Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 3, July 2026, pp: 1117-1132. e-ISSN: 2656-8632

RESEARCH ARTICLE OPEN ACCESS

Accuracy Enhancement of a Hybrid CNN–VGG16 Architecture through Dropout Regularization Strategy for Breast Cancer Histopathology Classification

Fawaidul Badri[1,2,3], Ilham Ari Elbaith Zaeni[1], Heru Wahyu Herwanto[1], Muhammad Khusairi Osman[4]

[1] Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Indonesia
[2] Department of Electrical Engineering, Universitas Islam Malang, Malang, Indonesia
[3] Department of Informatics, Universitas Islam Malang, Malang, Indonesia
[4] Electrical Engineering Studies, Universiti Teknologi MARA, Pulau Pinang, Malaysia.

Corresponding author: Ilham Ari Elbaith Zaeni. (e-mail: Ilham.ari.ft@um.ac.id), **Author(s) Email**: Fawaidul Badri (e-mail: fawaidul.badri.2505349@students.um.ac.id, fawaidulbadri@unisma.ac.id), Heru Wahyu Herwanto (e-mail: heru_wh@um.ac.id), Muhammad Khusairi Osman (khusairi@uitm.edu.my).

Abstract **Breast cancer remains the leading cause of cancer-related mortality among women globally, necessitating accurate diagnosis through histopathological image analysis. However, manual examination of these images is time-consuming and susceptible to inter-observer variability, highlighting the critical need for reliable automated computer-aided diagnostic (CAD) systems. This study was conducted to systematically evaluate and optimize convolutional neural network (CNN) architectures for automated classification of breast cancer histopathology images, with a focus on mitigating overfitting and enhancing diagnostic accuracy through hybrid deep learning methodologies. The principal innovation is the development of a CNN-VGG16 hybrid architecture that strategically integrates pre-trained feature extraction with a customized CNN framework, hypothesized to substantially improve classification accuracy and model generalization. Three model configurations were developed and comparatively analyzed: (1) baseline CNN, (2) CNN with dropout regularization, and (3) hybrid CNN-VGG16 model. Input images underwent preprocessing, including resizing to 150×150 pixels, normalization, and data augmentation. All models were trained with identical hyperparameters: an Adam optimizer with a learning rate of 0.001, a batch size of 32, and 10 epochs. Dropout regularization with a fixed rate of 0.5 was applied to fully-connected layers to mitigate overfitting. Model evaluation was conducted utilizing standard performance metrics. The proposed CNN-VGG16 hybrid model achieved superior performance: accuracy of 85.19%, precision of 87.16%, recall of 92.75%, and F1-score of 88.37%. These metrics represent significant improvements of 4.2% relative to baseline CNN and 3.4% compared to the dropout-regularized variant, indicating substantially enhanced diagnostic capability and reduced false-negative rates. Strategic integration of pre-trained feature extraction with customizable CNN architectures significantly improves generalization and classification performance in histopathological image analysis. Future investigations should incorporate larger heterogeneous datasets, attention mechanisms, and explainable artificial intelligence (XAI) to enhance clinical interpretability and strengthen practitioner confidence in digital pathology systems.**

Keywords **Breast Cancer; Histopathology Image Classification; Convolutional Neural Network (CNN); Transfer Learning; VGG16.**

I. Introduction

Breast cancer affects more than 2 million women annually worldwide, making it the most dangerous malignant disease for women [1], [2]. Early detection and accurate classification can significantly improve patient survival rates, with five-year survival rates for early-stage breast cancer reaching 80-90%, substantially higher than those for advanced-stage disease [3]. Pathologists still rely on manual microscopic analysis of tissue samples stained with hematoxylin-eosin (H&E) dye to diagnose breast cancer histopathology using traditional methods. This

Manuscript received 20 November 2025; Revised 5 January 2026; Accepted 3 June 2026; Available online 8 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1356

manual process is entirely dependent on the individual examiner's experience and is highly subjective in nature [4], [5]. This traditional approach presents significant limitations, including inter-observer variability reaching 10-15%, high diagnostic errors, and a shortage of experienced pathologists, particularly in remote areas [5], [6]. Therefore, we need an automated deep learning system that can accurately and efficiently analyze histopathology images to support clinical decision-making and improve diagnostic consistency [7].

In recent years, researchers have employed deep learning, particularly convolutional neural networks (CNNs), as the dominant paradigm for analyzing histopathology images and detecting breast cancer, due to its superior ability to extract complex features from high-dimensional visual data [8], [9]. Conventional CNN models achieve a baseline accuracy of 80-81% when researchers apply them for histopathology classification on breast cancer datasets [10]. To overcome overfitting and improve generalization, researchers have demonstrated that the dropout regularization strategy provides significant contributions in contemporary literature [11]. Empirical studies show that applying dropout regularization increases CNN accuracy by 0.75-2% compared to the baseline [12], [13]. Researchers have established transfer learning using pre-trained models such as VGG16 as the industry standard for medical image classification [14]. When researchers implement transfer learning with VGG16 to breast cancer histopathology, they achieve measurable accuracy improvements, although these results are highly dependent on dataset characteristics. Wakili et al. [15] report that various CNN-based transfer learning approaches achieve accuracy in the range of 82-84% on the BreakHis dataset. Chitumakuru et al. [16] demonstrate that a standalone CNN achieves 83% accuracy, while transfer learning approaches reach 84-85%. Rahman et al. [17] prove that transfer learning with fine-tuning can improve performance to 84-85%. Recent research confirms that applying standard transfer learning and deep learning models to histopathological images consistently achieves accuracies of 80-85% [18].

These findings reveal that researchers still have significant opportunities to improve accuracy through combinations of more sophisticated hybrid architectures, more advanced regularization strategies, and more careful optimization. The hybrid approach that combines multiple architectures has demonstrated significant potential in enhancing model performance [19]. Sureshkumar et al. [20] integrated a convolutional neural network and an extreme learning machine framework for automated breast cancer diagnosis, achieving significant accuracy improvements. This study introduces three main contributions compared to previous hybrid CNN-VGG16 models. First, we applied layer-specific dropout regularization with different rates (p=0.25 in early convolutional layers, p=0.3 in mid-level layers, and p=0.5 in classification layers). Second, we performed incremental analysis by comparing three models step-by-step: a baseline CNN, a CNN with dropout, and a hybrid CNN-VGG16, to clearly show the contribution of each component. Third, the custom CNN layers were specifically designed for histopathological images to handle cellular morphology, tissue texture, and staining variations.

Although researchers have made significant progress, they still face substantial research gaps. First, although VGG16-based transfer learning improves accuracy compared to the baseline, this improvement is modest, with accuracy ranging from 80-85% across various settings [15], [16], [17]. Specifically, while VGG16-based transfer learning improves accuracy, the mechanisms underlying this improvement and its synergies with regularization strategies remain underexplored. Particularly, prior research has not systematically investigated: (1) the comparative contribution of dropout regularization versus transfer learning features in isolation, (2) how integration of pre-trained VGG16 features specifically affects feature extraction capability, and (3) the synergistic effects when both strategies are combined. This study addresses these gaps through systematic comparative analysis of three progressive model configurations (baseline CNN, CNN+Dropout, and Hybrid CNN-VGG16), enabling quantification of each component's incremental contribution to classification performance. Second, while dropout regularization has been widely adopted, the efficacy of regularization dropout rates (particularly 0.5) strategically positioned in specific network layers for hybrid CNN-VGG16 architectures in histopathological image classification remains underexplored. Although previous research has investigated various dropout configurations, a comprehensive comparative analysis combining fixed dropout regularization with transfer learning approaches specifically for breast cancer histopathology classification requires further investigation. [11], [12], [13]. This study addresses this limitation by implementing a dropout regularization strategy with differentiated dropout rates tailored to each layer's position and function within the architecture. Third, researchers have not conducted a comprehensive investigation into optimizing the hybrid CNN-VGG16 architecture specifically to handle the unique characteristics of histopathology images. Fourth, most existing studies have not performed an in-depth analysis of the synergistic effects between the hybrid architecture and the dropout regularization strategy. Fifth, while most existing studies focus on

Manuscript received 20 November 2025; Revised 5 January 2026; Accepted 3 June 2026; Available online 8 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1356

overall classification metrics (accuracy, precision, recall), they lack a comprehensive comparative analysis examining the incremental contribution of each architectural component (baseline CNN vs. dropout regularization vs. transfer learning).

