Vol. 6 No. 3, 2025, pp. 254-263 **Jurnal Pepadun**

Two-Stage Convolutional Neural Network (CNN) Architectures for Breast Cancer Image Classification

[*1]Admi Syarif, [2]Adinda Aulia Sari, [3]Wartariyus, [4]Favorisen R. Lumbanraja, [5]Apri Candra

[1,2,3,4,5] Department of Computer Science, Universitas Lampung,
Jl. Prof. Dr. Ir. Sumantri Brojonegoro No.1, Bandar Lampung, Lampung, Indonesia, 35145
e-mail: *[1]**admi.syarif@fmipa.unila.ac.id**, [2]adindaaul90@gmail.com, [3]wartariyus@fmipa.unila.ac.id,
[4]favorisen.lumbanraja@fmipa.unila.ac.id, [5]apricandra1982@gmail.com

Abstract *- Breast cancer remains one of the most common and deadly diseases among women globally. Early detection significantly increases the chances of patient recovery. The main objective of this research is to evaluate the performance of three Convolutional Neural Network (CNN) architectures, namely ResNet50, VGG16, and DenseNet201, for breast cancer image classification. In this study, there are two classification stages used: the first is to differentiate between normal and abnormal images, and the second is to distinguish between benign and malignant tumors. The dataset was obtained through the Kaggle website. It was then pre-processed using normalization and augmentation through flipping and rotation. After each CNN model was trained using transfer learning, its performance was evaluated using accuracy, precision, recall, and F1 score. In the Normal and Abnormal classification task, the DenseNet201 model outperformed other models with an accuracy of 91%. Meanwhile, ResNet50 showed the most optimal results in the Benign and Malignant classification with an accuracy of 83%.*

Keywords*: Breast Cancer; CNN; DenseNet201; Image Classification; ResNet50.*

1. INTRODUCTION

Breast cancer is one of the most common diseases suffered by women worldwide. The abnormal and irregular growth of breast tissue that then develops into a structure resembling a tumor is known as breast cancer [1]. Based on Globocan data from 2020, more than 396,914 cancer cases occurred in Indonesia, with breast cancer ranking first with 68,858 cases or 16.6% of the total cancer cases in Indonesia [2]. Accurate predictions of breast cancer are very important for determining the appropriate treatment and improving patient survival rates [3]. However, the manual diagnosis process requires analysis of complex and large amounts of data, which certainly poses a significant challenge for medical professionals. Therefore, computer-based systems are increasingly being used to assist in the disease classification process, including breast cancer, in line with the development of Artificial Intelligence (AI) technology [4].

Artificial Intelligence (AI) performs well in improving consistency and efficiency in medical examinations, as well as reducing the rate of diagnostic errors [5]. The application of AI can help reduce the number of false negatives and false positives in breast cancer. This is because AI is not influenced by subjectivity and can reduce the workload of radiologists by 88% with a decrease in reading time per case of up to 11%. The implementation of AI can improve accuracy in mammography [6].

One of the machine learning algorithms used for image recognition is the Convolutional Neural Network (CNN). CNN automatically analyzes the visual representation of images by optimizing filters or kernels. As the name suggests, the main concept of CNN is the convolution operation, where filters are applied to the input image. This helps in pattern recognition at deeper layers [7]. The Convolutional Neural Network (CNN) method has the potential to produce beneficial results, thereby enhancing diagnostic capabilities [8]. Several methods, such as mammography and histopathological examination, can be used to detect breast cancer through medical imaging. Although histopathology is still performed manually by pathologists considering the cancer severity score, it is the gold standard examination in establishing a breast cancer diagnosis [9]. To

© 2025 The Authors | doi: 10.23960/pepadun.v6i3.292

improve the chances of recovery, early detection is very important, but traditional approaches are often not accurate enough to find malignant tumors at an early stage [10].

Previous studies have shown that CNNs can detect breast cancer more accurately than conventional approaches [11]. Research [12] used AlexNet and showed that its architecture in eye cancer classification using AlexNet achieved an accuracy of 98.37%. Research [13] used breast histopathology images obtained from a public dataset, successfully achieving an accuracy of 96% in oral cancer classification using the DenseNet201 architecture. The dataset must consist of clear and detailed images of oral cancer, so the model can learn unique patterns from each image. Previous research has used alternative CNN structures such as the VGG-16 architecture. Study [14] evaluated the model's performance in breast cancer classification using this structure. The evaluation results showed that the VGG-16 model achieved an accuracy of 78.87% based on the metrics of accuracy, precision, recall, and F1 score.

