ISSN 2723-6471

Improving Early Detection of Cervical Cancer Through Deep Learning-Based Pap Smear Image Classification

Nita Merlina[1,*], Arfhan Prasetyo[2], Ida Zuniarti[3], Nissa Almira Mayangky[4], Daning Nur Sulistyowati[5], Faruq Aziz[6]

[1,2,3,4,5,6]*Faculty of Technology Information, Universitas Nusa Mandiri, Jalan Raya Jatiwaringin No 2, Cipinag Melayu, Jakarta Timur*

(Received: December 9, 2024; Revised: December 19, 2024; Accepted: January 19, 2025; Available online: March 3, 2025)

Abstract

Cervical cancer is one of the leading causes of death in women worldwide, making early detection of the disease crucial. This study proposes a deep learning-based approach that has the advantage of leveraging pre-trained models to save data, time, and computation to classify Pap smear images without relying on segmentation, which is traditionally required to isolate key morphological features. Instead, this method leverages deep learning to identify patterns directly from raw images, reducing preprocessing complexity while maintaining high accuracy. The dataset used in this study is a public data repository from Nusa Mandiri University (RepomedUNM), which has a wider variety of data. This dataset is used to classify images into four categories: Normal, LSIL, HSIL, and Koilocytes. The dataset consists of 400 images evenly distributed, ensuring class balance during training. Transfer learning is applied using five Convolutional Neural Network (CNN) architectures: ResNet152V2, InceptionV3, ResNet50V2, DenseNet201, and ConvNeXtBase. To prevent overfitting, techniques such as data augmentation, dropout regularization, and class weight adjustment are applied. The evaluation results in this study showed the highest accuracy with a value of ResNet152V2 = 0.9025, InceptionV3 = 0.8953 and DenseNet201 = 0.8845. ResNet152V2 excelled in extracting complex features, while InceptionV3 showed better computational efficiency. The study also highlighted the clinical impact of misclassification between Koilocytes and LSIL, which may affect diagnostic outcomes. Data augmentation techniques, including horizontal and vertical flipping and normalization, improved the model's generalization to a wide variety of images. Specificity was emphasized as a key evaluation metric to minimize false positives, which is important in medical diagnostics. The findings confirmed that transfer learning effectively overcomes the limitations of small datasets and improves the classification accuracy of pap smear images. This approach shows potential for integration into clinical workflows to enable automated and efficient cervical cancer detection.

Keywords: CNN, Cervical Cancer Detection, Classification, Pap Smear, RepomedUNM

1. Introduction

Cervical cancer, especially in Indonesia, is a cancer with the largest number of sufferers, ranking second as one of the cancers that often attacks women after breast cancer [1], due to the presence of malignant cells in the cervical tissue. Early detection of this disease can be done through a Pap smear test, which aims to find cell abnormalities before they develop into cancer. However, the implementation of this test has challenges, such as the potential for errors due to human factors and a process that takes a long time [2]. In addition, cervical intraepithelial neoplasia is often small and its cells can overlap or be covered by blood mucus, making interpretation of Pap smear images difficult and the procedure prone to error [3].

In an automated cervical cancer screening system, the main thing to consider is determining whether the sample shows any indication of cancer cells [4]. The process of classifying cervical cancer image datasets is very important to support early screening for cancer cells in women. Analysis of cells is carried out by utilizing certain features to classify cells into normal or suspicious categories. This procedure is important because it provides essential initial information in detecting cancer or precancerous cells.

Classification of cervical cells in Pap smear images is challenging due to limited image resolution and the complexity of morphological changes in cell structure. This procedure is crucial because it provides initial information to detect

***Corresponding author: Nita Merlina (nita@nusamandiri.ac.id)**
DOI: https://doi.org/10.47738/jads.v6i2.576

cervical cancer as soon as possible “Cervical Cancer Classification from Pap Smear Using XCeption Model,”. Classification of Pap smear images is carried out through an analysis process to exploration of deep learning methods to perform classification. The dataset used in this study is the public data repository of Nusa Mandiri University (RepomedUNM), which has a wider variety of data [5]. This study began with image digitization through acquisition from RepoMedUNM, followed by the stage of analyzing the acquired images to identify cell nuclei and classify normal and abnormal images.

Deep learning technology [6], [7] is expected to be able to identify abnormal cells or precancerous lesions better than conventional methods. One of the advantages of the CNN model is that it eliminates the segmentation process because the model can learn from the original data (citations). Transfer learning uses a pre-trained model as a foundation for new tasks. In medical imaging, it leverages large datasets like ImageNet to adapt general features for specific tasks, such as Pap Smear classification, reducing training time and improving performance on small datasets [8]. With fine-tuning, the model quickly adapts to specific tasks like water meter recognition, even with limited data, and is easy to implement using modern frameworks. These models have learned general features that can be adapted for specific tasks, such as Pap Smear image classification. Using a pre-trained model reduces the training time and computational resources required, and improves model performance on datasets with limited size, such as the one used in this study. This research is expected to facilitate the task of pathologists in detecting Pap smear images, including identifying overlapping or clustered cell nuclei in microscopic images. By utilizing deep learning methods in Pap smear image classification, it is expected to improve early detection capabilities, diagnostic accuracy, and access to health services for the community globally [9].

2. Related Works

Several studies have compared different methods for treating cervical cancer [10]. Advances in computer vision have improved over traditional medical imaging, allowing automation and Artificial Intelligence prediction. Models learn image features to detect cells, with larger datasets yielding more accurate predictions [11].

