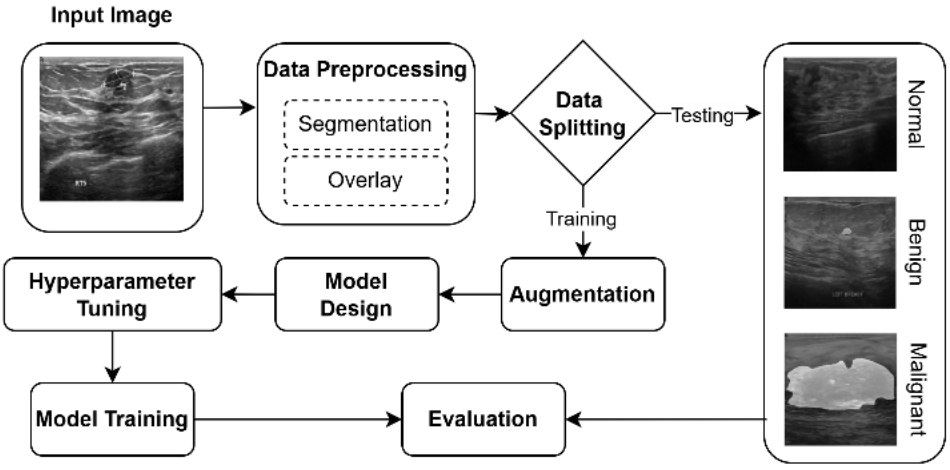


Fig. 1. Flow of research

Manuscript received March 8, 2024; Accepted May 5, 2025; date of publication May 9, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.779

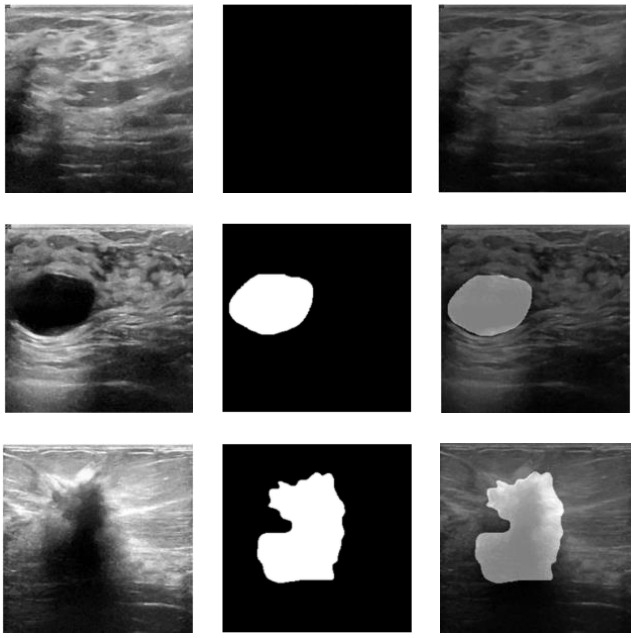
hyperparameter tuning, and training and evaluation schemes. Section III displays the results of training models and evaluation on testing set. This section also shows the result of interpretation using Grad-CAM. Section IV discusses the comparison of evaluation on testing set with other models and limitation from the proposed model. And the last section is section V, which rewrite the objectives, main findings, and future works.

II. Method

Fig. 1. illustrates the workflow of this study. This study consists of several procedures, including data processing, model design, hyperparameter tuning, model training, and evaluation.

A. Preprocessing Dataset

This study uses the Breast Ultrasound Image (BUSI) dataset [27]. The dataset consists of ultrasound images of breast cancer obtained from 600 female patients aged 25 - 75 years at Baheya Hospital, Egypt. The ultrasound devices used to collect the dataset were LOGIC E9 Ultrasound and LOGIC E9 Agile Ultrasound System. The format of this dataset is PNG with a size of 500x500 pixels. Each image is accompanied by a class label indicating the patient's clinical status, such as normal, benign, and malignant, with data distributions of 133, 437, and 210 images, respectively. For every image are resized to 256x256 pixel.



(a) (b) (c)

Fig. 2. Sample (a) original image, (b) segmented image, and (c) overlayed image.

Each image in each class is segmented to separate the disease object from the background. Segmentation is performed on the original images for each class using a U-Net framework [28]. The result of segmentation segmented image is overlaid on the original image to produce an overlay image. The merging process uses an alpha value of 0.5, corresponding to 50% transparency. This value is chosen to enhance the image representation without losing information from the original image. Fig shows the result of segmentation and overlay in each class.

The overlay images are split into 15% for the testing set and 85% for the training and validation sets. The training and validation sets undergo Stratified K-Fold Cross Validation with 5 folds to ensure a more balanced distribution of data across each class. The testing set consists of 20 images from the normal class, 66 images from the benign class, and 31 images from the malignant class. The detailed data distribution for the training and validation sets is presented in Table 1 Data augmentation is applied to the training set to increase data variety and prevent model overfitting. The transformations include rotation range, width shift range, height shift range, horizontal flip, and vertical flip for each dataset.

Table 1. Data distribution in training and validation sets.

Fold	Data	Class		
		Normal	Benign	Malignant
1	Training	91	296	143
	Validation	22	75	36
2	Training	90	297	143
	Validation	23	74	36
3	Training	90	297	143
	Validation	23	74	36
4	Training	90	297	144
	Validation	23	74	35
5	Training	91	297	143
	Validation	22	74	36

B. Model Design

In this study, the author proposes a neural network model called Dense-SASE, DenseNet-121 with Scaled-Dot Product Attention and Squeeze-and-Excitation. This architecture is based on transfer learning, utilizing DenseNet-121 as the backbone, originally developed by Huang [29]. This convolutional neural network is pretrained on the ImageNet dataset and is known for its dense connectivity, where each layer receives inputs from all previous layers in the same block. DenseNet-121 consists of four main dense blocks, each comprising multiple densely connected layers where every layer receives input from all preceding layers within the same block. This dense connectivity promotes feature reuse, improves gradient flow, and enhances model efficiency. The number of dense layers in each block is as follows: 6 layers in

Manuscript received March 8, 2024; Accepted May 5, 2025; date of publication May 9, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.779