However, the main challenge in breast cancer classification is the diversity of tumor types that can appear. The three main categories of breast tumors are Benign, Malignant, and Normal. To assess the effectiveness of the CNN model in that classification, two testing schemes were conducted. In the first classification, namely Normal and Abnormal, which aims to detect the early presence of abnormalities, and the second classification, namely Benign and Malignant, which focuses on the level of tumor malignant [15]. To assess the performance of the CNN model in breast cancer classification, evaluation metrics such as accuracy, precision, recall, and F_1 score will be used. By using these evaluation metrics, we can determine how well the model is able to distinguish between Normal, Benign, and Malignant images [16].

This study aims to evaluate a CNN-based breast cancer classification system using various architectures, including ResNet50, VGG16, and DenseNet201. The objective is to evaluate and compare their performance to identify the most effective model for breast cancer detection. The findings are expected to guide future improvements in deep learning-based diagnostic tools.

2. RESEARCH METHODOLOGY

This research was conducted through several stages to achieve optimal results. Figure 1 illustrates the sequential research phases, commencing from the data acquisition process through to the evaluation of model performance.

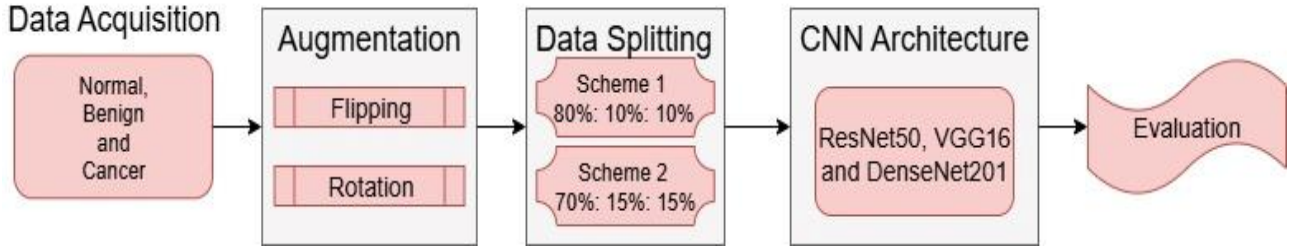
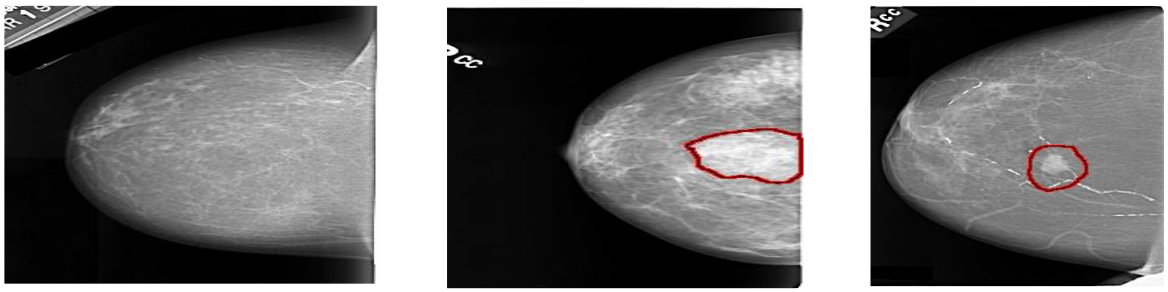


Figure 1. The steps of the research.

2.1. Data Acquisition

The data used was obtained from the Kaggle website https://www.kaggle.com/datasets/cheddad/miniddsm/data. The mammography images were sourced from publicly available datasets, with the original data collected in 2020. The breast cancer images are in PNG format and categorized into three classes: Normal, Benign, and Malignant. The classification process in this study was carried out using two schemes. Scheme 1 differentiates between normal and abnormal breast tissue as an initial screening step to identify whether there are any indications of cancer. Images classified as abnormal are then processed further in Scheme 2, which distinguishes between benign and Malignant. Tissue, this second scheme is more complex as it focuses on identifying specific characteristics within cancerous tissue. Class grouping can be seen in Figure 2.

 | doi: 10.23960/pepadun.v6i3.292



a. Normal. b. Benign. c. Malignant

Figure 2. The classification of breast cancer.

2.2. Augmentation

Data augmentation is a technique used to increase the amount of data to balance the dataset, help the model learn data variations, and prevent overfitting [17]. At this stage, augmentation adds data used to train the model and improve performance on unseen data [18]. The augmentation methods applied in this study are flipping and rotation. The number of datasets after augmentation is presented in Table 1 and Table 2.

