IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 1, February 2026, pp. 338~349
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i1.pp338-349

Deep learning-based cervical cancer detection via colposcopy images integrated into an Android mobile application

Retno Supriyanti[1], Arsil Kultura Anzil[1], Yogi Ramadhani[1], Suroso[1], Wahyu Widanarto[2], Muhammad Alqaaf[3], Kartika Dwi Hapsari[4], Futiat Diana Kartika[4]
[1]Department of Electrical Engineering, Faculty of Engineering, Jenderal Soedirman University, Purbalingga, Indonesia
[2]Department of Physics, Faculty of Mathematics and Natural Science, Jenderal Soedirman University, Purbalingga, Indonesia
[3]Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma, Japan
[4]Regional Technical Implementation Unit, Kalimanah Community Health Center, Purbalingga, Indonesia

Article Info

Article history:

Received Nov 27, 2024
Revised Oct 20, 2025
Accepted Jan 10, 2026

Keywords:

Cervical cancer
Convolutional neural networks
Early detection
Mobile application
Visual inspection

ABSTRACT

Cervical cancer is a form of cancer that develops in the cells of the cervix, the lower part of the uterus that connects the uterus to the vagina. Early detection is essential for improving the chances of recovery from cervical cancer. One method for early detection is colposcopy image analysis, a medical procedure that examines the cervix and captures images for evaluation. These images were analyzed to observe color changes after the visual inspection with acetic acid (VIA) process. However, this analysis requires experienced and specially trained medical personnel. To address this challenge, a system that can automatically classify cervical cancer images is needed. Therefore, researchers proposed designing and developing an Android mobile application to enable early detection of cervical cancer using the convolutional neural network (CNN) algorithm. The CNN model was tested using test data to evaluate its performance. The optimized CNN model utilizing the ResNet50 architecture achieved 86% test accuracy, 85% precision, and 87% recall. The test results indicate that the model's accuracy is consistent before and after its implementation on the mobile application, confirming the effectiveness of both the model and its implementation as diagnostic tools.

Corresponding Author:

Retno Supriyanti
Department of Electrical Engineering, Faculty of Engineering, Jenderal Soedirman University
Blater Campus, St. Mayjend Sungkono km 5, Blater, Purbalingga, Indonesia
Email: retno_supriyanti@unsoed.ac.id

1. INTRODUCTION

Cervical cancer is the uncontrolled growth of abnormal cells in the cervix, the lower part of the uterus connecting to the vagina. It is one of the most common cancer types in women [1]. On the other hand, a colposcopy image is a visual image produced from a colposcopy examination. Colposcopy is a medical procedure used to examine the condition of the cervix more closely. Colposcopy images help doctors detect precancerous lesions or cervical cancer at an early stage when they can still be treated more easily. So, it can be concluded that colposcopy images are a precious tool in the early detection of cervical cancer [2].

Detecting cervical cancer at an early stage is crucial for prevention. When caught early, it is one of the most treatable cancers. Colposcopy image analysis is one of the methods used for early detection. This procedure aims to capture an image of the cervix, which will be analyzed for color changes following the visual inspection with acetic acid (VIA) process. During the process, pre-cancerous (abnormal) cells change color to white, commonly called a lesion. Analyzing these abnormal cells requires skilled medical personnel

with specialized training. Therefore, there is a need for an automated system for detecting and classifying cervical cancer images to enhance both the accuracy and speed of colposcopy image classification.

Several individuals have researched cervical cancer. Ziyad *et al.* [3] conducted a study to identify the factors affecting successful cervical cancer screening among medical professionals in Ethiopia. Kabanda *et al*. [4] analyzed cross-sectional data to evaluate the effectiveness of community audio towers (CAT) as a health communication tool for promoting cervical cancer prevention. They surveyed women aged 21 to 60 about their intentions regarding cervical cancer screening. Generalized linear modeling, specifically using modified Poisson regression and backward variable elimination, was employed to identify prevalence ratios and determine the factors associated with intentions to undergo cervical cancer screening. Nartey *et al.* [5] utilized the nested multiplex polymerase chain reaction (NM-PCR) method to detect human papillomavirus (HPV) infections in cervical samples. To ascertain the relationship between the risk of cervical cancer and HPV infection, they applied logistic regression analysis. Gebremeskel and Gebretatios [6] conducted studies on human immunodeficiency virus (HIV)-infected women in several public hospitals in Tigray. Two specific studies were carried out: one focused on HIV-infected women who did not have access to cervical cancer screening facilities, and the other on those who did receive such screenings. The results from both groups were then analyzed using linear regression to compare the outcomes.

Alkhamis *et al*. [7] were conducted to measure the prevalence of cervical cancer screening and to identify predictive factors among sexually active women in Saudi Arabia. The target population included adult women aged 21 to 65 years who had been sexually active and had not undergone a hysterectomy. The researchers performed an analysis using multivariate logistic regression. Beyene *et al.* [8] were conducted on the role of men in encouraging their wives to undergo cervical cancer screening. This research took place in Ethiopia, where women in low-income countries often face resistance from their husbands regarding this screening. The research team gathered statistics on men's contributions, which were then analyzed using linear regression. The study's results informed the development of a strategic plan to involve husbands in supporting their partners to participate in cervical cancer screening. Bai *et al.* [9] conducted research utilizing 16S ribosomal (rRNA) sequencing technology and metagenomics to analyze vaginal microorganisms and screen for cervical cancer based on their bacterial presence.