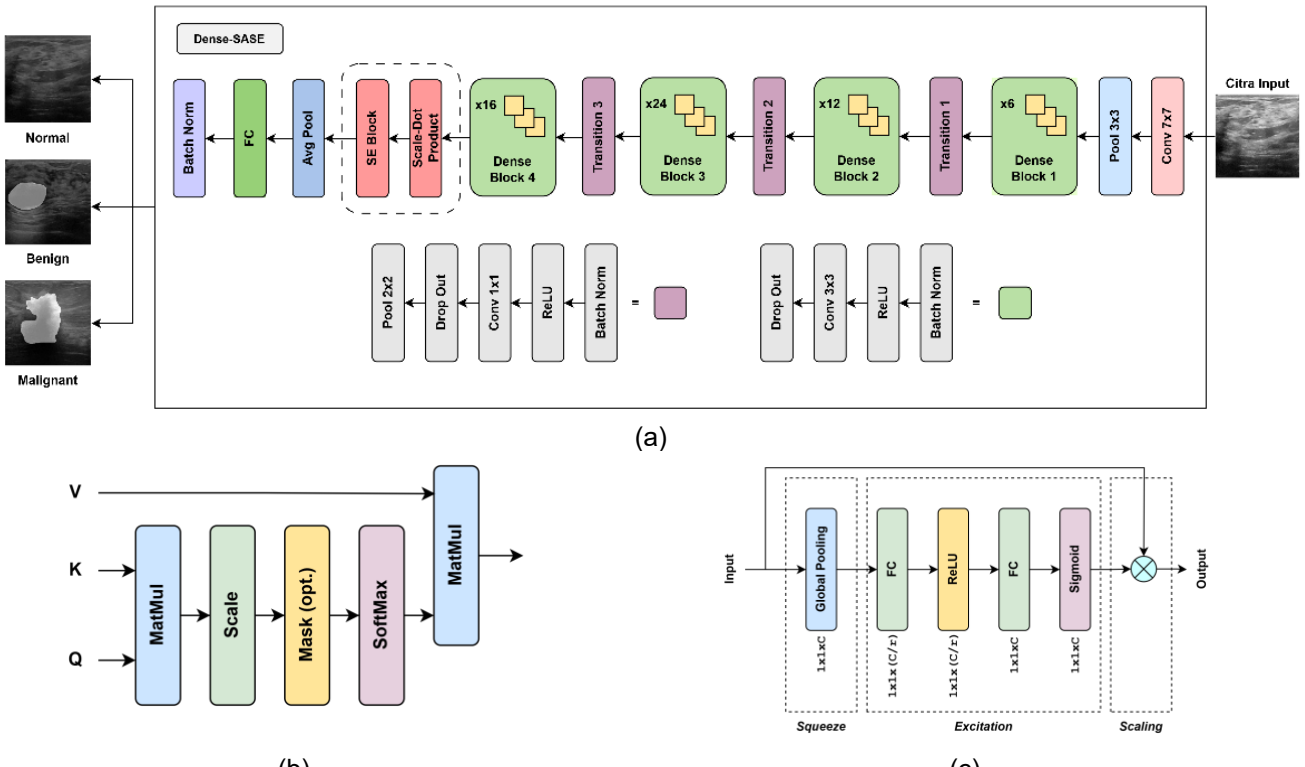


Fig. 3. Structure of (a) Dense-SASE, (b) Scaled Dot-Product Attention, and (b) SE Block.

Dense Block 1, 12 layers in Dense Block 2, 24 layers in Dense Block 3, and 16 layers in Dense Block 4, resulting in a total of 58 dense layers in the DenseNet-121 architecture.

In Dense-SASE, all layers of DenseNet-121 are kept frozen except the last five. This strategy leverages generic low-level features learned from ImageNet while allowing fine-tuning of high-level features specific to breast ultrasound classification. The model discards DenseNet's original fully connected layers and replaces them with customized attention and classification layers tailored to the target dataset. To better capture spatial relationships between features, the model integrates a Scaled-Dot Product Attention mechanism, originally introduced by Vaswani [30] in the transformer architecture. This attention block helps the model focus on spatially relevant features while reducing noise and redundant activations. Following spatial refinement, channel-wise attention is applied through an SE Block, which enhances the model's ability to focus on the most informative feature channels [22]. Fig. 3a. illustrates the architecture of Dense-SASE.

The output features from DenseNet-121 are projected into three representations query, key, and value using trainable weights. The attention scores are computed by performing a matrix multiplication between the query and the transpose of the key as formulated in Eq (1), resulting in an attention map that reflects the relationships between spatial locations within the feature map [30]. These scores are then normalized using the softmax function to ensure they represent probability values as illustrated in Eq. (2) [30]. The final output is computed as the weighted sum of V, enhancing important spatial features while maintaining original context. The attention output is then added back to the original input (residual connection), improving spatial expressiveness without discarding previously learned features as represented in Eq. (3) [30]. Fig. 3b. illustrates the structure of Scaled-Dot Product Attention.

$$Score\ (Q,K) = \frac{QK^T}{\sqrt{d_k}} \quad (1)$$

$$\alpha = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (2)$$

$$Attention\ (Q,K,V) = \alpha V = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

where Q represents the information being queried or sought after in the attention mechanism. K corresponds to the context or information that the model compares the query to in order to find relevant relationships. V contains the actual information that is passed through the attention mechanism and d_k means the

Manuscript received March 8, 2024; Accepted May 5, 2025; date of publication May 9, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.779

dimensionality of the key vectors and is used for normalization.

Following spatial refinement, channel-wise attention is applied through an SE Block, which enhances the model's ability to focus on the most informative feature channels. The input features are compressed into a 1D representation using Global Average Pooling (GAP) as formulated in Eq. (4) [22]. This compressed representation is then processed through two dense layers. The first dense layer applies the ReLU activation function to capture non-linear relationships between channels as illustrated in Eq. (5) and the second dense layer uses a sigmoid activation function to generate a channel attention map as illustrated in Eq. (6) [22]. This attention map is then multiplied by the input features, amplifying relevant channels while suppressing less important ones as represented in Eq. (7) [22]. This process allows the model to amplify important channels and suppress less relevant ones, making the classification process more robust, especially in subtle cases. Fig. 3. presents the illustration of the SE Block.

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{c,i,j} \quad (4)$$

$$s = \delta(W_1 z) \quad (5)$$

$$a = \sigma(W_2 s) \quad (6)$$

$$\tilde{X}_c = s_c \cdot X_c \quad (7)$$

where z_c represent the global context or average value of the feature map for a specific channel c. H means height and W means width of the map. Specifically, X_c refers to the feature map at c. δ represent ReLU and σ represent sigmoid.

After passing through the SE Block, the features are processed using GAP to reduce spatial dimensions, resulting in a one-dimensional vector that represents the overall feature set. This vector passes through a Dense layer with ReLU activation, followed by Batch Normalization and Dropout to improve training stability and prevent overfitting. The final output layer is a Dense layer with 3 neurons and softmax activation, producing class probabilities for the three target classes: normal, benign, and malignant.

