Comparison of CNN Architectures for Pre-Cancerous Cervical Lesion Classification Based on Colposcopy Images Using IARC and AnnoCerv Datasets

Sigit Prasetyo Noprianto[1]*, Siti Nurmaini[2], Dian Palupi Rini[3]
Department of Masters in Computer Science[1], [2], [3]
University of Sriwijaya
Palembang, Indonesia
noprisigit@gmail.com[1], sitinurmaini@gmail.com[2], dprini@unsri.ac.id [3]

***Abstract*— Cervical cancer represents a significant public health issue affecting women worldwide, and identifying the severity of lesions early on is crucial to selecting the right treatment. This research investigates and compares the effectiveness of various Convolutional Neural Network (CNN) models in classifying colposcopic images according to the severity of cervical lesions. The dataset used was obtained from the International Agency for Research on Cancer (IARC) and AnnoCerv, consisting of 452 colposcopy images categorized into four classes: Normal, CIN 1, CIN 2, and CIN 3. Five CNN architectures were evaluated: MobileNetV2, InceptionV3, Xception, VGG16, and DenseNet121. Experiments were conducted using default hyperparameters: batch size of 32, learning rate of 0.001, and 100 epochs. The results showed that MobileNetV2 achieved the highest accuracy at 67%, followed by DenseNet121 (60%), Xception (60%), InceptionV3 (55%), and VGG16 (42%). Based on these findings, MobileNetV2 is the most optimal model for classifying colposcopy images in this study. However, the study is limited by class imbalance and dataset size, which may affect model generalizability. Future work may explore ensemble learning techniques and larger, more diverse datasets for improved accuracy.**

Keywords*— *Cervical Cancer, Colposcopy Image Classification, CNN, CIN Classification

I. INTRODUCTION

Artificial Intelligence (AI) has rapidly advanced in recent decades with widespread applications across various domains, including healthcare. One of the most prominent branches of AI is deep learning, which has demonstrated outstanding performance in medical image analysis [1]. Deep learning models, particularly Convolutional Neural Networks (CNN), have been successfully applied to various disease classification tasks, such as brain tumors, breast cancer, and skin cancer, achieving high levels of accuracy [2][3][4].

Cervical cancer ranks as the fourth most common cancer affecting women and is one of the most feared diseases among them. WHO statistics reveal that globally in 2020, cervical cancer led to around 604,000 new cases and approximately 341,831 fatalities [5]. About 95% of these cases are caused by infection with the Human Papillomavirus (HPV) [6] [7]. WHO recommends several screening methods, including cytology tests (Pap smear), HPV tests, visual inspection with acetic acid (VIA), and colposcopy [8]. Among the available methods, VIA is regarded as one of the most cost-effective and can be applied in resource-limited healthcare settings. The procedure involves applying acetic acid to the cervix, followed by direct visual inspection to detect acetowhite lesions. The World Health Organization (WHO) classifies abnormal cervical lesions found through VIA or alternative screening tools as Cervical Intraepithelial Neoplasia (CIN), which is divided into three grades: CIN 1, CIN 2, and CIN 3.

In clinical practice, colposcopy is often performed after a VIA test to further examine suspicious areas of the cervix. Colposcopy utilizes an optical magnifying instrument (colposcope) to produce clearer and more detailed cervical images. These images reveal specific visual patterns such as mosaicism, punctation, and lesion borders that help determine CIN severity levels. These visual features serve as critical indicators for developing deep learning-based classification models to detect and differentiate CIN levels more accurately [9].

Previous studies on pre-cancerous cervical lesion classification have mostly focused on the Pap smear method [10][11][12], and predominantly employed binary classification approaches—distinguishing only between normal and abnormal images. Although this approach aids in initial screening, binary classification does not provide specific information on lesion severity, which is essential for determining proper medical follow-up. Some studies have attempted to classify cervical lesions from colposcopy images into categories such as normal, CIN1, CIN2, CIN3, and cancer. For instance, a study using the ResNet-152 model achieved an average accuracy of 51.7% for multi-class CIN classification, with an Area Under the Curve (AUC) of 0.781 to distinguish high-risk from low-risk lesions [13]. Another study developed an ensemble deep learning model named CYENET for classifying cervical cancer from colposcopy images, which improved classification accuracy to 92.3%, compared to 73.3% achieved by the VGG19 model [14]. In a different study, a CAD system was developed by integrating deep learning descriptors

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2361, Copyright ©2025
Submitted : May 6, 2025, Revised : May 13, 2025, Accepted : May 14, 2025, Published : May 26, 2025

such as ResNet50, ResNet101, and ResNet152 with dimensionality reduction techniques for colposcopy image classification. This approach achieved outstanding performance, ranging from 97%–100% in normal-abnormal classification and lesion type identification [15].

