Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 3, Juni 2025, Page. 1131-1146
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4397

Comparative Analysis of Augmentation and Filtering Methods in VGG19 and DenseNet121 for Breast Cancer Classification

I Kadek Seneng*[1], Putu Desiana Wulaning Ayu[2], and Roy Rudolf Huizen[3]

[1]Magister Program, Department of Magister Information System, Institut Teknologi dan Bisnis STIKOM Bali, Indonesia
[2,3]Department of Magister Information System, Institut Teknologi dan Bisnis STIKOM Bali, Indonesia
Email: [1] ikadekseneng13@gmail.com

Received: Jun 30, 2025; Revised: Mar 11, 2025; Accepted: Mar 12, 2025; Published: Jun 10, 2025

Abstract

Breast cancer is one of the most prevalent malignancies and a leading cause of mortality among women worldwide. Mammography plays a crucial role in early detection, yet challenges in manual interpretation have led to the adoption of Convolutional Neural Networks (CNNs) to improve classification accuracy. This study evaluates the performance of Visual Geometry Group (VGG19) and Densely Connected Convolutional Networks (DenseNet121) in mammogram classification. It examines the impact of data augmentation and image enhancement techniques, including Contrast-Limited Adaptive Histogram Equalization (CLAHE), Median Filtering, and Discrete Wavelet Transform (DWT), as well as the influence of varying epochs and learning rates. A novel approach is introduced by assessing data augmentation effectiveness and exploring model adaptations, such as layer incorporation and freezing during training. Classification performance is enhanced through fine-tuning strategies combined with image enhancement techniques, reducing reliance on data augmentation. These findings contribute to medical imaging and computer science by demonstrating how CNN modifications and enhancement methods improve mammogram classification, providing insights for developing robust deep learning-based diagnostic models. The highest performance was achieved using VGG19 with DWT, a learning rate of 0.0001, and 20 epochs, yielding 98.04% accuracy, 98.11% precision, 98% recall, and a 97.99% F1-score. Data augmentation did not consistently enhance results, particularly in clean datasets. Increasing epochs from 10 to 20 improved accuracy, but performance declined at 30 epochs. The confusion matrix showed high accuracy for Benign (100%) and Cancer (99.5%), with more misclassifications in the Normal class (94.5%).

Keywords: *Augmentation , Breast Cancer, Classification , Mammograms, VGG19*

1. INTRODUCTION

Cancer is a condition characterized by the abnormal growth of cells in the body, often caused by genetic mutations. This condition can arise due to various physical, chemical, or environmental factors[1]. One of the most concerning types of cancer is breast cancer, which often begins as either benign or malignant tumors. Benign tumors, such as fibroadenomas or cysts, do not spread to surrounding tissues and are generally not life-threatening. In contrast, malignant tumors are invasive and can spread to nearby tissues, posing a significant threat to life. Key risk factors for breast cancer have been identified in a study conducted at Dr. Moewardi Hospital, including age over 40 years, a family history of breast cancer, alcohol consumption, obesity, and exposure to cigarette smoke, all of which have been associated with an increased risk of developing the disease[2]. A study conducted at RSUD Prof. Dr. H. Aloei Saboe, Gorontalo, revealed that genetic factors, menarche before the age of 12, prolonged use of hormonal contraceptives, and early menopause are significantly correlated with breast cancer, emphasizing the importance of early detection and public education in its prevention[3].

Other studies also highlighted the relationship between dietary patterns and breast cancer risk has been highlighted in previous studies. It has been suggested that a high-carbohydrate diet combined with calorie restriction may be associated with a lower risk of breast cancer. Both genetic predisposition and

Vol. 6, No. 3, Juni 2025, Page. 1131-1146
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4397

dietary factors seem to play a role in breast cancer development, and adopting a low-calorie diet with a high carbohydrate ratio may help mitigate certain genetic risk factors[4]. According data from the Global Burden of Cancer (GLOBOCAN) 2020 indicate that breast cancer was the most frequently diagnosed cancer worldwide, with over 2.3 million new cases reported in 2020. This cancer was responsible for approximately 16% of all cancer-related deaths among women, with an estimated 685,000 fatalities recorded that year[5]. The International Agency for Research on Cancer, a part of the World Health Organization (WHO), estimated that in 2020, there were 65,858 new cases of breast cancer among women in Indonesia, resulting in 22,430 deaths due to the disease[6].

Mammography is a medical procedure that utilizes X-rays to detect abnormalities in breast tissue, including breast cancer. This procedure involves compressing the breast to produce clear and detailed images, which are then examined by radiologists to identify early signs of cancer, such as lumps or other abnormal changes[7]. Breast cancer mortality rates have been significantly reduced through the utilization of mammograms[8]. In developed countries, such as the United States, mammography screening is widely implemented, particularly among women aged 40 and older[9]. Early diagnosis significantly improves survival rates, reinforcing the importance of breast cancer screening programs[8].

Despite the effectiveness of mammography in early breast cancer detection, the manual interpretation of mammogram images presents several challenges. The complexity of mammographic images requires a significant amount of time and a high level of expertise from medical professionals to achieve an accurate diagnosis. Breast tumors and cancers are particularly challenging to detect and treat due to their variability[10], this has led to the development of artificial intelligence (AI)-based technologies to support medical image analysis. The potential of AI in prognosis and therapeutic response prediction has been demonstrated through the direct linkage of histological tumor features to analysis models[11]. By processing vast amounts of data, patterns and abnormalities that may not be detectable to the human eye can be identified by AI algorithms, facilitating earlier cancer detection, improved diagnostic accuracy, and the development of personalized treatment plans for patients[12].

Convolutional Neural Networks (CNNs) have been extensively utilized for medical image classification, including breast cancer detection. However, ongoing debates persist regarding the effectiveness of various CNN architectures, emphasizing the necessity of understanding the impact of data augmentation techniques on CNN model performance in enhancing classification accuracy. Numerous studies have explored the application of AI, particularly deep learning, in automatic mammogram classification using CNN architectures, incorporating data augmentation techniques and architecture comparisons. For instance, a study by Aditi Kajala, Sandeep Jaiswal, and Rajesh Kumar (2023), titled Comparative Analysis of CNN Architectures for Breast Histopathology Image Classification, aimed to compare the performance of various CNN architectures in classifying breast histopathology images[13]. Similarly, a study by Deshan Fonseka and Christos Chrysoulas (2020), titled Data Augmentation for Enhancing CNN Performance in Image Classification, discussed the application of CNNs on small datasets, exploring data augmentation, feature extraction, and fine-tuning to improve model accuracy[14]. Data augmentation plays a crucial role in improving model generalization and performance across different datasets and architectures[15].

Previous studies have implemented data augmentation techniques on the Digital Database for Screening Mammography (DDSM) dataset and its derivatives, such as Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM) and Miniature Digital Database for Screening Mammography (Mini-DDSM), to enhance accuracy in mammogram classification. Youness et al. (2024) applied augmentation techniques such as rotation, flipping, and zooming on the CBIS-DDSM dataset, consisting of 10,239 images, achieving 76.71% accuracy with the DenseNet121 model[16]. Sanassi et al. (2024) augmented the CBIS-DDSM dataset (4,560 images) using rotation at various angles and a 3×3 median filter, improving DenseNet121's accuracy to 94.36%[17]. Meanwhile, Subasish et al. (2022) utilized the Mini-DDSM dataset (9,752 images) with rotation, flipping, zooming, and a 3×3 median filter, but only achieved 65% accuracy with VGG16[18]. Conversely, Sumyat Thwin et al. (2024) employed rotation, flipping, and scaling on the DDSM dataset (55,890 images), successfully increasing the accuracy of the ECS-A-Bersih model to 96.50%[19]. However, it has been indicated by several studies that data augmentation should be carefully adapted to the specific characteristics of the images utilized, as accuracy improvements are not always achieve. While data augmentation can enhance average accuracy, its application can reduce accuracy in certain classes by up to 20%,

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 3, Juni 2025, Page. 1131-1146
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4397

particularly in ambiguous or highly similar classes, exacerbating misclassification errors[20]. Excessive augmentation in skin lesion images has been shown to cause models to learn irrelevant features, ultimately decreasing classification accuracy, highlighting that improper augmentation can introduce noise and artifacts that disrupt model learning[21].