Namutundu *et al.* [10] identify various barriers related to health facilities, community interactions, interpersonal relationships, and individual factors, enabling facilitators to comprehend and apply most effective methods for improving access to cervical cancer screening for the people of Uganda. Karena and Faldu [11] conducted a study to evaluate hospital nurses' knowledge and responses to cervical cancer screening in India. Kim *et al*. [12] researched glyoxalase-01 (GL01) expression for detoxifying methylglyoxal to assess the role of GL01 in cervical cancer metabolism. A study was conducted by Agénor *et al.* [13] to examine the death rate from cervical cancer and racism. The research highlighted the disparities in health services available to minority groups in the United States. Data analysis was performed using a thematic analysis approach. Lei *et al.* [14] researched the role of stem cells in developing cervical cancer. The study examined the effects of zinc oxide on cervical cancer stem cells and its relationship with ferroptosis, intending to determine whether zinc oxide can inhibit the proliferation of cervical cancer cells. Lartey *et al.* [15] researched HPV testing among Ghanaian women. A survey was conducted using personal telephones to gather information about cervical cancer. The analysis was performed using the Fischer test. Mantula and Toefy [16] examined male involvement in cervical cancer screening programs from the perspective of women in Zimbabwe. This qualitative study included input from both women and health service providers in the region. Data analysis utilized a thematic approach supported by ATLAS software.

Our research has extensively discussed the use of artificial intelligence-based images to enhance diagnoses, including the identification of leukocytes, Alzheimer's disease, and other conditions [17]–[20]. Recent studies have utilized advanced convolutional neural network (CNN) architectures and hybrid models to enhance the accuracy of cervical cancer classification. Models like EfficientNet and DenseNet, recognized for their computational efficiency and excellent feature extraction capabilities, have shown remarkable performance. In this context, our research positions itself by using ResNet50 as a strong and dependable baseline, making it a practical solution for mobile devices. We are currently researching diagnosing cervical cancer using artificial intelligence-based image processing techniques. In previous studies, we successfully explored features in the colposcopy images of cervical cancer patients [19]. This paper will focus on developing a mobile application that uses the CNN method to classify cervical cancer. This research is novel because it integrates rigorous scientific validation into a clinically viable prototype designed for real-world application. Unlike typical studies that focus solely on classifying images or developing applications, this paper aims to bridge the gap between often unrealistic academic research and poorly validated field implementations. The strength of this research lies in its proposal to create a system that is not only functional but also proven reliable through rigorous statistical methods. Furthermore, it is efficient on low-power mobile devices—a combination that is rarely demonstrated comprehensively in a single study.

2. METHOD

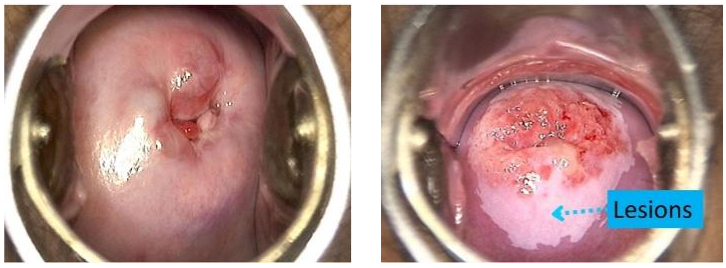
2.1. Initial data processing

This section outlines the data processing procedure, which consists of four steps: data collection, image segmentation, data augmentation, and data preprocessing. First, we will discuss the data sources and methods employed to collect data, ensuring sufficient diversity and representation of both colposcopy images, including normal and abnormal ones. Next, we will explain the image segmentation process and the techniques used. The data preprocessing stage will be elaborated on in detail, covering steps such as data splitting and augmentation for better generalization. Additionally, we will address the image resizing process to ensure that all photos are adjusted to a consistent resolution.

ResNet50 was selected as the base architecture for our model not only due to its outstanding performance in initial testing but also because of its considerable architectural advantages. A key feature of ResNet50 is its use of residual connections, which facilitate a smooth flow of information across multiple layers. This design effectively addresses the issues of vanishing gradients and accuracy degradation, which are common challenges in training deep networks. ResNet50 has demonstrated its reliability and robustness as a benchmark in various computer vision applications, including medical imaging, making it a logical choice for our colposcopy image classification task.

2.1.1. Data collection

The data used in this study were obtained from the International Agency for Research on Cancer (IARC) VIA Image Bank. It consisted of 180 cancer cases and 420 data images [21]. The dataset includes various images: untreated images, those treated with acetic acid (VIA), and those treated with Lugol's iodine. For this study, we focused on 164 images that had undergone VIA, including 73 abnormal and 91 normal images. This dataset is valuable because it clearly distinguishes between abnormal and normal images, particularly with lesions. Lesions appear as white areas on the cervix's surface following VIA treatment, indicating the potential for cancer in the abnormal cells present in that area. Figure 1 shows an example of the imagery used in this research. Figure 1(a) presents an image of the cervix in normal conditions, while Figure 1(b) displays an image of the cervix affected by cancer.