Table 1. Dataset for the first steps.

Class	Original	Augmented Dataset
Normal	2728	6956
Abnormal	6956	6956
Total	**9684**	**13912**

Table 2. Dataset for the second steps.

Class	Original	Augmented Dataset
Benign	3360	3596
Cancer	3596	3596
Total	**6959**	**7192**

2.3. Data Splitting

In this study, two testing schemes were applied to evaluate the performance of the models. Data distribution for data first split steps. In Step 1, there are 11.130 training samples, 1.391 validation samples, and 1.391 testing samples. In Scheme 2, the dataset includes 9.738 training samples, 2.087 for validation, and 2.087 for testing. Meanwhile, Table 5 presents the data distribution for the data split into the second step. In Step 2, 5.753 images were used for training, 719 for validation, and 719 for testing. In Scheme 2, the dataset consists of 5.034 training images, 1.078 validation images, and 1.078 testing images.

2.4. Architecture Convolutional Neural Network (CNN)

There are three architectures used in this study, as shown in Figure 3, Figure 4, and Figure 5.

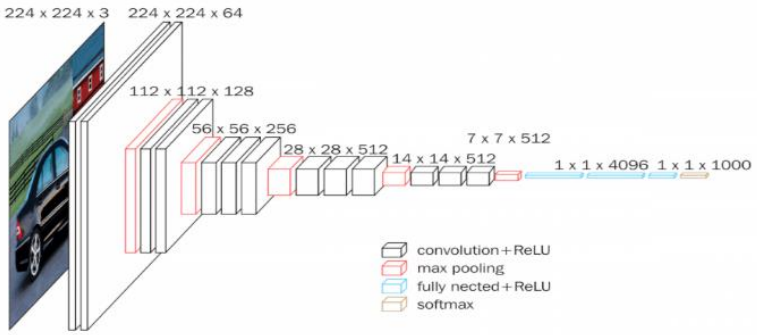


Figure 3. VGG16 architecture.

© 2025 The Authors | doi: 10.23960/pepadun.v6i3.292

Figure 3 illustrates the architecture of VGG16, which consists of 13 convolutional layers and 3 fully connected (FC) layers. The default input image size is 224 × 224 pixels. All convolutional layers in VGG16 use 3×3 kernels with a stride of 1 and 'same' padding, while the pooling layers use 2×2 kernels with a stride of 2. The first convolutional block contains 64 filters, followed by blocks with 128, 256, and increasing numbers of filters in deeper layers. After the convolutional layers, the network includes three fully connected layers. The first FC layer has 4,096 units, and the final FC layer performs classification for the ILSVRC dataset with 1,000 output channels, one for each class [19].

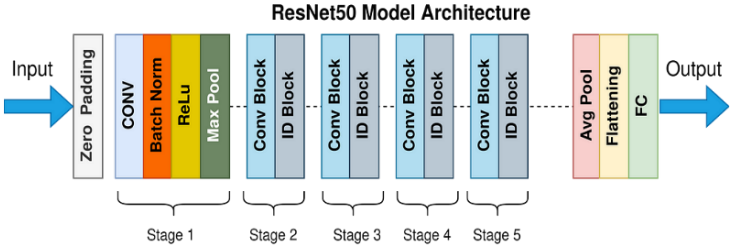


Figure 4. ResNet50 architecture.

The ResNet-50 architecture illustrated in Figure 4 was adapted for this study. The model begins with zero-padding and initial convolutional layers for feature extraction, followed by batch normalization and ReLU activation. It then advances through four stages composed of residual blocks with skip connections, which help address the vanishing gradient issue. Modifications include the addition of a Global Average Pooling 2D layer to compress spatial dimensions, and a dense layer with sigmoid activation to enable classification into three categories. These modifications align with transfer learning techniques designed for domain-specific applications [20].

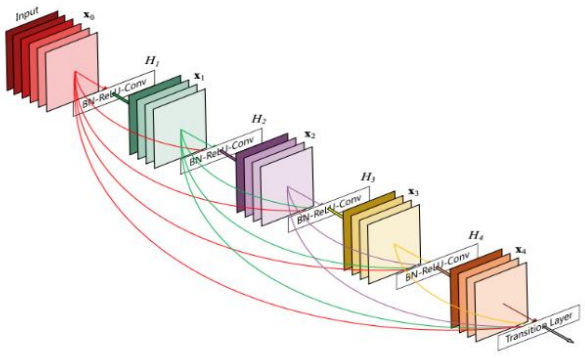


Figure 5. DenseNet201 architecture.