Image enhancement techniques also play a critical role in improving breast cancer classification accuracy, such as the application of CLAHE to enhance contrast while preserving image details. The use of this technique in mammogram preprocessing can significantly improve the performance of CNN-based CAD systems, making image enhancement methods crucial for achieving more accurate detection[22]. Denoising techniques are widely implemented to enhance accuracy, particularly in medical image analysis, as research has demonstrated that denoising can clarify important details, thereby supporting more precise and reliable diagnoses. For example, Rasheed et al. (2022) employed hybrid filtering that combines median filtering and statistical techniques to reduce salt-and-pepper noise, improving image quality and PSNR performance[23]. Christudhas & Fathima (2024) developed a Modified Min-Max Median (MMM) Filter using a Tritonic Sorter to detect and process noisy pixels, minimizing excessive smoothing while preserving important image details[24]. Meanwhile, Qu et al. (2024) optimized the Black Widow Optimization (BWOA) algorithm with Tent mapping and a combination of median, mean, Gaussian, and bilateral filters, significantly improving medical image quality[25].

This study investigates the implementation of two Convolutional Neural Network (CNN) architectures, VGG19 and DenseNet121, for mammogram classification in breast cancer detection. It provides an in-depth analysis of the effects of various data augmentation techniques both individually and in combination on the Mini-DDSM dataset, evaluating their impact on CNN model performance. Additionally, this study examines the role of image enhancement techniques, such as noise removal, in improving classification accuracy. The key contribution of this research lies in optimizing VGG19 and DenseNet121 architectures through layer freezing and adding new layers such as Flatten, Dense, and Dropout to enhance model generalization. This study also explores the impact of learning rate adjustments and increased epoch counts, conducting multiple experiments to identify the optimal parameter combinations for improving model performance.

This approach demonstrates that achieving high accuracy is not solely dependent on data augmentation but can also be attained through a combination of model refinement strategies, optimal image preprocessing techniques, and well-tuned training parameters. By utilizing the Mini-Digital Database for Screening Mammography (DDSM) dataset, this research aims not only to determine the best-performing architecture and training configurations but also to provide new insights into the application of deep learning in breast cancer classification.

1. METHOD

This study evaluates the performance of two Convolutional Neural Network (CNN) architectures, VGG19 and DenseNet121, in classifying mammogram images from the Mini-DDSM dataset. The model performance is comprehensively assessed using metrics such as accuracy, precision, sensitivity (recall), and F1-score to ensure more accurate and efficient classification results. The research workflow is illustrated in Figure 1.

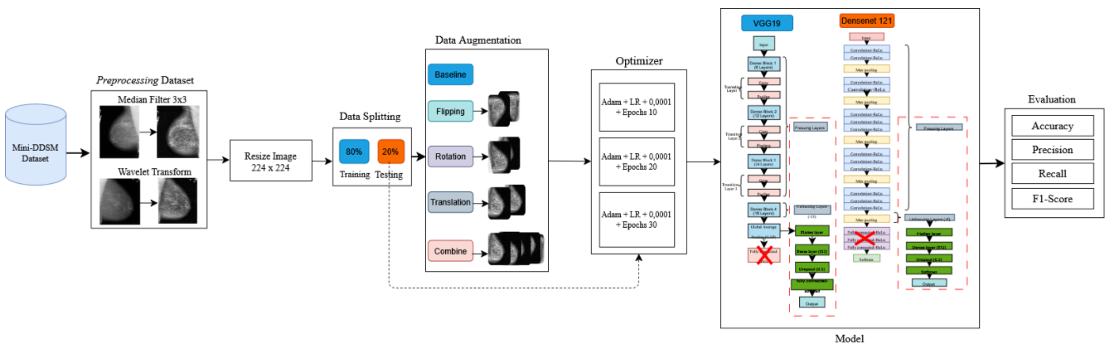


Figure 1. Research Methods

P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 3, Juni 2025, Page. 1131-1146
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4397

The study consists of several key stages for mammogram image classification: data collection, preprocessing, data augmentation, modeling, and evaluation, as depicted in Figure 1.

1.1. Mini DDSM Dataset

The data used in this study is derived from the Database for Screening Mammography (Mini-DDSM), a simplified version of the Digital Database for Screening Mammography (DDSM). Mini-DDSM is widely utilized in research on breast cancer detection and diagnosis using mammography. It was developed to improve accessibility and usability by reducing the size of the DDSM dataset without compromising data quality[26]. The Mini-DDSM dataset consists of three classes: Normal, Benign, and Cancer, as shown in Figure 2.

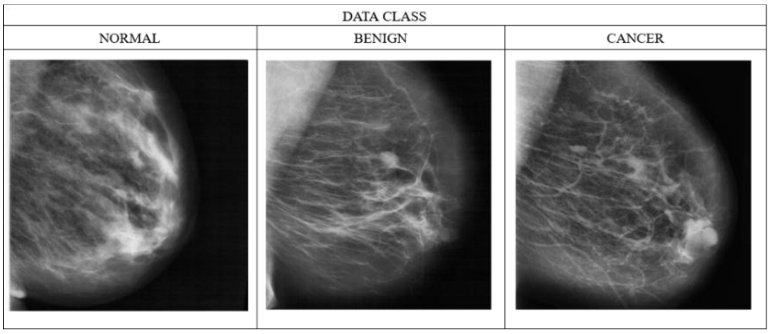


Figure 2. Mini-DDSM Class Label

This study only utilizes MLO (Medio-Lateral Oblique) images of the left and right breasts, with 200 images from each class, resulting in a total of 600 images. This approach was taken to reduce the dataset size, making it suitable for analyzing the effects of data augmentation.

1.2. Preprocessing

Preprocessing data refers to the process of transforming raw data into quality data through steps such as normalization, outlier detection, missing data imputation, and dimensionality reduction. This process aims to ensure data compatibility as input for AI models and improve the quality of analytical outcomes[27].

1.2.1. Filtering

Image preprocessing in this study employs Contrast-Limited Adaptive Histogram Equalization (CLAHE) to enhance image quality by improving local contrast and highlighting details in areas with low contrast. However, its application may also introduce noise[26]. The focus of this preprocessing step is on noise removal using Discrete Wavelet Transform (DWT) and a 3x3 Median Filter, which are combined with CLAHE and subsequently compared for analysis. This approach leverages CLAHE to adaptively enhance contrast, DWT for better noise separation and multi-resolution analysis, and the Median Filter to effectively reduce noise without blurring edges. The method provides a balance between enhancing image clarity and preserving important details, making it highly effective for comparative performance analysis in image enhancement tasks.

The formula for calculating CLAHE is defined as:

$$\beta = \frac{p}{q} 1 + \left(1 + \frac{a}{100} S_{max}\right) \quad (1)$$

Where β is the result of CLAHE, p is the number of pixels in each block, q is the dynamic range of the block, S_{max} is the maximum slope, and α is the clipping factor[28]

The formula for calculating DWT is defined as:

$$DWT(k) = \frac{1}{\sqrt{a_0^m}} + \sum_k f(k) \psi\left(\frac{k - n b_0 a_0^m}{a_0^m}\right) \quad (2)$$

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 3, Juni 2025, Page. 1131-1146
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4397

Where a_0 represents the scale parameter, and b_0 represents the translation parameter. Additionally, both *mm* and *nn* are integers belonging to the set of integers Z. The discrete ECG signal is then represented as *f(k)* [29].

