Volume 11 Issue 1 Year 2026 | Page 116-127 ISSN: 2527-9866
Received: ***18-12-2025*** | *Revised:* ***30-12-2025*** | *Accepted:* ***15-01-2025***

Performance Analysis of MobileNetV2 and GhostNetV2 in Classifying Cervical Cancer Images in the SIPaKMeD Dataset

Shela[1], Siska Devella[2]
Multi Data Palembang University, Palembang, South Sumatra, Indonesia, 30113
e-mail: shela_2226250087@mhs.mdp.ac.id[1], siskadevella@mdp.ac.id[2]
*Correspondence: shela_2226250087@mhs.mdp.ac.id

Abstract: *Cervical cancer remains a significant global health burden, largely due to limited screening coverage and the reliance on manual cytological interpretation. The intrinsic complexity of cervical cell morphology and constraints in clinical resources necessitate automated classification systems that are both accurate and computationally efficient. This study aims to evaluate and compare the performance of two lightweight CNN architectures, MobileNetV2 and GhostNetV2, for cervical cell image classification using the SIPaKMeD dataset. The dataset comprises 4,049 cell images, which were preprocessed through normalization, augmentation, and partitioning into training, validation, and testing sets. Both models were implemented using transfer learning and trained under comparable hyperparameter settings with basic data augmentation. Model performance was assessed using confusion matrices and standard evaluation metrics, including accuracy, precision, recall, and F1-score. Experimental results demonstrate that MobileNetV2 achieved superior performance with an accuracy of 98.50%, outperforming GhostNetV2, which attained a maximum accuracy of 97.60%. The consistent performance across metrics indicates robust and balanced classification capability. These findings suggest that MobileNetV2 offers an optimal trade-off between accuracy and computational efficiency, making it a promising candidate for deployment in resource-constrained and edge-based cervical cancer screening systems. Nevertheless, further external validation and clinical evaluation are required prior to real-world implementation.*

Keywords: *Cervical Cancer, GhostNetV2, Image Classification, MobileNetV2, SIPaKMeD*

1. Introduction

Cervical cancer is a malignancy that develops in the cervix and is generally caused by persistent infection with high-risk Human Papillomavirus (HPV), which triggers abnormal changes in the cervical epithelial cells [1] [2]. Persistent HPV infection can induce molecular alterations in the transformation zone that progress into precancerous lesions (CIN/SIL) and may evolve into invasive carcinoma within 10 to 20 years if left untreated [1] [2]. Globally, cervical cancer ranks as the fourth leading cause of cancer death among women, with more than 660,000 new cases and approximately 350,000 deaths each year [3].

Cervical cancer is in fact preventable and can be detected earlier through cytological examinations such as the Pap smear or liquid-based cytology (LBC). However, the high mortality in developing countries is attributable to limited access to screening, causing many patients to present at advanced stages [4]. Although conventional methods such as the Pap smear and biopsy have effectively reduced incidence in developed countries, manual detection errors can still reach 10 to 30% owing to complex microscopic interpretation that is time-consuming and highly dependent on the subjective expertise of pathologists [5] [6]. The complexity of cervical cells demands an automated deep-learning–based system to improve the consistency and accuracy of classification.

The COVID-19 pandemic exacerbated the situation by causing a significant decline in screening activities, which led to an increase in advanced-stage cases in subsequent years [7]. In practice, shortages of pathologists create sample backlogs and diagnostic waiting times that can reach 2 to 4 weeks [8] [9]. Prolonged diagnostic intervals allow cancer to progress and reduce opportunities for early detection, so a fast and accurate automated system is required to accelerate the cell classification process.

Advances in deep learning, particularly convolutional neural networks (CNNs), have enhanced medical image analysis and shown effectiveness in identifying normal and abnormal cells in cervical cancer [10] [11]. The SIPaKMeD dataset provides thousands of single-cell images across five main categories and has become an important benchmark in cervical cell classification research [12] [13]. However, most studies still employ heavy architectures such as VGG-16, ResNet, and Inception, while investigations of lightweight architectures like MobileNetV2 and GhostNetV2 on the SIPaKMeD dataset remain very limited.

MobileNetV2 uses depthwise separable convolutions and inverted residuals to reduce parameters without sacrificing feature-extraction capability [14]. This model has been shown to deliver high performance with low inference time on medical images [15]. GhostNetV2, with its Ghost Module and DFC Attention, generates features more efficiently and captures broader spatial relationships at low computational cost [16]. Research indicates MobileNetV2 is effective in breast histopathology [17], whereas GhostNetV2 can outperform EfficientNet while consuming fewer resources [16]. Nevertheless, no study has directly compared the two on SIPaKMeD cervical cell classification, creating an important research gap.