(a) (b)

Figure 1. Examples of image data used of (a) normal condition and (b) abnormal condition

The data imbalance, noted with 73 abnormal and 91 normal cases, highlighted the need for additional data collection. To address this, we gathered data from the IARC Colposcopy Image Bank [22], which included 200 cases and 913 images. Similar to the previous dataset, various images were included, such as untreated images, treated images with acetic acid (VIA), and images using a green filter. In this collection, we focused on images treated with acetic acid (VIA) to ensure consistency in the characteristics of the data. Ultimately, we used 90 normal images and 72 abnormal images. This resulted in 163 images for each class, bringing the overall total to 326 images.

2.1.2. Cropping image

At this stage, abnormal and normal images will be cropped to minimize noise in each data image. The cropping process will be performed manually, focusing on the region of interest (ROI). The ROI refers to the specific area that is the primary focus for feature extraction in an image processing model. In this study, the cervix will be used as the parameter for defining the ROI, as illustrated in Figure 2. The cervix was selected as the ROI because VIA treatment is applied to this area. Consequently, lesions representing abnormal cells are expected to appear on the surface of the cervix.

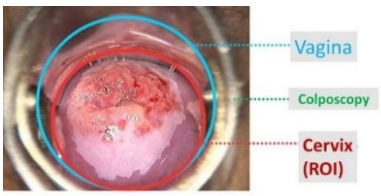


Figure 2. Colposcopy image parts

2.1.3. Pre-processing data

At this stage, two data pre-processing steps will be carried out. The first step involves dividing the dataset into training, validation, and test data. This division is essential to separate the data used during the training phase from that used for testing. The training and validation data will be employed in the training process. Specifically, the training data is used to train the deep learning model to identify similarities in image features. In contrast, the validation data evaluates the model's performance during training. The data division percentage is set at 80:20, as this approach is believed to yield a more effective model [23]. This research has 194 training samples, 64 validation samples, and 68 test samples. The relatively small number of training data makes it necessary to introduce data variation, which is achieved through data augmentation [24]. To overcome the limitations of a small dataset and minimize the risk of overfitting, we will use k-fold cross-validation with k set to 5 for model validation. This method provides more reliable and unbiased performance estimates compared to the simple hold-out approach, as it ensures that every sample in the dataset is utilized for both testing and training purposes.

Data augmentation is a technique used in image processing to increase the diversity of data by manipulating existing images. This study employs two augmentation techniques: horizontal flip and vertical flip. These techniques involve reversing the original image either horizontally or vertically. As a result, the images produced by these techniques are flipped to the right or turned upside down. Model performance was evaluated using 5-fold cross-validation. The model was trained and tested five times, and performance metrics were averaged to provide more reliable estimates. Table 1 describes the model performance summary using k-fold cross-validation.

Table 1. Model performance summary using k-fold cross-validation

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
1	85.0	83.4	86.1	84.7
2	87.5	89.2	86.8	88.0
3	86.3	85.0	87.5	86.2
4	88.7	87.1	90.2	88.6
5	84.0	82.5	85.6	84.0
Average (±SD)	86.3 (1.8%)	85.4 (2.6%)	87.2 (1.6%)	86.3 (2.0%)

2.2. Convolutional neural network modeling

This section will cover the design and selection of the CNN model used in this study, including the steps taken to determine the neural network's architecture and select the deep learning model's hyperparameters. When developing a CNN model, selecting the appropriate architecture is essential for achieving optimal classification results. The chosen architecture impacts both computation time and the extraction of features from images. This study's photos are complex and require a detailed approach to extract relevant features effectively. Therefore, choosing the exemplary architecture is crucial for the success of this research. In this research, we decided to pick four popular architectures, namely ResNet50, VGG19, InceptionV3, and LeNet. Then, the four architectures will be compared based on two primary factors: accuracy and computation time. These two factors will determine the final architecture for this study's CNN model. Comparison testing will be done by training the model using image data from the previous preprocessing stage.

2.3. Android mobile application development

This phase involves developing an Android mobile application for classifying colposcopy images related to cervical cancer. The development will consist of three stages: front-end development, back-end development (logical system), and testing the model's performance within the application. Additionally, during this stage, the deep learning model will be discussed.

The front end is the part of a website or application that directly interacts with the user. This section includes design, layout, and user interface elements. The development will be based on a wireframe with five-page layouts, each designed using extensible markup language (XML). XML is a markup language that employs tags as markers to categorize and provide detailed explanations of data [23]. The five pages include the home page, classification page, guide page, about page, and settings page.

The back-end is the part of the system that users do not directly see. It is responsible for processing data and implementing system logic to support the user interface on the front end. Development is carried out using the Kotlin programming language, along with several open-source libraries such as TensorFlow lite, Photo picker, and Vanniktech image cropper. In Android development using Android studio, each activity layout automatically includes a class designed for processing the logical aspects of the system, referred to as an activity. In this application, five activity classes have been created, with the home activity serving as the main class that runs when the application is first opened. The five courses include home activity, classification activity, guide activity, about activity, and settings activity.