C. Hyperparameter Tuning

Hyperparameter tuning is performed to find the optimal combination of parameters, aiming to maximize the model's performance in classification. Table 2. presents the initial hyperparameters used before the tuning process. Hyperparameter tuning is conducted on several components, including optimizer type, learning rate, and dropout rate. In the optimizer type, we will experiment using Adam, SGD, and RMSprop. The learning rate will be experiment for values of 0.001, 0.0001, and 0.00001. Meanwhile, the dropout value will be experiment for values of 0.3, 0.4, and 0.5. The method used is Bayesian Optimization because it is efficient in exploring the parameter space with fewer trials by using a probabilistic approach to predict the most promising combination of parameters based on previous results [31]. The best-performing parameters obtained from this process will be used during the training of the proposed model.

Table 2. Initial hyperparameter values.

Hyperparameter	Value
Batch size	32
Activation function	Softmax
Loss Function	Categorical Crossentropy
Epoch	30
Optimizer	Adam
Learning rate	0.001
Dropout rate	0.5

The training process will be conducted on the neural network model using the BUSI dataset. The training is performed on the training set with overlay images as input. To ensure the neural network remains in an optimal state, validation is carried out at the end of each epoch to prevent underfitting or overfitting.

D. Evaluation

Evaluation is conducted to measure the performance of the proposed model. A confusion matrix is employed to compare the classification algorithm's performance with the actual classification results [32].

1. Accuracy

Accuracy is a type of metric that represents the ratio of correctly classified data to the total amount of data. In medical diagnostics, a high accuracy does not necessarily indicate good performance in detecting malignant/benign cases. Clinically, this could lead to undetected cancer, delaying treatment and affecting prognosis. Eq. (8) represents the formula for accuracy in class i, while Eq. (9) represents the formula for macro accuracy [33]

$$Accuracy = \frac{\sum_i^m TP_i}{\sum_i^m (TP_i + FN_i + FP_i)} \quad (8)$$

$$Macro\ Accuracy = \frac{1}{m} \frac{\sum_i^m TP_i}{\sum_i^m (TP_i + FN_i + FP_i)} \quad (9)$$

where m means the total number of classes, TP_i (True Positive) means the number of correctly predicted samples that actually belong to class i. FN_i (False Negative) means the number of actual samples from class i that were incorrectly predicted as another class. FP_i (False Positive) means the number of samples from other classes that were incorrectly predicted as class i.

2. Precision

Precision is a metric used to measure how many of the positive predictions are actually correct compared to

Manuscript received March 8, 2024; Accepted May 5, 2025; date of publication May 9, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.779

the total positive predictions made. In a clinical, precision is particularly important for minimizing false positives. A high precision for the malignant/benign class implies that most predictions labeled as cancer are correct, reducing unnecessary patient anxiety, additional imaging, or biopsies. Thus, high precision enhances diagnostic trust and efficiency in healthcare systems. Eq. (10) represents the precision formula for class i. Eq. 11 represents the macro precision formula, which calculates the average precision across all classes [33].

$$Precision = \frac{TP_i}{TP_i + FP_i} \quad (10)$$

$$Macro\ Precision = \frac{\sum_i^m Precision_i}{m} \quad (11)$$

3. Recall

Recall is a metric used to measure how many correct positive predictions are made relative to the total actual positives. Clinically, recall is critical in ensuring that all cases of interest, especially cancer, are detected. High recall in the cancer category means the model can detect most cancer cases, minimizing the risk of false negatives. Therefore, in cancer screening and diagnosis, recall is often prioritized to ensure sensitivity and early intervention. Eq. (12) represents the recall formula for class i. Eq. (13) represents the macro recall formula, which averages the recall across all classes [33].

$$Recall = \frac{TP_i}{TP_i + FN_i} \quad (12)$$

$$Macro\ Recall = \frac{\sum_i^m Recall_i}{m} \quad (13)$$

4. F1-Score

F1-Score is a metric that calculates the harmonic mean of precision and recall. In a medical diagnosis context, the F1-score is essential when both false positives and false negatives must be minimized. A balanced F1-score indicates that the model maintains an effective trade-off between correctly identifying positive cases and avoiding incorrect alerts. This balance is crucial for clinical decision support systems where both underdiagnosis and overdiagnosis carry significant risks. Eq. (14) represents the F1-Score formula for class i and Eq. (15) represents the macro F1-Score formula, averaging the F1-Scores across all classes [33].

$$F1 - Score = 2 \cdot \frac{Precision_i \cdot Recall_i}{Precision_i + Recall_i} \quad (14)$$

$$Macro\ F1 - Score = \frac{\sum_i^m F1 - Score_i}{m} \quad (15)$$

III. Result

A. Hyperparameter Tuning

Hyperparameter tuning is performed before training to find the optimal combination of parameters, aiming to enhance model performance in classification tasks. This process focuses on improving model generalization by adjusting key parameters. The first experiment tests different optimizers to evaluate their effect on the proposed model's performance. Based on the results shown in Fig. 4, the choice of optimizer significantly impacts the model's accuracy and loss. The experiment reveals that SGD achieves the highest validation accuracy, reaching 96.97%, with a relatively low loss of 0.1415. The momentum mechanism in SGD helps the model overcome local minima, making it more effective in generalizing on validation data. Although SGD converges slower than Adam, it remains more stable during the final epochs, leading to better generalization performance.