To address these gaps, we propose a Hybrid CNN-VGG16 Architecture with Dropout Regularization Strategy that integrates advanced components to significantly and consistently improve the accuracy of breast cancer histopathology classification. Our research aims to develop an accurate and clinically reliable breast cancer histopathology classification system by combining a hybrid CNN-VGG16 architecture and a dropout regularization strategy. We contribute to the literature in four key ways: (1) we develop a hybrid CNN-VGG16 architecture systematically designed for histopathology classification, (2) we conduct systematic empirical investigation into the effectiveness of dropout regularization in improving model performance, (3) we perform comprehensive comparative analysis across various model development stages with detailed analysis of incremental contributions from each proposed strategy, and (4) we provide a comprehensive evaluation framework using confusion matrix and multiple metrics (Accuracy, Precision, Recall, F1-Score) to analyze the trade-off between sensitivity and specificity, offering practical insights into clinically critical metrics such as reducing false-negative rates in malignancy detection.

This paper is organized into four main sections: (1) a literature review (Section II), (2) the proposed methodology (Section III), (3) experimental findings along with comparative analysis against alternative approaches (Section IV), and (4) conclusions with directions for future research (Section V).

II. Literature Review

A. Application of CNN in Breast Cancer Classification

Convolutional neural network technology has demonstrated considerable efficacy as a leading solution for medical image classification, particularly for detecting breast cancer in histopathological samples [57]. The effectiveness of this methodology is attributed to its superior ability to perform automated, accurate extraction of morphological features, such as variations in cell shape, chromatin distribution, and nuclear structure, associated with breast cancer. Previous studies have demonstrated that CNNs can identify subtle patterns in histopathological images and have been successfully utilized to accurately differentiate benign lesions from malignant neoplasms, achieving over 90% on breast cancer datasets [21]. These studies also emphasize the importance of optimizing CNN architectures to achieve optimal performance in the high-complexity setting of histopathological images.

B. Transfer Learning to Improve the Performance of CNN

The integration of pre-trained learning models has become a standard approach for managing data insufficiency in medical image analysis. Pretrained models trained on large-scale datasets such as ImageNet can be adapted for specific classification tasks, such as breast cancer detection in histopathological images. The use of pretrained models like VGG16 and ResNet50 can achieve high accuracy even on limited datasets, as these models have already learned to recognize fundamental patterns present in medical images [22]. By fine-tuning the final layers, transfer learning can accelerate training and enhance model performance in breast cancer image classification.

C. Dropout as a Regularization Technique in CNN

Overfitting constitutes one of the paramount difficulties in developing convolutional neural networks when operating with limited medical image datasets. To counteract this issue, practitioners employ the Dropout regularization framework, a technique that functions through the selective and random deactivation of neuronal units during model training phases. By preventing the neural architecture from placing excessive reliance on specific feature patterns, this method significantly bolsters the model's capability to adapt and perform effectively on previously unseen data. Evidence from recent literature demonstrates that incorporating Dropout within CNN frameworks substantially mitigates overfitting while simultaneously improving diagnostic accuracy for breast cancer detection in histopathological tissue analysis, especially when datasets are sparse and exhibit class imbalance [23]. The results of these studies also demonstrate that Dropout can significantly reduce the model’s dependency on limited data, making it more robust to data variations.

D. Combination of CNN and Transfer Learning Models for Improved Classification Accuracy

Recent studies have begun combining CNNs, Dropout, and transfer learning to improve breast cancer classification accuracy. Several studies have shown that integrating CNN with VGG16-based transfer learning and Dropout can significantly enhance classification performance in histopathological breast cancer images [58]. In their experiments, this hybrid model demonstrated a 6% increase in accuracy compared to a standard CNN without transfer learning and Dropout. The researchers found that VGG16, with its numerous convolutional layers, was able to extract

Manuscript received 20 November 2025; Revised 5 January 2026; Accepted 3 June 2026; Available online 8 July 2026
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v8i3.1356

more complex features, while Dropout helped prevent overfitting on the limited dataset [24] [59].

E. Data Augmentation for Histopathological Images

Data augmentation is a methodology for artificially expanding training datasets by systematically generating image variants via computational transformations. In histopathological image classification, data augmentation is a pivotal mechanism that substantially enhances deep learning architectures' capacity to detect morphological patterns and diagnostic features in medical imagery, particularly given the inherent scarcity of comprehensive histopathological repositories. Prevalent methodologies encompass operations such as angular rotation, bilateral flipping (both horizontal and vertical axes), magnification scaling, chromatic modulation, and geometric displacement. The fundamental objective of such augmentation protocols is to synthesize a heterogeneous collection of images, thereby facilitating model comprehension of object recognition across multidirectional viewpoints and heterogeneous environmental conditions [25]. The impact of image augmentation on breast cancer classification models has been shown to be effective in improving accuracy and reducing overfitting. Rotation and zooming help the model learn to recognize patterns at various orientations and scales, while flipping enables the model to detect symmetry within histopathological tissues. Furthermore, augmentation helps address challenges such as class imbalance and lighting variability in histopathological images. Image augmentation techniques enhance the generalization ability of CNN models in breast cancer classification by reducing dataset bias. In the referenced study, rotation and zoom transformations were applied to enrich the variety of histopathological images used to train the model [26]. In addition, other studies have demonstrated that image augmentation improves the accuracy of deep learning models on histopathological cancer images by employing techniques such as flipping and contrast adjustment to enhance the quality of limited datasets.

III. Method

The primary research objective was to develop an integrated hybrid deep learning model synthesizing Convolutional Neural Network (CNN) architecture, Dropout regularization methodologies, and VGG16 Transfer Learning approaches for automated breast cancer classification in histopathological images. The investigative methodology encompassed distinct phases: data preprocessing and preparation procedures, model architecture construction and training, and comprehensive performance evaluation. Detailed exposition of the technical procedures and implementation frameworks for each methodological phase is presented in the subsequent sections.

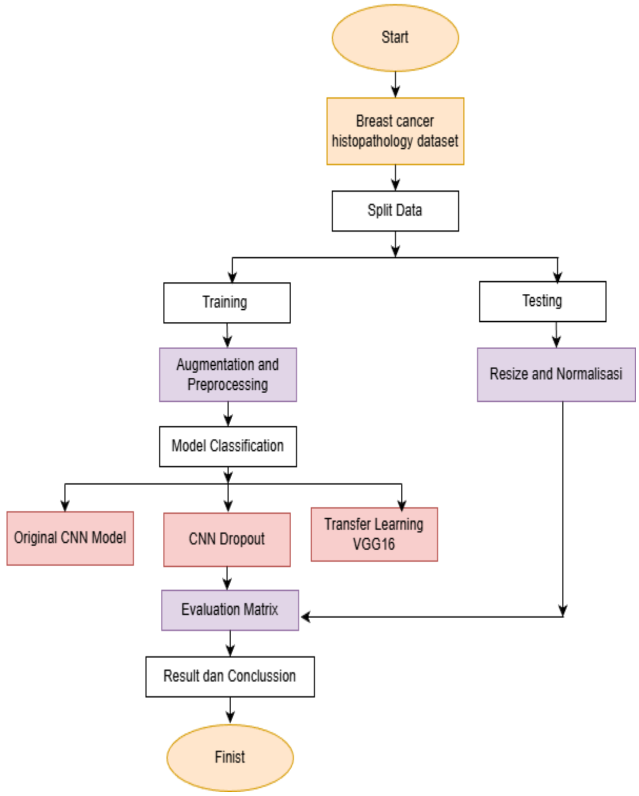

Fig. 1. The flowchart describes the breast cancer histopathology classification process, including data preprocessing, training using CNN and VGG16 models, evaluation, and result conclusion stages.

Fig. 1. This section illustrates the workflow of the breast cancer classification research using histopathological images with a deep learning approach. A histopathology imaging dataset encompassing breast tissue specimens was curated and dichotomized into benign and malignant classifications. The consolidated data repository was subsequently bifurcated into two distinct cohorts: a training cohort for model parameterization and a test cohort for empirical validation.

This methodological approach enabled rigorous model evaluation on held-out data, thereby affording an impartial assessment of model efficacy and generalization capacity across unseen samples. Subsequently, data augmentation and preprocessing are applied to the training data to enhance image diversity and quality through techniques such as rotation, shifting, zooming, and brightness adjustment, while the testing data are only normalized and resized. After the preprocessing stage, model training is conducted using several deep learning approaches.

Manuscript received 20 November 2025; Revised 5 January 2026; Accepted 3 June 2026; Available online 8 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1356

The methodological approach employed sequential CNN architectures. The baseline CNN consists of an input layer of 150×150×3 RGB images, followed by two convolutional blocks with 32 and 64 filters (3×3 kernel, ReLU activation), two max pooling layers (2×2), a flatten layer, a dense layer with 128 neurons (ReLU), and an output dense layer with 2 neurons (Sigmoid).

For the hybrid model, we used the frozen pre-trained VGG16 as the feature extractor. On top of it, we added a flatten layer, a dense layer with 256 neurons (ReLU), a dropout layer with a rate of 0.5, and a final output dense layer with 2 neurons (Sigmoid). The same hyperparameters were applied to all three models (baseline CNN, CNN with dropout, and hybrid CNN-VGG16) to ensure a fair comparison. The trained model was systematically evaluated through quantitative performance measurement employing multiple evaluation indices: accuracy, precision, recall, and F1-score. A confusion matrix visualization was additionally generated to provide a comprehensive representation of the classification outcomes and error patterns.