3. RESULTS AND DISCUSSION

We define the image and batch sizes in the initial data pre-processing stage. After that, we specify the file paths for the training, validation, and test datasets. Each image in the dataset is then read and subjected to horizontal and vertical flipping, which preserves the shape of the endocervix. In the context of image processing, particularly with colposcopy images used for detecting cervical cancer, the terms "horizontal flip" and "vertical flip" refer to data augmentation techniques that involve flipping images. These techniques are commonly employed to augment training data in deep learning and machine learning applications.

Figure 3 depicts an example of the final data results after the augmentation process. Figure 3(a) illustrates an image with a horizontal flip perspective. A horizontal flip involves flipping the image from left to right, similar to viewing a mirror image. For instance, if there is a lesion on the left side of the cervix, after a horizontal flip, it will appear on the right side. This technique can sometimes be helpful because the left and right sides of the cervix often exhibit symmetry. However, vertical flips should be used with caution or avoided altogether, as they can confuse medical interpretation. Figure 3(b) depicts an image with a vertical flip perspective, which means flipping the image from top to bottom. In this case, the cervical area, initially positioned at the top of the image, will move to the bottom after a vertical flip. In medical contexts such as colposcopy, vertical flips are less frequently used because they can disrupt the understanding of anatomical structures that have crucial orientations.

(a) (b) (c)

Figure 3. Augmented image of (a) original image, (b) horizontal flip, and (c) vertical flip

It includes 582 training data images, with each class containing 291 images. After the data is added, the images of each data type will be resized to 224×224 pixels. This adjustment helps ensure that the image dimensions are not too large, reducing the computational burden during the model training process. Additionally, having consistent height and width dimensions makes it easier for the model to train effectively. The value of 224 was chosen because it is commonly used in deep learning model training. We define the seed value, batch size, and class mode for our training process. The seed value is consistently set to 123 for all three data types to ensure uniformity in data distribution. The batch size determines the data processed in each training or testing cycle. We have chosen a batch size of 4 to prevent an excessive training load during any single process. Additionally, the class mode specifies the type of final classification result, which, in this study, is binary classification.

In CNN modeling, the ResNet50, VGG19, and InceptionV3 architectures utilize pre-trained models from an open-source library called Keras. A pre-trained model has already been trained on a dataset,

possessing weights and biases that signify its features [25]. These features will be leveraged in learning from the colposcopy image dataset through a concept known as transfer learning. This approach enhances the pre-trained model's ability to discern specific features related to the colposcopy images. In contrast, the LeNet architecture was built from scratch, as Keras does not yet offer a pre-trained model. Following the construction of the model, it will be trained over multiple iterations (epochs) precisely five times. Figures 4 and 5 illustrate the results comparing the accuracy and computation time of the four architectures.

ResNet50

Train	Validation	Train
96%	74%	63%

VGG19

Train	Validation	Train
93%	71%	43%

InceptionV3

Train	Validation	Train
60%	58%	42%

LeNet

Train	Validation	Train
91%	71%	53%

Figure 4. Comparison of the accuracy of four different architectures

ResNet50

Time
14 minutes and 49 seconds

VGG19

Time
38 minutes and 16 seconds

InceptionV3

Time
10 minutes and 37 seconds

LeNet

Time
13 minutes and 48 seconds

Figure 5. Comparison of computation time across four architectures

In Figure 4, the model based on the ResNet50 architecture achieves the highest accuracy, with a training accuracy of 96%, validation accuracy of 74%, and test accuracy of 63%. The VGG19 and LeNet architectures yield accuracy results relatively close to those of ResNet50. In contrast, the InceptionV3 architecture shows significantly lower accuracy, making it less suitable for the given image data. Figure 5 demonstrates that InceptionV3 has the fastest computation time, completing its tasks in 10 minutes and 37 seconds. This is followed closely by LeNet and ResNet50, which differ by only one minute. VGG19, however, has the longest computation time, taking 38 minutes and 16 seconds to complete. Based on these comparisons, the researcher chose the ResNet50 architecture as the foundation for the model because it achieved superior accuracy and has a computation time that remains competitive with the other architectures.

Next, we adjust the model and hyperparameters to train the CNN model. These adjustments are necessary because the previous model was still overfitting; the validation and test data accuracy values were significantly lower than those for the training data. Overfitting occurs when a model performs well on training data but fails to predict new, unseen data accurately. We modified the ResNet50 Keras model to address this issue and reduce overfitting. We made these adjustments by configuring the dense layer and incorporating regularization techniques. Regularization helps improve model generalization by reducing overfitting. For this purpose, we have chosen to use Keras Tuner, an open-source library for automatic model and hyperparameter optimization.

We implement two types of adjustments: model adjustments and hyperparameter tuning. For model adjustments, we will modify four parameters: the dropout value, the activation function, the number of units in the dense layer, and the kernel regularize value. We adjust three parameters for hyperparameter tuning: the optimizer, learning rate, and loss function. We utilize the random search method to identify the optimal values for each parameter by iterating through model training with various combinations of these predefined values. This method focuses on the accuracy achieved during each experiment iteration, with a maximum of 100 iterations performed. The combination of parameters that yields the highest accuracy will be selected as the final values for each tuning parameter. In Figure 6, a combination of parameters shows an accuracy rate of 100% after three training iterations. The final results of the model and hyperparameters after adjustment are also displayed in Figure 6. Subsequently, the model was re-trained using these adjusted parameters, achieving an accuracy of 100% with a validation accuracy of 89%.