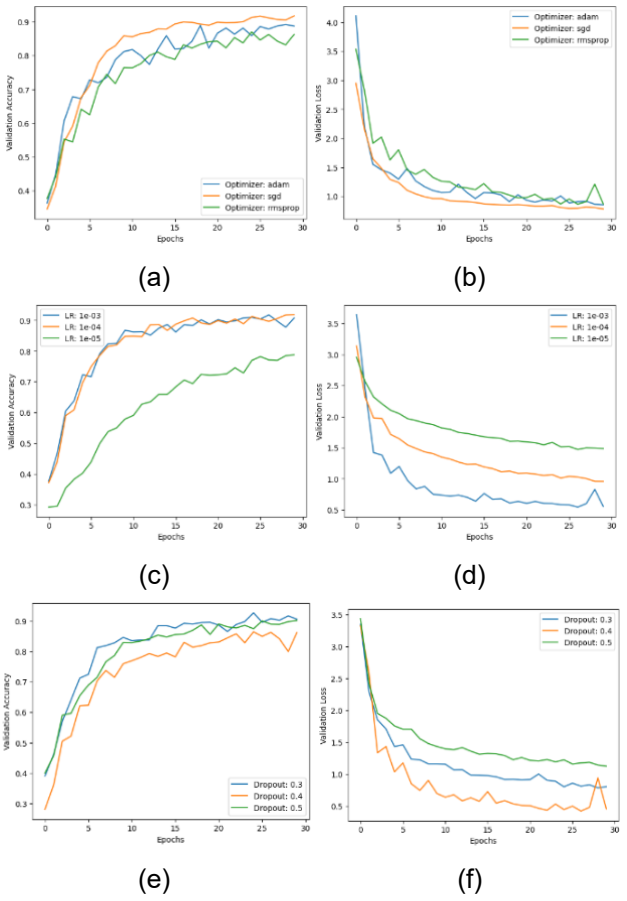


Fig. 4. Graph of accuracy and loss for three experiments: (a) and (b) optimizer; (c) and (d) learning rate; (e) and (f) dropout rate.

The second experiment was conducted to evaluate the effect of different learning rates. Based on the results shown in Fig. 4., a learning rate of 0.001 provided the best performance in terms of accuracy and loss. This value is large enough to accelerate learning while maintaining stable convergence, resulting in high accuracy and low loss.

Manuscript received March 8, 2024; Accepted May 5, 2025; date of publication May 9, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.779

The third experiment aimed to evaluate the effect of different dropout values on the proposed model. Based on the results shown in Fig. 4, dropout rate of 0.3 provided the best accuracy with lower and more stable loss. The model achieved high validation accuracy with relatively fast convergence due to a balance between learning important features and regularization. In the loss graph, the model also showed more stability and a consistent downward trend over time, indicating that the model generalizes well without overfitting.

After hyperparameter tuning, the model was trained using the best hyperparameters obtained. Table 3. presents the hyperparameters resulting from the tuning process.

Table 3. Hyperparameter tuning.

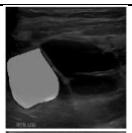
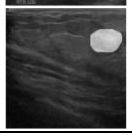
Hyperparameter	Value
Batch size	32
Activation function	Softmax
Loss Function	*Categorical Crossentropy*
Epoch	100
Optimizer	SGD
Learning rate	0.001
Dropout rate	0.3

B. Performance of the Dense-SASE Model

The training process of the Dense-SASE model takes 9 seconds per epoch. Each epoch includes a validation set to evaluate the model during training. Based on the Fig. 5. , the accuracy and loss graphs produced during the training process demonstrate good performance. The training accuracy increased from 38.79% in the first epoch to 93.41% by the 150th epoch. Validation accuracy also showed an upward trend, rising from 34.09% in the first epoch to 95.45% in the final epoch. This indicates that the model can generalize well to new data. Meanwhile, the training loss steadily decreased from 1.4054 at the start to 0.2566 by the last epoch and the validation loss dropped from 1.5659 to 0.1781, suggesting that the model successfully avoided overfitting.

Evaluation was also conducted on the testing dataset, with results shown in Table 5. and confusion matrix shown in Fig. 6. The Dense-SASE model demonstrated excellent performance, achieving an accuracy of 98.20%, indicating that the model accurately classified almost all samples in the dataset. From a medical perspective, the high precision (97.97%) minimizes false positives, reducing patient anxiety and unnecessary interventions. The high recall (98.98%) ensures most true cases are detected, critical for early and accurate diagnosis. The balanced F1-score (98.44%) indicates the model is both safe and effective for clinical decision support. In a clinical context, such performance implies a system that can significantly support physicians in early and accurate disease detection particularly vital in conditions like cancer, where early intervention can drastically improve prognosis.

Table 4. False prediction from Dense-SASE.

Label	Prediction	Image
Benign	Malignant	
Benign	Malignant	

In this study, two samples from the benign class were incorrectly predicted as malignant, shown in Table 4. This type of error can have significant implications in a clinical setting, potentially leading to unnecessary stress for patients and unneeded medical procedures. A closer examination of these misclassified images may reveal common visual features such as irregular textures or blurred boundaries that resemble malignant characteristics.

C. Analysis of Attention Mechanism

Experiments were also conducted to evaluate the impact of using attention mechanisms. The experiments involved four model configurations: (1) DenseNet-121; (2) DenseNet-121 with Scale-Dot Product Attention; (3) DenseNet-121 with Squeeze-and-Excitation (SE) Block; and (4) DenseNet-121 with both Scale-Dot Product Attention and Squeeze-and-Excitation (SE) Block. The comparison aimed to analyze how the addition of attention mechanisms improves model performance. The training process was carried out on all four models using the same parameters to ensure fair and objective results.

It can be seen in Fig. 5. that the Dense-SASE model demonstrates faster convergence compared to other models and a smaller gap between training and validation accuracy. This dual-attention setup not only improves feature extraction but also reduces redundancy in learned representations. Consequently, the model avoids over-relying on less meaningful features, contributing to better generalization and mitigating overfitting. Therefore, it can be concluded that incorporating attention mechanisms into the DenseNet-121 model improves performance. Fig. 5 also shows a comparison of the loss curves during training for the four tested models. The graph reveals that the loss remains persistently fluctuating in all models despite continuous improvement in accuracy.

Manuscript received March 8, 2024; Accepted May 5, 2025; date of publication May 9, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.779