A. Dataset Preparation

The dataset used in this study consists of histopathological images of breast cancer obtained from the publicly available BreakHis (Breast Cancer Histopathological Images) dataset, which is accessible on Kaggle. This dataset contains labeled images categorized into two main classes: benign (non-cancerous) and malignant (cancerous) tumors. The BreakHis dataset comprises a total of 7,909 histopathological images (4,521 benign and 3,388 malignant samples) collected from human breast tissue samples that were processed using a microscope at different magnification levels (40X, 100X, 200X, and 400X). For this study, the complete dataset was partitioned into two distinct cohorts: (1) Training dataset: 80% of samples (6,327 images) for model training and optimization; and (2) Testing dataset: 20% of samples (399 images) for independent evaluation and performance validation. This data-split strategy ensures rigorous model evaluation on held-out data, thereby providing an objective assessment of model generalization capability. Each image has been labeled according to the tumor's classification as either benign or malignant.

The images in the dataset exhibit variations in resolution, tissue staining intensity, and microscope magnification levels, resulting in visual diversity that may influence the model's classification performance. To improve data quality and prevent overfitting, this study applies image augmentation during preprocessing. The augmentation techniques include rotation, zooming, flipping, and color adjustment, which aim to enrich the diversity of the training data so that the model can better learn and generalize patterns, even in the presence of differences in orientation, scale, and illumination. The BreakHis dataset can be accessed on Kaggle at the link:https://www.kaggle.com/datasets/waseemalastal/breakhis-breast-cancer-histopathological-dataset.

B. Split Data

Data partitioning is crucial in machine learning to ensure model training on one dataset and evaluation on an independent, unseen dataset. The dataset is commonly segregated into two primary components: a training partition and a test partition. This stratification mitigates the risk of model overfitting, a phenomenon wherein the model excessively memorizes training data patterns and noise, subsequently compromising its ability to generalize effectively to novel instances. Mathematically, if we have a dataset D consisting of N samples, Eq. (1) [5].

$$D = \{(x_1, y_1), (x_2, y_2), \ldots\ldots\ldots\ldots, (X_N, Y_N),\} \quad (1)$$

where x_i is a feature (for example, a histopathological image) and y_i is a label (for example, the benign or malignant class), then we divide the dataset D into two subsets: the training data D_{train} and the testing data D_{test}. The division is carried out with a certain proportion, for example, 80% for training and 20% for testing. Mathematically, this division can be expressed in the following equation Eq. (2) [3].

$$D_{train} = \{(x_i, y_i) | i \in \text{test set}\} \mid D_{train}| = 0.8\, x \quad (2)$$

$$D_{test} = \{(x_i, y_i) | i \in \text{test set}\} \mid D_{test}| = 0.2\, x\, N \quad (3)$$

Equation Eq. (2) illustrates the division of the dataset into two main subsets, namely the training data (D_{train}) and the testing data (D_{test}). In this division, a dataset containing N samples are divided into 80% for training and 20% for testing. The first equation explains that the training data (D_{train}) consists of 80% of the total dataset, namely $0.8 \times N$, used to train the model. This data includes feature pairs x_i and label y_i, where i indicates that the data is included in the training set. Meanwhile, the equation Eq. (3) describes the distribution of test data (D_{test}), which consists of 20% of the total dataset, namely $0.2 \times N$, which is used to test the performance of the model after training. The test data also consists of feature pair x_i and labels y_i, where i This time, it's included in the test set. The data partitioning approach ensures that the model receives training on a comprehensive training dataset of sufficient size, while performance evaluation is conducted on independent test data representing previously unseen instances, thereby enabling an objective assessment of model generalization performance.

C. Augmentation and Preprocessing data

In the image preprocessing stage, histopathological images are processed to ensure a consistent format suitable for a … deep learning model input. Large and varied images are resized to a uniform dimension to

Manuscript received 20 November 2025; Revised 5 January 2026; Accepted 3 June 2026; Available online 8 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1356

maintain input consistency within the model. Each image is resized to 150×150 pixels, enabling the model to process data of consistent size and ensuring stability during training. This resizing also reduces computational complexity and accelerates the model training process [27],[28]. In addition to resizing, another important step is image normalization. Image normalization is performed using min-max normalization, where pixel values are divided by 255 according to the formula: normalized_value = original_pixel_value / 255, thereby transforming the pixel values into the range [0, 1]. Normalization accelerates convergence during training, reduces data irregularities, and ensures stable learning [29]. Data augmentation was applied to the training dataset with specific parameters: (1) rotation range of 20 degrees; (2) zoom range of 0.2 (20% scaling); (3) horizontal and vertical flipping (flip probability 1.0); and (4) color jittering to handle staining variations. [30]. These techniques enrich training data diversity and improve generalization. Augmentation was applied only to the training dataset, while the testing dataset underwent only normalization and resizing to maintain evaluation integrity. [31]. Therefore, normalization is an essential step in ensuring dataset quality and model stability in deep learning tasks for medical image classification [32].

D. CNN Model Architecture

Convolutional Neural Networks constitute a distinctive class of neural architectures optimized for visual content processing. They function through hierarchical feature learning: initial layers detect basic features (edges, textures), while deeper layers capture complex features (cellular morphology, tissue structures). [33], [34]. Convolutional Neural Networks are constructed through the integration of convolutional blocks, nonlinear activation units, spatial downsampling layers, and terminal dense layers for classification purposes. In the convolutional layer, a convolution operation is performed between the input image and a kernel or filter. This two-dimensional convolution operation is formulated as follows. Eq. (4)[7].

$$Y(i,j,k) = \sum_{m=1}^{K_h} \sum_{n=1}^{K_w} \sum_{c=1}^{C_{in}} W(m,n,c,k).X(i+m-1, j+n-1, c) + b_k \quad (4)$$

Eq. (4) illustrates the basic process of the convolution operation in a Convolutional Neural Network (CNN), where the output value at a given position (i,j) and channel to-k, namely $Y(i,j,k)$, obtained from the result of the directional addition between the kernel (filter) and the corresponding input image patch. Kernel $W(m,n,c,k)$ sized $K_h \times K_w$ and works on all input channels C_{in}, so that to produce one output value, it is necessary to add three indices: the kernel row index m, kernel column index n, and input channels c. The kernel value is multiplied one by one by the pixel value in the input. X on site $(i+m-1, j+n-1, c)$, that is, a small area of the image that the kernel is currently "scanning" [35]. The results of the element-by-element multiplication are then added together to form a single feature representation value at a spatial position. (i,j). Next, a bias b_k added for each output channel, which functions to shift activation values to make the model more flexible in learning data distribution. Thus, this equation represents how CNNs extract local patterns (such as texture, edges, and shape) through a process of weighting and directional summation, where the kernel moves across the image to produce an information-rich feature map. In practice, this operation is known as cross-correlation and is at the heart of CNNs' ability to recognize visual structure in images. [36], [37].

Whereas X is the input image, W is the kernel weight, and b_k is bias. The output size of the convolution result is calculated using Eq. (5) [8].

$$O = \left| \frac{I-K+2P}{S} \right| + 1 \quad (5)$$

Equation Eq. (5) is used to calculate the output size of a convolution operation in one dimension (height or width). The value I epresents the input size, K is the kernel (filter) size, P denotes the amount of padding added to the edges of the input, and S is the stride, which is the step size by which the kernel moves during each convolution. Essentially, this formula determines how many positions the kernel can occupy over the input after accounting for padding and stride. The addition of "+1" arises because the initial placement of the kernel produces one output value [38]. The larger the padding, the more the output size tends to remain constant or increase, whereas a larger stride reduces the output size since the kernel skips more positions [39]. Following the convolution operation, a non-linear activation mechanism such as the Rectified Linear Unit (ReLU), characterized by the formulation $\mathrm{ReLU}(x) = \max(0,x)$, to avoid the vanishing gradient problem and accelerate learning [40].