2.1. Detection of Cervical Cells

Some studies have focused on detecting specific object sizes, such as prominent, small, or very small objects [12], [13]. In research, detection networks can be either one-stage or two-stage [14], with one-stage networks predicting location in one step. Advances in deep learning have improved general object detection [15], using CNNs for both location and classification tasks. Karasu Benyes et al. [16] compared four models to classify Sure Path images, and achieved an accuracy of over 90%. For ThinPrep Pap classification, the accuracy increased to 92.65% with domain adaptation using Deep CORAL and ResNet101. To improve scalability, developed a faster RCNN-FPN model with deformable convolutional layers, but was limited by the use of Whole Slide Images [3]. Elakkiya et al. [17] achieved the best accuracy with small object detection neural network to identify cervical points. The 3cDe-Net model [17] showed superior performance with a MAP of 50.4%, detecting cells at the image level across a wide range of sizes. In [16], FasterRCNN and EfficientNet B4 detected cervical cells with an IoU score of 0.623. Xiang et al. [18]used YOLOv3 and InceptionV3 for efficient cervical cell classification, with near-perfect accuracy and sensitivity but low specificity.

2.2. Classification of Cervical Cells

Studies [19] and [20] used transfer learning (TL) techniques, including ResNet-50, DenseNet-121, and DenseNet-169, for binary classification to overcome data limitations. Study [20] proposed an acetowhite mask-based image processing method, which enhances the focus on relevant regions. Study [19] used a pre-trained DenseNet to classify lesion extent in cervical images, where DenseNet-169 outperformed DenseNet-121 in terms of accuracy and sensitivity, indicating a positive relationship between network depth and sensitivity. DenseNet also outperformed a SVM classifier trained with custom features and could analyze 600 images in less than a minute.

Huang et al. [19]introduced the classification of cervical biopsy tissue images using LASSO method and EL-SVM ensemble, which reduced the optimization time to 35.87 seconds while maintaining the classification accuracy. They used LASSO technique for feature selection and serial fusion, classifying 468 biopsy tissue images. ROC and error curves evaluated the performance of the classifier, which showed the best categorization results. They successfully completed the two-step feature selection process for individual cell differentiation.

 ISSN 2723-6471

This study outperforms previous studies using models such as ResNet50V2, ResNet152V2, InceptionV3, DenseNet201 and ConvNeXtBase. For example, when compared to a study [14] using these five CNN models.. This research focuses on classifying Pap smear images to detect conditions such as HSIL, LSIL, Koilocytes, and Normal, where the images exhibit overlapping nuclei and cytoplasm, along with some inflammation resembling but not merging with the nuclei.

This study focuses on Pap smear image classification, detecting conditions such as HSIL, LSIL, Koilocytes, and Normal, with images showing overlapping nuclei and cytoplasm, and some inflammations that resemble but do not merge with the nucleus (as seen in figure 1).

(a) (b)

Figure 1. Example of an image with inflammation (a) Normal image with Koilocyt present (b)

In figure 1(a), there are two nuclei, one outside the image that is inflammatory and resembles a nucleus. In figure 1(b) although there are 6 normal nuclei and one Koilocyte, but for a higher classification in the koilocyte class.

3. Methodology

This section will explain the dataset and methodology used. This study tested and compared five CNN models with transfer learning using the public dataset RepoMedUNM for the classification of cervical cancer cells into four categories.

The dataset used comes from RepoMedUNM, which provides a variety of images, including ThinPrep and non-ThinPrep (conventional) Pap smear images, with a total of 400 Dataset yang digunakan dalam penelitian. For our analysis, we used a subset consisting of 400 images, evenly divided into four classes: Normal (100 images), Koilocyt (100 images), LSIL (100 images), and HSIL (100 images). This balanced distribution ensures that each class is equally represented, which is crucial for training a robust model. Subsequently, this dataset underwent random augmentation, resulting in 1,934 training images, 828 testing images, and validation data.Pap smear cell images. The dataset images evenly distributed across four classes, ensuring balance and robustness during training. The images in the RepoMedUNM dataset were obtained through a digitization process using OLYMPUS CX33RTFS2 and X52-107BN microscopes and a Logitech camera. Figure 2 shows an example of the RepoMedUNM dataset divided into four classes: normal cells, and three classes of abnormal cells, namely L-Sil, H-Sil, and koilocytes.

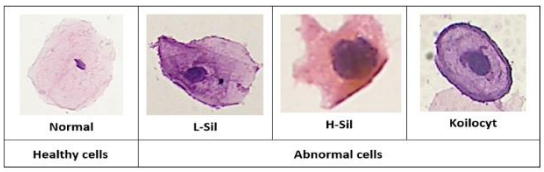

Figure 2. Example image from RepomedUNM with four cell types

The RepoMedUNM database was retrieved and converted into an array according to the input data format. However, the high variation of the data makes it less than ideal for model convergence. Therefore, the input was normalized with a range of values between 0 and 1 to help the model in the generalization process and improve its performance. To prevent overfitting, the data was divided into three parts: training data (70%), validation data (20%), and test data (10%). While the dataset size is relatively small, the use of data augmentation and transfer learning mitigates the risk of overfitting and enhances generalization. To prevent overfitting, we used regularization techniques such as data augmentation, dropout (rate 0.5), L2 regularization, and early stopping after 10 epochs without validation loss improvement.

ResNet152V2 and InceptionV3 outperformed DenseNet201 and ConvNeXtBase due to their superior feature extraction. ResNet152V2's deep residual network addresses the vanishing gradient problem, while InceptionV3's

 ISSN 2723-6471

inception module captures multi-scale features, making them well-suited for classifying cervical cells of varying shapes and sizes.