The 2D median filter is applied on I^2 as:

$$I^2(i,j) = median\{I^2(i,j)|(i,j) \in w\} \quad (3)$$

where, I^2 is outcome of the median filtering and w is the size of the window. The neighbour's pattern is decided by the size of the window, the window size of a 3 × 3 neighborhood has been used inthis work[30].

Numerous studies have employed DWT and median filters as primary methods for noise removal in digital images. The median filter is particularly effective in addressing salt-and-pepper noise, as it preserves important details such as edges and corners. On the other hand, DWT allows for the separation of frequency components, reducing noise without sacrificing image detail quality[31]. Furthermore, CLAHE has been applied in various studies to enhance image quality, especially in medical imaging, by improving contrast, making previously difficult-to-see details clearer[32]. Examples of images processed with CLAHE, DWT, and the Median Filter are shown in Figure 3.

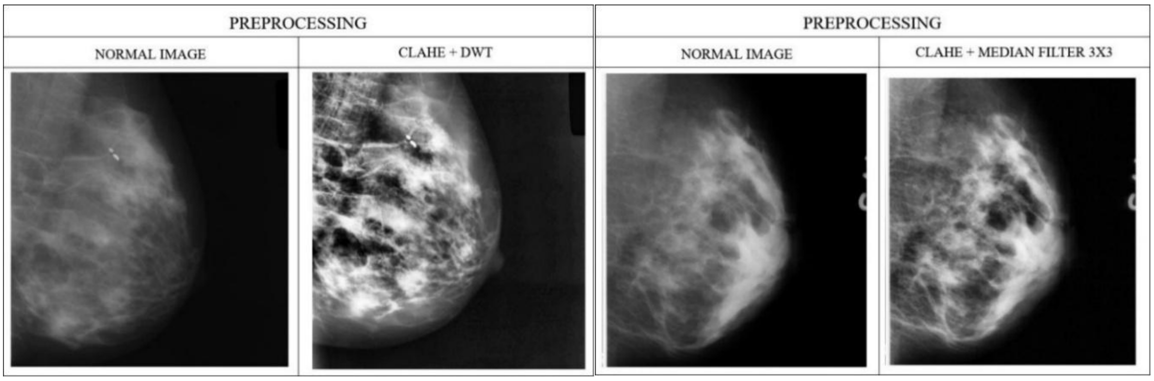


Figure 3. Daubechies Wavelet Transform Output and Median Filter 3x3

1.2.2. Resize Image

In image classification studies using CNN architectures such as VGG19 and DenseNet121, resizing images to 224x224 pixels is performed to ensure compatibility with model architectures designed to accept inputs of these dimensions. Additionally, this size aligns with the parameters of pre-trained models on large datasets such as ImageNet, enabling the use of optimal initial weights. This resizing process also reduces computational load, particularly for mammogram images, which typically have high resolutions, without losing critical information. By standardizing the image size, the model can consistently recognize patterns and features across the dataset, thereby improving training efficiency and accuracy in detecting breast cancer[33]. This makes resizing a crucial step in preparing data for processing by CNNs. This study adopted a data-splitting approach with 80% allocated for training and 20% for testing. Such a strategy is widely applied by researchers in various studies[34]. One of the main advantages of this method is its ability to provide more accurate estimations for the testing data compared to using the entire dataset solely for training. Furthermore, the 80%-20% split ratio has been demonstrated to yield better performance compared to other partitioning proportions[35].

1.2.3. Augmentation

Data augmentation is a technique used to increase the diversity of training data by modifying the original data without changing its labels. By employing data augmentation, the model can learn from a wider variety of images, thereby improving classification accuracy[18]. Several studies have emphasized that data augmentation is crucial in image-based research with small datasets, as it helps enhance training data diversity, reduce overfitting, and improve model generalization. Techniques such as rotation, flipping, zooming, and adding noise generate new variations[36],[37].

Vol. 6, No. 3, Juni 2025, Page. 1131-1146
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4397

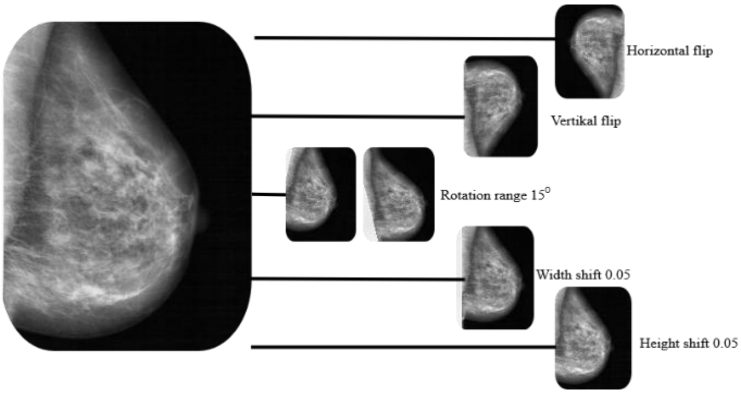


Figure 4. Data Augmentation

Figure 4 illustrates the three data augmentation techniques used in this study: flip augmentation, where images are flipped horizontally and vertically to provide variations in perspective, allowing the model to learn a broader range of patterns; rotation by 15°, which simulates variations in image orientation and reflects potential differences in capture angles in real-world scenarios; and translation augmentation, which shifts the image to ensure the model can still recognize objects even when their relative positions change within the frame. Additionally, a combined augmentation approach incorporating all three techniques is used to evaluate whether the combination enhances model performance more effectively than applying each technique individually.

1.2.4. Modification of VGG 19 and DenseNet121 model

VGG-19 is a neural network architecture in deep learning consisting of 19 layers, including 16 convolutional layers designed to extract image features such as patterns and textures, and 3 fully connected layers for classifying images based on the extracted features. Max-pooling layers are employed to reduce data dimensions, improve computational efficiency, and prevent overfitting without compromising essential information. This combination of layers enables VGG-19 to analyze and classify images with high accuracy, making it widely popular for various image analysis applications[38],[39]. In this study, the VGG-19 model was modified by freezing most layers to ensure that their parameters remained unchanged during training. Additional layers were also incorporated, to further enhance the model's performance.

In this implementation, only the last six layers are left unfrozen, while the remaining layers are frozen. The frozen layers include the first 13 convolutional layers (from five blocks) and all pooling layers. In addition to freezing the layers, several new layers were added after the feature extractor part of VGG-19, as shown in Figure 5, these additions include:

1. Flatten layer
2. Dense layer (512 neurons)
3. Dropout layer (rate 0.5)
4. Dense layer (output)

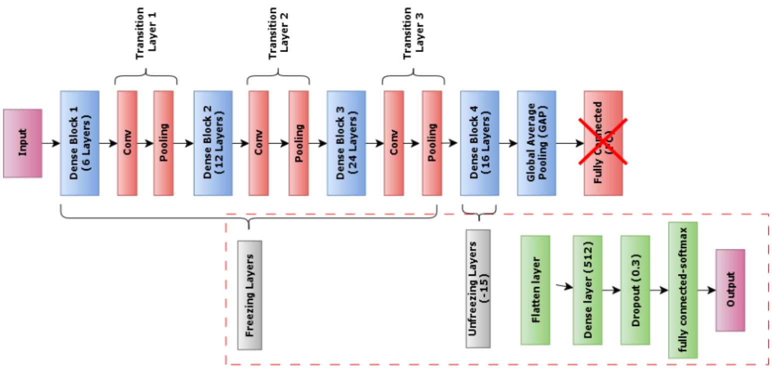


Figure 5. Modified VGG19 Model

Vol. 6, No. 3, Juni 2025, Page. 1131-1146
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4397

DenseNet121 is a deep learning neural network architecture consisting of 121 layers, including 108 convolutional layers that function to extract image features by connecting each layer to all previous layers. This connectivity allows for feature reuse and reduces the number of parameters required[40],[41]. In this study, DenseNet121 was modified by freezing most of the layers, with only the last four layers left unfrozen.