The subsequent stage involves the pooling operation, which reduces the spatial dimensionality of the extracted features whilst retaining essential information. Maximum pooling represents the predominant pooling methodology, functioning by extracting the peak value from designated spatial regions, thereby generating a condensed feature representation and mitigating the likelihood of model overfitting [41]. The concluding network stage involves feature flattening followed by transmission through a dense layer, wherein the softmax activation function transforms feature representations into probability

Manuscript received 20 November 2025; Revised 5 January 2026; Accepted 3 June 2026; Available online 8 July 2026
Digital Object Identifier (DOI): https://doi.org/10.35882/jeeemi.v8i3.1356

distributions across output classes [42]. The softmax activation mechanism constitutes a specialized mathematical function deployed in the terminal layer of neural network architectures to generate probability distributions across multiple classification categories. Given a vector of scores $z = (z_1, z_2, \ldots, z_C)$, then the softmax function will transform each value z_i into probability values through an equation. The softmax transformation is mathematically represented by the equation Eq. (6) [9].

$$\text{Softmax(z)\_i} = \frac{e^{z\_i}}{\sum_{j=1}^{c} e^{z\_j}} \tag{6}$$

The softmax function expresses the exponential value of each score so that all values become positive, then normalizes them by the sum of all exponentials so that the total result equals 1. Thus, the value softmax$(z)_i$ indicates how likely an input is to belong to the class i compared to other classes. The higher the score z_i, the greater the probability value produced. The equation Eq. (6) This function is important because it allows the network to produce classification decisions that can be interpreted as probabilities and can be directly used with loss functions such as categorical cross-entropy for the equation training process. Eq. (7) [11].

$$L = \sum_{I=1}^{C} Y_i \, log(y_i) \tag{7}$$

value y_i is the ground truth label in the form of one-hot encoding, where only the correct class has a value of 1 and the other classes 0. Meanwhile, $\hat{y}_i$ is the predicted probability of the model for the the-i class, usually obtained from the softmax function [43]. This loss function works by penalizing the model when the probability $\hat{y}_i$for the correct class, the value is low; the lower the probability of the correct class, the greater the resulting loss value. Lower loss values result from high confidence predictions directed toward the true class label. In other words, this function encourages the model to make predictions with a high probability for the class that corresponds to the actual target, so that the model learns to improve its accuracy during the training process. [44]. This equation is the Categorical Cross-Entropy loss function, which is used to measure how well the model predicts the classification problem written in the equation Eq. (7).

E. CNN Dropout

To address the limitation of using fixed dropout rates in hybrid architectures, this study implemented layer-specific dropout regularization. A dropout rate of 0.25 was applied to the early convolutional layers (Conv1 and Conv2) to preserve low-level morphological features. The mid-level convolutional layer (Conv3) employed a dropout rate of 0.3, while the dense classification layers used a higher rate of 0.5 to mitigate overfitting. This differentiated dropout strategy tailors the regularization intensity according to each layer's position and susceptibility to overfitting. Dropout is a regularization technique employed to mitigate overfitting in deep neural networks, including CNNs. It operates stochastically, inactivating neurons with a specified probability during training, preventing their participation in forward and backward propagation. In this study, a dropout rate of 0.5 was applied after the Flatten layer and before the dense classification layers. This placement targets the fully connected layers, which are most prone to overfitting when trained on limited data. The fixed dropout rate was chosen to ensure reproducibility, computational efficiency, and consistency throughout the training process. Mathematically, dropout can be defined by the equation Eq. (8) [13].

$$\hat{h} = h \odot r \tag{8}$$

h represents the output of neurons in a layer, r is a binary-valued mask vector generated from a Bernoulli distribution $r \sim Bernoulli(p)$, with p is the probability that the neuron is maintained, and $\odot$ represents an element-by-element multiplication operation. During training, neurons that have a value of 0 in the mask r will be removed, while neurons with a value of 1 will remain activated in the equation Eq. (8). At the time of inference (testing), the activation value is scaled by a factor p to maintain consistency of output magnitude, namely the equation Eq. (9) [15].

$$h_{test} = p.\hat{h} \tag{9}$$

With this mechanism, the model is forced to be independent of specific neurons so that the network learns more robust feature representations and is able to generalize better to new data. Dropout significantly improves network performance especially on complex and large datasets [45]. This approach has proven valuable in advancing the accuracy metrics of diagnostic image classification tasks, encompassing histopathology examination scenarios [46].

F. Transfer Learning VGG16

Transfer learning in the VGG16 architecture is a technique that utilizes pre-trained weights. Models initially trained on large repositories like ImageNet are then optimized through adaptation to specialized, task-specific datasets, such as histopathological images [47]. In general, VGG16 consists of a series of 3×3 convolutional and pooling layers that function as feature extractors, and fully connected layers that act as the classifier. In transfer learning, the feature extraction part is retained, while the classifier part is modified and retrained according to the number of target classes. This process can be mathematically formulated in Equation Eq. (10) [22].

$$f = F(X; \theta_{pre}) \tag{10}$$

where X denotes the input image, F(·) represents the pre-trained VGG16 network, and θ_{pre} refers to the weights obtained from the initial training. These

Manuscript received 20 November 2025; Revised 5 January 2026; Accepted 3 June 2026; Available online 8 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1356

extracted feature representations are then input to a newly instantiated classification component, as shown in the equation. Eq. (11) [22].

$$\hat{y} = softmax(W\,f\; +\; b) \tag{11}$$

The parameters W and b are the classification parameters that will be trained on the new dataset [49]. During the fine-tuning process, some of the weights in the deeper convolutional layers can be unfrozen and updated using the gradient derived from Equation Eq. (12) [22].

$$\theta_{new} = \theta_{pre} - \eta . \frac{\partial \mathcal{L}}{\partial \theta_{pre}} \tag{12}$$

where η denotes the learning rate and $\mathcal{L}$ represents the loss function (e.g., cross-entropy). This approach is particularly effective when the dataset size is small, as the network does not need to learn low-level features from scratch, thereby reducing the risk of overfitting and accelerating the training process. Recent studies have shown that VGG16 with transfer learning can significantly improve the classification accuracy of histopathological and medical images. [50].

G. Model Evaluation

Classification efficacy is quantified through fundamental performance metrics encompassing Accuracy, Precision, Recall, and F1-Score, which are systematically computed from Confusion Matrix constituents. The binary classification framework utilizes a four-element matrix architecture, Model predictions result in four distinct classification outcomes: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The cross-tabulation matrix summarizing these prediction categories is presented below (Table 1.)

Table 1. Confusion Matrix Table for Evaluating the Performance of Breast Cancer Histopathology Image Classification Model.

Actual	Prediction	
	Negative	Positive
Negative	TN	FP
Positive	FN	TP

Mathematically, Accuracy is a measure that indicates how many of the model's predictions are correct compared to the total number of tested data. Accuracy calculates the proportion of all correctly classified instances, including both positive and negative predictions [51]. The metric encompasses the overall effectiveness of classification decisions made by the model. Its mathematical representation is defined in the following formulation. Eq. (13) [36].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{13}$$

Classification outcomes are delineated through four categorical instances: True Positive (TP) captures positive exemplars subject to successful positive predictions; True Negative (TN) captures negative exemplars subject to successful negative predictions; False Positive (FP) encompasses negative exemplars exhibiting erroneous positive predictions; and False Negative (FN) comprises positive exemplars experiencing erroneous negative predictions. The Accuracy metric, which ranges from 0 to 1, reflects the proportion of correct predictions out of all predictions made values near 1 signify higher classification effectiveness. Nevertheless, this metric may not always provide an accurate representation of model performance in cases of class imbalance, as a classifier might achieve high accuracy merely by favoring the majority class. To overcome this limitation, accuracy is typically analyzed alongside additional evaluation measures such as Precision, Recall, and the F1-Score, which together give a more balanced performance assessment.

In imbalanced scenarios where one class significantly outweighs the other, the accuracy metric can introduce bias. This happens because a model may seem effective simply by predominantly predicting the majority category, thereby diminishing its ability to correctly identify minority class instances. Consequently, supplementary metrics like Precision and Recall are employed. Precision quantifies the correctness of positive predictions, representing the fraction of samples predicted as positive that are indeed positive in reality. Mathematically, Precision can be expressed as follows: Eq. (14) [36].

$$Precision = \frac{TP}{TP+FP} \tag{14}$$

True Positive (TP) represents positive instances correctly predicted as positive, while False Positive (FP) represents negative instances incorrectly predicted as positive. Higher Precision indicates fewer false positive predictions. This performance indicator assesses what proportion of true positive instances are successfully captured by the model's detection mechanism. In high-risk scenarios, particularly within medical diagnostic contexts, insufficient performance on this metric causes many legitimate positive cases to remain unidentified (False Negatives). Mathematically, Recall is formulated as follows: Eq. (15) [36].

$$Recall = \frac{TP}{TP+FN} \tag{15}$$

Negative (FN) represents the number of positive samples incorrectly classified as negative [52]. A high Recall value indicates that the model is able to detect most of the positive samples, thereby demonstrating good performance in terms of sensitivity to positive class identification. To balance Precision and Recall, the F1-Score metric is used, which is the harmonic mean of the two. This metric addresses the evaluation challenge posed by asymmetrical False Positive and

Manuscript received 20 November 2025; Revised 5 January 2026; Accepted 3 June 2026; Available online 8 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1356

Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 3, July 2026, pp: 1117-1132. e-ISSN: 2656-8632

False Negative distributions through balanced measurement. It becomes indispensable in applications characterized by class imbalance, such as clinical image analysis. The harmonic mean foundation ensures stringent scoring: elevated results mandate concurrent high performance in both Precision and Recall, unlike traditional arithmetic mean approaches. Thus, the F1-Score serves as a reliable indicator for evaluating models that are expected to maintain a balanced performance between prediction accuracy for positive cases and the ability to correctly detect positive classes. Mathematically, the F1-Score is expressed as follows: Eq. (16) [36].

$$F1 - Score = 2.\frac{Precision \cdot Recall}{Precision + Recall} \quad (16)$$

The F1-Score serves as an important evaluation metric for classification models because it captures the trade-off between Precision and Recall. It is defined as the harmonic mean of these two measures, which implies that a high F1-Score can only be achieved when both Precision and Recall attain strong values. This makes the F1-Score particularly useful for assessing models trained on imbalanced datasets, as it reflects not only the accuracy of positive predictions but also the model's ability to correctly identify all relevant positive samples. This relationship is expressed in the following equation: Eq. (16).