The DenseNet201 architecture illustrated in Figure 5 was adapted for this study. The pretrained DenseNet201 model on the ImageNet dataset was utilized as the backbone in this study. All convolutional layers of the base model were frozen to retain the initial weights and avoid retraining the fundamental features already learned. To adapt the top of the network, a GlobalAveragePooling2D layer was added to convert the spatial dimensions of the backbone output into a one-dimensional vector. This was followed by a Dense layer with 256 units and a ReLU activation function to enhance feature representation and reduce the risk of overfitting [21].

2.5. Performance Evaluation

Model evaluation is the process of measuring how well a trained model can generate accurate predictions. The model's performance will be assessed using evaluation metrics such as accuracy, precision, recall, and F_1-score [22]. The evaluation of the model's performance is also assisted by a confusion matrix to illustrate the prediction results and classification errors for each class. This helps to see how well the model can recognize disease patterns in breast cancer.

 | doi: 10.23960/pepadun.v6i3.292

A confusion matrix is a method widely used in machine learning to evaluate or visualize how a model performs in the context of classification [23]. The matrix used includes accuracy, precision, recall, and F_1-score. This method is widely used for classification in CNN architectures. Table 3 shows the confusion matrix for classification.

Table 3. Confusion matrix binary classification.

Actual/Predicted	Negative Prediction	Positive Prediction
Actual Negative	TN	FP
Actual Positive	FN	TP

According to [24], There are 4 terms commonly used as representations of results in the confusion matrix classification process. The four terms are as follow:

1. True Negative (TN): Occurs when the predicted class is negative, and the prediction is correct. For example, the system predicts benign breast cancer, and it is indeed true that the cancer is in a benign condition.
2. False Negative (FN): Occurs when the predicted class is negative, and the prediction is incorrect. For example, the system predicts benign breast cancer, whereas the cancer is malignant.
3. False Positive (FP): Occurs when the predicted class is positive, but the prediction is incorrect. For example, the system predicts that breast cancer is malignant, but in reality, the cancer is benign.
4. True Positive (TP): Occurs when the predicted class is positive, and the prediction is correct. For example, the breast cancer detection system is affected by malignant cancer, and it is true that the cancer is indeed malignant.

Accuracy reflects the model's ability to correctly classify instances. In classification tasks, accuracy refers to the proportion of both true positive and true negative predictions among the total number of cases. However, in imbalanced datasets, a model that mostly predicts the majority class (positive) may achieve high accuracy but fail to detect the minority class (negative). Therefore, while accuracy provides a general measure of performance, it may not always reflect the model's effectiveness in identifying minority classes [25]. Calculation of accuracy is presented in Equation 1.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (1)$$

Precision measures how well the model predicts the positive class among all instances it has labeled as positive. In other words, it evaluates the proportion of true positive predictions relative to the total predicted positives. Precision is useful in scenarios where false positives carry a high cost [25]. Calculation of precision is presented in Equation 2.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

Recall, also known as sensitivity, indicates how effectively the model identifies actual positive cases. It is the ratio of true positive predictions to the total number of actual positives in the data. A high recall means the model successfully detects most of the positive cases [25]. Calculation of recall is presented in Equation 3.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (3)$$

The F1-score provides a harmonic mean of precision and recall, offering a balance between the two metrics. It is particularly useful when a balance between false positives and false negatives is needed. A high F_1-score indicates that the model performs well in terms of both precision and recall [25]. Calculation of F_1-score is presented in Equation 4.

$$\text{F}_1\text{Score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall+Precision}} \quad (4)$$

 | doi: 10.23960/pepadun.v6i3.292

3. NUMERICAL EXPERIMENTS & RESULTS

This study focuses on breast cancer classification using 3 CNN architectures, namely ResNet-50, DenseNet201, and VGG16, using the "Breast Cancer" dataset, which consists of various resolutions. All images of the dataset initially have varying sizes, but to train the CNN model, the images are resized to 224 x 224 pixels using the downscaling method. This process aims to simplify the model input without losing important information. The model training process in this study was carried out using a batch size of 64, which means that the model processes 64 images at once before updating its parameters. To optimize model performance, the Adam optimizer was chosen because of its efficiency and adaptive learning capabilities. Training was carried out for 30 epochs, allowing the model to iteratively learn from the data through multiple passes. In addition, a learning rate of 0.0001 was used to ensure gradual and stable updates to the model weights, helping to prevent overfitting during optimization and contributing to more accurate convergence.