In addition to freezing the layers, as shown in Figure 6 several new layers were added, including:

1. Flatten layer
2. Dense layer (512 neurons)
3. Dropout layer (rate 0.5)
4. Dense layer (output)

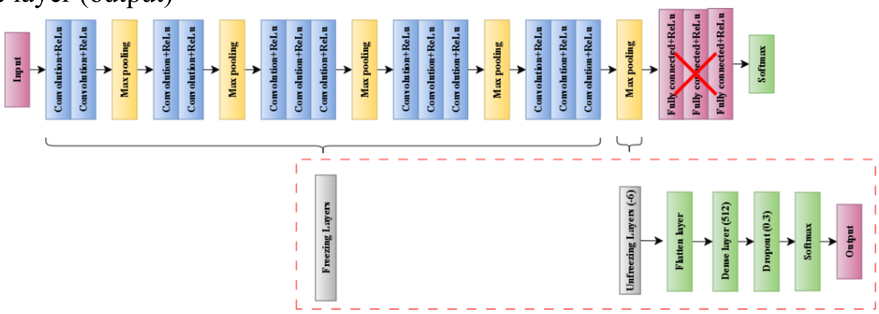


Figure 6. Modified DenseNet121 Model

At this stage, the model's performance is evaluated using a confusion matrix. Several parameters are employed, including accuracy, recall, precision, and F1 score. This evaluation aims to identify the best augmentation techniques, image quality enhancement methods, and architectures for breast cancer classification on mammogram images. These evaluation metrics provide diverse approaches to assessing the performance of classification models. Accuracy offers an overall measure of the model's correct predictions but is sensitive to class imbalance. Recall assesses the model's ability to capture all instances of a specific class, making it crucial for minimizing omission errors. Precision evaluates the accuracy of positive predictions, helping to avoid an excessive number of false positives. F1 score, a harmonic combination of precision and recall, is ideal for assessing model performance on imbalanced datasets. Together, these metrics facilitate a comprehensive analysis of the model's strengths and weaknesses across various application contexts[42].

2. RESULT

The testing was conducted progressively based on combinations of models (VGG19 or DenseNet121), noise removal methods (DWT or Median Filter), and data augmentation techniques. The initial hyperparameter settings included 10 and 20 epochs with a learning rate of 0.0001 for all scenarios, including the baseline and all types of augmentation. The best-performing combinations were further tested by increasing the number of epochs to 30 and the learning rate to 0.0005 to optimize model performance. This approach ensured that every combination of preprocessing, augmentation, and model parameters was comprehensively evaluated, providing valuable insights into the contribution of each factor to the model's accuracy in breast cancer classification.

2.1.1. Testing Model

Testing began with the baseline, which involved no augmentation, to understand the model's basic performance with preprocessing methods such as DWT and the 3x3 Median Filter. Subsequently, the flip augmentation technique was applied, where the images were flipped horizontally and vertically to provide important variations in perspective, allowing the model to learn broader patterns. A 15° rotation was performed to simulate variations in image orientation, reflecting potential differences in capture angles in real-world scenarios. Finally, translation augmentation was applied to ensure the model could still recognize objects, even if their relative position shifted within the frame.

Table 1 and Table 2, present the results of the model combination testing, which consists of eight main scenarios. These model combinations involve two CNN architectures, VGG19 and DenseNet121, with two different preprocessing methods: DWT and Median Filter 3x3. Each combination was tested

Vol. 6, No. 3, Juni 2025, Page. 1131-1146
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4397

using a learning rate (LR) of 0.0001 and two epoch settings, 10 and 20, resulting in eight primary testing configurations:

1. VGG19 + DWT + LR 0.0001 + Epoch 10
2. VGG19 + DWT + LR 0.0001 + Epoch 20
3. VGG19 + Median Filter 3x3 + LR 0.0001 + Epoch 10
4. VGG19 + Median Filter 3x3 + LR 0.0001 + Epoch 20
5. DenseNet121 + DWT + LR 0.0001 + Epoch 10
6. DenseNet121 + DWT + LR 0.0001 + Epoch 20
7. DenseNet121 + Median Filter 3x3 + LR 0.0001 + Epoch 10
8. DenseNet121 + Median Filter 3x3 + LR 0.0001 + Epoch 20

Additionally, these models were tested on data with five different approaches: augmentation (Rotation, Flip, and Translation), no augmentation (baseline), and combinations of augmentations. Each main scenario was tested with three augmentation approaches, bringing the total number of test scenarios to 40.

Table 1. Test Results of the VGG19 Architecture

Architecture	Image Enhancement	Augmentation	Epoch	LR	Accuracy	precision	Recall	F1 Score
VGG19	DWT	Baseline	10	0.0001	97.71	97.77	97.67	97.66
VGG19	DWT	Flip	10	0.0001	93.96	93.96	93.90	93.90'
VGG19	DWT	Rotate 15°	10	0.0001	96.08	96.27	96.04	96.02
VGG19	DWT	Translation	10	0.0001	95.92	96.05	95.90	95.89
VGG19	DWT	Combined	10	0.0001	95.92	95.98	95.90	95.89
VGG19	**DWT**	**Baseline**	**20**	**0.0001**	**98.04**	**98.11**	**98.00**	**97.99**
VGG19	DWT	Flip	20	0.0001	97.76	97.76	97.68	97.67
VGG19	DWT	Rotate 15°	20	0.0001	97.59	97.59	97.52	97.51
VGG19	DWT	Translation	20	0.0001	96.85	96.85	96.84	96.84
VGG19	DWT	Combined	20	0.0001	96.26	96.26	96.18	96.20'
VGG19	Median 3x3	Baseline	10	0.0001	96.57	96.63	96.53	96.54
VGG19	Median 3x3	Flip	10	0.0001	91.68	91.61	91.63	91.62
VGG19	Median 3x3	Rotate 15°	10	0.0001	93.80	93.84	93.75	93.77
VGG19	Median 3x3	Translation	10	0.0001	93.31	93.78	93.19	93.23
VGG19	Median 3x3	Combined	10	0.0001	78.79	82.53	78.91	78.99
VGG19	Median 3x3	Baseline	20	0.0001	95.43	95.38	95.38	95.38
VGG19	Median 3x3	Flip	20	0.0001	92.16	92.23	92.03	91.99
VGG19	Median 3x3	Rotate 15°	20	0.0001	96.08	96.22	96.03	96.04
VGG19	Median 3x3	Translation	20	0.0001	95.75	95.87	95.72	95.73
VGG19	Median 3x3	Combined	20	0.0001	94.45	94.44	94.41	94.4

Table 2. Test Results of the DenseNet121 Architecture

Architecture	Image Enhancement	Augmentation	Epoch	LR	Accuracy	precision	Recall	F1 Score
Densenet121	DWT	Baseline	10	0.0001	96.41	96.43	96.34	96.34
Densenet121	DWT	Flip	10	0.0001	89.23	89.89	89.1	89.05
Densenet121	DWT	Rotate 15°	10	0.0001	96.08	96.04	93.87	93.87
Densenet121	DWT	Translation	10	0.0001	94.61	95.02	96.00'	96.01
Densenet121	DWT	Combined	10	0.0001	92.16	91.57	91.34	91.31
Densenet121	DWT	Baseline	20	0.0001	96.73	96.71	96.68	96.68
Densenet121	DWT	Flip	20	0.0001	96.41	96.38	96.36	96.35