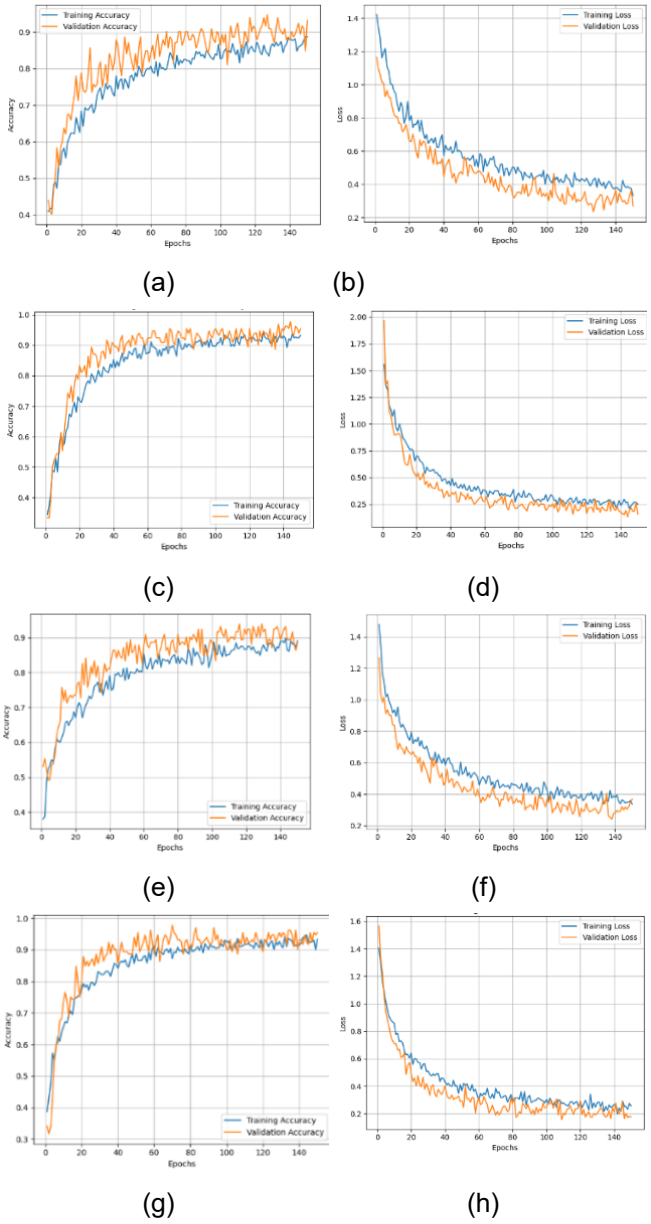


Fig. 5. Comparison of accuracy and loss graph for four models: (a) and (b) represent Model 1; (c) and (d) represent Model 2; (e) and (f) represent Model 3; (g) and (h) represent Model 4.

This prolonged fluctuation indicates that the models face challenges in achieving stable convergence, which may be driven by several technical factors. Nevertheless, it can be observed that the Dense-SASE model demonstrates the best performance, with a faster decline in loss and the lowest final loss among all models. The amplitude of fluctuations in the last few epochs is also smaller and more stable. This proves that integrating two types of attention mechanisms helps the model endure instability and continue learning from data.

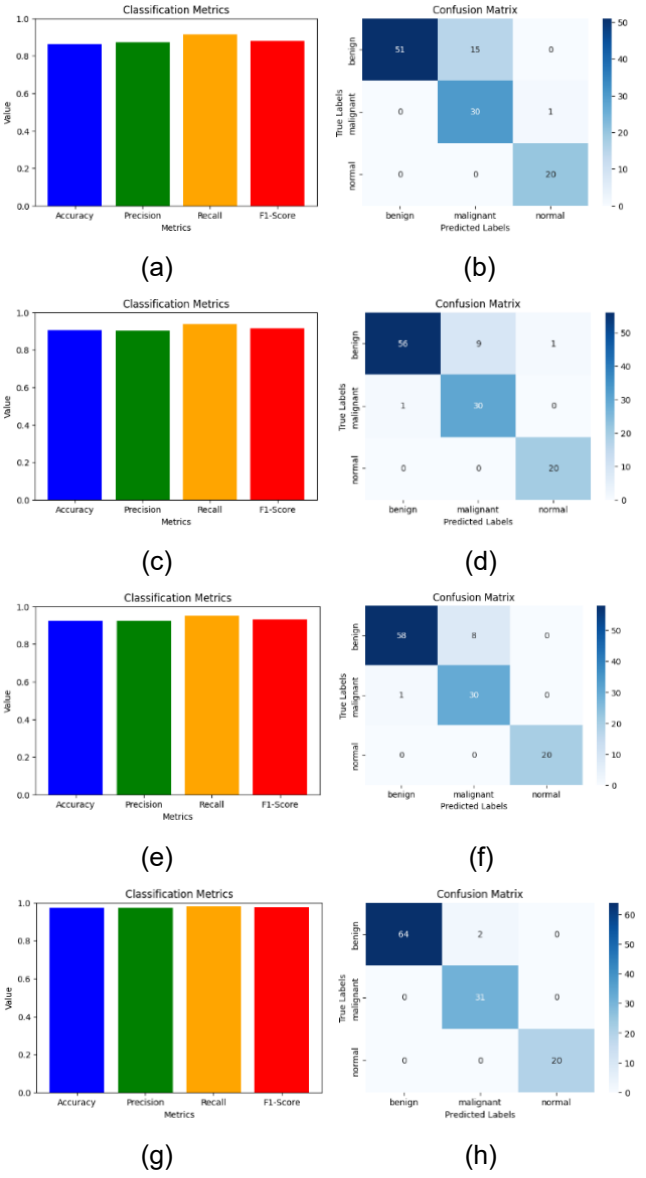


Fig. 6. Comparison of classification metrics and confusion matrix for four models: (a) and (b) represent Model 1; (c) and (d) represent Model 2; (e) and (f) represent Model 3; (g) and (h) represent Model 4.

After the training process, each model was tested using the BUSI dataset testing set. Based on the evaluation results in Table 5. the Dense-SASE model achieved the best performance among all tested models. The proposed model also achieved an optimal balance between precision and recall. The integration of Scaled-Dot Product Attention and Squeeze-and-Excitation (SE) Block plays a significant role in this improvement. The Scaled-Dot Product Attention allows the model to focus on the most relevant spatial features by dynamically adjusting the importance of different

Manuscript received March 8, 2024; Accepted May 5, 2025; date of publication May 9, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.779

regions within the feature map, while the SE Block adaptively recalibrates feature importance across channels. This ensures that even subtle yet crucial patterns are highlighted, improving the model's ability to differentiate between classes an essential factor in medical image classification tasks, such as distinguishing between normal, benign, and malignant breast ultrasound images.

To statistically verify whether there are significant performance differences among the evaluated models, a one-way ANOVA test was conducted on the validation F1-scores. The analysis resulted in an F-statistic of 8.98 and a p-value of 7.97×10^{-6}, indicating a statistically significant difference ($p < 0.05$) among the models. This finding implies that the variations in model design, including attention and recalibration modules, contribute to measurable changes in classification matrix, warranting further exploration through post-hoc analysis to identify which models differ significantly from others.