While several studies have attempted colposcopy-based classification, the often use limited datasets, focus on binary classification, or apply only one CNN model, making it difficult to assess comparative model performance. In addition, limited publicly available colposcopy datasets, such as IARC and AnnoCerv, pose challenges related to data imbalance dan image variability.

Despite these promising results, challenges such as class imbalance, varying image quality, and differences in colposcopy equipment across institutions remain. Therefore, this study aims to evaluate and compare several popular CNN architectures—MobileNetV2, InceptionV3, Xception, VGG16, and DenseNet121—for CIN lesion classification based on colposcopy images. MobileNetV2 is known for its high efficiency [16], InceptionV3 is effective in capturing complex visual features from medical images [17], Xception improves upon Inception using separable convolutions for higher accuracy in image classification tasks [18] , VGG16, despite its large number of parameters, remains a strong baseline due to its simple yet effective convolutional structure [19], and DenseNet121 allows for efficient feature and gradient propagation through dense layer connections, showing strong performance in image-based disease classification [20]. This evaluation is expected to provide insights into model performance and its potential applications in clinical settings, particularly for early detection of cervical cancer.

This study contributes to the field by evaluating the classification effectiveness of five different CNN models for multi-class classification of cervical lesion severity using colposcopy images. To the best our knowledge, no existing studies have performed a direct evaluation and comparison of these five CNN architectures using both IARC and AnnoCerv datasets for four-class cervical lesion classification from colposcopy images. This positions our work as a novel contribution that can guide future model selection and deployment in clinical decision support systems.

II. RESEARCH METHODOLOGY

This research began with problem identification, followed by a literature review to build upon previous related studies. The next step involved collecting data aligned with analytical requirements. The research methodology includes data preprocessing, architectural design, model training and evaluation, and finally, documentation and reporting. The research stages are described as follows.

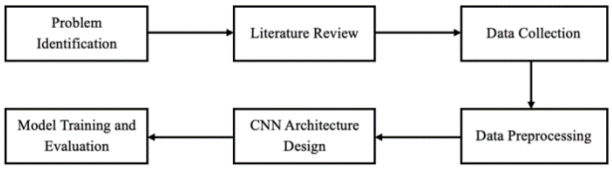


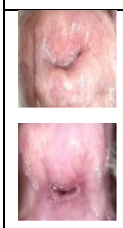
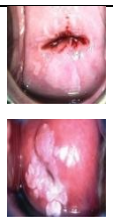
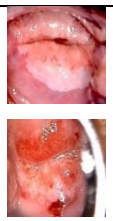
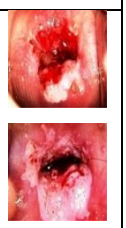
Fig. 1. Research flow

A. Problem Identification

The initial stage of this study involved identifying the specific research problem. The focus of this study is the classification of pre-cancerous cervical lesions resulting from visual inspection with acetic acid (VIA) into four categories: Normal, CIN 1, CIN 2, and CIN 3. This research aims to solve the problem by implementing a system based on different CNN models, which are then compared to find the one that performs best in detecting pre-cancerous cervical lesions.

As illustrated in Table 1, the dataset used in this study comprises colposcopic images categorized into four classes based on lesion severity: “Normal,” “CIN 1” (mild dysplasia), “CIN 2” (moderate dysplasia), and “CIN 3” (severe dysplasia). The “Class” row defines the lesion grade, while the “Image” row displays representative samples of each category, highlighting distinct visual characteristics such as changes in epithelial texture, color, and vascular patterns. These annotated samples serve as input for training and evaluating the proposed deep learning models aimed at automated classification of cervical lesion severity.