Vol. 6, No. 3, Juni 2025, Page. 1131-1146
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4397

Densenet121	DWT	Rotate 15°	20	0.0001	96.24	96.29	96.01	96.01
Densenet121	DWT	Translation	20	0.0001	96.24	96.23	94.52	94.53
Densenet121	DWT	Combined	20	0.0001	94.94	92.14	92.05	92.03
Densenet121	**Median 3x3**	**Baseline**	**10**	**0.0001**	**96.90**	**96.91**	**96.87**	**96.85**
Densenet121	Median 3x3	Flip	10	0.0001	90.70	90.98	90.70'	90.71
Densenet121	Median 3x3	Rotate 15°	10	0.0001	94.94	94.98	94.89	94.90'
Densenet121	Median 3x3	Translation	10	0.0001	94.94	94.97	94.93	94.92
Densenet121	Median 3x3	Combined	10	0.0001	89.88	90.12	89.82	89.82
Densenet121	Median 3x3	Baseline	20	0.0001	96.83	96.9	96.83	96.83
Densenet121	Median 3x3	Flip	20	0.0001	95.16	95.38	95.16	95.18
Densenet121	Median 3x3	Rotate 15°	20	0.0001	95.99	96.17	96	96.01
Densenet121	Median 3x3	Translation	20	0.0001	96.66	96.78	96.66	96.66
Densenet121	Median 3x3	Combined	20	0.0001	95.33	95.37	95.33	95.32

Based on the test results in Table 1 and Table 2, the baseline model (without augmentation) consistently showed the best performance compared to models using individual or combined augmentation techniques. The model with the configuration VGG19 + DWT + LR 0.0001 at 20 epochs achieved the highest accuracy of 98.04%, with supporting metrics of precision 98.11%, recall 98%, and F1 score 97.99%. This indicates that data augmentation does not always improve model performance, particularly when the dataset is already clean and sufficiently representative. The superior performance of the baseline model may be due to the minimal additional interference from augmentation, allowing the model to focus on learning the patterns within the original data. The increase in epochs from 10 to 20 also contributed to the improvement in accuracy. Given the excellent results, the next step in this testing will involve increasing the number of epochs and raising the learning rate to 0.0005 to evaluate whether these adjustments can further enhance the model's performance without causing overfitting. The results of these tests are presented in Table 3.

Table 3. Further Testing for the Best Model

Architecture	Image Enhancement	Augmentation	Epoch	LR	Accuracy	precision	Recall	F1 Score
VGG19	**DWT**	**Baseline**	**20**	**0.0001**	**98.04**	**98.11**	**98.00**	**97.99**
VGG19	DWT	Baseline	30	0.0001	97	97.05	97	96.99
VGG19	DWT	Baseline	20	0.0005	94.99	95	95	94.99
VGG19	DWT	Baseline	30	0.0005	96.16	96.23	96.16	96.16

Table 3. shows a decline in model performance after 30 epochs compared to 20 epochs, as indicated by the decrease in Accuracy, Precision, Recall, and F1 score. At epoch 20 with a learning rate of 0.0001, the model achieved an accuracy of 98.04%. However, when the number of epochs was increased to 30, accuracy dropped to 97%, precision decreased from 98.11% to 97.05%, recall declined from 98.00% to 97%, and F1 score reduced from 97.99% to 96.99%. The decreases were 1.06% for Accuracy, 1.08% for Precision, 1.02% for Recall, and 1.02% for F1 score. In the experiment with a learning rate of 0.0005 and 20 epochs, the model achieved an accuracy of 94.99%, precision of 95.00%, recall of 95.00%, and F1 score of 94.99%, showing a decline compared to the previous experiment. The decrease was 3.05% in Accuracy, 3.11% in Precision, 3.00% in Recall, and 3.00% in F1 score. This result suggests that a higher learning rate can accelerate convergence but risks reducing the final model performance.

In the experiment with a learning rate of 0.0005 and 30 epochs, the model achieved an accuracy of 96.16%, precision of 96.23%, recall of 96.16%, and F1 score of 96.16%, showing a performance decline. The decrease was 0.84% in Accuracy, 0.82% in Precision, 0.84% in Recall, and 0.83% in F1 score. This indicates that although increasing the number of epochs provides more time for the model to learn, a higher learning rate can still hinder overall model performance. The best-performing model

Vol. 6, No. 3, Juni 2025, Page. 1131-1146
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4397

proposed is the combination of VGG19 + DWT + Learning Rate 0.0001, which achieved the highest accuracy of 98.04% at epoch 20. Additionally, the model's performance was supported by other evaluation metrics: precision of 98.11%, recall of 98%, and F1 score of 97.99%, demonstrating the model's excellent ability to handle the classification task.

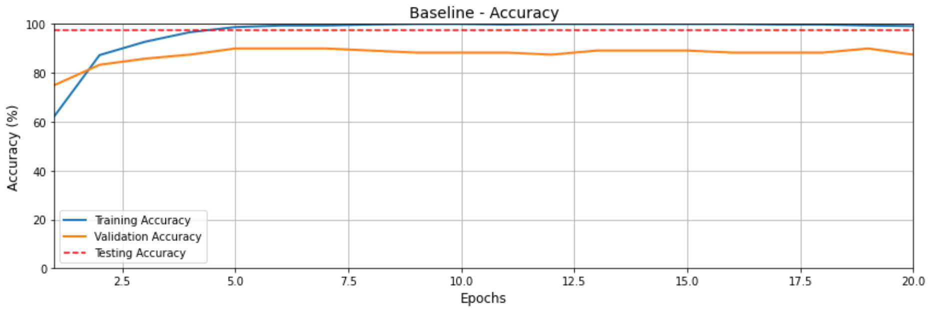


Figure 7. Graph of the Best Model

Figure 7 shows the model's performance during the training process based on the Accuracy metric. The blue line represents Training Accuracy, indicating how well the model learns patterns from the training data. The orange line represents Validation Accuracy, which measures the model's performance on validation data to evaluate its generalization ability. Meanwhile, the dashed red line represents Testing Accuracy, reflecting the model's performance on unseen test data.

At baseline, the model demonstrates excellent performance, with Training Accuracy reaching 99.99%, indicating that the model almost perfectly recognizes patterns in the training data. Validation Accuracy reaches 89.16%, suggesting that the model maintains stability without significant signs of overfitting. Furthermore, the high Testing Accuracy of 98.04% shows that the model generalizes very well to new data, making it reliable for classification tasks. This graph shows a consistent improvement in the early stages of training, followed by stability in subsequent epochs, indicating that the model has achieved good convergence. Overall, these results reflect a good balance between the model's learning from training, validation, and testing data. The observed decline may occur due to several factors, such as changes in model parameters or the model's reduced ability to generalize well to more complex data. The model may have reached its optimal point in earlier epochs, and further training could result in a decrease in performance on validation or test data.

2.1.2. Accuracy Improved Analysis

The accuracy comparison in this study focuses on analyzing the performance of the best model, which is VGG19 with the application of DWT as a noise removal technique, a learning rate (LR) of 0.0001, and varying epoch counts. This model serves as a reference to evaluate the impact of factors such as the increase in epoch count, the application of median filters, DWT noise reduction, and the combination of all three on improving accuracy and other performance metrics, such as precision, recall, and F1 score, as seen in Table 4.