IV. Discussion

This study utilizes Grad-CAM (Gradient-weighted Class Activation Mapping) developed by Selvaraju [34] to interpret how the model makes classification decisions by visualizing key areas that influence predictions. Grad-CAM generates an activation map in the form of a heatmap, highlighting the regions of the image that contribute most to the classification outcome. Based on the Grad-CAM results shown in Fig. 7., the proposed model successfully highlights key areas around abnormal tissue in the ultrasound images. In the malignant class, the model effectively focuses on regions with irregular textures and indistinct boundaries characteristics typical of malignant tissue. The heatmap is sharper and more concentrated on the core area of the abnormality. Thus, the Grad-CAM interpretation not only increases trust in the model's predictions but also demonstrates that the Dense-SASE model has stronger feature understanding, aligning more closely with the clinical reasoning of doctors when analyzing breast cancer ultrasound images

Karhtik et al [27] propose a Stacking Ensemble with custom CNN architectures. The model has achieved a misclassification rate of 7.85%. Another study by [28] used VGG-16 to create high-speed classification model, resulting an accuracy of 90.12 with a loss of 0.2641. The model's performance varies significantly across different classes, suggesting an imbalance in learning or insufficient feature representation for certain categories. A study by [29] propose an Ensemble Deep Convolutional Neural Network (EDCNN) that combines MobileNet and Xception models. However, the model still exhibits moderate accuracy levels across datasets, indicating room for improvement in its generalization and class-wise stability.

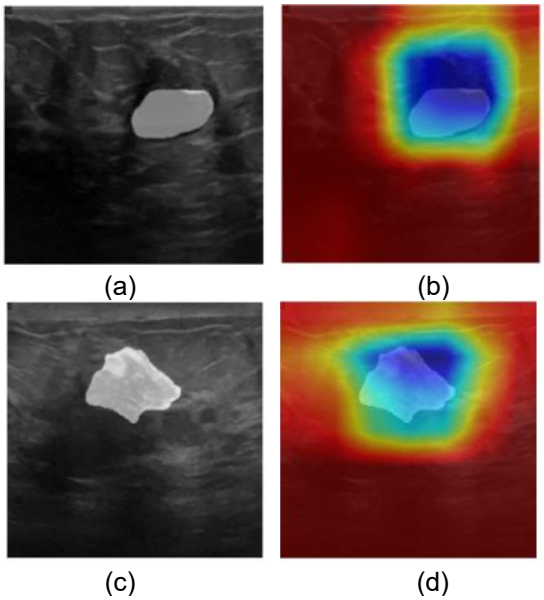


Fig. 7. Overlay image and Grad-CAM visualization for two labels: (a) and (b) benign; (c) and (d) malignant.

[30] propose a hybrid CAD system combining AlexNet, Boruta-SHAP, and Random Forest for breast cancer classification. However, the use of pre-trained features still limits class discrimination, particularly for subtle differences in ultrasound images. A study by [31] uses deep hybrid CNN, particularly the ShuffleNet-ResNet scheme. Nevertheless, despite the strong performance, the model's accuracy still depends on the quality and size of the dataset, and there is room for improvement in generalizing to larger and more diverse datasets. Another study by [32] proposes model using meta-learning framework with multiple CNN models, achieving 90% accuracy on the BUSI dataset. However, the model's complexity stemming from a high number of trainable parameters—and the limited size and diversity of the dataset represent notable limitations.

Based on Table 5. the proposed model demonstrates superior performance compared to previous methods. It achieves an accuracy of 98.29%, precision of 97.97%, recall of 98.98%, and an F1-score of 98.44%, consistently outperforming other models. Overall, the Dense-SASE model proves to be more effective and efficient in capturing patterns from ultrasound images. It also maintains a balanced performance across precision, recall, and F1-score, making it a robust solution for breast cancer

Manuscript received March 8, 2024; Accepted May 5, 2025; date of publication May 9, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.779

Table 5. The comparison of evaluation metrics between Dense-SASE and other model.

Model	Metric (%)			
	Accuracy	Precision	Recall	F1-Score
CNN [26]	88.90	91.00	88.80	89.20
Ensemble CNN [35]	92.15	92,26	92,17	92,21
VGG16 + CNN [36]	90.12	81,00	77,00	79,00
Deep CNN [37]	87.82	87,33	85,33	86,00
AlexNet + Random Forest [38]	96.10	96.30	96.20	96.20
ResNet + ShuffleNet [39]	92.10	90.10	91.20	90.60
DenseNet + Inception + ResNet50 [40]	90.00	90.00	89.50	89.50
DenseNet-121	86.32	87.30	91.34	87.95
DenseNet-121 + Scaled-Dot Product Attention	90.59	90.12	93.87	91.44
DenseNet-121 + SE Block	92.30	92.41	94.88	93.25
Dense-SASE (Proposed Model)	**98,29**	**97,97**	**98,98**	**98,44**

classification. The Grad-CAM interpretation provides deeper insight into the key areas the model focuses on during classification. Based on the generated heatmap visualization, the Dense-SASE model tends to pay more attention to dense tissue areas and distinctive textures commonly associated with malignant tumors. This indicates that the model not only memorizes surface-level patterns but also identifies essential features contributing to classification decisions. These findings strengthen the argument that the Dense-SASE approach offers more transparent and trustworthy interpretability.

However, despite its strong performance, several limitations should be addressed. One key limitation lies in the reliance on the BUSI dataset, which, although publicly available, may not fully represent the diversity of clinical cases encountered in real-world scenarios. The dataset's relatively limited size and class imbalance could introduce bias and affect the model's ability to generalize. Additionally, the model architecture, while powerful, may be sensitive to variations in image acquisition settings, such as contrast and noise levels inherent in ultrasound imaging. Another limitation observed during training is the fluctuating trend in the loss and accuracy curves, suggesting that the model may still experience instability during optimization. This could be attributed to the learning rate settings, data variability, or the complexity of the architecture. Smoother convergence could potentially be achieved by fine-tuning hyperparameters or employing learning rate schedulers. Furthermore, the interpretability of the model is currently limited to Grad-CAM visualizations. While Grad-CAM provides valuable spatial insights, relying solely on this method may restrict the depth of understanding regarding the model's decision-making process. Future work could incorporate additional interpretability techniques, such as SHAP or LIME, to provide more comprehensive explanations and enhance trust in clinical applications

Therefore, future work should consider external validation using datasets from different populations and imaging devices to assess the model's robustness. Moreover, incorporating more advanced data augmentation techniques, exploring ensemble methods, or experimenting with lighter architectures could further enhance performance. Addressing these limitations openly not only improves the transparency and credibility of the study, but also sets a solid foundation for future research to advance breast cancer detection technology.