Table 4. Comparison of the confusion matrix in normal and abnormal classification.

Scheme	Model	True Positive (TP)	True Negative (TN)	False Positive (FP)	False negative (FN)	Accuracy
1	DenseNet201	660	609	87	36	0.9116
	VGG16	626	582	114	70	0.8678
	ResNet50	606	485	211	90	0.7838
2	**DenseNet201**	978	938	106	938	**0.9155**
	VGG16	911	939	105	133	0.8867
	ResNet50	911	760	284	133	0.7969

Table 4 presents a comparison of the confusion matrix results in the classification of breast cancer images categorized as normal and abnormal for two testing schemes. In scheme 1, the DenseNet201 model showed the best performance with an accuracy of 91.16% (TP=660, TN=609, FP=87, FN=36), VGG16 with an accuracy of 86.78%, and ResNet50 with an accuracy of 78.38%. Meanwhile, in scheme 2, DenseNet201 again achieved the highest accuracy of 91.55% (TP=978, TN=938, FP=106, FN=83), followed by VGG16 with an accuracy of 88.67%, and ResNet50 with an accuracy of 79.69%. These results indicate that DenseNet201 consistently provides the best performance in both classification schemes.

Table 5. Comparison of accuracy in normal and abnormal classification.

Scheme	Model	Precision	Recall	F1- Score
1	DenseNet201	0.9139	0.9116	0.9115
	VGG16	0.8693	0.8678	0.8677
	ResNet50	0.7926	0.7838	0.7821
2	**DenseNet201**	**0.9128**	**0.9180**	**0.9151**
	VGG16	0.8888	0.8848	0.8864
	ResNet50	0.8040	0.7953	0.7948

Table 5 presents a comparison of evaluation results across all CNN architectures tested in this study. After conducting multiple experiments with various data schemes and hyperparameters, the best performance for each architecture was identified. The highest accuracy in the classification of Normal and Abnormal categories was achieved by the DenseNet201 architecture using Scheme 2, with a data split of 70:15:15 and 30 training epochs. This model reached a maximum accuracy of 91%. Figures 6 show the plot charts for both schemes in the Normal and Abnormal classification.

Table 6. Comparison of the confusion matrix in benign and malignant classification.

Scheme	Model	True Positive (TP)	True Negative (TN)	False Positive (FP)	False negative (FN)	Accuracy
1	DenseNet201	233	287	73	127	0.7222
	VGG16	295	271	89	65	0.7819
	ResNet50	494	376	164	46	0.8278

 | doi: 10.23960/pepadun.v6i3.292

Scheme	Model	True Positive (TP)	True Negative (TN)	False Positive (FP)	False negative (FN)	Accuracy
2	**DenseNet201**	419	398	142	121	0.7565
	VGG16	411	426	92	77	0.7750
	ResNet50	463	448	114	129	**0.8343**