Model

Dropout	Units	Activation	Kernel Regulizer
0.01	160	relu	0.1

Hyperparameter

Optimizer	Learning Rate	Loss Function
AdamW	0.0001	binary_crossentropy

Figure 6. Results of model fitting and hyperparameter tuning

The model will be evaluated to determine its effectiveness in classifying data. This evaluation is performed by calculating the accuracy value and generating a classification report, which includes precision, recall, and F1-scores derived from the confusion matrix. A confusion matrix is a table used to assess the performance of machine learning and deep learning models. It consists of four components: i) true positive (TP): this indicates that the model's prediction and actual value are also correct; ii) true negative (TN): this indicates that the model's prediction is correct while the actual value is incorrect; iii) false positive (FP): this indicates that the model's prediction is incorrect, but the actual value is correct; and iv) false negative (FN): this indicates that the model's prediction and actual value are also incorrect. This framework allows for a comprehensive assessment of the model's performance. Figure 7 shows two leading labels: "true," which indicates the actual values, and "predicted," which represents the results generated by the model. The labels "abnormal" and "normal" also classify the image data. The confusion matrix results show TP value of 29, TN value of 30, FP value of 5, and FN value of 4. We use the accuracy formula shown in (1) to evaluate the accuracy of the model's predictions. By substituting the label values into (1), we obtain an accuracy result of 86%. The model has classified the test data images accurately.

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

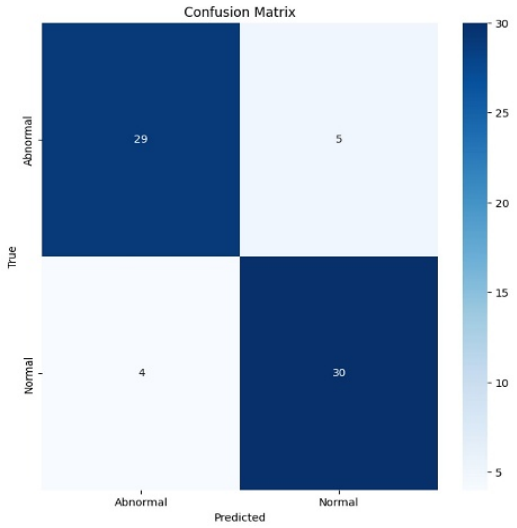


Figure 7. Confusion matrix

Next, we calculate the precision value to determine how many optimistic predictions are accurate from all positive predictions using (2). By substituting the label values in (2), a precision value of 85% is derived. The precision value results indicate that the model can effectively predict positive images using actual values. Next, we calculate the recall value to determine how many positive instances are correctly predicted from all the actual positive instances, utilizing (3).

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

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

By entering the label value in (3), we get a recall value of 87%. The recall value results indicate that the model can predict positive images well from each positive image. We also measure the value of the F1-score, which is a combined measure of precision and recall, to describe the balance between the two using (4). By inputting the label values into (4), we obtain an F1-score of 85%. This result indicates that there is no imbalance issue in the model. Thus, based on the accuracy and classification report values, we can conclude that the CNN using the ResNet-50 architecture performs well, achieving an accuracy of 86%. Therefore, the model is ready for implementation in the Android mobile application.

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \quad (4)$$

Failure analysis is essential in diagnostic applications. Out of the nine identified errors, six were FN (abnormal images incorrectly classified as usual) and three were FP (normal images incorrectly classified as abnormal). Examination of the FN cases revealed that these errors often occurred in images with very small or poorly defined lesions. In contrast, the FP cases frequently displayed visual artifacts, such as uneven lighting or shadow areas, which the model misinterpreted as signs of abnormalities. This analysis highlights the need for future enhancements to improve the model's reliability in complex clinical scenarios. The analysis of the FN cases indicated that these errors often occurred in images with very small or poorly defined lesions. On the other hand, the FP cases frequently featured visual artifacts, such as uneven lighting or shadowed areas, which the model misinterpreted as signs of abnormalities. This analysis identifies key areas for future improvement, aiming to enhance the model's reliability in complex clinical scenarios.

The performance of the model implemented in the mobile app was assessed using device-relevant metrics. The model's inference time (latency) on a mid-range Android device was 120 ms, and its additional memory usage was 45 MB. This performance allows the app to deliver real-time diagnoses at the point of care. The implementation using TensorFlow lite ensures that the model can function effectively on resource-constrained devices without requiring an internet connection, making it an ideal solution for regions with inadequate infrastructure. We performed model performance testing on the mobile application we developed. This testing was conducted manually by inputting all test data individually and classifying each entry. Subsequently, we measured the accuracy using (1). The results of each test case can be found in Table 2.