IV. Result

A. Confusion Matrices

The confusion matrix is one of the most widely utilized evaluation techniques for measuring the effectiveness of classification models, especially within machine learning applications. It provides a detailed comparison between the predicted outputs generated by the model and the actual ground-truth labels, thereby offering insights into the model's capability to correctly identify each class. The confusion matrix was instrumental in providing a granular assessment of the model's classification accuracy and discriminative capacity regarding benign versus malignant breast tissue characterization within histopathological specimen examination. By analyzing this matrix, it becomes possible to determine the number of accurate predictions as well as classification errors for each category. These values then serve as the basis for computing key performance indicators, including accuracy, precision, recall, and the F1-score. Consequently, the confusion matrix delivers a more holistic understanding of the model's overall strengths and limitations in classification performance.

1. Confusion Matrices CNN

The Dropout-enhanced CNN correctly classified 236 malignant and 87 benign cases out of 399 test samples. It produced 38 false negatives and 38 false positives. False negatives mainly came from well-differentiated carcinomas, while false positives came from complex benign lesions such as sclerosing adenosis and fibroadenomas. The 150×150 pixel resolution also limited the capture of fine nuclear details.

The baseline CNN correctly classified 87 benign and 236 malignant samples, with a sensitivity of 86.13% for detecting malignant cases. The symmetric misclassification (38 FP and 38 FN) indicates room for improvement through better regularization and higher-resolution inputs. Future work should use higher image resolution (at least 224×224 pixels) and multi-scale feature extraction to better differentiate morphological subtypes (Table 2).

Table 2. Confusion Matrix Table for Evaluating the Performance of Breast Cancer Histopathology Image Classification Original Model CNN

Actual	Prediction	
	Negative	Positive
Negative	87	38
Positive	38	236

2. Confusion Modified Model with Dropout

The confusion matrix for the Dropout-regularized CNN model reveals accurate classification of 75 benign and 251 malignant tissue samples. However, the model misclassified 50 benign samples as malignant (false positives) and 23 malignant samples as benign (false negatives). Confusion matrix table for evaluating the performance of breast cancer histopathology image classification modified model with dropout can be seen at Table 3.

3. Confusion matrix for Hybrid CNN-VGG16

The confusion matrix for the hybrid CNN-VGG16 model shows correct classification of 95 benign and 285 malignant specimens out of 399 test samples. There were 30 false positives and 24 false negatives. Compared to the Dropout model, false positives decreased by 40% (from 50 to 30). The hybrid model also achieved a sensitivity of 84.07%. These results indicate that combining pre-trained VGG16 features with custom classification layers improves diagnostic accuracy for breast cancer detection (Table 4).

Table 4. Confusion Matrix Table for Evaluating the Performance of Breast Cancer Histopathology Image 3 Confusion matrix for Hybrid CNN-VGG16

Actual	Prediction	
	Negative	Positive
Negative	95	30
Positive	24	285

Manuscript received 20 November 2025; Revised 5 January 2026; Accepted 3 June 2026; Available online 8 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1356

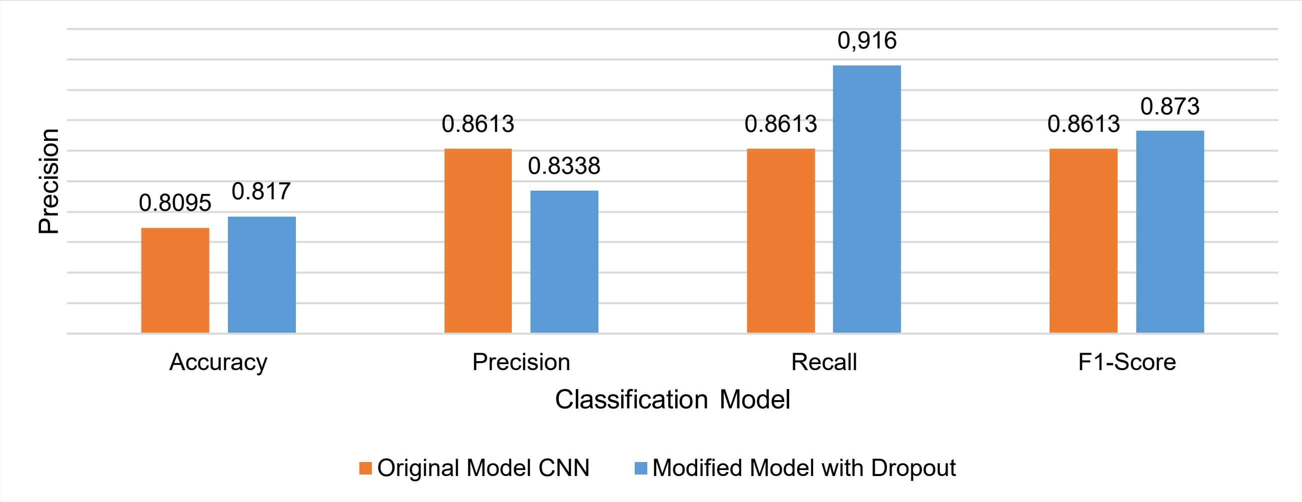

Fig. 2. Comparison of Classification Performance Between Original CNN and Dropout-Enhanced CNN Model

B. Accuracy

This study compared three models for classifying breast cancer histopathological images from the BreakHis dataset: a baseline CNN, a CNN with dropout regularization, and the proposed hybrid CNN-VGG16 model. Performance was evaluated using accuracy, precision, recall, and F1-score. A comprehensive comparison of all three evaluated architectures is presented in Table 5, which provides a detailed assessment of the baseline CNN, CNN with Dropout regularization, and the proposed Hybrid CNN-VGG16 model. The baseline model with the original CNN architecture demonstrates basic performance with an Accuracy of 0.809, and Precision, Recall, and F1-Score values of 0.861 each. This model is capable of recognizing fundamental tissue texture patterns but still shows limitations in generalization. Following architectural augmentation through the incorporation of Dropout regularization, measurable enhancements in model efficacy were observed. The CNN architecture enhanced with Dropout regularization demonstrated quantitative improvements, achieving an Accuracy of 0.817, a precision of 0.833, a recall of 0.916, and an F1-Score of 0.873. The substantial elevation in Recall performance indicates that the application of Dropout regularization successfully mitigated overfitting phenomena and augmented the model's capacity to accurately identify an increased proportion of positive instances. A detailed comparative assessment between the baseline CNN architecture and the Dropout-regularized variant is presented in Table 5. For a more detailed view, the performance comparison between the Original CNN model and the modified CNN model with Dropout regularization can be seen in Fig. 2. As shown in Fig. 2, Performance comparison between the Original CNN model and the modified CNN model with Dropout regularization across four evaluation metrics (Accuracy, Precision, Recall, and F1-Score) in percentage (%). The Modified CNN with Dropout demonstrates notable improvements, particularly in Recall and F1-Score metrics compared to the baseline model, although a slight decrease is observed in precision. From these results, it can be concluded that the addition of a Dropout layer improves the model's ability to recognize a greater number of positive samples without relying too heavily on specific training data. This finding confirms that the regularization strategy using Dropout is effective in enhancing the stability and generalization capability of

Table 5. Classification Performance Comparison of Original CNN and Modified CNN with Dropout Mode

Model	Accuracy	Precision	Recall	F1-Score
Original Model CNN	0.8095	0.8613	0.8613	0.8613
Modified Model with Dropout	0.8170	0.8338	0.9160	0.8730
Hybrid CNN-VGG16	85.10	87.16	92.75	88.37

Manuscript received 20 November 2025; Revised 5 January 2026; Accepted 3 June 2026; Available online 8 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1356

the CNN model for breast cancer histopathology image classification tasks. Consequently, the Modified CNN Model with Dropout demonstrates more reliable and adaptive performance across data variations, as illustrated in Fig. 2.