This study uses augmentation techniques to prevent overfitting and assigns class weights inversely proportional to their frequency, increasing the error penalty for minority classes. Additionally, balanced batches ensure uniform class distribution in each batch, preventing the model from focusing too much on the majority class. Some literature shows that the most commonly used classification method is CNN. This model has been trained for a specific task and is often used as a basis for other tasks, because the learning outcomes from the previous task can be utilized for new tasks. The flow of the trained model is able to generalize images from other learning data and significantly improve image classification [21], [22].

In this study, the transfer learning technique was used to classify cervical classes. This technique allows the application of knowledge and skills gained from previous tasks to new tasks. Building a CNN model from scratch requires a large amount of training data because the model must learn millions of weights. Therefore, the use of pre-trained models has become a popular approach to automatically transfer features to new datasets. Each fully connected layer in the pre-trained model is replaced with a modified fully connected layer, which has four output nodes to represent the four cervical classes. Transfer learning leverages pre-trained models on large datasets, significantly reducing training time and improving performance on smaller datasets such as RepomedUNM. Figure 3 shows the complete flow of cervical cancer classification in this study and figure 4 shows the workflow from data collection, preprocessing, model training, evaluation, to fine-tuning and final classification.

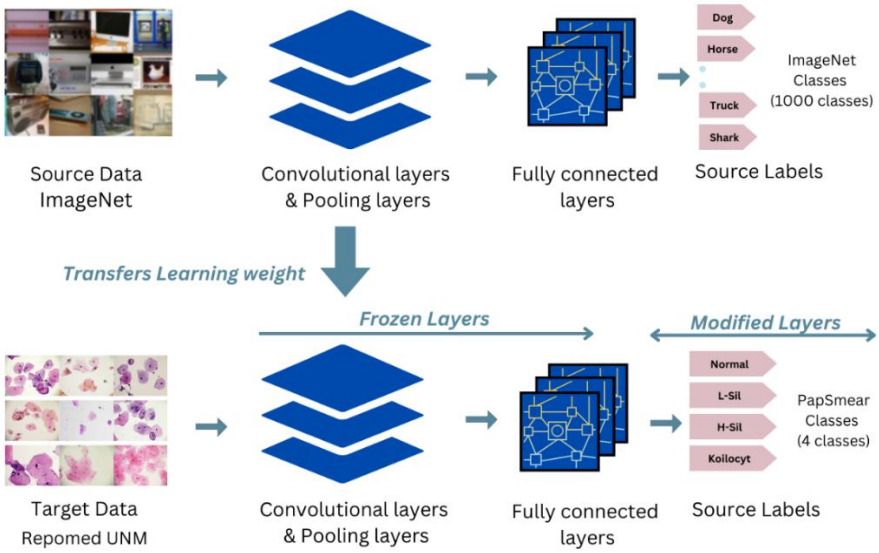

Figure 3. CNN work stages for cervical cancer classification using transfer learning

This process ensures that five leading transfer learning models namely: ResNet152V2, InceptionV3, ResNet50V2, DenseNet201, and ConvNeXtBase are selected for evaluation, so that they can provide accurate and optimal predictions based on the existing data. In the context of cervical cancer diagnosis, authors [23], [24]describe how deep learning technologies can be applied and impactful in analyzing complex cervical images and addressing observation bias. This current research encourages exploration of various CNN architectures commonly used for image processing, such as ConvNeXtBase. These models were chosen due to their superior performance on various classification tasks.

 ISSN 2723-6471

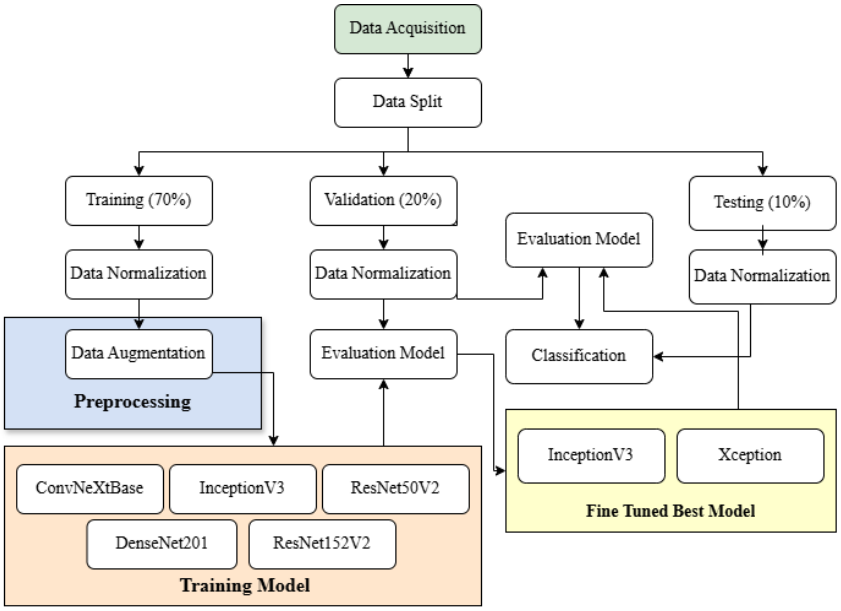

Figure 4. Workflow of the CNN model used in this research

To improve model performance and generalization, data augmentation was performed using rotation, width and height shift, shear, zoom, horizontal and vertical flip, and rescale (normalization) techniques. Horizontal and vertical flips simulate realistic image variations caused by microscope rotation, while zoom and shifts improve model robustness against spatial distortions. This augmentation adds variety to the dataset so that the model can learn from various scenarios that may occur in real-world data. Data augmentation increases dataset variation and helps the model generalize better. Techniques like horizontal and vertical flips are effective since Pap Smear images have no fixed orientation, and flipping doesn't change class information. Rotation, zoom, and shift simulate real image variations, making the model more robust to changes in cell position and perspective, thus improving generalization on unseen test data.