TABLE I. CERVICAL PRE-CANCEROUS LESIONS COLPOSCOPY IMAGE DATASET

Normal	CIN 1	CIN 2	CIN 3
			

B. Data Collection Method

This stage includes the process of gathering cervical lesion image data to build the dataset used in this classification research. The dataset was obtained from two primary sources: the International Agency for Research on Cancer (IARC) and AnnoCerv [21]. The IARC dataset comprises 913 colposcopy images from 200 case examinations, while the AnnoCerv dataset contains 527 images from 100 cases. The image formats are .jpg for IARC and a combination of .jpg and .png for AnnoCerv; however, only the .jpg format was used.

Colposcopic images in the IARC dataset are collected for each case after applying a sequence of diagnostic fluids—namely, normal saline, vinegar acid solution with and without a green filter, and an iodine-based solution. Meanwhile, the AnnoCerv dataset provides images after acetic acid and Lugol’s iodine applications. These datasets were used for both training and testing in developing a machine learning-based image classification model to detect and identify pre-cancerous

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2361, Copyright ©2025
Submitted : May 6, 2025, Revised : May 13, 2025, Accepted : May 14, 2025, Published : May 26, 2025

cervical lesion severity.

C. Pre-Processing

The preprocessing stage was conducted to prepare the data for training and evaluating the cervical lesion image classification models. The datasets from IARC and AnnoCerv were first grouped into four classes based on lesion severity: Normal, CIN 1, CIN 2, and CIN 3. Label assignment referred to the accompanying medical diagnostic information for each image. Only acetic acid-applied images were used in this study, as they are medically considered the most relevant for identifying pathological changes in cervical tissue. The grouping resulted in the following distribution: 89 images for Normal, 148 for CIN 1, 105 for CIN 2, and 110 for CIN 3, as detailed in Table 2.

TABLE II. DISTRIBUTION OF IARC AND ANNOCERV DATASET PER CLASS

Dataset	Normal	CIN 1	CIN 2	CIN 3
IARC	35	22	36	59
Annocerv	54	126	69	51
Total	89	148	105	110

Each image was resized to 224×224 pixels to ensure compatibility with the expected input format of the CNN frameworks. The original resolution of IARC images was 800×600 pixels, while AnnoCerv images were 2976×1984 pixels.

To address class imbalance, image augmentation was applied. Augmentation strategies involved applying rotations, horizontal and vertical displacements, horizontal mirroring, and brightness adjustments to the images. The applied augmentations contributed to balancing class distributions, enhancing dataset variability, and mitigating overfitting. After augmentation, each class contained 148 images, resulting in a final dataset of 592 images. Following data preprocessing, a split was performed where 70% of the dataset was utilized for training and 30% for validation.

D. Model Architecture

This study conducted training and evaluation for several CNN model architectures, including MobileNetV2, InceptionV3, Xception, VGG16, and DenseNet121. Using a 70:30 training validation split, 412 images were used for training and 180 for validation.

E. Evaluation of Architectural Design Result

A confusion matrix was employed in the evaluation process of the trained models, offering performance indicators including accuracy, precision, recall, and the F1-score to judge model effectiveness. The formulas used in this evaluation process are as follows:

$$Accuracy = \frac{\sum_{i=1}^{N} TP_i}{Total\ Data} \quad (1)$$

$$Precision = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i} \quad (2)$$

$$Recall = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N}(TP_i + FN_i)} \quad (3)$$

$$F1\ Score = 2\ x\ \frac{Precision\ x\ Recall}{Precision + Recall} \quad (4)$$

III. RESULTS AND DISCUSSION

This research evaluates the classification accuracy of several CNN models in identifying pre-cancerous cervical lesions, employing MobileNetV2, InceptionV3, Xception, VGG16, and DenseNet121 in the experimentation process. All models were trained under the same settings, using a batch size of 32, a learning rate of 0.001, and 100 epochs. A total of 412 data points were used for training, while 180 were allocated for validation. The following are the results obtained from each model:

A. MobileNetV2

The classification outcomes of the MobileNetV2 model, represented as a confusion matrix, are depicted in Figure 2, which provides detailed insight into the classification performance across the four classes: Normal, CIN1, CIN2, and CIN3.