Table 4. Percentage of Accuracy Improvement from Each Factor

Architecture	Factor	Epoch	Accuracy	precision	Recall	F1 Score
VGG19	Model Modification	10	95.24	95.6	95.23	95.24
VGG19	Model Modification + Epoch Improvement	20	95.99	96.1	96	96
VGG19	Model Modification + Median 3x3	10	96.08	96.24	96.12	96.08
VGG19	Model Modification + DWT	10	97.83	97.93	97.83	97.83
VGG19	Model Modification + DWT + Epoch Improvement	20	98.04	98.11	98	97.99

Vol. 6, No. 3, Juni 2025, Page. 1131-1146
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4397

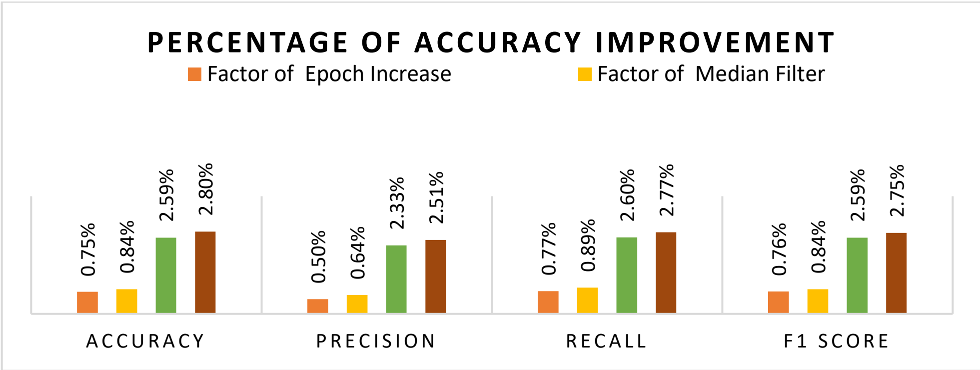


Figure 8. Percentage of Accuracy Improvement

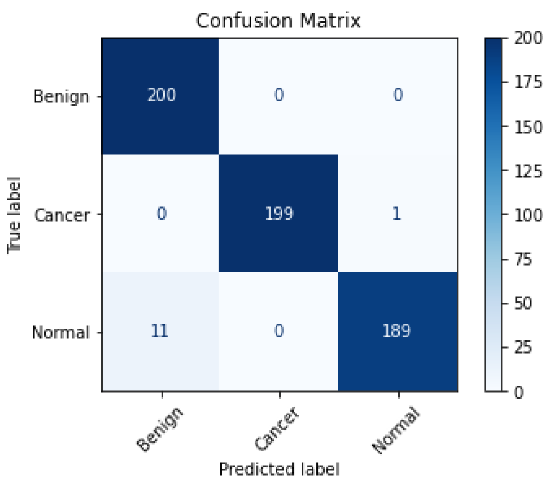
Figure 8 illustrates the percentage improvement in the model's accuracy, influenced by several factors, including the increase in the number of epochs, the application of the median filter, noise reduction using DWT, and the combination of all these factors. The measurements were taken by comparing the performance of the initial model (without applying noise removal and increasing the number of epochs) with the performance of the model after applying each factor individually as well as in combination. The baseline model used the VGG19 architecture with a learning rate of 0.0001 and 10 epochs.

1. Increase in Epochs: Contributed between 0.75% and 0.84% to the performance metrics, such as accuracy, precision, recall, and F1 score.
2. Median Filter: Had a more significant impact on precision, with an improvement of 2.93%.
3. DWT: Showed consistent contributions to performance improvement, with the highest increase in recall, at 2.77%.
4. Combination of All Factors: Delivered the best results, with a 2.80% improvement in accuracy and a 2.79% increase in F1 score.

These results emphasize that the combination of these factors had the most significant impact on enhancing model performance compared to the application of individual factors.

2.1.3. Testing Class

After obtaining the best model based on overall accuracy during testing, the next step is to use the model to analyze accuracy for each class. This process is crucial for understanding how well the model performs in classifying each class, particularly in datasets with varying class distributions or levels of complexity. The proposed model achieves an overall accuracy of 98.04%, with remarkably high per-class accuracies: 100% for the Benign class, 99.5% for the Cancer class, and 94% for the Normal class. These results indicate that the overall accuracy is closely aligned with the per-class accuracies, demonstrating the robustness of the model in classifying each class effectively.



Vol. 6, No. 3, Juni 2025, Page. 1131-1146
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4397

Figure 9. Confusion Matrix for Each Class

Figure 9 illustrates the performance of the classification model with three classes: Benign, Cancer, and Normal. The model demonstrates excellent performance, with the following results:

1. For the Benign class, all 200 samples were correctly classified without any errors.
2. For the Cancer class, 199 samples were correctly classified, but 1 sample was misclassified as Normal.
3. For the Normal class, 189 samples were correctly classified, but 11 were misclassified as Benign.

Overall, the model shows high accuracy, with minimal errors in classifying the Cancer and Normal classes, particularly for the Normal class, which has more significant misclassifications compared to the other courses. The confusion matrix reveals that the dataset for each class contains an equal number of samples, indicating that the dataset is balanced. With a balanced dataset, the model can be evaluated fairly without any bias towards a specific class, both for the baseline and for experiments with augmentation.

3. DISCUSSIONS

Table 5 presents several previous studies that utilized the DDSM dataset and its derivatives, such as CBIS-DDSM and Mini-DDSM, as benchmarks for analyzing model accuracy. These studies employed various augmentation techniques, including rotation, flipping, zooming, scaling, and filtering, to enhance model performance. The models used in these studies include DenseNet121, VGG16, and ECS-A-Net, achieving accuracy rates ranging from 65% to 96.50%. In contrast, the proposed model in this study utilizes the Mini-DDSM dataset with a significantly smaller number of images (600 samples) and does not incorporate additional augmentation techniques. Instead, it leverages image enhancement methods such as CLAHE and DWT for data preprocessing. These approaches resulted in a higher accuracy of 98.04% using DenseNet121, making it a superior baseline compared to previous studies. This research provides valuable insights into the effectiveness of augmentation techniques and image enhancement methods for developing detection models based on the DDSM dataset.

Table 5. Previous Research

Previous Research	Dataset	Augmentation	Image enhancement	Accuracy
Youness, et al.2024[16]	CBIS-DDSM (10.239 Image)	Rotation, flipping, and zooming.		DenseNet121: 76.71%
Sannasi, et al. 2024[17]	CBIS-DDSM (4.560 Image)	Rotation 45°, 90°, 135°, 180°, 235°, 270°, and flipping.	Median Filter 3x3	DenseNet121: 94.36%
Subasish, et al. 2022[18]	Mini-DDSM (9752 Image)	Rotation, flipping, and zooming	Median Filter 3x3	VGG16: 65%
Sumyat Thwin, et al. 2024[19]	DDSM (55.890 Image)	Rotation, flipping, and scaling		ECS-A-Net: 96.50%
Proposed Model	**Mini-DDSM (600 Image)**	**Baseline**	**DWT**	**VGG19: 98.04%**

The limitations of this study include several aspects. First, the research only compares two architectures, DenseNet121 and VGG19, leaving the potential of other, possibly more optimal architectures unexplored. Second, the Mini-DDSM dataset used is a simplified version of DDSM, which may not fully capture the diversity and complexity of real-world data. Third, the model's performance evaluation is limited to metrics such as Accuracy, Precision, Recall, and F1 Score, without incorporating additional metrics like ROC-AUC, which could provide a more comprehensive analysis. Fourth, this study focuses solely on developing a model without extending it into a system that can be directly implemented, limiting its practical application in clinical settings. Lastly, the preprocessing methods,

Vol. 6, No. 3, Juni 2025, Page. 1131-1146
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4397

such as DWT and a 3x3 Median Filter, are relatively basic and may not be sufficient for handling more complex noise.