Raw images have pixel values between 0 and 255. Normalization changes this scale to a range of 0 to 1 (by dividing the pixel values by 255), making the data consistent and more manageable for the model. This helps maintain a consistent data distribution between training and test data, which is important for model generalization. In the image preprocessing stage, we use quality enhancement techniques to ensure high-quality data. Noise reduction removes unwanted artifacts, while contrast enhancement clarifies cellular details for better feature extraction. These techniques ensure clean, informative input, improving the model's classification performance.

To increase the variation of training data, data augmentation was performed using parameters as in table 1 and After augmentation, the total number of training images increased to 1,934, while the validation data remained at 828 images. The use of parameters to train the CNN model after conducting an in-depth literature review can be seen in table 1.

Table 1. Parameters for training data variations

Parameter	Value
Horizontal flip	True
Vertical flip	True
Rotation range	30 degrees
Zoom range	0.3
Width shift range	0.3
Height shift range	0.3
Shear range	0.2

 ISSN 2723-6471

The model was trained for 50 epochs using Categorical Crossentropy. To prevent overfitting, a dropout of 0.5 was applied to certain layers in the network. The optimizer used was Adam with a default learning rate of 0.001. In addition, an early stopping mechanism with a patience level of 10 epochs was applied to stop training if there was no significant improvement in model performance. In the last layer (dense last layer), there are 1024 neurons with ReLU activation function, followed by an output layer with softmax activation function to perform classification.

Different deep learning model architectures were employed to detect health conditions based on images classified as HSIL, LSIL, Koilocytes, and Normal. Each input image was resized to 224 x 224 pixels with three color channels (RGB) to preserve critical details. Image normalization and contrast enhancement were applied to ensure consistent data quality and improve feature extraction. During training, a batch size of 32 was used, balancing memory usage and convergence speed, while the Adam optimizer, selected for its adaptive learning rate capabilities, was employed with a learning rate of 1e-5 to ensure stable weight updates. Early stopping with a patience of 10 epochs was implemented to prevent overfitting by halting training when no improvement was observed on the validation data. The categorical cross-entropy loss function, suitable for multi-class classification, was used to optimize the models. These hyperparameters, chosen based on iterative experimentation, provided an effective balance between computational efficiency and convergence stability, particularly given the relatively small dataset. The models were trained for 50 epochs with preprocessing and data augmentation to enhance performance. Evaluation revealed that the two best models, ResNet152V2 and InceptionV3, achieved high accuracy and performance after fine-tuning. These findings confirm that carefully designed deep learning models can effectively classify Pap smear images, serving as a valuable tool for early detection of related health conditions.

Hyperparameter selection was done through experimentation and cross-validation. An initial learning rate of 0.0001 was chosen after several trials showed that this value provided a good balance between stability and convergence speed. A batch size of 32 was chosen based on available GPU memory and training efficiency. Other hyperparameters such as the number of epochs and dropout rate were also adjusted based on performance on the validation set, ensuring that the model did not overfit.It would be useful to discuss how the results of this study compare with previous findings, especially when the study mentions a lack of evaluation for some models.

The Softmax activation function is used in the output layer to convert the results into probabilities for each class. In equation (1) the Softmax activation function will be presented.

$$\text{Softmax (xi)} = \frac{\exp(\text{xi})}{\Sigma_{\text{j}} \exp (\text{xj})} \tag{1}$$

Exp (xi) denotes the exponential function applied to the input vector, while (*exp xj*) denotes the exponential function applied to the output vector. The loss function, optimizer, and evaluation matrix are used to evaluate the model. The loss is calculated using sparse categorical cross-entropy, and better training and testing results are indicated by a decrease in the loss function value. Equation (2) shows the loss function used in this study.

$$\text{Loss } = -\sum_{i=0}^{N} \text{yi} \log(pi) \tag{2}$$

N is the number of classes, yi is the output value (1 or 0), and pi is the predicted value.

The output of five publicly available deep CNN models is then analyzed. Unknown test data is given to the classification process, then trained with the training data to check whether the samples can be classified accurately. This study uses accuracy, precision, recall, specificity, and F1-score as evaluation measures because they are effective for testing various class classifications. To calculate accuracy, specificity, recall, sensitivity, and F1-score values, Equations (3), (4), (5), and (6) can be used.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FN+FP+TN}} \text{ x } 100\% \tag{3}$$

$$\text{recision } = \frac{\text{TP}}{\text{TP+FP}} \text{ x } 100\% \tag{4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \text{ x } 100\% \tag{5}$$

$$\text{F1} - \text{score } = \frac{(2*\text{Recall}*\text{Presisi})}{\text{Recall+Presisi}} \text{ x } 100\% \tag{6}$$

TP = True Positive, TN = True Negative, FP = False Positive and FN = False Negative.

Specificity is emphasized as it reduces false positives, which is critical in medical diagnoses to avoid unnecessary treatments. The combination of accuracy, precision, recall, and F1-score provide a comprehensive evaluation of model performance.

4. Results and Discussion

4.1. Results

In this study, five CNN models utilizing transfer learning will be compared. Table 2 shows the performance comparison of all models.