This comparison is relevant because the two architectures employ different approaches to efficiency: MobileNetV2 optimizes convolutional operations, while GhostNetV2 exploits synthetic features. Evaluation on cervical cell images, which exhibit subtle morphological variations, can reveal the relative effectiveness of each architecture for medical classification tasks. In addition to accuracy, metrics such as precision, recall, and F1-score are important to ensure the model is not only accurate but also reliable in detecting abnormalities. Additional challenges, such as staining and illumination variations, necessitate the use of data augmentation to prevent overfitting [18].This study also contributes by evaluating the role of attention mechanisms in GhostNetV2 within the context of complex cytological images. Using the public SIPaKMeD dataset and transfer learning from ImageNet enables reproducibility and training efficiency. A comparative analysis of MobileNetV2 and GhostNetV2 is expected to identify the optimal architecture for rapid, efficient diagnosis and to serve as a reference for developing edge-computing–based medical AI systems to support mass screening for cervical cancer.

2. Literature Review

Numerous studies have investigated the use of deep learning to improve the accuracy and efficiency of cervical cancer classification from Pap smear images, with the SIPaKMeD dataset being one of the most widely used benchmarks. Prior works employing CNN-based methods, transfer learning, and ensemble strategies have reported high classification accuracy, typically ranging from approximately 94% to above 99%, depending on the network architecture, optimization strategy, and preprocessing techniques applied [19] [20] [21]. Several of these studies also highlight the importance of data augmentation, class-imbalance handling, and fine-tuning sometimes supported by GAN based approaches to enhance model generalization and robustness, particularly in medical imaging scenarios where data variability is high [19].

Beyond standard CNN architectures, other research has explored hybrid and multi-stage approaches that integrate deep learning with traditional machine-learning techniques or advanced image-processing methods prior to classification. Strategies such as multi-

architecture feature extraction, hybrid deep feature fusion, and the use of preprocessing techniques including PMD and CLAHE have been shown to improve classification performance without a proportional increase in model complexity. Additionally, feature-optimization methods such as Particle Swarm Optimization (PSO) and Ant Lion Optimization (ALO), as well as the incorporation of segmentation and augmentation pipelines, underscore the importance of achieving a balance between computational efficiency and performance stability in medical image–based diagnostic systems.

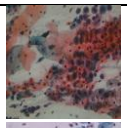
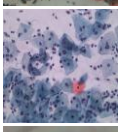
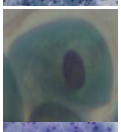
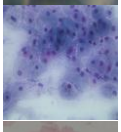
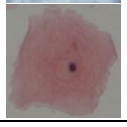
Despite these advances, many existing studies on the SIPaKMeD dataset continue to rely on large, computationally intensive architectures such as VGG-16, ResNet, and Inception [13]. while direct comparisons involving lightweight models remain limited. Recent developments in efficient architectures, including MobileNetV2 and GhostNet, demonstrate that lightweight networks leveraging depthwise separable convolutions, synthetic feature generation, and attention mechanisms can achieve competitive performance with significantly reduced computational cost. However, comparative evaluations of MobileNetV2 and GhostNetV2 on cervical cell images are still scarce, particularly in the context of subtle morphological variations inherent in Pap smear analysis, which motivates the present study to assess and contrast these two architectures on the SIPaKMeD dataset.

3. Methods

A. Data Description

The SIPaKMeD dataset used in this study was obtained from the Kaggle platform under the title "Cervical Cancer Largest Dataset (SIPaKMeD)", uploaded by Prahladmehandiratta. This dataset consists of two-dimensional images of cervical cancer cells and comprises five classes, namely Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial–Intermediate, as shown in Table 1. Each class in the dataset is divided into three subsets which is Training data, Validation data and Testing data.

Table 1. Sample of Dataset

Class Name Dataset	Dataset Sample	Count of Dataset
Dyskeratotic		813
Koilocytotic		825
Metaplastic		793
Parabasal		787
Superficial-Intermediate		831

Details regarding the number of samples allocated to each subset for every class are presented in Table 2.

Table 2. Detail of Dataset

Subset	Dataset Class	Proportion	Sample Count	Total Samples
Training	Dyskeratotic	80%	650	3239
	Koilocytotic	80%	660	
	Metaplastic	80%	634	
	Parabasal	80%	630	
	Superficial-Intermediate	80%	665	
Validation	Dyskeratotic	10%	81	404
	Koilocytotic	10%	82	
	Metaplastic	10%	79	
	Parabasal	10%	79	
	Superficial-Intermediate	10%	83	
Testing	Dyskeratotic	10%	82	406
	Koilocytotic	10%	83	
	Metaplastic	10%	80	
	Parabasal	10%	78	
	Superficial-Intermediate	10%	83	

B. Researh Methods and Evaluation

The system design scheme in this study has a similar structure for both architectures, namely MobileNetV2 and GhostNetV2. The main difference lies only in the fourth stage, which is the model training process that employs different CNN architectures. The comparison of these two architectures was chosen because each offers a distinct approach to model efficiency. MobileNetV2 focuses on reducing the number of convolutional operations, whereas GhostNetV2 emphasizes the generation of synthetic features to mimic the outputs of complex convolutions.