V. Conclusion

In this study, the primary aim was to develop a robust deep learning model for breast cancer classification using ultrasound images by enhancing DenseNet-121 with Scaled-Dot Product Attention and Squeeze-and-Excitation, resulting in the Dense-SASE model. The main findings show that Dense-SASE outperforms other models, achieving an accuracy of 98.29%, precision of 97.97%, recall of 98.98%, and an F1-score of 98.44%, indicating excellent classification capability. An additional finding is that Grad-CAM visualizations revealed the model's ability to focus on clinically relevant tissue areas, enhancing its interpretability. Despite these results, the training process showed some fluctuations, suggesting room for improvement in training stability. For future work, expanding the dataset, applying more advanced augmentation techniques, and exploring alternative attention mechanisms may further enhance model performance and generalizability in real-world clinical settings.

Manuscript received March 8, 2024; Accepted May 5, 2025; date of publication May 9, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.779

Acknowledgment

The authors would like to express sincere gratitude to Universitas Sebelas Maret for providing research funding through the Hibah Riset Group scheme as stipulated in contract No. 371/UN27.22/PT.01.03/2025. We would also like to thank the Faculty of Information Technology and Data Science for the invaluable support and resources provided throughout this research. The facilities, academic environment, and continuous encouragement from faculty members have significantly contributed to the completion of this work. This study would not have been possible without the institution's dedication to fostering research and innovation in the field of information technology and data science

References

[1] International Agency for Research on Cancer (IARC), "Global Cancer Observatory: Cancer Today," https://gco.iarc.fr/today/en/dataviz/pie?mode=cancer&group_populations=1&cancers=39&types=0.

[2] P. B. Gordon, "The Impact of Dense Breasts on the Stage of Breast Cancer at Diagnosis: A Review and Options for Supplemental Screening," *Current Oncology*, vol. 29, no. 5, pp. 3595–3636, May 2022, doi: 10.3390/curroncol29050291.

[3] M. Rawashdeh *et al.*, "Breast density awareness and cancer risk in the UAE: Enhancing Women's engagement in early detection," *Radiography*, vol. 31, no. 1, pp. 350–358, Jan. 2025, doi: 10.1016/j.radi.2024.12.012.

[4] Z. He *et al.*, "A review on methods for diagnosis of breast cancer cells and tissues," *Cell Prolif*, vol. 53, no. 7, Jul. 2020, doi: 10.1111/cpr.12822.

[5] W. A. Berg, "Reducing Unnecessary Biopsy and Follow-up of Benign Cystic Breast Lesions," *Radiology*, vol. 295, no. 1, pp. 52–53, Apr. 2020, doi: 10.1148/radiol.2020200037.

[6] S. A. Alshoabi, A. A. Alareqi, F. H. Alhazmi, A. A. Qurashi, A. M. Omer, and A. M. Hamid, "Utility of Ultrasound Imaging Features in Diagnosis of Breast Cancer," *Cureus*, Apr. 2023, doi: 10.7759/cureus.37691.

[7] R. Iacob *et al.*, "Evaluating the Role of Breast Ultrasound in Early Detection of Breast Cancer in Low- and Middle-Income Countries: A Comprehensive Narrative Review," *Bioengineering*, vol. 11, no. 3, p. 262, Mar. 2024, doi: 10.3390/bioengineering11030262.

[8] A. A. Bhatt, D. H. Whaley, and C. U. Lee, "<scp>Ultrasound-Guided</scp> Breast Biopsies," *Journal of Ultrasound in Medicine*, vol. 40, no. 7, pp. 1427–1443, Jul. 2021, doi: 10.1002/jum.15517.

[9] O. Díaz, A. Rodríguez-Ruíz, and I. Sechopoulos, "Artificial Intelligence for breast cancer detection: Technology, challenges, and prospects," *Eur J Radiol*, vol. 175, p. 111457, Jun. 2024, doi: 10.1016/j.ejrad.2024.111457.

[10] C. Trepanier, A. Huang, M. Liu, and R. Ha, "Emerging uses of artificial intelligence in breast and axillary ultrasound," *Clin Imaging*, vol. 100, pp. 64–68, Aug. 2023, doi: 10.1016/j.clinimag.2023.05.007.

[11] J. Egger *et al.*, "Medical deep learning—A systematic meta-review," *Comput Methods Programs Biomed*, vol. 221, p. 106874, Jun. 2022, doi: 10.1016/j.cmpb.2022.106874.

[12] M. Chaieb, M. Azzouz, M. Ben Refifa, and M. Fraj, "Deep learning-driven prediction in healthcare systems: Applying advanced CNNs for enhanced breast cancer detection," *Comput Biol Med*, vol. 189, p. 109858, May 2025, doi: 10.1016/j.compbiomed.2025.109858.

[13] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.

[14] Y. Wang, E. J. Choi, Y. Choi, H. Zhang, G. Y. Jin, and S.-B. Ko, "Breast Cancer Classification in Automated Breast Ultrasound Using Multiview Convolutional Neural Network with Transfer Learning," *Ultrasound Med Biol*, vol. 46, no. 5, pp. 1119–1132, May 2020, doi: 10.1016/j.ultrasmedbio.2020.01.001.

[15] T. Choudhary, V. Mishra, A. Goswami, and J. Sarangapani, "A transfer learning with structured filter pruning approach for improved breast cancer classification on point-of-care devices," *Comput Biol Med*, vol. 134, p. 104432, Jul. 2021, doi: 10.1016/j.compbiomed.2021.104432.

[16] Z. Cao, L. Duan, G. Yang, T. Yue, and Q. Chen, "An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures," *BMC Med Imaging*, vol. 19, no. 1, p. 51, Dec. 2019, doi: 10.1186/s12880-019-0349-x.

[17] T. Zhou, X. Ye, H. Lu, X. Zheng, S. Qiu, and Y. Liu, "Dense Convolutional Network and Its Application in Medical Image Analysis," *Biomed Res Int*, vol. 2022, no. 1, Jan. 2022, doi: 10.1155/2022/2384830.

[18] X. Li *et al.*, "Deep Learning Attention Mechanism in Medical Image Analysis: Basics and Beyonds," *International Journal of Network Dynamics and*

Manuscript received March 8, 2024; Accepted May 5, 2025; date of publication May 9, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.779

Intelligence, pp. 93–116, Mar. 2023, doi: 10.53941/ijndi0201006.

[19] J. Li, J. Shi, J. Chen, Z. Du, and L. Huang, "Self-attention random forest for breast cancer image classification," *Front Oncol*, vol. 13, Feb. 2023, doi: 10.3389/fonc.2023.1043463.