However, the proposed study demonstrates that model accuracy can still be improved without employing data augmentation techniques and by utilizing a smaller dataset. This improvement was achieved through model modifications during the training process with several key focuses. One of the techniques applied is DWT, which effectively reduces noise in the images, improving the quality of the data used without requiring additional augmentation processes. Additionally, this study optimizes critical aspects of model training, such as adaptively adjusting the learning rate and determining the optimal number of epochs. This approach enables the model to achieve a good balance between accuracy and training efficiency, avoiding overfitting and maximizing performance even with a limited amount of data.

This research contributes significantly to the field of computer science, particularly in the domain of medical image analysis and deep learning. By demonstrating the effectiveness of DWT in enhancing image quality and optimizing training strategies without relying on extensive data augmentation, this study provides valuable insights for developing more efficient deep learning models for medical applications. Moreover, the findings emphasize the importance of balancing model complexity and data preprocessing techniques to achieve high accuracy with limited data resources. These results can serve as a foundation for future research aimed at improving automated breast cancer detection systems, potentially advancing clinical decision support tools, and aiding radiologists in early diagnosis. The implications of this study extend beyond mammogram classification, offering a framework for applying similar optimization strategies to other medical imaging tasks, thus contributing to the broader field of artificial intelligence in healthcare.

4. CONCLUSION

This study demonstrates that data augmentation does not always enhance the performance of CNN models. While augmentation techniques can be beneficial in certain cases to increase the diversity of training data, the findings of this study reveal that a non-augmentation approach or improper augmentation techniques can lead to lower performance. The results show that the baseline VGG19 model, with the best configuration (learning rate of 0.0001, 20 epochs, and image quality enhancement using DWT), achieved an overall Accuracy of 98.04%, Precision of 98.11%, Recall of 98%, and F1 Score of 97.99%. This confirms that augmentation is not always necessary when image quality and model architecture are already optimized.

The findings also indicate that DWT combined with CLAHE is more effective than augmentation techniques in supporting the performance of CNN models for breast cancer classification. Applying DWT to the dataset resulted in high per-class Accuracy: 100% for the Benign class, 99.5% for the Cancer class, and 94% for the Normal class. This study suggests that enhancing image quality through DWT significantly contributes to the model's ability to recognize complex patterns in medical images, surpassing the contributions of data augmentation techniques.

The comparison between VGG19 and DenseNet121 architectures shows that VGG19 outperforms DenseNet121 in breast cancer classification tasks under optimal settings. Through fine-tuning and layer modifications, VGG19 achieves higher Accuracy and evaluation metrics compared to DenseNet121. Furthermore, the study reveals that excessive epochs and learning rates can lead to decreased model performance, highlighting the importance of selecting appropriate hyperparameters to ensure optimal performance for both architectures. Overall, this study demonstrates that an approach focusing on higher image quality (using CLAHE+DWT), combined with fine-tuning and architectural modifications (VGG19), is the most effective strategy for achieving superior performance in breast cancer classification tasks.

REFERENCES

[1] H. Sumarti, S. R. Anggita, Hartono, F. R. Pratama, and A. N. Tasyakuranti, “Texture-Based Classification of Benign and Malignant Mammography Images using Weka

Vol. 6, No. 3, Juni 2025, Page. 1131-1146
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4397

Machine Learning: An Optimal Approach," *Evergreen*, vol. 10, no. 3, pp. 1570–1580, Sep. 2023, doi: 10.5109/7151705.

[2] A. Nurrohmah *et al.*, "Risk Factors of Breast Cancer," *GASTER JOURNAL OF HEALTH SCIENCE*, vol. 20, no. 1, pp. 1–10, 2022, doi: 10.30787/gaster.v20i1.

[3] A. B. Khaerunnisa, K. L. Latief, F. I. Syahruddin, I. Royani, and and R. P. Juhamran, "Hubungan Tingkat Pengetahuan dan Sikap terhadap Deteksi Dini Kanker Payudara pada Pegawai Rumah Sakit Ibnu Sina Makassar," no. Vol. 3 No. 9, Sep. 2023, doi: https://doi.org/10.33096/fmj.v3i9.291.

[4] J. Liu *et al.*, "Association of dietary carbohydrate ratio, caloric restriction, and genetic factors with breast cancer risk in a cohort study," *Sci Rep*, vol. 15, no. 1, p. 6263, Feb. 2025, doi: 10.1038/s41598-025-90844-0.

[5] M. Arnold *et al.*, "Current and future burden of breast cancer: Global statistics for 2020 and 2040," *Breast*, vol. 66, pp. 15–23, Dec. 2022, doi: 10.1016/j.breast.2022.08.010.

[6] C. Laza-Vásquez *et al.*, "Views of health professionals on risk-based breast cancer screening and its implementation in the Spanish National Health System: A qualitative discussion group study," *PLoS One*, vol. 17, no. 2 February, Feb. 2022, doi: 10.1371/journal.pone.0263788.

[7] M. Heinig, W. Schäfer, I. Langner, H. Zeeb, and U. Haug, "German mammography screening program: adherence, characteristics of (non-)participants and utilization of non-screening mammography—a longitudinal analysis," *BMC Public Health*, vol. 23, no. 1, Dec. 2023, doi: 10.1186/s12889-023-16589-5.

[8] T. Uematsu, "Rethinking screening mammography in Japan: next-generation breast cancer screening through breast awareness and supplemental ultrasonography," *Breast Cancer*, vol. 31, no. 1, pp. 24–30, Jan. 2024, doi: 10.1007/s12282-023-01506-w.

[9] L. Choridah *et al.*, "Knowledge and Acceptance Towards Mammography as Breast Cancer Screening Tool Among Yogyakarta Women and Health Care Providers (Mammography Screening in Indonesia)," *Journal of Cancer Education*, vol. 36, no. 3, pp. 532–537, Jun. 2021, doi: 10.1007/s13187-019-01659-3.

[10] K. Ishaq and M. Mustagis, "Computer Aided Detection and Classification of mammograms using Convolutional Neural Network," *Electrical Engineering and Systems Science*, pp. 1–16, Sep. 2024, doi: 10.48550/arXiv.2409.16290.

[11] S. Shafi and A. V. Parwani, "Artificial intelligence in diagnostic pathology," *BMC Res Notes*, vol. 18, no. 1, pp. 1–12, Dec. 2023, doi: 10.1186/s13000-023-01375-z.

[12] N. Nafissi *et al.*, "The Application of Artificial Intelligence in Breast Cancer," *Eurasian J Med Oncol*, vol. 8, no. 3, pp. 235–244, 2024, doi: 10.14744/ejmo.2024.45903.

[13] A. Kajala, S. Jaiswal, and R. Kumar, "Comparative Analysis of CNN Architectures for Classification of Breast Histopathological Images," vol. 2023, no. 8, pp. 4696–4705, Oct. 2023.

[14] D. Fonseka and C. Chrysoulas, "Data Augmentation to Improve the Performance of a Convolutional Neural Network on Image Classification," in *2020 International Conference on Decision Aid Sciences and Application, DASA 2020*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 515–518. doi: 10.1109/DASA51403.2020.9317249.

[15] T. Kumar, A. Mileo, R. Brennan, and M. Bendechache, "Image Data Augmentation Approaches: A Comprehensive Survey and Future directions," *IEEE Access*, vol. 12, pp. 187536–187571, Jan. 2024, doi: 10.1109/ACCESS.2024.3470122.

[16] Y. Khourdifi, A. El Alami, M. Zaydi, Y. Maleh, and O. Er-Remyly, "Early Breast Cancer Detection Based on Deep Learning: An Ensemble Approach Applied to Mammograms," *BioMedInformatics*, vol. 4, no. 4, pp. 2338–2373, Dec. 2024, doi: 10.3390/biomedinformatics4040127.

[17] S. Chakravarthy *et al.*, "Multi-class Breast Cancer Classification Using CNN Features Hybridization," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, Dec. 2024, doi: 10.1007/s44196-024-00593-7.