The first stage of the design scheme begins with the dataset preparation process, in which the researchers search for a dataset relevant to cervical cancer classification. After exploring datasets from internet sources, the researchers identified a suitable dataset, namely **"SIPaKMeD",** consisting of 4.049 cervical cancer cell images that had been manually cropped. The second stage involves preprocessing the research dataset, during which the cervical cancer cell images are cleaned by retaining only images in .bmp format. In addition, each cervical cancer cell image undergoes a resizing process to ensure consistency and compatibility with the CNN model input size for both MobileNetV2 and GhostNetV2 architectures, which is 224 × 224 pixels. The third stage is the dataset splitting process, in which the dataset is divided into three subsets training data, validation data, and testing data. The fourth stage is the CNN model training process using two different architectures, namely MobileNetV2 and GhostNetV2. Both models are trained separately by applying commonly used basic data augmentation techniques, such as rotation, flipping, cropping, and scaling [22], to enrich the diversity of the training data. The same hyperparameter settings are applied to both models, including the number of epochs, batch size, learning rate, and optimizer type. Details of the hyperparameter configurations used in this study are presented in Table 3.

Table 3. Hyperparameter Configuration

Hyperparamater	Configuration Values
Epoch	30 and 50
Batch Size	32 and 64
Learning Rate	0.0001 and 0.00001
Optimizer	AdamW

Before training, the cervical cell images are first converted into tensors using the *toTensor()* function so they can be processed by the *PyTorch* library. Next, normalization is performed using the mean and standard deviation values from the ImageNet dataset [0.485, 0.456, 0.406] for the mean and [0.229, 0.224, 0.225] for the standard deviation [23]. This normalization step aims to align the input data distribution with the model, accelerate model convergence, and prevent any single variable from dominating the training process [24]. The fifth stage is evaluation, in which the trained CNN models are used to classify the testing data that were previously split. To assess the model's performance, a testing/evaluation process is conducted using 406 test images (10% of the total dataset). This performance evaluation is carried out using a confusion matrix, which provides a comprehensive overview of the model's ability to classify each class. The evaluation metrics include accuracy, precision, recall, and F1-score, enabling a thorough objective, and measurable performance analysis. The formulas for accuracy, precision, recall, and F1-Score can be seen in equations 1, 2, 3, and 4.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (4)$$

Formula Description:
TP: True Positive; TN: True Negative; FN: False Negative; FP: False Positive.

The sixth and final stage of the design is selecting the best model for this study. The best model is chosen based on comparative performance between the two models. If one model with a particular architecture and hyperparameter configuration demonstrates superior performance, that model will be selected and saved as the best model.

4. Results and Discussion

In this study, Convolutional Neural Network (CNN) models with MobileNetV2 and GhostNetV2 architectures for cervical cancer image classification were developed on the Kaggle Notebook platform using the Python 3.11.13 programming language, with the support of a T4 × 2 GPU accelerator. The objective of this model training was to compare the performance of two CNN architectures which is MobileNetV2 and GhostNetV2 in classifying cervical cancer cell images from the SIPaKMeD dataset. Each model was trained using uniform hyperparameter configurations, including the number of epochs (30 and 50), batch size (32 and 64), learning rate (0.0001 and 0.00001), and optimizer (AdamW). After the classification models were fully trained, the training and validation results were presented in the form of graphical summaries and evaluation tables detailing key performance metrics, namely accuracy, precision, recall, and F1-score.

A. MobileNetV2

Technically, MobileNetV2 is defined by a sequence of inverted residual blocks that operate sequentially. First, a 1 × 1 expansion convolution expands the channel dimensions from a

narrow space to a wider space, next, a 3 × 3 depthwise convolution performs spatial processing independently on each channel; and finally, a 1 × 1 projection convolution reduces the channel dimensions back to the original bottleneck. When the spatial dimensions and the number of input and output channels are the same, a residual (shortcut) connection is applied [25].

In this section, the training process focuses on the MobileNetV2 model, resulting in eight different training scenarios. These eight scenarios arise from combinations of the hyperparameter variations used, namely the number of epochs (30 and 50), batch size (32 and 64), and learning rate (0.0001 and 0.00001). Each combination produces one training scenario, enabling a more comprehensive evaluation of model stability, generalization capability, and the performance of MobileNetV2 in classifying cervical cancer cell images. The results of these scenarios are presented in the form of graphs and evaluation tables to facilitate performance analysis, particularly in comparing evaluation metrics such as accuracy, precision, recall, and F1-score. Through this presentation, the performance of each hyperparameter combination can be observed more clearly and systematically.