V. Discussion

A. Classifier

The study compared three deep learning models for classifying breast cancer histopathological images: a baseline CNN, a CNN with dropout regularization, and a hybrid CNN–VGG16 model. The hybrid CNN–VGG16 achieved the highest accuracy of 85.20%, compared to 80.95% for the baseline CNN and 81.70% for the dropout CNN. It also recorded a precision of 87.16%, a recall of 92.75%, and an F1-score of 88.37%. This improvement comes mainly from VGG16's pre-trained features, which help the model better recognize patterns in tissue images. More deeply, the hybrid architecture leverages the strength of VGG16 as a powerful feature extractor that has already learned rich hierarchical representations from millions of natural images. When combined with custom CNN layers, it enables the model to capture both low-level textures (edges and color variations due to staining) and higher-level morphological patterns (nuclear shape, chromatin distribution, and tissue architecture) that are essential for distinguishing benign from malignant histopathological samples. The application of layer-specific dropout regularization further enhances this capability by preventing the network from overly relying on specific neurons during training, thereby improving generalization on the limited BreakHis dataset. The notable increase in recall to 92.75% indicates that the model becomes more sensitive in detecting malignant cases, which is clinically vital because reducing false negatives can prevent delayed diagnosis and improve patient prognosis. Meanwhile, the moderate precision of 87.16% reflects the inherent difficulty in histopathological images where some benign lesions (such as fibroadenomas or sclerosing adenosis) share visual similarities with low-grade malignancies. Overall, these results demonstrate a synergistic effect between transfer learning and dropout regularization, where each component contributes incrementally to both quantitative performance and the model's robustness against overfitting and domain variability. The hybrid model showed a 4.25% increase in accuracy over the baseline CNN and a 3.50% increase over the dropout CNN. Compared to previous studies, the proposed model reached 85.20% accuracy. This is higher than Chauhan and Kumar (2025) at 76.92% (VGG16-SVM), Wakili et al. (2024) and Diyasa et al. (2025) at 83.33%, and Heikal et al. (2024) at 84% before optimization. The model provides better performance than standard VGG16 implementations without needing complex optimization methods (Table 6). summarizes this comparative performance.

Table 6 presents a comparative analysis of the proposed hybrid CNN-VGG16 model against several recent studies on breast cancer histopathology classification. The proposed model achieved an accuracy of 85.19%, which is higher than all comparative studies: Chauhan & Kumar (2025) at 76.92%, Wakili et al. (2024) at 83.33%, Heikal et al. (2024) at 84%, and Diyasa et al. (2025) at 83.33%. This result shows similarities with previous works in the use of the BreakHis dataset and VGG16-based transfer learning. However, there are significant differences in the architectural approach. While previous studies generally used single models or simple combinations, this study proposes a hybrid architecture with custom CNN layers and layer-specific dropout regularization. This combination yields better performance, especially in Recall (92.75%), which is clinically crucial for reducing false negatives in cancer detection. Nevertheless, the achieved accuracy remains below 90%, which contradicts some recent trends reporting accuracies above 90% using more advanced architectures or higher image resolutions. This indicates that there is still room for further improvement in future work.

Table 6. Comparative Analysis of Classification Performance Between the Proposed CNN–VGG16 Method and Previous Studies.

Study	Model Used	Accuracy (%)
Chauhan & Kumar, 2025. [54]	VGG16 + SVM	76.92
Wakili et al., 2024. [55]	DenseNet (Transfer Learning)	83.33
Heikal et al., 2024. [56]	Custom CNN (without optimization)	84.00
Diyasa et al., 2025. [57]	VGG16 + WGAN	83.33
Proposed Method	**CNN + VGG16 (Transfer Learning)**	**85.19**

Although the results are promising, this study has several limitations. First, the use of a relatively small public dataset (BreakHis) with only 7,909 images, even after an 80/20 split, may limit the model's ability to generalize to more diverse real-world clinical samples

Manuscript received 20 November 2025; Revised 5 January 2026; Accepted 3 June 2026; Available online 8 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1356

that vary in staining protocols, scanner types, and patient demographics. Second, resizing all images to 150×150 pixels resulted in the loss of fine nuclear details, which contributed to the 38 false negatives and 38 false positives observed in the confusion matrices, particularly when differentiating subtle low-grade malignancies from complex benign lesions. Third, the study only evaluated VGG16 as the transfer learning backbone and did not compare it with other state-of-the-art architectures such as ResNet, EfficientNet, or Vision Transformers, which might offer different trade-offs between accuracy and computational efficiency. Additionally, the model still operates as a black-box system, lacking mechanisms to explain which specific histopathological features influence its decisions, thereby potentially reducing trust from clinical pathologists.

Future work should use higher-resolution images (at least 224×224 pixels), a larger, more diverse dataset, and more advanced models. Nevertheless, the proposed hybrid CNN–VGG16 model offers a good balance between accuracy and computational efficiency, making it suitable for computer-aided diagnosis in digital pathology. Future work should also involve larger and more diverse datasets, model fusion strategies, and explainable AI approaches to enhance both interpretability and clinical reliability. Overall, this study demonstrates that the hybrid CNN-VGG16 model can significantly improve the classification accuracy of breast cancer histopathological images. Despite limitations related to dataset size and image resolution, this approach provides important implications for the development of more accurate, efficient, and clinically ready computer-aided diagnosis systems in digital pathology. Technically, the findings offer a practical and reproducible framework for researchers working with limited medical imaging datasets by showing how hybrid architectures with strategic regularization can deliver meaningful performance gains without high computational cost. Clinically, the high recall achieved in this study holds potential to assist pathologists in reducing inter-observer variability and missed malignant cases, which is especially valuable in settings with limited numbers of experienced specialists. In the broader context of healthcare in developing countries, where breast cancer continues to be a leading cause of mortality among women, such efficient CAD tools could support faster screening, earlier intervention, and more equitable access to quality diagnostics. Thus, the proposed method represents a meaningful step toward bridging the gap between deep learning research and practical digital pathology applications.

VI. Conclusion

The main objective of this study was to evaluate and enhance the performance of CNN architectures for breast histopathology image classification. Three model variations were tested: an original CNN, a dropout-modified CNN, and a hybrid CNN combined with the pretrained VGG16 network. The experimental results revealed that the proposed hybrid CNN–VGG16 achieved the best overall performance with 85.19% accuracy, 87.16% precision, 92.75% recall, and an F1-score of 88.37%, outperforming the baseline CNN (80.95%) and dropout CNN (81.70%) by 4.2% and 3.4%, respectively. In addition, dropout regularization effectively reduced overfitting and improved recall (from 86.13% to 91.61%), though with a slight decrease in precision. The hybrid model provided the most balanced and robust performance, confirming that combining pretrained feature extraction with a customized CNN enhances classification accuracy and generalization. Future research should address identified limitations through targeted enhancements. Larger heterogeneous datasets directly mitigate false-negative errors caused by underrepresented well-differentiated malignancy morphotypes. Attention mechanisms illuminate which tissue regions drive the 38 false-negative and 38 false-positive errors, revealing whether failures stem from missing nuclear details or from overweighting cellularity features. XAI techniques (Grad-CAM, LIME, SHAP) map predictions to interpretable histopathological features (nuclear pleomorphism, glandular architecture), thereby directly building clinical confidence by demonstrating that models rely on pathologically meaningful rather than spurious features. Higher input resolution (224×224+ pixels) compensates for the identified loss of fine-grained nuclear details, potentially reducing false negatives on subtle malignancies. Together, these enhancements directly address the specific performance gaps identified in this work.

Acknowledgment

The authors would like to express their sincere gratitude to the supervisors for their valuable guidance, insightful suggestions, and continuous support throughout the completion of this research. The authors also extend their appreciation to all individuals and institutions who contributed directly or indirectly to the successful completion of this study. This research was supported by the PDDI Scholarship under the Beasiswa PDDI and PPAPT Kemdiktisaintek. Doctoral Scholarship Program for Indonesian Lecturers, Center for Higher Education Funding and Assessment, and Ministry of Higher Education, Science and Technology of the Republic of Indonesia.

Manuscript received 20 November 2025; Revised 5 January 2026; Accepted 3 June 2026; Available online 8 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1356

Funding
This research was supported by the PDDI Scholarship (Beasiswa PDDI) and PPAPT Kemdiktisaintek under the Doctoral Scholarship Program for Indonesian Lecturers, Center for Higher Education Funding and Assessment, Ministry of Higher Education, Science and Technology of the Republic of Indonesia.

Data Availability
The dataset used in this study was obtained from the publicly available BreakHis (Breast Cancer Histopathological Image Classification) dataset. The dataset is available through Kaggle and can be accessed at:
https://www.kaggle.com/datasets/waseemalastal/breakhis-breast-cancer-histopathological-dataset

Author Contribution
Fawaidul Badri contributed to data acquisition, data preprocessing, model implementation, experimental investigation, data analysis, visualization, and manuscript drafting. Ilham Ari Elbaith Zaeni contributed to conceptualization, methodology, supervision, validation, interpretation of results, manuscript review, and editing. Heru Wahyu Herwanto contributed to validation, critical review of the manuscript, and supervision of the research process. Muhammad Khusairi Osman contributed to validation, critical manuscript review, and final approval of the published version. All authors read and approved the final manuscript and agreed to be accountable for all aspects of the work.