Table 2. Results of model testing on mobile applications

No	Image name	Exact value	Expected value	Result	No	Image name	Exact value	Expected value	Result
1	AACY	Abnormal	Abnormal	True	35	ADI	Normal	Normal	True
2	AACW	Abnormal	Abnormal	True	36	ADL	Normal	Abnormal	False
3	AADD	Abnormal	Normal	False	37	ADP	Normal	Normal	True
4	AACZ	Abnormal	Normal	False	38	ADU	Normal	Normal	True
5	AADB	Abnormal	Abnormal	True	39	ADV	Normal	Normal	True
6	AADI	Abnormal	Abnormal	True	40	ADW	Normal	Normal	True
7	AADK	Abnormal	Normal	False	41	AEA	Normal	Normal	True
8	AADL	Abnormal	Normal	False	42	AEI	Normal	Normal	True
9	AADJ	Abnormal	Abnormal	True	43	AEG	Normal	Normal	True
10	AADW	Abnormal	Abnormal	True	44	AEV	Normal	Normal	True
11	AADM	Abnormal	Abnormal	True	45	AEW	Normal	Normal	True
12	AADX	Abnormal	Abnormal	True	46	AFC	Normal	Normal	True
13	AAEU	Abnormal	Abnormal	True	47	AFE	Normal	Normal	True
14	AAEJ	Abnormal	Abnormal	True	48	AFI	Normal	Normal	True
15	AAEL	Abnormal	Abnormal	True	49	AFH	Normal	Normal	True
16	AAFR	Abnormal	Abnormal	True	50	AFK	Normal	Abnormal	True
17	AAGY	Abnormal	Abnormal	True	51	AGA	Normal	Normal	True
18	AAEM	Abnormal	Abnormal	True	52	AFJ	Normal	Normal	True
19	AAEZ	Abnormal	Abnormal	True	53	AFL	Normal	Normal	True
20	AAHN	Abnormal	Abnormal	True	54	AHA	Normal	Normal	True
21	AAHO	Abnormal	Abnormal	True	55	AHD	Normal	Normal	True
22	AAFS	Abnormal	Abnormal	True	56	AHQ	Normal	Normal	True
23	AAHP	Abnormal	Abnormal	True	57	AHR	Normal	Normal	True
24	AAHS	Abnormal	Abnormal	True	58	AHT	Normal	Normal	True
25	AAIW	Abnormal	Abnormal	True	59	AHS	Normal	Normal	True
26	AAHQ	Abnormal	Normal	False	60	AIB	Normal	Normal	True
27	AAJA	Abnormal	Abnormal	True	61	AIF	Normal	Normal	True
28	AAHR	Abnormal	Abnormal	True	62	AJL	Normal	Abnormal	False
29	AAHT	Abnormal	Abnormal	True	63	AID	Normal	Normal	True
30	AAHU	Abnormal	Abnormal	True	64	AIH	Normal	Normal	True
31	AAJB	Abnormal	Abnormal	True	65	AIO	Normal	Normal	True
32	AAJI	Abnormal	Abnormal	True	66	AJC	Normal	Normal	True
33	AAJK	Abnormal	Abnormal	True	67	AJG	Normal	Normal	True
34	AAJJ	Abnormal	Normal	False	68	AJE	Normal	Normal	True

Table 2 shows 68 test samples with two parameters: the actual value and the predicted value. The exact value indicates the original classification of each image, while the expected value reflects the model's classification result. In this table, we observe nine errors in the model's predictions. These errors include six instances where the actual class is "abnormal." Still, the model incorrectly predicted them as "normal," and three cases where the actual class is "normal," but the model predicted them as "abnormal." Using the data from these test samples, we calculated the accuracy using (1), resulting in an accuracy rate of 86%. This accuracy is consistent with the model's performance before its implementation in the mobile application. These results demonstrate that the CNN model developed can function effectively using TensorFlow lite. While initial testing shows strong technical performance, real-world adoption relies significantly on usability and acceptance by medical professionals. Therefore, we propose conducting an initial usability study to evaluate the app's effectiveness, efficiency, and user satisfaction. This study could employ methods such as the think-aloud protocol to identify user interaction issues and gather valuable qualitative feedback directly.

4. CONCLUSION

Deep learning models are generally less effective when applied to small datasets. This is primarily due to their tendency to overfit, meaning they perform well only on data the model has already encountered. Data augmentation techniques can be employed to mitigate overfitting in deep learning models with small datasets. This approach is supported by findings from this study, which demonstrate that the accuracy of validation and test datasets improved to nearly match that of the training dataset. When developing a deep learning model, applying hyperparameter tuning is crucial for selecting the optimal configuration for the dense layer and other hyperparameters to achieve maximum accuracy. In this study, TensorFlow lite was identified as the best choice for designing Android mobile applications, as evidenced by the consistent accuracy of the model's results on test data. The model's ability to operate on-device using TensorFlow lite makes it an up-and-coming solution for deployment in rural or resource-constrained areas. By eliminating the need for internet connectivity and external servers, this application can serve as an effective screening tool in places where it is most needed.

ACKNOWLEDGMENTS

We would like to thank the Ministry of Higher Education, Research, and Technology of the Republic of Indonesia for the opportunity to obtain an Applied Research Grant for this research.

FUNDING INFORMATION

This research was funded by the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia through an applied research scheme grant (decision letter no. 054/E5/PG.02.00.PL/2024) and contract no. 20.78/UN23.35.5/PT.01.00/VI/2024.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Retno Supriyanti	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓		✓
Arsil Kultura Anzil		✓	✓			✓	✓		✓					
Yogi Ramadhani	✓			✓				✓	✓					✓
Suroso	✓			✓					✓				✓	✓
Wahyu Widanarto	✓			✓					✓	✓	✓	✓	✓	
Muhammad Alqaaf	✓	✓								✓	✓			
Kartika Dwi Hapsari		✓		✓						✓		✓		
Futiat Diana Kartika		✓		✓						✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

The research related to human use has been complied with all the relevant national regulations and institutional policies in accordance with the tenets of the Helsinki Declaration and has been approved by the authors' institutional review board or equivalent committee.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [RS], upon reasonable request.