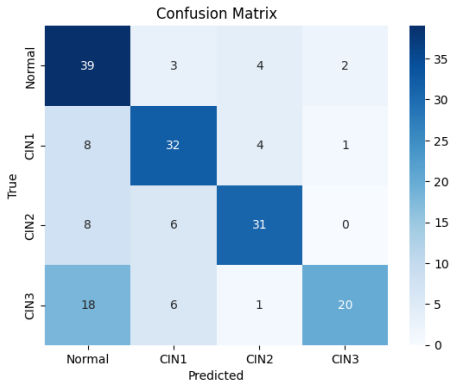


Fig. 2. Confusion Matrix of MobileNetV2

TABLE III. MOBILENETV2 MODEL EVALUATION

	Precision	Recall	F1-Score
Normal	53%	81%	64%
CIN 1	68%	71%	70%
CIN 2	78%	69%	73%
CIN 3	87%	44%	59%

The model demonstrates strong performance in detecting Normal cases, correctly classifying 39 out of 48 samples, which corresponds with the high recall value of 81% reported in Table 3. However, the relatively low precision of 53% for this class indicates a high number of false positives, as seen in the misclassification of several CIN 1, CIN 2, and CIN 3 cases as

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2361, Copyright ©2025
Submitted : May 6, 2025, Revised : May 13, 2025, Accepted : May 14, 2025, Published : May 26, 2025

Normal. A total of 32 samples from the CIN 1 class were correctly identified by the model, out of 45 instances, reflecting a balanced performance with a recall of 71% and precision of 68%. Similarly, the CIN 2 class shows reliable classification results with 31 correct predictions out of 45, consistent with its 69% recall and 78% precision. In contrast, the CIN 3 class reveals a notable weakness of the model: although it achieves a high precision of 87%, indicating that most predictions labeled as CIN 3 are accurate, it only correctly classifies 20 out of 45 actual CIN 3 cases. This results in a low recall of 44%, suggesting that many true CIN 3 cases are not being detected. The overall pattern observed in the confusion matrix confirms the results summarized in Table 3 and highlights the model's strength in precision for higher-grade lesions (e.g., CIN 3) while revealing its limited sensitivity in detecting all true cases within that class.

B. InceptionV3

The classification outcomes of the InceptionV3 model, represented as a confusion matrix, are depicted in Figure 3, offering a detailed breakdown of the classification outcomes across the four diagnostic categories: Normal, CIN 1, CIN 2, and CIN 3. The model correctly identified 27 Normal cases, out of what appears to be 48 actual Normal samples, resulting in a recall of approximately 56%, which matches the value reported in Table 4.

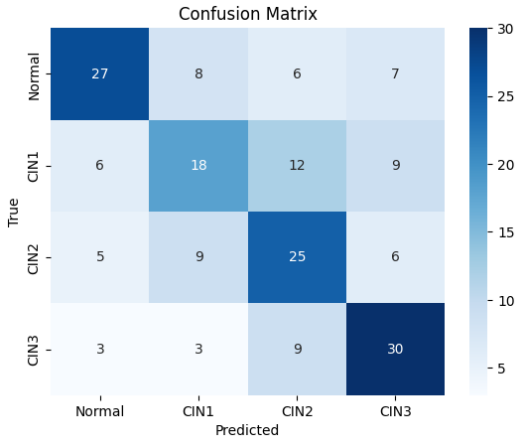


Fig. 3. Confusion Matrix of Inception V3 Model

TABLE IV. INCEPTIONV3 MODEL EVALUATION

	Precision	Recall	F1-Score
Normal	66%	56%	61%
CIN 1	47%	40%	43%
CIN 2	48%	56%	52%
CIN 3	58%	67%	62%

The precision for the Normal class was 66%, suggesting that when the model predicted a case as Normal, it was correct most of the time. For the CIN 1 class, only 18 cases were correctly classified out of an estimated 45, yielding a recall of about 40%. This class also had a low precision of 47%, which is consistent with its significant overlap with other categories, particularly CIN 2, as reflected in the confusion matrix. The CIN 2 class showed a relatively balanced performance, with 25 out of 45 samples correctly classified, leading to a recall of 56% and a precision of 48%, aligning with the reported metrics. In contrast, the CIN 3 class exhibited the highest recall at 67%, as the model correctly classified 30 of the 45 CIN 3 samples. However, the precision for CIN3 was lower (58%), indicating that a number of other class instances (e.g., CIN 2) were misclassified as CIN 3. This pattern suggests that the model was more sensitive in detecting CIN 3 but lacked specificity. Overall, the confusion matrix confirms the findings in Table 4 and supports the observation that InceptionV3, while able to capture a broader range of CIN 3 cases, struggled with class separability—particularly between CIN 1 and CIN 2—likely due to the subtle visual differences among these lesion grades.