Declarations
Ethical Approval
This study used publicly available and anonymized secondary data from the BreakHis breast cancer histopathological image dataset. No direct human participants, clinical intervention, or identifiable personal data were involved; therefore, formal ethical approval was not required.

Consent for Publication Participants.
Not applicable, as this study used publicly available secondary image data and did not involve identifiable human participants.

Competing Interests
The authors declare that they have no competing interests.

References
[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, and A. Jemal, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer J. Clin.*, vol. 72, no. 6, pp. 601–649, Nov. 2022. https://doi.org/10.3322/caac.21834

[2] American Cancer Society, *Breast Cancer Statistics 2024*, 2024. [Online]. Available: https://www.cancer.org

[3] S. Alimuddin, et al., "Application of VGG16 transfer learning for breast cancer detection," Information, vol. 16, no. 3, pp. 227, 2023. https://doi.org/10.3390/info16030227

[4] M. Ramasamy, et al., "Performance analysis of breast cancer histopathology image classification using transfer learning," Int. J. Electrical & Computer Eng., vol. 72, pp. 24, 2024. http://doi.org/10.11591/ijece.v14i5.pp6006-6015

[5] A. Bagchi, P. Pramanik, and R. Sarkar, “A multi-stage approach to breast cancer classification using histopathology images,” Diagnostics, vol. 13, no. 1, p. 126, Dec. 2022, doi: 10.3390/diagnostics13010126.

[6] M. Yusuf, A. F. Donfack Kana, M. A. Bagiwa, and M. Abdullahi, "Multi-classification of breast cancer histopathological image using enhanced shallow convolutional neural network," J. Eng. Appl. Sci., vol. 72, pp. 24, 2025. doi: 10.1186/s44147-025-00589-w.

[7] X. Zhao, T. Wang, Z. Liu, and H. Chen, "Breast Cancer Histopathological Image Classification Based on High-Order Modeling and Multi-Branch Receptive Fields," Applied Sciences, vol. 15, no. 11, pp. 6085, 2025. doi: 10.3390/app15116085.

[8] A. Sani, M. Royat, and M. Baghshah, "Classification of breast cancer histopathology images using a modified supervised contrastive learning method," Journal of Imaging, vol. 45, no. 3, pp. 156-168, 2025. doi: 10.1007/s11517-024-03245-6.

[9] T. A. Smrity, M. Muntaqim, and et al., "Recent advancements in machine learning and deep learning for early detection of breast cancer: A comprehensive review," Intelligence-Based Medicine, 2026, doi: 10.1016/j.ibreh.2026.100050.

[22] V. Sureshkumar et al., "Breast Cancer Detection and Analytics Using Hybrid CNN and RNN Model," Journal of Personalized Medicine, vol. 14, no. 8, Art. no. 792, 2024. DOI: https://doi.org/10.3390/jpm14080792

[23] R. B. Eshun et al., "A deep convolutional neural network for the classification of imbalanced breast cancer dataset," Intelligence-Based

Manuscript received 20 November 2025; Revised 5 January 2026; Accepted 3 June 2026; Available online 8 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1356

Medicine, 2024, Art. no. 100270, doi: 10.1016/j.ibmed.2024.100270.

[24] S. P. Jakkaladiki and F. Maly, "Integrating hybrid transfer learning with attention-enhanced deep learning models to improve breast cancer diagnosis," PeerJ Computer Science, vol. 10, Art. no. e1850, 2024. doi: 10.7717/peerj-cs.1850.

[25] T. E. Komolafe, Y. Tian, O. J. Awoniya, S. Chen, and X. Yang, "Comparative analysis of cone-beam breast computed tomography and digital breast tomosynthesis for breast cancer diagnosis: A comprehensive study on reconstruction algorithms," International Journal of Imaging Systems and Technology, vol. 34, no. 3, Art. no. e23071, 2024. DOI: https://doi.org/10.1002/ima.23071

[26] N. Brancati and M. Frucci, "Improving Breast Tumor Multi-Classification from High-Resolution Histological Images with the Integration of Feature Space Data Augmentation," Information, vol. 15, no. 2, Art. no. 98, Feb. 2024, doi: 10.3390/info15020098.

[27] F. Schwarzhans et al., "Image normalization techniques and their effect on the robustness and predictive power of breast MRI radiomics," European Journal of Radiology, vol. 187, Art. no. 112086, Jun. 2025, doi: 10.1016/j.ejrad.2025.112086.

[28] Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). *A Dataset for Breast Cancer Histopathological Image Classification*. IEEE Transactions on Biomedical Engineering. doi: 10.1109/TBME.2015.2496264.

[29] LeCun, Y., Bottou, L., Orr, G. B., & Müller, K. R. (2012). *Efficient BackProp*. In Neural Networks: Tricks of the Trade. Springer. doi: 10.1007/978-3-642-35289-8_3.

[30] Komura, D., & Ishikawa, S. (2018). *Machine Learning Methods for Histopathological Image Analysis*. Computational and Structural Biotechnology Journal. doi: 10.1016/j.csbj.2018.01.001.

[31] M. Fatima, R. Zia, and I. A. Usmani, "A hybrid framework integrating deep learning and XAI for breast cancer detection with stratified cross-validation," Systems and Soft Computing, vol. 7, Art. no. 200412, Dec. 2025, doi: 10.1016/j.sasc.2025.200412.

[32] S. Bayard et al., "Screening mammography mitigates breast cancer disparities through early detection of triple negative breast cancer," Clinical Imaging, vol. 80, pp. 430–437, Dec. 2021, doi: 10.1016/j.clinimag.2021.08.013.

[33] M. Krichen, "Convolutional neural networks: A survey," Computers, vol. 12, no. 8, p. 151, Aug. 2023, doi: 10.3390/computers12080151.

[34] D. Dave, T. P. Shah, and A. Chavada, "A modified hybrid CNN-based feature selection and ANN classification framework for patient-level breast cancer histopathology analysis," Franklin Open, vol. 15, Art. no. 100578, 2026, doi: 10.1016/j.fraope.2026.100578.

[35] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint arXiv:1409.1556, 2015, doi: 10.48550/arXiv.1409.1556.

[36] A. S. Wadood et al., "M3-Net++: A multi-scale, multi-level, multi-stream network for nuclei segmentation in breast cancer histopathology using hierarchical context extraction and hybrid loss optimization," Computers in Biology and Medicine, vol. 196, Part B, Art. no. 110804, 2025, doi: 10.1016/j.compbiomed.2025.110804.

[37] Z. Iklima et al., "Reduction of Feature Extraction for COVID-19 CXR using Depthwise Separable Convolution Network," Journal of Electronics, Electromedical Engineering, and Medical Informatics (JEEEMI), vol. 4, no. 4, pp. 204–209, 2022, doi: 10.35882/jeeemi.v4i4.255.

[38] W. Li et al., "Radiomics analysis combining gray-scale ultrasound and mammography for differentiating breast adenosis from invasive ductal carcinoma," Frontiers in Oncology, vol. 14, Art. no. 1390342, 2024, doi: 10.3389/fonc.2024.1390342.

[39] A. H. M. Z. Karim, K. B. O. Mahmud, and C. Shahnaz, "A hybrid feature extraction based ensemble model for breast cancer detection and classification using different medical images," Intelligence-Based Medicine, vol. 12, Art. no. 100290, 2025, doi: 10.1016/j.ibmed.2025.100290.

[40] I. Pacal and O. Attallah, "InceptionNeXt-Transformer: A novel multi-scale deep feature learning architecture for multimodal breast cancer diagnosis," Biomedical Signal Processing and Control, vol. 110, Part B, Art. no. 108116, 2025, doi: 10.1016/j.bspc.2025.108116.

[41] A. Ben Atitallah, J. Kamoun, M. D. Alanazi, T. M. Alanazi, M. Albekairi, and K. Kaaniche, "An advanced medical diagnosis of breast cancer histopathology using convolutional neural networks," Computers, Materials and Continua, vol. 83, no. 3, pp. 5761–5779, 2025, doi: 10.32604/cmc.2025.063634.

[42] F. Badri et al., "Systematic literature review: Research ethics in artificial intelligence development for breast cancer histopathology

Manuscript received 20 November 2025; Revised 5 January 2026; Accepted 3 June 2026; Available online 8 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1356

detection–data privacy, informed consent, and accountability," Informatics, Electrical and Electronics Engineering (Infotron), vol. 5, no. 2, pp. 107–120, 2025, doi: 10.33474/infotron.v5i2.24677.