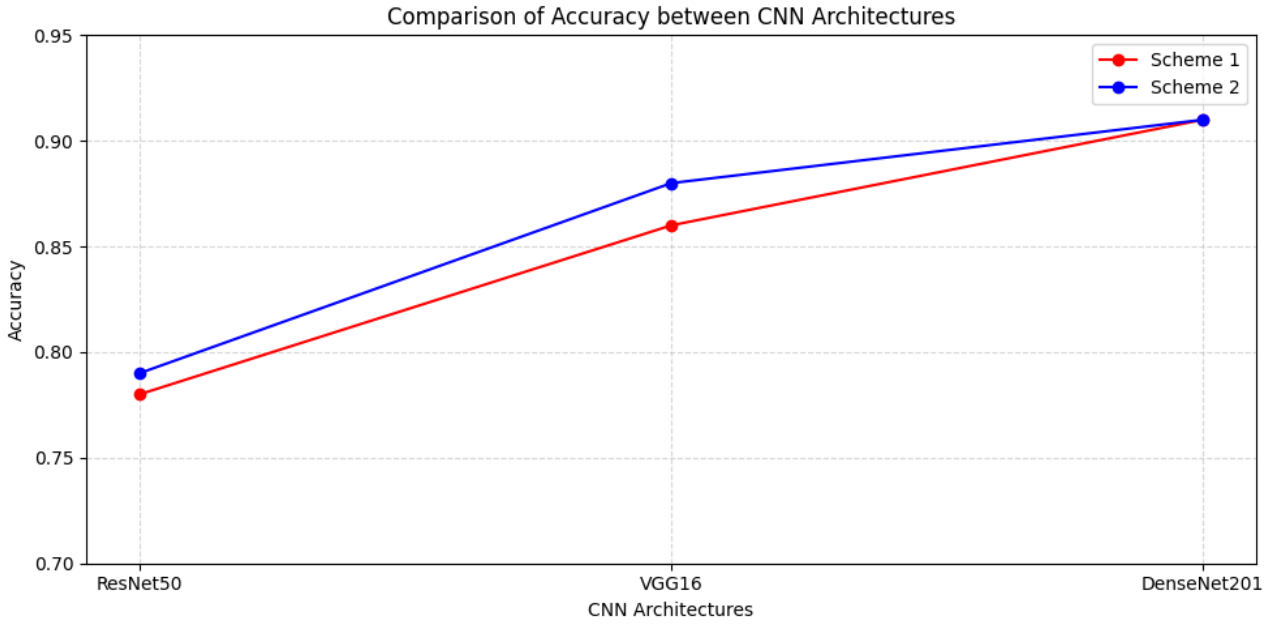


Figure 6. Accuracy graph of normal and abnormal CNN architectures.

Table 6 shows a comparison of the confusion matrix results for the classification of benign and malignant categories under two testing schemes. In scheme 1, the ResNet50 model achieved the highest accuracy of 82.78% (TP=494, TN=376, FP=164, FN=46), VGG16 with an accuracy of 78.19%, and DenseNet201 with an accuracy of 72.22%. In scheme 2, ResNet50 again showed the best performance with an accuracy of 83.43% (TP=463, TN=448, FP=114, FN=129), followed by VGG16 with an accuracy of 77.50%, and DenseNet201 with an accuracy of 75.65%. These findings indicate that, unlike normal-abnormal classification, in benign-malignant classification, the ResNet50 model outperforms other architectures.

Table 7. Comparison of accuracy in benign and malignant classification.

Scheme	Model	Precision	Recall	F1-Score
1	DenseNet201	0.7273	0.7222	0.7207
	VGG16	0.7826	0.7819	0.7818
	ResNet50	0.8311	0.8278	0.8273
2	DenseNet201	0.7569	0.7565	0.7564
	VGG16	0.7752	0.7750	0.7750
	ResNet50	**0.8345**	**0.8343**	**0.8342**

Table 7 presents a comparison of evaluation results for all CNN architectures used in this study. After conducting several experiments with different data schemes and hyperparameter settings, the best performance was identified for each tested architecture. The highest accuracy in classifying Benign and Malignant categories was achieved by the ResNet50 architecture under both Scheme 2, with data splits of 70:15:15, respectively, and 30 training epochs. This model reached a maximum accuracy of 83%. Figures 7 show the plot charts for both schemes in the Benign and Malignant classification.

The classification results in Step 1, Normal and Abnormal, tend to be better compared to Step 2, Benign and Malignant, because the differences in visual characteristics between the classes in Step 1 are clearer and easier

 | doi: 10.23960/pepadun.v6i3.292

for the model to recognize. At this stage, the model only needs to distinguish between healthy breast tissue (Normal) and tissue showing abnormalities (Abnormal), where the differences in texture, shape, or structure between the two are usually quite striking.