REFERENCES

[1] A. Mathis, U. D. Smith, V. Crowther, T. Lee, and S. Suther, "An epidemiological study of cervical cancer trends among women with human immunodeficiency virus," *Healthcare*, vol. 12, no. 12, 2024, doi: 10.3390/healthcare12121178.

[2] V. Chandran *et al.*, "Diagnosis of cervical cancer based on ensemble deep learning network using colposcopy images," *BioMed Research International*, vol. 2021, no. 1, 2021, doi: 10.1155/2021/5584004.

[3] Y. A. Ziyad *et al.*, "Determinants of cervical cancer screening among female health professionals in Harar Town, Eastern Ethiopia: a cross-sectional study," *Infectious Diseases in Obstetrics and Gynecology*, vol. 2024, no. 1, 2024, doi: 10.1155/2024/1430978.

[4] R. Kabanda, A. Kiconco, A. Ronald, K. M. M. Beyer, and S. A. John, "Correction: correlates of intention to screen for cervical cancer among adult women in Kyotera District, Central Uganda: a community based cross-sectional study," *BMC Women's Health*, vol. 24, no. 1, 2024, doi: 10.1186/s12905-024-03155-3.

[5] Y. Nartey *et al.*, "Human papillomavirus genotype distribution among women with and without cervical cancer: implication for vaccination and screening in Ghana," *PLOS ONE*, vol. 18, no. 1, 2023, doi: 10.1371/journal.pone.0280437.

[6] T. G. Gebremeskel and M. Z. Gebretatios, "Determinants of cervical cancer screening utilization among HIV-positive women, in public general hospitals of Central Zone, Tigray, Ethiopia, 2020: case-control study," *PLOS ONE*, vol. 18, no. 12, 2023, doi: 10.1371/journal.pone.0289042.

[7] F. H. Alkhamis, Z. A. S. Alabbas, J. E. Al Mulhim, F. F. Alabdulmohsin, M. H. Alshaqaqiq, and E. A. Alali, "Prevalence and predictive factors of cervical cancer screening in Saudi Arabia: a nationwide study," *Cureus*, vol. 358, no. 11, 2023, doi: 10.7759/cureus.49331.

[8] D. A. Beyene, S. G. Ayele, H. D. Wubneh, and A. W. Tsige, "Male support for cervical cancer screening in Debre Berhan City Ethiopia a community based cross sectional survey," *Scientific Reports*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-69439-8.

[9] B. Bai, G. Tuerxun, A. Tuerdi, R. Maimaiti, Y. Sun, and A. Abudukerimu, "Analysis of vaginal flora diversity and study on the role of Porphyromonas asaccharolytica in promoting IL-1β in regulating cervical cancer," *Scientific Reports*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-73146-9.

[10] J. Namutundu *et al.*, "Barriers and facilitators of cervical cancer screening literacy among rural women with HIV attending rural public health facilities in East Central Uganda: a qualitative study using the integrated model of health literacy," *BMC Women's Health*, vol. 24, no. 1, 2024, doi: 10.1186/s12905-024-03340-4.

[11] Z. V. Karena and P. S. Faldu, "A cross-sectional study on knowledge, attitude, and practices related to cervical cancer screening among the nursing staff in a Tertiary Care Hospital in the Western Region of India," *Cureus*, vol. 16, no. 1, 2024, doi: 10.7759/cureus.51566.

[12] J.-Y. Kim *et al.*, "Glyoxalase 1: emerging biomarker and therapeutic target in cervical cancer progression," *PLOS ONE*, vol. 19, no. 6, 2024, doi: 10.1371/journal.pone.0299345.

[13] M. Agénor *et al.*, "Barriers to and opportunities for advancing racial equity in cervical cancer screening in the United States," *BMC Women's Health*, vol. 24, no. 1, 2024, doi: 10.1186/s12905-024-03151-7.

[14] J.-Y. Lei, S.-X. Li, F. Li, H. Li, and Y.-S. Lei, "Zinc oxide nanoparticle regulates the ferroptosis, proliferation, invasion and steaminess of cervical cancer by miR-506-3p/CD164 signaling," *Cancer Nanotechnology*, vol. 13, no. 1, 2022, doi: 10.1186/s12645-022-00134-x.

[15] A. A. A.-Lartey *et al.*, "Effectiveness of a culturally tailored text messaging program for promoting cervical cancer screening in Accra, Ghana: a quasi-experimental trial," *BMC Women's Health*, vol. 24, no. 1, 2024, doi: 10.1186/s12905-023-02867-2.

[16] F. Mantula and Y. Toefy, "Women and health providers' perspectives on male support for cervical cancer screening in Gwanda district, Zimbabwe," *PLOS ONE*, vol. 18, no. 10, 2023, doi: 10.1371/journal.pone.0282931.