[20] Z. Li, L. Yuan, H. Xu, R. Cheng, and X. Wen, "Deep Multi-Instance Learning with Induced Self-Attention for Medical Image Classification," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, Dec. 2020, pp. 446–450. doi: 10.1109/BIBM49941.2020.9313518.

[21] J. Deng, Y. Ma, D. Li, J. Zhao, Y. Liu, and H. Zhang, "Classification of breast density categories based on SE-Attention neural networks," *Comput Methods Programs Biomed*, vol. 193, p. 105489, Sep. 2020, doi: 10.1016/j.cmpb.2020.105489.

[22] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2018, pp. 7132–7141. doi: 10.1109/CVPR.2018.00745.

[23] K. Fukitani *et al.*, "3D object detection using improved PointRCNN," *Cognitive Robotics*, vol. 2, pp. 242–254, 2022, doi: 10.1016/j.cogr.2022.12.001.

[24] X. Zhang *et al.*, "SERNet: Squeeze and Excitation Residual Network for Semantic Segmentation of High-Resolution Remote Sensing Images," *Remote Sens (Basel)*, vol. 14, no. 19, p. 4770, Sep. 2022, doi: 10.3390/rs14194770.

[25] K. Munishamaiaha *et al.*, "Robust Spatial–Spectral Squeeze–Excitation AdaBound Dense Network (SE-AB-Densenet) for Hyperspectral Image Classification," *Sensors*, vol. 22, no. 9, p. 3229, Apr. 2022, doi: 10.3390/s22093229.

[26] W. K. Moon, Y.-W. Lee, H.-H. Ke, S. H. Lee, C.-S. Huang, and R.-F. Chang, "Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks," *Comput Methods Programs Biomed*, vol. 190, p. 105361, Jul. 2020, doi: 10.1016/j.cmpb.2020.105361.

[27] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data Brief*, vol. 28, p. 104863, Feb. 2020, doi: 10.1016/j.dib.2019.104863.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[29] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jul. 2017, pp. 2261–2269. doi: 10.1109/CVPR.2017.243.

[30] A. Vaswani *et al.*, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, in NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, pp. 6000–6010.

[31] V. Nguyen, "Bayesian Optimization for Accelerating Hyper-Parameter Tuning," in *2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, IEEE, Jun. 2019, pp. 302–305. doi: 10.1109/AIKE.2019.00060.

[32] I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis, and N. Doulamis, "Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem," *Technologies (Basel)*, vol. 9, no. 4, p. 81, Nov. 2021, doi: 10.3390/technologies9040081.

[33] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, Jan. 2021, doi: 10.1016/j.aci.2018.08.003.

[34] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *Int J Comput Vis*, vol. 128, no. 2, pp. 336–359, Feb. 2020, doi: 10.1007/s11263-019-01228-7.

[35] R. Karthik, R. Menaka, G. S. Kathiresan, M. Anirudh, and M. Nagharjun, "Gaussian Dropout Based Stacked Ensemble CNN for Classification of Breast Tumor in Ultrasound Images," *IRBM*, vol. 43, no. 6, pp. 715–733, Dec. 2022, doi: 10.1016/j.irbm.2021.10.002.

[36] S. Armoogum, K. Motean, D. A. Dewi, T. B. Kurniawan, and J. Kijsomporn, "Breast Cancer Prediction Using Transfer Learning-Based Classification Model," *Emerging Science Journal*, vol. 8, no. 6, pp. 2373–2384, Dec. 2024, doi: 10.28991/ESJ-2024-08-06-014.

[37] M. R. Islam *et al.*, "Enhancing breast cancer segmentation and classification: An Ensemble Deep Convolutional Neural Network and U-net approach on ultrasound images," *Machine Learning with Applications*, vol. 16, p. 100555, Jun. 2024, doi: 10.1016/j.mlwa.2024.100555.

[38] F. Taheri and K. Rahbar, "Improving breast cancer classification in fine-grain ultrasound images through feature discrimination and a transfer learning approach," *Biomed Signal*

Manuscript received March 8, 2024; Accepted May 5, 2025; date of publication May 9, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.779

Process Control, vol. 106, p. 107690, Aug. 2025, doi: 10.1016/j.bspc.2025.107690.
[39] A. Sahu, P. K. Das, and S. Meher, "High accuracy hybrid CNN classifiers for breast cancer detection using mammogram and ultrasound datasets," *Biomed Signal Process Control*, vol. 80, p. 104292, Feb. 2023, doi: 10.1016/j.bspc.2022.104292.
[40] M. D. Ali *et al.*, "Breast Cancer Classification through Meta-Learning Ensemble Technique Using Convolution Neural Networks," *Diagnostics*, vol. 13, no. 13, p. 2242, Jun. 2023, doi: 10.3390/diagnostics13132242.

Author Biography

Hanina Nafisa Azka is an undergraduate student at Universitas Sebelas Maret, Surakarta, majoring in Informatics. Her academic journey focuses on machine learning, computer vision, and medical image processing. She has been actively involved in research projects, including the development of a breast cancer classification model utilizing ultrasound images, enhanced with attention mechanisms to improve performance. Hanina has also contributed as a teaching assistant in several courses, such as Digital Systems, Operating Systems, and Software Engineering. Her work extends to participating in innovation programs like PKM, reflecting her passion for technological advancements and product development in the healthcare domain.

Wiharto is a senior lecturer in the Informatics Department, Faculty of Information Technology and Data Science at Universitas Sebelas Maret, Surakarta. He has a strong academic background and extensive experience in the field of Computational Science & Engineering. Wiharto earned his doctoral degree from Universitas Gadjah Mada, where he specialized in BioMedical Informatics. His research primarily focuses on Bio-Medical Informatics, Artificial Intelligence, and Computational Intelligence, with an emphasis on solving real-world problems through innovative computational approaches. His contributions extend to mentoring students, publishing impactful research, and collaborating on interdisciplinary projects to advance the fields of health informatics and artificial intelligence.

Esti Suryani is a dedicated researcher in the Computational Science and Engineering research group. She holds a Master's degree in Computer Science from Universitas Gadjah Mada, where she cultivated a strong foundation in advanced computing methodologies. Her research interests span across various domains, including data analysis and the development of computational models to solve complex, real-world engineering problems. With a passion for interdisciplinary research, she collaborates with experts from diverse fields to explore innovative solutions, particularly focusing on optimizing algorithms and leveraging data-driven approaches. She also involved in mentoring students and guiding research projects, fostering the next generation of computational scientists.

Manuscript received March 8, 2024; Accepted May 5, 2025; date of publication May 9, 2025
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v7i3.779