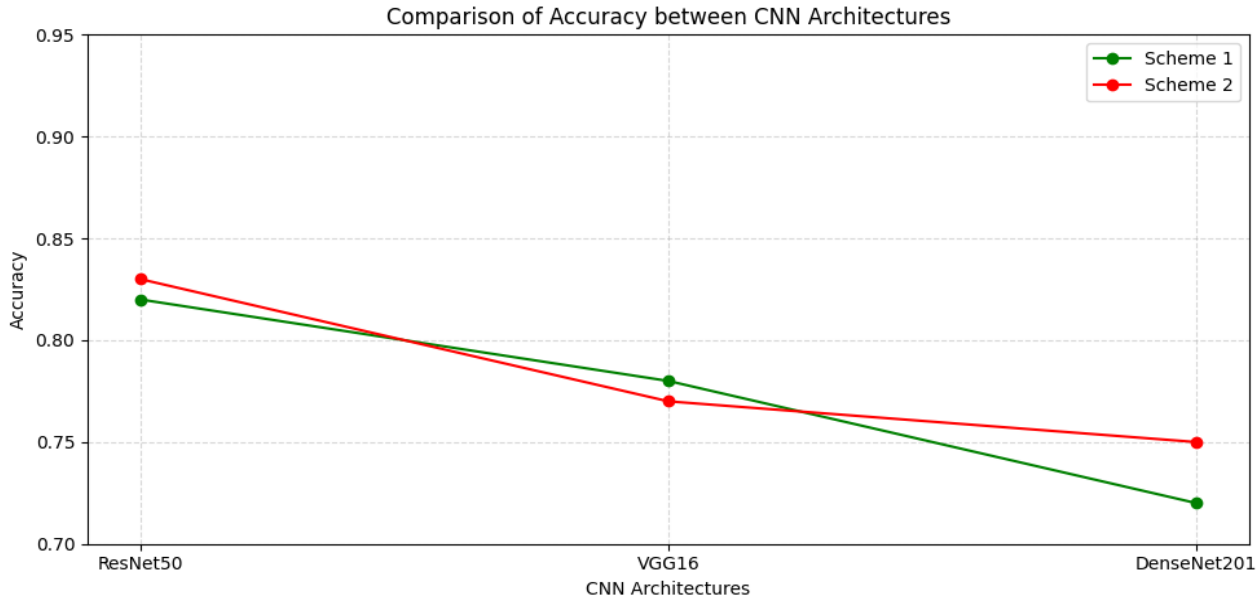


Figure 7. Accuracy graph of benign and malignant CNN architectures.

In contrast, at Step 2, the model must classify the detected abnormalities as benign or malignant. At this stage, the challenge is much greater because benign and malignant tissues often have high visual similarity, and the differences between them can be very subtle even for experienced medical professionals. Therefore, the model requires more complex feature representations and larger data sets to accurately distinguish between the two classes. As a result, the model's performance in Step 2 is generally lower compared to Step 1.

4. CONCLUSIONS

This study confirms that a two-stage classification approach is more effective than direct multi-class classification. By separating the tasks first, identifying Normal and Abnormal images, then distinguishing Benign from Cancer, the model is better able to learn distinct visual features, leading to improved accuracy and reduced misclassification. Experimental results show DenseNet201 achieved the best performance in Scheme 2 with 91% accuracy using a 70:15:15 split, while ResNet50 performed best in Scheme 2 with 83% accuracy under a 70:15:15 split, indicating its strength in handling more complex patterns. The use of additional metrics such as Precision, Recall, and F1-Score is also essential for evaluating performance, especially with imbalanced data. Future work is encouraged to explore ensemble learning, improved preprocessing techniques, noise reduction, contrast enhancement, and more diverse datasets to further enhance model reliability and generalization in medical image classification.

ACKNOWLEDGMENTS

The authors express sincere gratitude to the Department of Computer Science, Faculty of Mathematics and Natural Sciences, University of Lampung, for providing the facilities and support required for this research.

LITERATURE

[1] T. T. M. M. Khan, “Breast Tumor Detection Using Robust and Efficient Machine Learning and Convolutional Neural Network Approaches,” *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi:

10.1155/2022/6333573.

[2] P. F. Wiliyanarti, "Kualitas Hidup Pasien Kanker Payudara "Pendekatan Health Belief Model " - Pipit Festi Wiliyanarti - Google Buku," 2021.

[3] M. Diwakaran and D. Surendran, "Breast Cancer Prognosis Based on Transfer Learning Techniques in Deep Neural Networks," *Inf. Technol. Control*, vol. 52, no. 2, pp. 381–396, 2023, doi: 10.5755/j01.itc.52.2.33208.

[4] A. Cruz-Roa, "Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks," in *Medical Imaging 2014: Digital Pathology*, 2014. doi: 10.1117/12.2043872.

[5] S. Pacilè, J. Lopez, P. Chone, T. Bertinotti, J. M. Grouin, and P. Fillard, "Improving breast cancer detection accuracy of mammography with the concurrent use of an artificial intelligence tool," *Radiol. Artif. Intell.*, vol. 2, no. 6, 2020, doi: 10.1148/ryai.2020190208.

[6] J. Liu, "Mammography diagnosis of breast cancer screening through machine learning: a systematic review and meta-analysis," *Clin. Exp. Med.*, vol. 23, no. 6, 2023, doi: 10.1007/s10238-022-00895-0.

[7] M. Busaleh, M. Hussain, H. A. Aboalsamh, and Fazal-E-amin, "Breast mass classification using diverse contextual information and a convolutional neural network," *Biosensors*, vol. 11, no. 11, 2021, doi: 10.3390/bios11110419.

[8] V. Ulagamuthalvi, G. Kulanthaivel, A. Balasundaram, and A. K. Sivaraman, "Breast Mammogram Analysis and Classification Using Deep Convolution Neural Network," *Comput. Syst. Sci. Eng.*, vol. 43, no. 1, pp. 275–289, 2022, doi: 10.32604/csse 2022.023737.