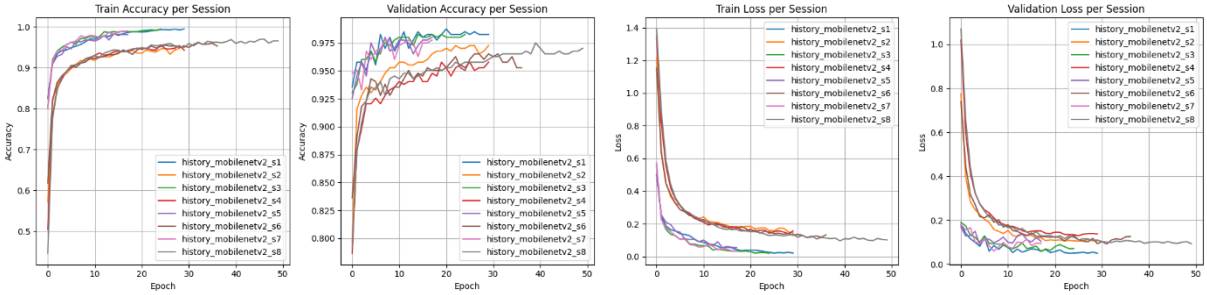


Figure 1. Training Accuracy, Validation Accuracy, Training Loss, and Validation Loss Graphs – MobileNetV2

The graphs in Figure 1 illustrate the performance progression of the MobileNetV2 model across the eight training scenarios distinguished by different hyperparameter combinations. In general, both training and validation accuracy increase as the number of epochs grows, with several scenarios reaching stability faster than others. The patterns observed in the training loss and validation loss graphs also show a consistent decrease, indicating that the learning process is effective. Although each scenario exhibits slightly different curve tendencies, the overall results in Figure 1 demonstrate that MobileNetV2 adapts well to hyperparameter variations and exhibits relatively stable performance across various training configurations.

Table 4. Evaluation Results of MobileNetV2 Architecture

Ep	BS	LR	Opt	Acc	Prec	Rec	F1
30	**32**	**0.0001**	**AdamW**	**98.50%**	**98.60%**	**98.50%**	**98.50%**
30	32	0.00001	AdamW	96.60%	96.80%	96.60%	96.60%
30	64	0.0001	AdamW	97.60%	97.60%	97.60%	97.60%
30	64	0.00001	AdamW	96.60%	96.70%	96.60%	96.60%
50	32	0.0001	AdamW	97.10%	97.10%	97.10%	97.10%
50	32	0.00001	AdamW	96.10%	96.20%	96.10%	96.10%
50	64	0.0001	AdamW	97.80%	97.80%	97.80%	97.80%
50	64	0.00001	AdamW	97.80%	97.80%	97.80%	97.80%

Description:
Ep: Epoch; BS: Batch Size; LR: Learning Rate; Opt: Optimizer; Acc: Accuracy; Prec: Precision; Rec: Recall; F1: F1-Score

The experimental results presented in Table 4 indicate that the best configuration was achieved in the first scenario, with a combination of 30 epochs, a batch size of 32, a learning rate of 0.0001, and the AdamW optimizer. This configuration produced the highest performance, with

an accuracy of 98.50%, precision of 98.60%, recall of 98.50%, and an F1-score of 98.50%. This hyperparameter combination provides an optimal balance between weight update stability, learning depth, and the model's generalization capability. A batch size of 32 yields stable gradient estimation, the learning rate of 0.0001 enables smooth and controlled learning, and AdamW plays a critical role in effective regularization, allowing the model to accurately recognize cytological features without overfitting. The consistency across all evaluation metrics also indicates that the model successfully minimizes errors in both false positives and false negatives, resulting in highly reliable classification performance.

These results demonstrate that MobileNetV2 with the specified configuration operates at an optimal point for cervical cell classification on the dataset used. The high performance across all four key metrics confirms that the model is not only accurate but also stable and efficient in distinguishing between cell classes with very subtle morphological differences. To further validate the quality of these predictions, the analysis was extended by presenting a confusion matrix derived from the testing results of the best-performing scenario (Scenario 1) using 10% of the testing data. This visualization provides a clearer depiction of the distribution of correct and incorrect predictions across each class. The confusion matrix is shown in Figure 2.