Table 2. Model Evaluation Results

No	Model	Test Accuracy	Test Loss	Time per Step	Precision	Recall	F1-Score
1	ResNet152V2	9.025	3.728	141s	0.9	0.9	0.9
2	InceptionV3	8.953	3.549	36s	0.89	0.89	0.89
3	ResNet50V2	8.881	5.388	51s	0.89	0.89	0.89
4	DenseNet201	8.845	4.953	79s	0.88	0.88	0.88
5	ConvNeXtBase	8.225	1.3083	78s	0.82	0.82	0.82

ResNet152V2's deeper architecture allows superior feature extraction, particularly for complex morphological structures. Meanwhile, InceptionV3's efficient architecture balances computation speed and accuracy, outperforming DenseNet201 and ConvNeXtBase in this context. Parameters were tuned for each model to find the optimal combination with the best performance. After training and evaluation, two superior models were found with accuracy above 90%, namely ResNet152V2 and InceptionV3, which are marked in bold for easy identification.

This study achieved higher accuracies for ResNet152V2 (90.25%) and InceptionV3 (89.53%) compared to prior studies using similar datasets. For example, ResNet50V2, which was less evaluated in existing research, performed slightly lower (88.81%) but still exceeded benchmarks from studies on Herlev and CRIC datasets. The superior performance of ResNet152V2 is attributed to its deeper architecture, which allows more robust feature extraction, while InceptionV3 balances accuracy with computational efficiency, making it suitable for real-time applications. In the image preprocessing stage, we apply several quality enhancements shows better results than DenseNet201 and ConvNeXtBase. This can be attributed to their architectural strengths in feature extraction. ResNet152V2, with its deep residual network, effectively addresses the vanishing gradient problem, allowing the model to learn more complex features. The incepion module in InceptionV3 allows the model to capture multi-scale features, which is very useful for analyzing various shapes and sizes of cervical cells. These characteristics likely contribute to their superior performance in our classification task.

Fine-tuning was applied to the ResNet152V2 and InceptionV3 models by freezing the pre-trained early layers and retraining only the final layers. For ResNet152V2, we froze the first 140 layers and retrained the subsequent layers with a lower learning rate to avoid overfitting. For InceptionV3, layers 249 and below were frozen and the remaining layers were retrained. Parameter adjustments such as an initial learning rate of 0.0001 and the use of the Adam optimizer were chosen to maintain training stability while ensuring effective convergence. It can be explained in this study we only use augmentation techniques related to preventive overfitting and also

The evaluation of these two models was further refined to achieve optimal accuracy and performance. Parameter selection was carried out carefully to balance accuracy, computational efficiency, and prevent overfitting. This proves that the deep learning model is effective for pap smear image classification and contributes to early detection of health conditions.

The confusion matrix shows that both models are able to classify the Normal and HSIL classes with excellent precision, but experience slight confusion between the koilocyt and LSIL classes. Misclassifications between Koilocytes and

 ISSN 2723-6471

LSIL could lead to delayed or inappropriate treatment. Further investigation into these errors is critical to minimize clinical risks. The confusion matrix shows several misclassifications between the Koilocyt and LSIL classes. These errors have significant clinical implications, as Koilocyt represent cellular changes associated with HPV infection that can progress to precancerous lesions (LSIL). Classifying Koilocyt as LSIL may lead to overtreatment and unnecessary patient anxiety. Conversely, misclassifying LSIL as Koilocyt may lead to undertreatment, resulting in precancerous lesions not being treated in a timely manner. Therefore, accuracy in differentiating these two classes is critical to determining appropriate clinical steps.

This study uses accuracy, precision, recall, F1-score, and specificity as evaluation metrics. Accuracy gives a general view of performance, but in medical classification, precision and recall are more crucial. Precision shows the proportion of true positives, helping to avoid over-treatment, while recall measures the ability to detect all positive cases, ensuring no cases are missed. F1-score balances precision and recall. Specificity, which measures true negatives, is also important to prevent misdiagnosis and unnecessary treatment.

4.2. Discussion

Evaluation of pre-trained CNN models was conducted through an analysis of the evaluation matrix of the five models used in this study. Particular attention was given to addressing class imbalance in Pap smear image classification to understand its impact on model performance and classification results. Several key steps were implemented to tackle this issue. Data augmentation was applied to increase the number of samples, while class weighting was incorporated into the loss function during training, assigning higher weights to underrepresented classes. This ensured the model prioritized learning these classes without bias toward majority classes. Beyond these methods, balanced batch sampling was employed to maintain proportional class distributions during each training epoch, further mitigating imbalance. Validation and evaluation were carefully conducted using metrics such as precision, recall, F1-score, and the confusion matrix to ensure accurate and reliable predictions while providing in-depth insight into the advantages and limitations of each CNN model in the context of this study.

The performance improvement stage is carried out on the two best models found by fine-tuning. The fine-tuning process involves retraining the last few layers of the model with a low learning rate and using regularization to prevent overfitting. Fine-tuning was applied to the top-performing models, ResNet152V2 and InceptionV3, by unfreezing the last few layers while keeping earlier layers frozen to retain pre-trained knowledge. For both models, a low learning rate (1e-5) was used to ensure stable weight updates.

This section will discuss the model with the best performance results, namely ResNet152V2 with Fine-tuning results: Test Accuracy: 0.9025 and Test Loss: 0.3728. The results of Fine Tuning can be seen in figure 5 and the Results of the Confusion Matrix ResNet152V2 can be seen in figure 6.