[9] E. Susilowati, A. T. Hapsari, M. Efendi, and P. Edi, "Diagnosa Penyakit Kanker Payudara Menggunakan Metode K - Means Clustering," *J. Sist. Informasi, Teknol. Inform. dan Komput.*, vol. 10, no. 1, pp. 27–32, 2019.

[10] G. A. R. Dyanti and N. L. P. Suariyani, "faktor-faktor keterlambatan penderita kanker payudara dalam melakukan pemeriksaan awal ke pelayanan kesehatan," *J. Kesehat. Masy.*, vol. 11, no. 2, 2016, doi: 10.15294/kemas.v11i2.3742.

[11] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A Dataset for Breast Cancer Histopathological Image Classification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, 2016, doi: 10.1109/TBME.2015.2496264.

[12] F. N. Cahya, N. Hardi, D. Riana, and S. Hadiyanti, "Klasifikasi Penyakit Mata Menggunakan Convolutional Neural Network (CNN)," *SISTEMASI*, vol. 10, no. 3, 2021, doi: 10.32520/stmsi.v10i3.1248.

[13] M. Fauzan Novriandy, B. Rahmat, and A. Junaidi, "Klasikasi Citra Pada Penyakit Kanker Mulut Menggunakan Arsitektur Densenet201 Menggunakan Optimasi Adam Dan Sgd," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 4, pp. 6132–6140, 2024, doi: 10.36040/jati.v8i4.10077.

[14] I. Idawati, D. P. Rini, A. Primanita, and T. Saputra, "Klasifikasi Kanker Payudara Menggunakan Metode Convolutional Neural Network (CNN) dengan Arsitektur VGG-16," *J. Sist. Komput. dan Inform.*, vol. 5, no. 3, p. 529, 2024, doi: 10.30865/json.v5i3.7553.

[15] R. Nima and F. Shila, "Crack classification in rotor-bearing system by means of wavelet transform and deep learning methods: an experimental investigation," *J. Mech. Eng. Autom. Control Syst.*, vol. 1, no. 2, 2020, doi: 10.21595/jmeacs 2020.21799.

[16] K. Liu, G. Kang, N. Zhang, and B. Hou, "Breast Cancer Classification Based on Fully-Connected Layer First Convolutional Neural Networks," *IEEE Access*, vol. 6, 2018, doi: 10.1109/ACCESS.2018.2817593.

[17] Y. Huang, R. Li, X. Wei, Z. Wang, T. Ge, and X. Qiao, "Evaluating Data Augmentation Effects on the Recognition of Sugarcane Leaf Spot," *Agric.*, vol. 12, no. 12, 2022, doi: 10.3390/agriculture12121997.

 | doi: 10.23960/pepadun.v6i3.292

[18] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, "A review of medical image data augmentation techniques for deep learning applications," 2021. doi: 10.1111/1754-9485.13261.

[19] A. S. Almryad and H. Kutucu, "Automatic identification for field butterflies by convolutional neural networks," *Eng. Sci. Technol. an Int. J.*, vol. 23, no. 1, pp. 189–195, 2020, doi: 10.1016/j.jestch.2020.01.006.

[20] S. R. and J. S. K. He, X. Zhang, "'Deep Residual Learning for Image Recognition,' 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778. doi: 10.1109/CVPR.2016.90," *2016 IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016.

[21] N. D. Miranda, L. Novamizanti, and S. Rizal, "convolutional neural network pada klasifikasi sidik jari menggunakan resnet-50," *J. Tek. Inform.*, vol. 1, no. 2, 2020, doi: 10.20884/1.jutif.2020.1.2.18.

[22] K. Bhosle and V. Musande, "Evaluation of CNN model by comparing with convolutional autoencoder and deep neural network for crop classification on hyperspectral imagery," *Geocarto Int.*, vol. 37, no. 3, 2022, doi: 10.1080/10106049.2020.1740950.

[23] O. Caelen, "A Bayesian interpretation of the confusion matrix," *Ann. Math. Artif. Intell.*, vol. 81, no. 3–4, 2017, doi: 10.1007/s10472-017-9564-8.

[24] V. Kotu and B. Deshpande, *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. 2014. doi: 10.1016/C2014-0-00329-2.

[25] A. Vanacore, M. S. Pellegrino, and A. Ciardiello, "Fair evaluation of classifier predictive performance based on binary confusion matrix," *Comput. Stat.*, vol. 39, no. 1, 2024, doi: 10.1007/s00180-022-01301-9.