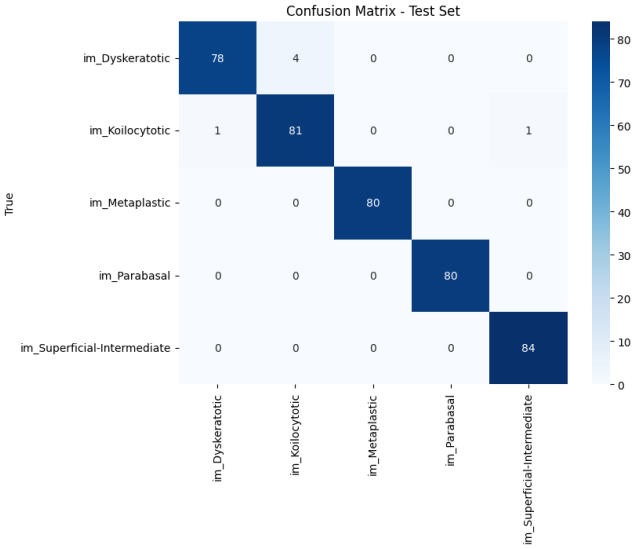


Figure 2. Confusion Matrix Test - MobileNetV2

Based on the confusion matrix in Figure 2, misclassifications are predominantly concentrated between the Dyskeratotic and Koilocytotic classes, as evidenced by a small number of reciprocal errors. This behavior is expected due to the overlapping nuclear characteristics shared by these two classes, which pose challenges for fine-grained visual discrimination. In contrast, the Metaplastic and Parabasal classes achieve perfect classification, with all samples correctly predicted and no confusion with other categories. The Superficial–Intermediate class also demonstrates very high classification performance, although it still receives a minimal misclassification from another class. These results indicate that MobileNetV2 effectively captures discriminative morphological features such as cell size, cytoplasmic thickness, and nucleus-to-cytoplasm ratio enabling clear separation for most classes. Overall, the confusion matrix confirms that classification errors are limited, biologically interpretable, and primarily occur between morphologically similar cell types, while the remaining classes are recognized with high robustness and consistency.

B. GhostNetV2

GhostNetV2 is an evolution of GhostNet designed to produce rich features at low computational cost through the Ghost Module [26], and it is further enhanced with DFC Attention to capture spatial relationships more effectively [16]. The plots in Figure 3 present the performance

development of GhostNetV2 across eight training scenarios, providing insight into learning stability, convergence patterns, and the model's sensitivity to different configurations.

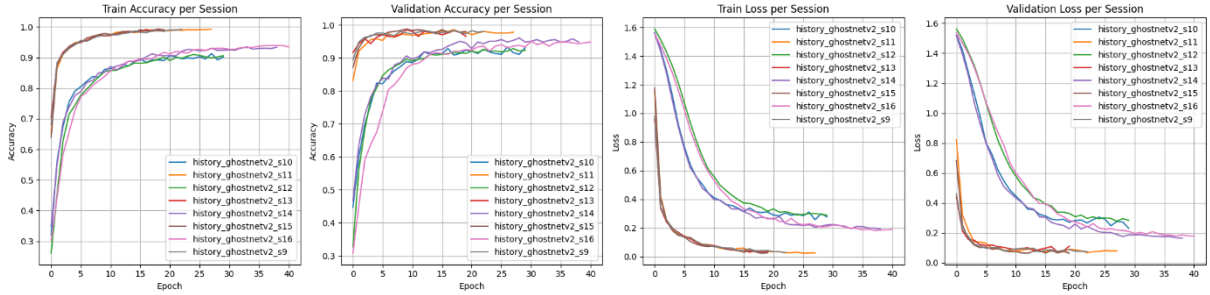


Figure 3. Training Accuracy, Validation Accuracy, Training Loss, and Validation Loss: GhostNetV2

The plots in Figure 3 show the performance progression of the GhostNetV2 model over eight training scenarios with varying hyperparameter combinations. Overall, both the training and validation accuracy curves exhibit a consistent upward trend as epochs increase, with some scenarios reaching convergence faster than others. This indicates that GhostNetV2 is capable of learning progressively and stably across all configurations. Meanwhile, the training-loss and validation-loss curves display steadily improving downward trends, suggesting that the optimization process is effective and that there are no significant error spikes during training.

Although there are minor differences in curve shapes between scenarios particularly in convergence rate and stability toward the final epochs the overall results indicate that GhostNetV2 adapts well to hyperparameter variation. The relatively stable and consistent patterns observed across all four plots confirm that the model can maintain strong performance under a range of training configurations, providing a solid basis for subsequent testing-stage evaluation. To further strengthen the understanding of the model's behavior during training, the analysis then turns to Table 5, which presents the final test results for the GhostNetV2 architecture across the various configuration scenarios.

Table 5. Evaluation Results of GhostNetV2 Architecture

Ep	BS	LR	Opt	Acc	Prec	Rec	F1
30	32	0.0001	AdamW	96.30%	96.60%	96.30%	96.40%
30	32	0.00001	AdamW	93.40%	93.50%	93.40%	93.40%
30	**64**	**0.0001**	**AdamW**	**97.60%**	**97.60%**	**97.60%**	**97.60%**
30	64	0.00001	AdamW	92.40%	92.60%	92.40%	92.40%
50	32	0.0001	AdamW	95.80%	96.10%	95.80%	95.90%
50	32	0.00001	AdamW	95.10%	95.20%	95.10%	95.10%
50	64	0.0001	AdamW	97.30%	97.50%	97.30%	97.40%
50	64	0.00001	AdamW	95.40%	95.40%	95.40%	95.40%

Description:
Ep: Epoch; BS: Batch Size; LR: Learning Rate; Opt: Optimizer; Acc: Accuracy; Prec: Precision; Rec: Recall; F1: F1-Score