Classification Report: ResNet152V2

	precision	recall	f1-score	support
Normal	0.957	0.900	0.928	50
Koilocyt	0.868	0.920	0.893	50
LSIL	0.865	0.900	0.882	50
HSIL	0.938	0.900	0.918	50
accuracy			0.905	200
macro avg	0.907	0.905	0.905	200
weighted avg	0.907	0.905	0.905	200

Figure 5. ResNet152V2 Classification Report

The report of ResNet152V2 Classification can be explained as follows: This model shows high ability in terms of precision, with very high precision values for the Normal class (0.957) and the HSIL class (0.938), indicating the model's ability to avoid classifying positive samples as negative. In terms of recall, this model managed to get the highest value for the Koilosit class (0.920), while the other classes have the same value of 0.900, indicating its ability to find all positive samples. F1-Score, which is the average between precision and recall, was obtained with a value of 0.928 for the Normal class and 0.918 for the HSIL class. With an overall average of 0.905, this model managed to classify 90% of all test samples correctly.

 ISSN 2723-6471

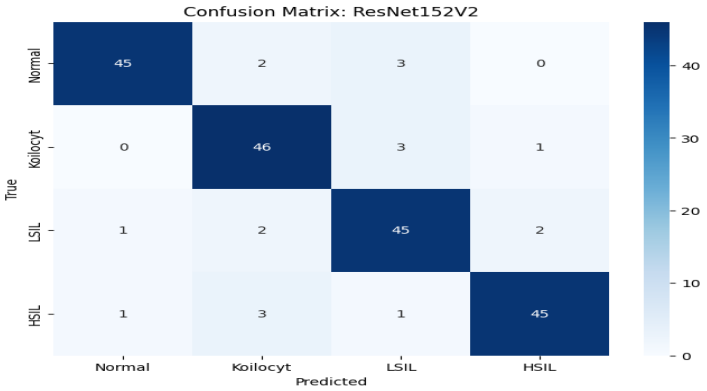

Figure 6. Confusion Matrix ResNet152V2

Next, we will discuss the second-best model after ResNet152V2, namely inceptionV3 with the following Fine-tuning results: Test Accuracy: 0.8953 and Test Loss: 0.3549. The results of the Fine Tuning can be seen in figure 7 and the results of the inceptionV3 Confusion Matrix can be seen in figure 8.

Classification Report: InceptionV3

	precision	recall	f1-score	support
Normal	0.845	0.980	0.907	50
Koilocyt	0.922	0.940	0.931	50
LSIL	0.894	0.840	0.866	50
HSIL	0.932	0.820	0.872	50
accuracy			0.895	200
macro avg	0.898	0.895	0.894	200
weighted avg	0.898	0.895	0.894	200

Figure 7. inceptionV3Classification Report

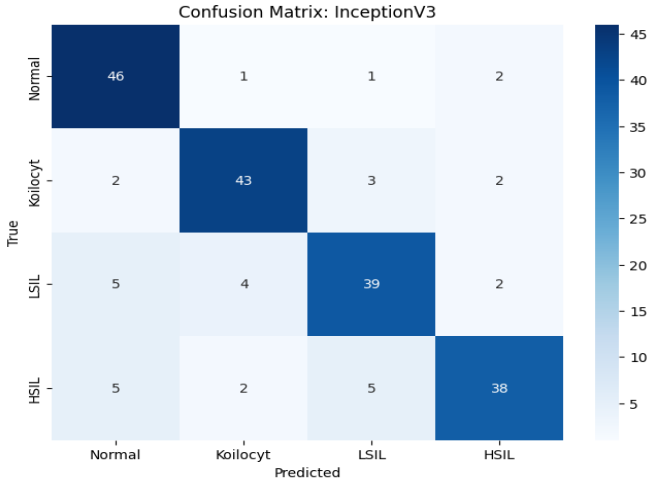

Figure 8. Confusion Matrix inceptionV3

The InceptionV3 classification report highlights key performance metrics: Precision, which measures the model's ability to avoid misclassifying positive samples as negative, is notably high for the HSIL class (0.932) and the koilosit class (0.922). Recall, representing the model's capability to detect all positive samples, is highest for the Normal class (0.980) and the koilosit class (0.940). The F1-Score, which balances precision and recall, is reported for the ResNet152V2 model, with values of 0.931 for the koilosit class and 0.907 for the Normal class. Lastly, the model achieves an accuracy of 0.895, correctly classifying 89% of all test samples.

 ISSN 2723-6471

In general, it can be concluded that ResNet152V2 has the best accuracy and F1-score performance, but the time per step is much longer than InceptionV3. In contrast, InceptionV3 offers an optimal balance between high accuracy and inference speed, making it more suitable for real-time applications.

This study uses the Adam optimizer for 50 epochs, while previous studies [14], [25], [26], only used 30 epochs. This adjustment was made to match the latest trends and facilitate comparison with the results in table 3. Some parameters in this study are the latest versions of previous studies, such as ResNet50V2 and ResNet152V2 with analysis using the RepoMedUNM public dataset.

It should be noted that although these parameters are not optimal and may require further adjustment, the initial results are sufficient as a basis for further research. The results show that the ResNet152V2, InceptionV3 and DenseNet201 models perform better than the results in [14] and were not evaluated in [25],[26], so a direct comparison cannot be made. The authors of [14] used the Herlev dataset with different optimizer and model configurations, which may have caused the difference in results.