C. Xception

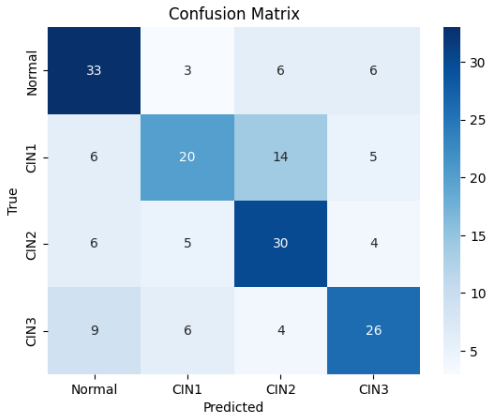


Fig. 4. Confusion Matrix of Xception

TABLE V. XCEPTIOON MODEL EVALUATION

	Precision	Recall	F1-Score
Normal	61%	69%	65%
CIN 1	59%	44%	51%
CIN 2	56%	67%	61%
CIN 3	63%	58%	60%

This class also achieved a precision of 61%, indicating a fair trade-off between correctly identifying Normal cases and minimizing false positives. For the CIN 1 class, only 20 out of an estimated 45 instances were correctly classified, leading to the lowest recall among the classes (44%) and a precision of 59%. This supports the observation that CIN 1 remains the most difficult class to distinguish, likely due to its visual similarity to both Normal and CIN 2 cases, as also seen in the confusion with 14 CIN 1 samples misclassified as CIN 2. The CIN 2 class showed good classification ability, with 30 correct predictions and a recall of 67%, while achieving a precision of 56%. For the CIN 3 class, the model correctly predicted 26 samples with a recall of 58% and a precision of 63%, reflecting a moderate but consistent detection performance. Overall, the confusion matrix supports the evaluation metrics in Table 5, confirming that Xception maintains a more even balance between sensitivity and specificity compared to previous models. While its performance does not dominate in any single class, it offers

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2361, Copyright ©2025
Submitted : May 6, 2025, Revised : May 13, 2025, Accepted : May 14, 2025, Published : May 26, 2025

stable results across all categories, making it a promising candidate for general-purpose classification in this task.

D. VGG16

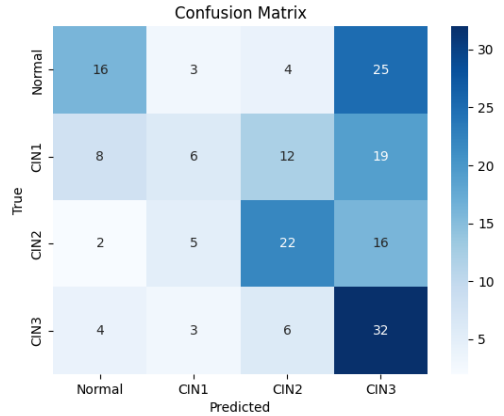


Fig. 5. Confusion Matrix of VGG16

TABLE VI. VGG16 MODEL EVALUATION

	Precision	Recall	F1-Score
Normal	53%	33%	41%
CIN 1	35%	13%	19%
CIN 2	50%	49%	49%
CIN 3	35%	71%	47%