The experimental results presented in Table 5 indicate that training using GhostNetV2 was conducted across eight training scenarios with different hyperparameter combinations. Among all configurations, Scenario 11 using 30 epochs, a batch size of 64, a learning rate of 0.0001, and the AdamW optimizer achieved the best performance with an accuracy of 97.60%, precision of 97.60%, recall of 97.60%, and an F1-score of 97.60%. The identical values across these four metrics indicate that the model is not only accurate in predicting the correct classes but also well balanced in minimizing both false positive and false negative errors. The use of a batch size of 64 enables more stable gradient estimation for a lightweight architecture such as GhostNetV2, while a learning rate of 0.0001 provides a smooth learning process without causing overshooting of the loss function minima. In this context, AdamW plays a key role in

maintaining effective regularization, allowing the model to optimally capture microscopic features without sacrificing generalization. These combined factors make Scenario 11 the most efficient and stable configuration for producing reliable predictions on cervical cytology data.

The high performance observed in these quantitative metrics requires further validation through evaluation of the prediction distribution at the class level. Therefore, the analysis was extended by examining the confusion matrix to gain a detailed understanding of how the model classifies each class, including the identification of areas prone to misclassification. The confusion matrix provides insight into whether the high accuracy is achieved uniformly across all classes or is influenced by dominance in predictions for certain classes. The confusion matrix shown in Figure 4 is derived from testing results using the best configuration, namely Scenario 11, with 10% of the total dataset allocated as test data. This visualization serves as an essential basis for confirming that the high performance recorded in accuracy, precision, recall, and F1-score [27] genuinely reflects the model's consistent ability to recognize each cell category.

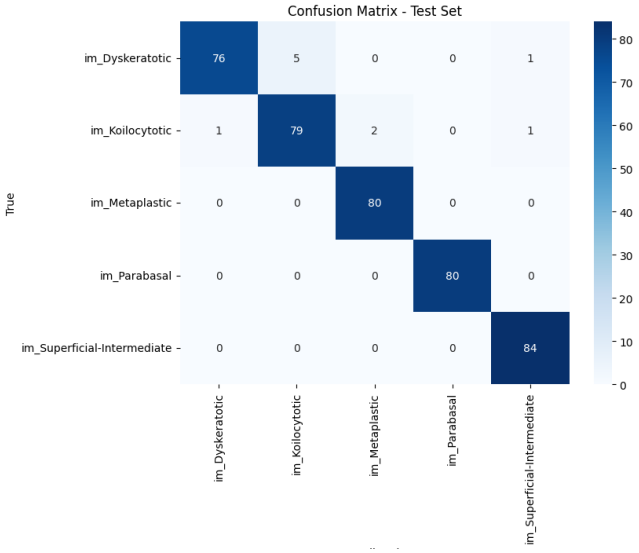


Figure 4. Confusion Matrix - GhostNetV2

Based on the confusion matrix in Figure 4, although GhostNetV2 achieves high overall accuracy, the confusion matrix reveals a higher level of confusion among morphologically similar abnormal classes, particularly between Dyskeratotic and Koilocytotic cells. Misclassifications predominantly occur as mutual predictions between these two classes, reflecting their overlapping nuclear irregularities and cytoplasmic features. The Parabasal class is the only category that achieves perfect classification, with all samples correctly predicted and no confusion with other classes. In contrast, the Metaplastic and Superficial–Intermediate classes demonstrate high but not perfect performance, as they still receive a small number of misclassified samples from other categories. This pattern indicates that while GhostNetV2 is computationally efficient, its synthetic feature generation may be less sensitive to subtle morphological variations required for precise discrimination between closely related abnormal cell types. Overall, the confusion matrix confirms that classification errors are systematic and

C. Discussion

Based on both training results, MobileNetV2 in Scenario 1 (30 epochs, batch size 32, learning rate 0.0001) outperforms the best GhostNetV2 result found in Scenario 11 (30 epochs, batch size 64, learning rate 0.0001). According to the final metrics, MobileNetV2 achieved Accuracy 98.50%, Precision 98.60%, Recall 98.50%, and F1-score 98.50%, whereas the best GhostNetV2 reached Accuracy 97.60% with Precision, Recall, and F1 each at 97.60%. The approximately 0.9 percentage-point difference in accuracy, together with the high consistency across metrics for MobileNetV2, indicates a practical advantage: this model not only produces correct

predictions more frequently but also balances false positives and false negatives effectively a critical property in medical screening where the costs of the two error types differ.

From an architectural perspective, MobileNetV2 demonstrates superior performance due to its inverted residual structure and depthwise separable convolutions, which preserve fine-grained spatial information while maintaining parameter efficiency. These characteristics are particularly advantageous for cervical cytology images, where subtle differences in nuclear shape, chromatin texture, and cytoplasm boundaries play a critical role in class discrimination. In contrast, GhostNetV2 relies on synthetic feature generation, which, while computationally efficient, may be less effective in capturing such subtle morphological cues. Compared to recent studies using heavier architectures or ensemble models on the SIPaKMeD dataset, which typically report accuracies between 94% and 99% [19] [20] [21], the proposed MobileNetV2-based approach achieves competitive performance while using a significantly lighter architecture. This indicates that high classification accuracy can be achieved without relying on computationally expensive models, making the proposed method more suitable for real-time and resource-constrained screening applications.