Table 3. Comparison of Model Evaluation Results

Model	Accuracy (This Study)	Accuracy [14]	Accuracy [25]	Accuracy [26]
ResNet50V2	0.8881	0.7145	-	0.8937
ResNet152V2	0.9025	0.7496	-	-
InceptionV3	0.8953	0.8486	0.701	-
DenseNet201	0.8845	0.8702	-	-
ConvNeXtBase	0.8225	-	-	-

Overall, the pre-trained model proved to be very effective and inspiring, indicating that our proposal for automatic diagnosis of cervical cancer without segmentation methods or artificial features is very practical. Segmentation is often needed in Pap Smear analysis to isolate cells for feature extraction, which can be time-consuming and error-prone. However, our approach using a pre-trained CNN model bypasses segmentation by learning directly from raw images, simplifying the process and reducing errors.

Despite the positive results, the study has limitations. The small dataset size may limit the model's generalizability, and performance could decrease with imbalanced data or underrepresented image types. Future research should focus on using larger, more diverse datasets to enhance model robustness.

5. Conclusion

This study aims to detect cervical cancer or pre-cancer by testing various trained CNN models, enhancing the accuracy of diagnosis and aiding pathologists in reducing misdiagnosis, which facilitates faster and more effective treatment. The strength of this research lies in the comparison of five CNN models using the complex RepomedUNM dataset, classifying cervical cells into four categories. The results demonstrated strong performance, with ResNet152V2 achieving the highest accuracy. However, the study acknowledges the limitation of using a small dataset (400 images), which may hinder generalization to diverse clinical scenarios. Further investigation is needed to assess the models' performance on imbalanced or low-quality data. The findings suggest that the proposed models can be integrated into clinical workflows as decision support tools for automated cervical cancer detection, reducing manual effort and improving diagnostic consistency. Implementing these models in clinical settings would involve incorporating them into digital cytology platforms for real-time analysis of Pap smear images. Future research should focus on expanding the dataset, addressing class imbalance, and evaluating the models in real-world clinical environments to improve their robustness and reliability for clinical use.

The findings of this study have the potential to be implemented in clinical settings for automated detection of cervical cancer. The developed model can be integrated into a digital Pap Smear screening system providing decision support to pathologists by automatically detecting abnormal cells. This implementation can improve the efficiency and

accuracy of diagnosis, as well as reduce the workload of pathologists. In addition, this model can be used in areas with limited medical resources, assisting in mass screening and early detection of cervical cancer.

6. Declarations

6.1. Author Contributions

Conceptualization: N.M., A.P., I.Z., N.A.M., D.N.S., and F.A.; Methodology: D.N.S.; Software: N.M.; Validation: N.M., D.N.S., and F.A.; Formal Analysis: N.M., D.N.S., and F.A.; Investigation: N.M.; Resources: D.N.S.; Data Curation: D.N.S.; Writing Original Draft Preparation: N.M., D.N.S., and F.A.; Writing Review and Editing: D.N.S., N.M., and F.A.; Visualization: N.M. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

This research was funded through the National Competitive Applied Research Grant, Directorate General of Higher Education, Ministry of Education, Culture, Research, and Technology.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] I. Mauludiyah, S. Andarini, Holipah, L. Muflikhah, and L. Zuhriyah, "Cervical Cancer awareness and screening in Malang City, Indonesia: Sociodemographic determinants," *J. Med. Pharm. Chem. Res.*, vol. 7, no. 4, pp. 714–724, 2025, doi: 10.48309/jmpcr.2025.466820.1323.

[2] W. William, A. Ware, A. H. Basaza-Ejiri, and J. Obungoloch, "Cervical cancer classification from Pap-smears using an enhanced fuzzy C-means algorithm," *Informatics Med. Unlocked*, vol. 14, no. 2, pp. 23–33, 2019, doi:10.1016/j.imu.2019.02.001.

[3] X. Li, Z. Xu, X. Shen, Y. Zhou, B. Xiao, and T. Q. Li, "Detection of cervical cancer cells in whole slide images using deformable and global context aware faster rcnn-fpn," *Curr. Oncol.*, vol. 28, no. 5, pp. 3585–3601, 2021, doi: 10.3390/curroncol28050307.

[4] A. Tareef *et al.*, "Optimizing the cervix cytological examination based on deep learning and dynamic shape modeling," *Neurocomputing*, vol. 248, no. 7, pp. 28–40, 2017, doi:10.1016/j.neucom.2017.01.093.

[5] N. Merlina, E. Noersasongko, P. N. Andono, M. A. Soeleman, D. Riana, and J. Na, "Medical Image Registration at Pap Smear for Early Identification of Cervical Cancer," *TEM J.*, vol. 12, no. 2, pp. 726–731, 2023, doi: 10.18421/TEM122.

[6] N. Youneszade, M. Marjani, and C. P. Pei, "Deep Learning in Cervical Cancer Diagnosis: Architecture, Opportunities, and Open Research Challenges," *IEEE Access*, vol. 11, no. 1, pp. 6133–6149, 2023, doi: 10.1109/ACCESS.2023.3235833.

[7] M. Kalbhor, S. Shinde, D. E. Popescu, and D. J. Hemanth, "Hybridization of Deep Learning Pre-Trained Models with Machine Learning Classifiers and Fuzzy Min–Max Neural Network for Cervical Cancer Diagnosis," *Diagnostics*, vol. 13, no. 7, pp. 1–16, 2023, doi: 10.3390/diagnostics13071363.

[8] A. Sohail, "Transfer Learning" for Bridging the Gap Between Data Sciences and the Deep Learning," *Ann. Data Sci.*, vol. 11, no. 1, pp. 337–345, 2024, doi:10.1007/s40745-022-00384-x.