The classification outcomes of the VGG16 model, represented as a confusion matrix, are depicted in Figure 5, which reveals the weakest overall performance among the tested models, aligning with its lowest reported accuracy of 42%. The Normal class shows a poor classification outcome, with only 16 correctly predicted samples out of an estimated 48, resulting in a recall of approximately 33% and a precision of 53%, as stated in Table 6. This indicates a high number of false negatives, with many Normal cases misclassified as CIN 3. The CIN 1 class performed the worst overall, with only 6 correct predictions and a recall of just 13%, reflecting the model's significant struggle to identify CIN 1 instances. Its precision of 35% further emphasizes the high degree of confusion with other classes, particularly CIN 2 and CIN 3, as evident in the matrix. CIN 2 showed slightly better results with 22 correct predictions, yielding a recall of 49% and a matching precision of 50%, demonstrating somewhat balanced but modest performance. Notably, CIN 3 achieved the highest recall at 71%, with 32 true positives. However, its precision was just 35%, suggesting that while the model is sensitive to CIN 3 cases, it frequently misclassifies samples from other classes—especially Normal and CIN 1—as CIN 3. This imbalance indicates that the model heavily overpredicts CIN 3, potentially due to its limited capacity to distinguish between lower-grade lesions. Overall, the confusion matrix confirms that VGG16 suffers from overgeneralization and class confusion, especially in distinguishing Normal and CIN 1 cases, possibly due to the architecture's rigidity and lack of deeper adaptive feature extraction, as suggested in Table 6.

E. DenseNet121

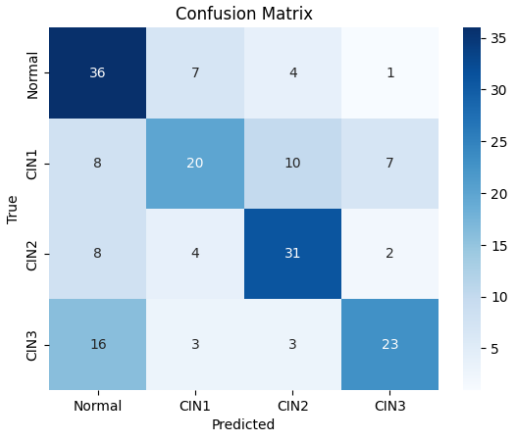


Fig. 6. Confusion Matrix of DenseNet21

TABLE VII. DENSENET21 MODEL EVALUATION

	Precision	Recall	F1-Score
Normal	53%	75%	62%
CIN 1	59%	44%	51%
CIN 2	65%	69%	67%
CIN 3	70%	51%	59%

Figure 6 shows the confusion matrix for the DenseNet121 model, which provides insight into the classification behavior of its densely connected architecture. The model achieved a strong performance in identifying Normal cases, correctly predicting 36 out of approximately 48 samples, corresponding to a recall of 75% and a precision of 53%, as reported in Table 7. This indicates a high sensitivity toward Normal class detection, although some misclassifications still occurred, particularly into CIN 1. The CIN 1 class had a lower recall of 44% with 20 correct predictions, and a precision of 59%, suggesting continued challenges in differentiating this class from neighboring grades like CIN 2 and CIN 3. CIN 2 was the best-performing class, with 31 out of around 45 instances correctly identified, yielding a recall of 69% and a precision of 65%. This strong performance supports the claim that DenseNet121 is well-suited for mid-stage lesion classification, likely due to its ability to retain and combine multi-level features effectively. For CIN 3, the model correctly predicted 23 samples, achieving a recall of 51% and the highest precision among the classes at 70%. While this indicates that most CIN 3 predictions were accurate, the relatively low recall means that a notable portion of CIN 3 cases were missed, often misclassified as Normal or CIN 2. Overall, the confusion matrix validates the evaluation results in Table 7, showing that DenseNet121 offers well-rounded performance, particularly excelling in CIN 2 classification. However, further optimization—such as enhanced regularization—may be required to boost its

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2361, Copyright ©2025
Submitted : May 6, 2025, Revised : May 13, 2025, Accepted : May 14, 2025, Published : May 26, 2025

generalization in more challenging class boundaries.

F. Model Comparison Summary

TABLE VIII. EVALUATION METRIC COMPARISON

Model	Accuracy	Average Precision	Average Recall	Average F1-Score
MobileNetV2	67%	71,5%	66,25%	66,5%
InceptionV3	55%	54,75%	54,75%	54,%
Xception	60%	59,75%	59,5	59,25
VGG16	42%	43,25%	41,5%	39%
DenseNet121	60%	61,75%	59,75	59,75%