[43] G. Chyrmang, B. Barua, K. Bora, G. N. Ahmed, A. K. Das, L. Kakoti, B. Lemos, and S. Mallik, "Self-HER2Net: A generative self-supervised framework for HER2 classification in IHC histopathology of breast cancer," *Pathology - Research and Practice*, vol. 270, Art. no. 155961, 2025, doi: 10.1016/j.prp.2025.155961.

[44] H. A. Helaly, M. Badawy, E. M. El-Gendy, and A. Y. Haikal, "Early breast cancer detection, affected cell classification, and segmentation framework," Engineering Applications of Artificial Intelligence, vol. 162, Part D, Art. no. 112598, 2025, doi: 10.1016/j.engappai.2025.112598.

[45] M. Shirdel, M. Di Mauro, dan A. Liotta, "Worthiness Benchmark: A novel concept for analyzing binary classification evaluation metrics," *Information Sciences*, vol. 678, art. 120882, 2024.

[46] S. Szabó, I. J. Holb, V. É. Abriha-Molnár, G. Szatmári, S. K. Singh, and D. Abriha, "Classification assessment tool: A program to measure the uncertainty of classification models in terms of class-level metrics," Applied Soft Computing, vol. 155, Art. no. 111468, 2024, doi: 10.1016/j.asoc.2024.111468.

[47] J. H. Cabot and E. G. Ross, "Evaluating prediction model performance," Surgery, vol. 174, no. 3, pp. 723–726, 2023, doi: 10.1016/j.surg.2023.05.023.

[49] V. N. Thatha et al., "Histopathological image based breast cancer diagnosis using deep learning and bio inspired optimization," Scientific Reports, vol. 15, Art. no. 19034, 2025, doi: 10.1038/s41598-025-04136-8.

[50] Z. Wang et al., "Illuminating breast cancer malignancies: Lightweight histopathology computer vision classifier for precise breast cancer screening," EngMedicine, vol. 2, no. 2, Art. no. 100053, 2025, doi: 10.1016/j.engmed.2024.100053.

[51] N. Erlianita, M. Itqan Mazdadi, T. H. Saragih, M. Reza Faisal, and Muliadi, "Implementation of SMOTE and Whale Optimization Algorithm on Breast Cancer Classification Using Backpropagation," *J. Electron. Electromedical Eng. Med. Informat.*, vol. 5, no. 4, pp. 303-313, Oct. 2023, doi:10.35882/jeeemi.v5i4.334.

[52] I. Nurcahyati, T. H. Saragih, A. Farmadi, D. Kartini, and Muliadi, "Classification of lung disease in X-ray images using gray level co-occurrence matrix method and convolutional neural network," J. Electron. Electromedical Eng. Med. Informat., vol. 6, no. 4, pp. 332–342, Oct. 2024, doi: 10.35882/jeeemi.v6i4.457.

[53] A. Desiani, "Data augmentation and U-Net CNN for accurate nuclei cells segmentation," J. Electron. Electromedical Eng. Med. Informat., vol. 6, no. 3, pp. 264–275, Jul. 2024, doi: 10.35882/jeeemi.v6i3.442.

[54] A. Chauhan and I. Kumar, "Deep feature extraction and optimized VGG16-SVM architecture for breast cancer characterization," *Discover Computing*, vol. 28, art. no. 208, 2025. DOI: 10.1007/s10791-025-09736-6.

[55] M. A. Wakili, N. Garg, A. Sinha, and M. Gupta, "Classification of Breast Cancer Histopathological Images Using DenseNet and Transfer Learning," *Procedia Computer Science*, vol. 237, pp. 1079–1086, 2024. DOI: 10.1016/j.procs.2024.08.189.

[56] A. Heikal, A. El-Ghamry, S. Elmougy, and M. Z. Rashad, "Fine-tuning deep learning models for breast tumor classification," *Scientific Reports*, vol. 14, no. 10753, pp. 1–26, 2024.

[57] I. G. S. M. Diyasa, S. Humairah, E. Y. Puspaningrum, F. D. Durry, W. D. Lestari, and W. Caesarendra, "Improving classification accuracy of breast ultrasound images using Wasserstein GAN for synthetic data augmentation," Journal of Robotics and Control (JRC), vol. 6, no. 4, pp. 1962–1971, 2025, doi: 10.18196/jrc.v6i4.25075.

[58] F. Badri, S. Patmanthara, and I. A. E. Zaeni, "Integration of knowledge-based CNN model for breast cancer histopathology image classification," ILKOMNIKA, vol. 7, no. 3, pp. 340–353, 2025, doi: 10.28926/ilkomnika.v7i3.801.

[59] A. S. Hyperastuty, A. N. Handayani, and H. W. Herwanto, "Grading breast cancer with deep learning method in the pathology image," in Proc. 2025 2nd Beyond Technology Summit on Informatics International Conference (BTS-I2C), 2025, pp. 272–275, doi: 10.1109/BTS-I2C67944.2025.11399422.

Manuscript received 20 November 2025; Revised 5 January 2026; Accepted 3 June 2026; Available online 8 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1356

Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 3, July 2026, pp: 1117-1132. e-ISSN: 2656-8632

Author Biography

Fawaidul Badri, S.Kom., M.T. received his Bachelor's degree in Computer Science (S.Kom.) from the Informatics Engineering Study Program, Universitas Trunojoyo Madura, in 2013. His undergraduate research focused on face recognition using the Fisherface method. During his studies, he served as a teaching assistant under the supervision of Prof. Arif Muntasa. He then pursued a Master's degree in Electrical Engineering at Institut Teknologi Sepuluh Nopember (ITS) Surabaya, specializing in Intelligent Multimedia Networks. His Master's thesis explored 3D image registration based on RGB-D data using the Scale Invariant Feature Transform (SIFT) method under the guidance of Dr. Eko Mulyanto Yuniarno. He obtained his Master of Engineering (M.T.) degree in 2016. Currently, he is pursuing a Doctoral degree in Electrical and Informatics Engineering at Universitas Negeri Malang. His research interests include image processing, computer vision, and deep learning applications for breast cancer histopathology image classification.

Ilham Ari Elbaith Zaeni, S.T., M.T., Ph.D. is a Lecturer at the Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Indonesia. He received his Bachelor of Engineering (S.T.) degree in Electrical/Electronics Engineering from Brawijaya University, Malang, in 2005, and his Master of Engineering (M.T.) degree in Control Systems and Electronics from the same university in 2012. He earned his Doctor of Philosophy (Ph.D.) degree in Biomedical Engineering from Southern Taiwan University of Science and Technology in 2017.

His research interests include biomedical electronic signal processing, machine learning, microcontroller design, and their applications in biomedical systems and healthcare technology. Dr. Ilham actively develops intelligent medical devices, embedded systems for real-time physiological signal monitoring, and machine learning models for early disease detection. His research focuses on integrating microcontroller technology with biomedical signal processing to create affordable and portable healthcare solutions. Throughout his academic career, he has been involved in numerous research projects, scientific publications, and collaborative initiatives that bridge engineering and medicine, making significant contributions to the advancement of healthcare technology.

Dr. Heru Wahyu Herwanto is a Lecturer in the Department of Electrical Engineering and Informatics at Universitas Negeri Malang. He completed his Bachelor of Engineering (B.Eng.) degree in Electrical and Electronics Engineering from Brawijaya University in 1995, his Master's degree in Electrical Engineering from Institut Teknologi Sepuluh Nopember (ITS) Surabaya in 2007, and earned his Doctorate (Ph.D.) in Electrical Engineering from Universitas Negeri Malang in 2020.

His research interests and academic expertise cover various fields, including informatics, image processing, computer vision, artificial intelligence, and electronics. Dr. Heru focuses on developing intelligent systems that integrate artificial intelligence with digital image analysis and processing. Many of his research projects are in the biomedical field and medical imaging, where he develops computer vision applications to support medical diagnosis, medical image processing, and automatic detection systems on radiological and pathological images. His work explores practical applications of computer vision in education, industry, smart technology, and healthcare.

Muhammad Khusairi Osman obtained his B.Eng. degree in Electrical and Electronic Engineering in 2000 and his M.Sc. degree in Electrical and Electronic Engineering in 2004 from Universiti Sains Malaysia (USM). In 2014, he earned his Ph.D. in Medical Electronic Engineering from Universiti Malaysia Perlis (UniMAP), Malaysia. He is currently serving as a Senior Lecturer at the Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), Malaysia.

His research interests include Image Processing, Pattern Recognition, and Artificial Intelligence. Dr. Muhammad Khusairi actively explores advanced digital image analysis techniques, intelligent pattern classification, and the application of machine learning algorithms to solve various real-world engineering problems. His research strongly emphasizes medical image processing, automatic disease detection systems, and the development of intelligent diagnostic tools. Throughout his career, he has made significant contributions through numerous research projects, academic publications, and collaborative initiatives that integrate artificial intelligence with medical electronics.

Manuscript received 20 November 2025; Revised 5 January 2026; Accepted 3 June 2026; Available online 8 July 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i3.1356