[9] M. MOMENZADEH, M. SEHHATI, A. M. DEHNAVI, A. TALEBI, and H. RABBANI, "Automatic Diagnosis of Vulvovaginal Candidiasis from Papsmear Images . Segmentasi dan Clasification," *Microscopy*, vol. 267, no. 3, pp. 299–

308, 2017, doi:10.1111/jmi.12566.

[10] K. P. Battula and B. S. Chandana, "Deep Learning based Cervical Cancer Classification and Segmentation from Pap Smears Images using an EfficientNet," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 9, pp. 899–908, 2022, doi: 10.14569/IJACSA.2022.01309104.

[11] O. Holmström *et al.*, "Point-of-Care Digital Cytology with Artificial Intelligence for Cervical Cancer Screening in a Resource-Limited Setting," *JAMA Netw. Open*, vol. 4, no. 3, pp. 1–13, 2021, doi: 10.1001/jamanetworkopen.2021.1740.

[12] S. L. Tan, G. Selvachandran, W. Ding, R. Paramesran, and K. Kotecha, "Cervical Cancer Classification From Pap Smear Images Using Deep Convolutional Neural Network Models," *Interdiscip. Sci. – Comput. Life Sci.*, vol. 16, no. 1, pp. 16–38, 2023, doi: 10.1007/s12539-023-00589-5.

[13] W. W. Q. L. H. F. J. S. H. L. R. Yang and A. Authors, "Salient Object Detection in the Deep Learning Era: An In-Depth Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 3239–3259, 2021, doi: 10.1109/TPAMI.2021.3051099.

[14] M. Kalbhor, S. Shinde, P. Wajire, and H. Jude, "CerviCell-detector: An object detection approach for identifying the cancerous cells in pap smear images of cervical cancer," *Heliyon*, vol. 9, no. 11, pp. 1–16, 2023, doi: 10.1016/j.heliyon.2023.e22324.

[15] X. Wu, D. Sahoo, and S. C. H. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, no. 1, pp. 39–64, 2020, doi: 10.1016/j.neucom.2020.01.085.

[16] Y. Karasu Benyes, E. C. Welch, A. Singhal, J. Ou, and A. Tripathi, "A Comparative Analysis of Deep Learning Models for Automated Cross-Preparation Diagnosis of Multi-Cell Liquid Pap Smear Images," *Diagnostics*, vol. 12, no. 8, pp. 1–20, 2022, doi: 10.3390/diagnostics12081838.

[17] R. E. V. S. V. V. A. Mahanti, "Cervical Cancer Diagnostics Healthcare System Using Hybrid Object Detection Adversarial Networks," *IEEE J. Biomed. Heal. Informatics*, vol. 26, no. 4, pp. 1464–1471, 2021, doi: 10.1109/JBHI.2021.3094311.

[18] P. C. et al Xiang Y, Sun W, "A novel automation-assisted cervical cancer reading method based on convolutional neural network," *Biocybern Biomed Eng*, vol. 40, no. 2, pp. 611–623, 2020, doi:10.1016/j.bbe.2020.01.01.

[19] P. Huang *et al.*, "Classification of cervical biopsy images based on LASSO and EL-SVM," *IEEE Access*, vol. 8, no. 1, pp. 24219–24228, 2020, doi: 10.1109/ACCESS.2020.2970121.

[20] L. P. et al Zhang T, Luo Y, "Cervical precancerous lesions classification using pre-trained densely connected convolutional networks with colposcopy images," *Biomed Signal Process Control*, vol. 55, no. 8, pp. 1–11, 2020, doi:10.1016/j.bspc.2019.101566.

[21] J. Song, S. Gao, Y. Zhu, and C. Ma, "A survey of remote sensing image classification based on CNNs," *Big Earth Data*, vol. 3, no. 3, pp. 232–254, 2019, doi: 10.1080/20964471.2019.1657720.

[22] U. Usman, F. Yunita, and M. R. Ridha, "Improving Classification Accuracy of Local Coconut Fruits with Image Augmentation and Deep Learning Algorithm Convolutional Neural Networks (CNN)," *J. Appl. Data Sci.*, vol. 6, no. 1, pp. 1–19, 2025, doi:10.47738/jads.v 6 i 1 . 389.

[23] P. Wang, L. Wang, Y. Li, Q. Song, S. Lv, and X. Hu, "Automatic cell nuclei segmentation and classification of cervical Pap smear images," *Biomed. Signal Process. Control*, vol. 48, no. 2, pp. 93–103, 2019, doi:10.1016/j.bspc.2018.09.008.

[24] M. M. Rahaman *et al.*, "A Survey for Cervical Cytopathology Image Analysis Using Deep Learning," *IEEE Access*, vol. 8, no. 3, pp. 61687–61710, 2020, doi: 10.1109/ACCESS.2020.2983186.

[25] P. B. J. M.B. Bijoy, S.M. Akondi, S. Abdul Fathaah, A. Raut, P.N. Pournami, "Cervix type detection using a self-supervision boosted object detection technique," *Int. J. Imag. Syst. Technol*, vol. 32, no. 5, pp. 1–6, 2022, doi:10.1002/ima.22696.

[26] E. Hussain, L. B. Mahanta, C. R. Das, and R. K. Talukdar, "A comprehensive study on the multi-class cervical cancer diagnostic prediction on pap smear images using a fusion-based decision from ensemble deep convolutional neural network," *Tissue Cell*, vol. 65, no. 2, pp. 1–8, 2020, doi: 10.1016/j.tice.2020.101347.