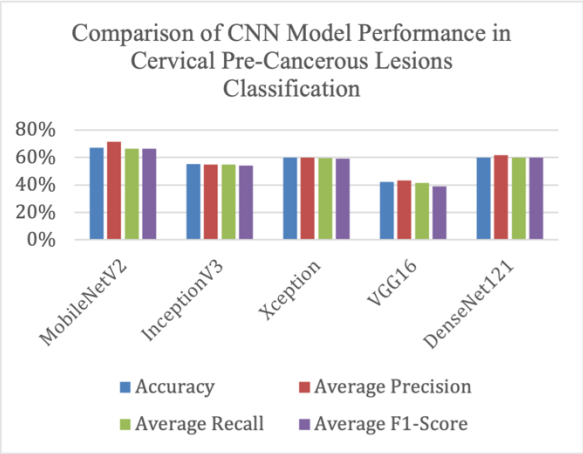


Fig. 7. CNN Model Performance Compqrison

Based on Table 8 and Figure 1, MobileNetV2 outperformed all other models in this study with the highest accuracy of 67%, while VGG16 had the lowest performance with an accuracy of 42%. MobileNetV2 excelled across all metrics, making it the best overall choice. DenseNet121 demonstrated comparable performance, particularly in precision and F1-score. Xception offered a well-balanced result across classes, while InceptionV3 showed inconsistent metrics, especially in F1-score. VGG16 was the least suitable model for this classification task.

G. Discussion and Limitations

While MobileNetV2 yielded the highest performance among tested CNN architectures, the overall accuracy across models (42% to 67%) highlights the difficulty of the classification task. This can be attributed to the relatively small dataset size and class imbalance, despite augmentation efforts. Models particularly struggled with distinguishing CIN 1, which may visually ensemble both Normal and CIN 2 cases. Futuru work should explore data balancing strategies, larger and more diverse datasets, and potentially ensemble-based models to enhance classification performance.

IV. CONCLUSIONS AND SUGGESTIONS

Based on the research findings and analysis, several conclusions can be drawn. The development of a classification model for pre-cancerous cervical lesions based on colposcopy images into four classes—Normal, CIN 1, CIN 2, and CIN 3—involved several stages, including problem identification, data collection and class labeling, image preprocessing, model architecture selection (MobileNetV2, InceptionV3, Xception, VGG16, and DenseNet121), and model evaluation.

Using identical training parameters—batch size of 32, learning rate of 0.001, and 100 epochs—the best accuracy achieved by each architecture was as follows: MobileNetV2: 67% InceptionV3: 55% Xception: 60% VGG16: 42% DenseNet121: 60%.

From these results, MobileNetV2 achieved the highest accuracy of 67%, indicating that it performed better than the other proposed architectures in classifying pre-cancerous cervical lesions based on colposcopy images across the four classes (Normal, CIN 1, CIN 2, and CIN 3). Meanwhile, VGG16 recorded the lowest accuracy at 42%, making it the least suitable architecture among those evaluated in this study.

Overall, the obtained accuracy levels indicate that the classification performance still leaves room for improvement. Therefore, future research should consider several enhancements: Increasing the dataset size and diversity, to ensure broader generalization and better applicability in real-world clinical settings. Optimizing training configurations to improve model performance. Applying more advanced data augmentation techniques or transfer learning strategies to boost model robustness. Exploring ensemble methods or attention-based CNNs to enhance classification accuracy and generalizability. It is expected that these strategies will result in models that are both dependable and suitable for clinical use in the early identification of cervical cancer using colposcopy images. Despite the promising result, further improvements in data quantity and model optimization are essential before clinical implementation can be considered.

REFERENCES

[1] A. Chaddad, J. Peng, J. Xu, and A. Bouridane, “Survey of Explainable AI Techniques in Healthcare,” Jan. 01, 2023, *MDPI*. doi: 10.3390/s23020634.

[2] N. A. Zebari *et al.*, “A deep learning fusion model for accurate classification of brain tumours in Magnetic Resonance images,” *CAAI Trans Intell Technol*, vol. 9, no. 4, pp. 790–804, Aug. 2024, doi: 10.1049/cit2.12276.

[3] E. O. Simonyan, Joke. A. Badejo, and J. S. Weijin, “Histopathological breast cancer classification using CNN,” *Mater Today Proc*, vol. 105, pp. 268–275, 2024, doi: https://doi.org/10.1016/j.matpr.2023.10.154.