The implications of these findings for the problem outlined in the introduction are direct and significant. An automated model with high recall and precision can reduce the number of missed cases and lower false alarms, helping to address shortages of pathology experts, shorten diagnostic wait times, and increase mass-screening capacity particularly if the model can be deployed on low-power edge platforms. However, further work is required: external validation, per-class confusion-matrix analysis to detect class bias, robust testing against staining and illumination variations, and measurement of real-world latency and resource consumption. Thus, although MobileNetV2 in Scenario 1 is the best result according to these experimental metrics, subsequent steps should include external evaluation and prospective studies to ensure clinical safety and effectiveness.

5. Conclusions

The experimental results show that MobileNetV2 (Scenario 1 with Epoch 30, Batch Size 32, Learning Rate 0.0001, Optimizer AdamW) outperformed the best GhostNetV2 configuration (Scenario 11 with Epoch 30, Batch Size 64, Learning Rate 0.0001). MobileNetV2 achieved an accuracy of 98.50%, precision of 98.60%, recall of 98.50%, and an F1-score of 98.50%, whereas the best GhostNetV2 attained an accuracy of 97.60% with precision, recall, and F1-score each at 97.60%, representing an accuracy gap of approximately 0.9 percentage points. MobileNetV2's advantage is not only in the higher frequency of correct predictions but also in the balance between false positives and false negatives, which is critical in the context of medical screening. Technically, the combination of the number of epochs, a low learning rate, a moderate batch size, and the AdamW optimizer appears to support stable convergence and reliable extraction of morphological features. Nevertheless, before clinical deployment can be considered, external validation, per-class confusion-matrix analysis on the dataset to identify potential bias, robust testing against variations in staining and illumination, and measurement of latency and resource consumption under real-world conditions are required. Therefore, although MobileNetV2 produced the best results in this experiment, further studies including external evaluation and prospective trials are necessary to ensure safety, reliability, and readiness for clinical implementation.

References

[1] V. Novalia, "Kanker Serviks," *Galen. J. Kedokt. dan Kesehat. Mhs. Malikussaleh*, vol. 2, no. 1, pp. 45–56, 2023.

[2] H. H. Xu, W. H. Yan, and A. Lin, "The Role of HLA-G in Human Papillomavirus Infections and Cervical Carcinogenesis," *Front. Immunol.*, vol. 11, no. June, pp. 1–9, 2020, doi: 10.3389/fimmu.2020.01349.

[3] WHO, "Cervical cancer," *who*, 2024. https://www.who.int/news-room/fact-sheets/detail/cervical-cancer (accessed Nov. 01, 2025).

[4] O. Farajimakin, "Barriers to Cervical Cancer Screening: A Systematic Review," *Cureus*, vol. 16, no. 7, 2024, doi: 10.7759/cureus.65555.

[5] M. S. Pramono and A. P. Wibowo, "Penerapan Convolutional Neural Network Untuk Identifikasi Penyakit Pada Tanaman Padi Dari Citra Daun Menggunakan Model Resnet-101," *Djtechno J. Teknol. Inf.*, vol. 5, no. 3, pp. 415–430, 2024, doi: 10.46576/djtechno.v5i3.5098.

[6] J. D. Pallua, A. Brunner, B. Zelger, M. Schirmer, and J. Haybaeck, "The future of pathology is digital," *Pathol. Res. Pract.*, vol. 216, no. 9, p. 153040, 2020, doi: 10.1016/j.prp.2020.153040.

[7] T. R. Martins *dkk.*, "Impact of the COVID-19 Pandemic on Cervical Cancer Screening in São Paulo State, Brazil," *Acta Cytol.*, vol. 67, no. 4, pp. 388–394, 2023, doi: 10.1159/000529249.

[8] E. Walsh and N. M. Orsi, "The current troubled state of the global pathology workforce: a concise review," *Diagnostic Pathol.* , vol. 19, no. 1, 2024, doi: 10.1186/s13000-024-01590-2.

[9] K. W. Noh *dkk.*, "Effect of Waiting Time from Pathological Diagnosis to Definitive Concurrent Chemoradiation for Cervical Cancer on Overall Survival," *Cancer Res. Treat.*, vol. 54, no. 1, pp. 245–252, 2021, doi: 10.4143/CRT.2021.023.

[10] P. K. Mall *dkk.*, "A Comprehensive Review of Deep Neural Networks for Medical Image Processing: Recent Developments and Future Opportunities," *Healthc. Anal.*, vol. 4, no. April, p. 100216, 2023, doi: 10.1016/j.health.2023.100216.