Vol. 6, No. 3, Juni 2025, Page. 1131-1146
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4397

[18] S. Mohapatra, S. Muduly, S. Mohanty, J. V. R. Ravindra, and S. N. Mohanty, "Evaluation of deep learning models for detecting breast cancer using histopathological mammograms Images," *Sustainable Operations and Computers*, vol. 3, pp. 296–302, Jan. 2022, doi: 10.1016/j.susoc.2022.06.001.

[19] S. M. Thwin, S. J. Malebary, A. W. Abulfaraj, and H. S. Park, "Attention-Based Ensemble Network for Effective Breast Cancer Classification over Benchmarks," *Technologies (Basel)*, vol. 12, no. 2, Feb. 2024, doi: 10.3390/technologies12020016.

[20] P. Kirichenko *et al.*, "Understanding the Detrimental Class-level Effects of Data Augmentation," in *NIPS'23: 37th International Conference on Neural Information Processing Systems New Orleans LA USA*, Curran Associates Inc.57 Morehouse LaneRed HookNYUnited States, Dec. 2023. doi: https://dl.acm.org/doi/10.1145/3659620.

[21] V. Auxilia Osvin Nancy, M. S. Arya, and N. Nitin, "Impact of Data Augmentation on Skin Lesion Classification Using Deep Learning," in *Proceedings - 2022 5th International Conference on Information and Computer Technologies, ICICT 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 67–72. doi: 10.1109/ICICT55905.2022.00020.

[22] K. Alshamrani, H. A. Alshamrani, F. F. Alqahtani, and B. S. Almutairi, "Enhancement of Mammographic Images Using Histogram-Based Techniques for Their Classification Using CNN," *Sensors*, vol. 23, no. 1, pp. 1–22, Jan. 2023, doi: 10.3390/s23010235.

[23] A. Rasheed, E. I. Essa, and B. F. Jumaa, "New hybrid Filtering Methods to remove noise from Digital Images," *Journal of Advanced Sciences and Nanotechnology*, vol. 1, no. 4, pp. 88–101, Oct. 2022, doi: 10.55945/joasnt.2022.1.4.88-101.

[24] C. Christudhas and A. Fathima, "VLSI implementation of a modified min-max median filter using an area and power competent tritonic sorter for image denoising," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-80053-6.

[25] H. Qu, K. Liu, and L. Zhang, "Research on improved black widow algorithm for medical image denoising," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-51803-3.

[26] S. N. Nia and F. Y. Shih, "Medical X-Ray Image Enhancement Using Global Contrast-Limited Adaptive Histogram Equalization," *Intern J Pattern Recognit Artif Intell*, vol. 38, no. 12, pp. 1–17, Sep. 2024, doi: 10.1142/S0218001424570106.

[27] A. Tawakuli, B. Havers, V. Gulisano, D. Kaiser, and T. Engel, "Survey:Time-series data preprocessing: A survey and an empirical analysis," *Journal of Engineering Research (Kuwait)*, pp. 1–38, Mar. 2024, doi: 10.1016/j.jer.2024.02.018.

[28] S. Saifullah, "ANALISIS PERBANDINGAN HE DAN CLAHE PADA IMAGE ENHANCEMENT DALAM PROSES SEGMENASI CITRA UNTUK DETEKSI FERTILITAS TELUR," *Jurnal Nasional Pendidikan Teknik Informatika*, vol. 9, no. 1, pp. 134–145, Mar. 2020, doi: https://doi.org/10.23887/janapati.v9i1.23013.

[29] M. Lin, Y. Hong, S. Hong, and S. Zhang, "Discrete Wavelet Transform based ECG classification using gcForest: A deep ensemble method," *Technology and Health Care*, vol. 32, pp. S95–S105, May 2024, doi: 10.3233/THC-248008.

[30] S. Goyal and R. Singh, "Detection and classification of lung diseases for pneumonia and Covid-19 using machine and deep learning techniques," *J Ambient Intell Humaniz Comput*, vol. 14, no. 4, pp. 3239–3259, Apr. 2023, doi: 10.1007/s12652-021-03464-7.

[31] N. Cao and Y. Liu, "High-Noise Grayscale Image Denoising Using an Improved Median Filter for the Adaptive Selection of a Threshold," *Applied Sciences (Switzerland)*, vol. 14, no. 2, pp. 1–21, Jan. 2024, doi: 10.3390/app14020635.

[32] S. Saifullah, A. P. Suryotomo, R. Drezewski, R. Tanone, and T. Tundo, "Optimizing Brain Tumor Segmentation Through CNN U-Net with CLAHE-HE Image Enhancement," 2024, pp. 90–101. doi: 10.2991/978-94-6463-366-5_9.

[33] S. S. Verma, A. Prasad, and A. Kumar, "CovXmlc: High performance COVID-19 detection on X-ray images using Multi-Model classification," *Biomed Signal Process Control*, vol. 71, no. 10, pp. 1–7, Jan. 2022, doi: 10.1016/j.bspc.2021.103272.

Vol. 6, No. 3, Juni 2025, Page. 1131-1146
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.3.4397

[34] M. Matin and M. Azadi, "Effect of Training Data Ratio and Normalizing on Fatigue Lifetime Prediction of Aluminum Alloys with Machine Learning," *International Journal of Engineering, Transactions A: Basics*, vol. 37, no. 7, pp. 1296–1305, Jul. 2024, doi: 10.5829/ije.2024.37.07a.09.

[35] D. A. Arafa, H. E. D. Moustafa, H. A. Ali, A. M. T. Ali-Eldin, and S. F. Saraya, "A deep learning framework for early diagnosis of Alzheimer's disease on MRI images," *Multimed Tools Appl*, vol. 83, no. 2, pp. 3767–3799, Jan. 2024, doi: 10.1007/s11042-023-15738-7.

[36] E. Goceri, "Medical image data augmentation: techniques, comparisons and interpretations," *Artif Intell Rev*, vol. 56, no. 11, pp. 12561–12605, Nov. 2023, doi: 10.1007/s10462-023-10453-z.

[37] A. Kebaili, J. Lapuyade-Lahorgue, and S. Ruan, "Deep Learning Approaches for Data Augmentation in Medical Imaging: A Review," *J Imaging*, vol. 9, no. 4, pp. 1–28, Apr. 2023, doi: 10.3390/jimaging9040081.

[38] A. Kokkula, P. C. Sekhar, and T. M. Praneeth Naidu, "Modified VGG-19 Deep Learning Strategies for Parkinson's Disease Diagnosis-A Comprehensive Review and Novel Approach," *Journal of Angiotherapy*, vol. 8, no. 3, 2024, doi: 10.25163/angiotherapy.839559.

[39] M. A. Rajab, F. A. Abdullatif, and T. Sutikno, "Classification of grapevine leaves images using VGG-16 and VGG-19 deep learning nets," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 22, no. 2, pp. 445–453, Apr. 2024, doi: 10.12928/TELKOMNIKA.v22i2.25840.

[40] R. Sujatha, J. M. Chatterjee, A. Angelopoulou, E. Kapetanios, P. N. Srinivasu, and D. J. Hemanth, "A transfer learning-based system for grading breast invasive ductal carcinoma," *IET Image Process*, vol. 17, no. 7, pp. 1979–1990, May 2023, doi: 10.1049/ipr2.12660.

[41] H. P. Hadi, E. H. Rachmawanto, and R. R. Ali, "Comparison of DenseNet-121 and MobileNet for Coral Reef Classification," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 23, no. 2, pp. 333–342, Mar. 2024, doi: 10.30812/matrik.v23i2.3683.

[42] J. Opitz, "A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice," *Trans Assoc Comput Linguist*, vol. 12, pp. 820–836, Jun. 2024, doi: 10.1162/tacl.