[4] A. Faghihi, M. Fathollahi, and R. Rajabi, “Diagnosis of skin cancer using VGG16 and VGG19 based transfer learning models,” *Multimed Tools Appl*, vol. 83, no. 19, pp. 57495–57510, 2024, doi: 10.1007/s11042-023-17735-2.

[5] H. Sung *et al.*, “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA Cancer J Clin*, vol. 71, no. 3, pp. 209–249, May 2021, doi: https://doi.org/10.3322/caac.21660.

[6] P. A. Cohen, A. Jhingran, A. Oaknin, and L. Denny, “Cervical Cancer,” *The Lancet*, vol. 393, no. 10167, pp. 169–182, Jan. 2019, doi: 10.1016/S0140-6736(18)32470-X.

[7] J. M. M. Walboomers *et al.*, “Human papillomavirus is a necessary cause

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2361, Copyright ©2025
Submitted : May 6, 2025, Revised : May 13, 2025, Accepted : May 14, 2025, Published : May 26, 2025

of invasive cervical cancer worldwide," *J Pathol*, vol. 189, no. 1, pp. 12–19, Sep. 1999, doi: https://doi.org/10.1002/(SICI)1096-9896(199909)189:1<12::AID-PATH431>3.0.CO;2-F.

[8] J. Liu, L. Li, and L. Wang, "Acetowhite Region Segmentation in Uterine Cervix Images Using a Registered Ratio Image," *Comput Biol Med*, vol. 93, pp. 47–55, 2018, doi: 10.1016/j.compbiomed.2017.12.009.

[9] E. Hussain, L. B. Mahanta, K. A. Borbora, H. Borah, and S. S. Choudhury, "Exploring explainable artificial intelligence techniques for evaluating cervical intraepithelial neoplasia (CIN) diagnosis using colposcopy images," *Expert Syst Appl*, vol. 249, p. 123579, 2024, doi: https://doi.org/10.1016/j.eswa.2024.123579.

[10] N. A. Alias *et al.*, "Pap Smear Images Classification Using Machine Learning: A Literature Matrix," Dec. 01, 2022, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/diagnostics12122900.

[11] P. B. Shanthi, K. S. Hareesha, and R. Kudva, "Automated Detection and Classification of Cervical Cancer Using Pap Smear Microscopic Images: A Comprehensive Review and Future Perspectives," 2022, *Engineered Science Publisher*. doi: 10.30919/es8d633.

[12] O. Yaman and T. Tuncer, "Exemplar pyramid deep feature extraction based cervical cancer image classification model using pap-smear images," *Biomed Signal Process Control*, vol. 73, p. 103428, 2022, doi: https://doi.org/10.1016/j.bspc.2021.103428.

[13] B. J. Cho *et al.*, "Classification of Cervical neoplasms on Colposcopic Photography using Deep Learning," *Sci Rep*, vol. 10, no. 1, pp. 1–10, Aug. 2020, doi: 10.1038/s41598-020-70490-4.

[14] V. Chandran *et al.*, "Diagnosis of Cervical Cancer based on Ensemble Deep Learning Network using Colposcopy Images," *Biomed Res Int*, vol. 2021, no. 1, May 2021, doi: 10.1155/2021/5584004.

[15] S. Saini, K. Ahuja, S. Chennareddy, and K. Boddupalli, "Deep Learning Descriptor Hybridization with Feature Reduction for Accurate Cervical Cancer Colposcopy Image Classification," *Pattern Recognit Lett*, May 2024, doi: 10.48550/arXiv.2405.01600.

[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474.

[17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826. doi: 10.1109/CVPR.2016.308.

[18] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions."

[19] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *ICLR*, Apr. 2015, [Online]. Available: http://arxiv.org/abs/1409.1556

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

[21] D. A. Minciună *et al.*, "AnnoCerv: A new dataset for feature-driven and image-based automated colposcopy analysis," *Acta Universitatis Sapientiae, Informatica*, vol. 15, no. 2, pp. 306–329, Dec. 2023, doi: 10.2478/ausi-2023-0019.

p-ISSN 2301-7988, e-ISSN 2581-0588
DOI : 10.32736/sisfokom.v14i2.2361, Copyright ©2025
Submitted : May 6, 2025, Revised : May 13, 2025, Accepted : May 14, 2025, Published : May 26, 2025