[11] P. Jiang *dkk., A systematic review of deep learning-based cervical cytology screening: from cell identification to whole slide image analysis*, vol. 56, no. s2. Springer Netherlands, 2023. doi: 10.1007/s10462-023-10588-z.

[12] M. E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, and A. Charchanti, "Sipakmed: A New Dataset for Feature and Image Based Classification of Normal and Pathological Cervical Cells in Pap Smear Images," *Proc. - Int. Conf. Image Process. ICIP*, no. October, pp. 3144–3148, 2018, doi: 10.1109/ICIP.2018.8451588.

[13] H. Kaur, R. Sharma, and J. Kaur, "Comparison of Deep Transfer Learning Models For Classification of Cervical Cancer From Pap Smear Images," *Sci. Rep.*, vol. 15, no. 1, pp. 1–18, 2025, doi: 10.1038/s41598-024-74531-0.

[14] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A Review of Convolutional Neural Networks in Computer Vision," *Artif. Intell. Rev.*, vol. 57, no. 4, pp. 1–43, 2024, doi: 10.1007/s10462-024-10721-6.

[15] Nirupama and Virupakshappa, "MobileNet-V2: An Enhanced Skin Disease Classification by Attention and Multi-Scale Features," *J. Imaging Informatics Med.*, vol. 38, no. 3, pp. 1734–1754, 2025, doi: 10.1007/s10278-024-01271-y.

[16] Y. Tang, K. Han, J. Guo, C. Xu, C. Xu, and Y. Wang, "GhostNetV2: Enhance Cheap Operation with Long-Range Attention," *Adv. Neural Inf. Process. Syst.*, vol. 35, no. NeurIPS, pp. 1–12, 2022.

[17] M. Güler, G. Sart, Ö. Algorabi, A. N. Adıguzel Tuylu, and Y. S. Türkan, "Breast Cancer Classification with Various Optimized Deep Learning Methods," *Diagnostics*, vol. 15, no. 14, pp. 1–22, 2025, doi: 10.3390/diagnostics15141751.

[18] T. Islam, M. S. Hafiz, J. R. Jim, M. M. Kabir, and M. F. Mridha, "A Systematic Review of Deep Learning Data Augmentation in Medical Imaging: Recent Advances and Future Research Directions," *Healthc. Anal.*, vol. 5, no. December 2023, p. 100340, 2024, doi: 10.1016/j.health.2024.100340.

[19] R. Austin and R. Parvathi, "CNN Based Method For Classifying Cervical Cancer Cells In Pap Smear Images," *Sci. Rep.*, vol. 15, no. 1, pp. 1–21, 2025, doi: 10.1038/s41598-025-10009-x.

[20] G. Shandilya *dkk.*, "Enhancing Advanced Cervical Cell Categorization With Cluster-Based Intelligent Systems By A Novel Integrated CNN Approach With Skip Mechanisms And GAN-Based Augmentation," *Sci. Rep.*, vol. 14, no. 1, pp. 1–18, 2024, doi: 10.1038/s41598-024-80260-

1.

[21] J. Gangrade, R. Kuthiala, S. Gangrade, Y. P. Singh, M. R, and S. Solanki, "A Deep Ensemble Learning Approach For Squamous Cell Classification In Cervical Cancer," *Sci. Rep.*, vol. 15, no. 1, pp. 1–15, 2025, doi: 10.1038/s41598-025-91786-3.

[22] A. Mumuni and F. Mumuni, "Data Augmentation: A Comprehensive Survey of Modern Approaches," *Array*, vol. 16, no. August, p. 100258, 2022, doi: 10.1016/j.array.2022.100258.

[23] D. Maulana, N. S. Azzahrah, R. A. Syach, and S. Syauri, "Identifikasi Konten Visual Buatan AI Dengan Resnet dan Fine-Grained Feature Extraction," *JITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 13, no. 3, 2025, doi: http://dx.doi.org/10.23960/jitet.v13i3S1.8056.

[24] E. C. Wibowo and A. D. Cahyono, "Analisis Perbandingan Algoritma Regresi Linier dengan Neural Network untuk Prediksi Harga Saham," *Sist. J. Sist. Inf.*, vol. 14, no. 4, pp. 2540–9719, 2025, [Online]. Available: http://sistemasi.ftik.unisi.ac.id

[25] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and Liang-Chieh Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks Mark," *Convolutional Neural Networks with Swift Tensorflow*, pp. 99–107, 2019.

[26] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More Features From Cheap Operations," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1577–1586, 2020, doi: 10.1109/CVPR42600.2020.00165.

[27] Tedyyana, A., Ghazali, O., & W. Purbo, O. (2024). Enhancing intrusion detection system using rectified linear unit function in pigeon inspired optimization algorithm. IAES International Journal of Artificial Intelligence (IJ-AI), 13(2), 1526-1534. doi:http://doi.org/10.11591/ijai.v13.i2.pp1526-1534