[17] M. Alqaaf *et al.*, "Discovering natural products as potential inhibitors of SARS-CoV-2 spike proteins," *Scientific Reports*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-024-83637-4.

[18] M. Alqaaf *et al.*, "Neural network-based analysis of mammography images for identifying breast cancer histological subtypes," in *2024 IEEE International Conference on Future Machine Learning and Data Science (FMLDS)*, Sydney, Australia: IEEE, 2024, pp. 474–480, doi: 10.1109/FMLDS63805.2024.00088.

[19] R. Supriyanti, A. S. Aryanto, M. I. Akbar, E. Sutrisna, and M. Alqaaf, "The effect of features combination on coloscopy images of cervical cancer using the support vector machine method," *IAES International Journal of Artificial Intelligence*, vol. 13, no. 3, pp. 2614–2622, 2024, doi: 10.11591/ijai.v13.i3.pp2614-2622.

[20] R. Supriyanti, S. L. Dzihniza, M. Alqaaf, M. R. Kurniawan, Y. Ramadhani, and H. B. Widodo, "Morphological features of lung white spots based on the Otsu and Phansalkar thresholding method," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 1, pp. 530–539, 2024, doi: 10.11591/ijeecs.v33.i1.pp530-539.
[21] S. Mittal and P. Basu, *Atlas of visual inspection of the cervix with acetic acid for screening, triage, and assessment for treatment*. Lyon, France: International Agency for Research on Cancer, 2020.
[22] P. Basu and R. Sankaranarayanan, *Atlas of colposcopy: principles and practice*. Lyon, France: International Agency for Research on Cancer, 2017.
[23] A. Gholamy, V. Kreinovich, and O. Kosheleva, "Why 70/30 Or 80/20 relation between training and testing sets : a pedagogical explanation," *Technical Report*. The University of Texas at El Paso, El Paso, Texas.
[24] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, "A review of medical image data augmentation techniques for deep learning applications," *Journal of Medical Imaging and Radiation Oncology*, vol. 65, no. 5, pp. 545–563, 2021, doi: 10.1111/1754-9485.13261.
[25] H. Wang, J. Li, H. Wu, E. Hovy, and Y. Sun, "Pre-trained language models and their applications," *Engineering*, vol. 25, pp. 51–65, 2023, doi: 10.1016/j.eng.2022.04.024.

BIOGRAPHIES OF AUTHORS

Retno Supriyanti is a professor at Department of Electrical Engineering, Jenderal Soedirman University, Indonesia. She received her Ph.D. in March 2010 from Nara Institute of Science and Technology, Japan. She also received her M.S. and bachelor's degrees in 2001 and 1998, respectively, from the Department of Electrical Engineering, Universitas Gadjah Mada, Indonesia. Her research interests include image processing, computer vision, pattern recognition, biomedical applications, e-health, telehealth, and telemedicine. She can be contacted at email: retno_supriyanti@unsoed.ac.id.

Arsil Kultura Anzil received his bachelor's degree from the Department of Electrical Engineering at Jenderal Soedirman University, Indonesia. His research interest is embedded systems. He can be contacted at email: arsil.anzil@gmail.com.

Yogi Ramadhani is an academic staff at Department of Electrical Engineering, Jenderal Soedirman University, Indonesia. He received his M.S. degree from Universitas Gadjah Mada, Indonesia, and his bachelor's degree from Jenderal Soedirman University, Indonesia. His research interest including computer network, decision support system, telemedicine, and medical imaging. He can be contacted at email: yogi.ramadhani@unsoed.ac.id.

Suroso received the B.Eng. degree in Department of Electrical Engineering, from Universitas Gadjah Mada, Indonesia in 2001, and the M.Eng. degree in Electrical and Electronics Engineering from Nagaoka University of Technology, Japan in 2008. He was a research student at Department of Electrical Engineering, Tokyo University, Japan from 2005 to 2006. He earned the Ph.D. degree in Department of Energy and Environment Engineering, Nagaoka University of Technology, Japan in 2011. He was a visiting researcher at Department of Electrical and Electronics Engineering, Shizuoka University, Japan from 2009 to 2011. Currently, he is a Professor at Department of Electrical Engineering, Jenderal Soedirman University, Purwokerto, Jawa Tengah, Indonesia. His research interest includes static power converters, and its application in renewable energy conversion system. He can be contacted at email: suroso.te@unsoed.ac.id.

Wahyu Widanarto completed his undergraduate degree majoring in Physics at Brawijaya University, his master degree majoring in Physics at the Institute of Technology Bandung, and his Ph.D. at the Universitaet der Bundeswehr Munich Germany in the field of Institut fuer Physik, Mikrosystemtechnik. His current research is on the development of carbon materials for sensors. He can be contacted at email: wahyu.widanarto@unsoed.ac.id.

Muhammad Alqaaf received his bachelor degree from Department of Electrical Engineering, Jenderal Soedirman University, Indonesia; M.S. degree in Nara Institute of Science and Technology, Japan. Currently, he is a doctor student in Nara Institute of Science and Technology, Japan. His research interest are image processing field and bioinformatics. He can be contacted at email: muhammad.alqaaf_subandoko.mb5@is.naist.jp.

Kartika Dwi Hapsari received her bachelor degree from the Department of Midwifery, Yogyakarta Ministry of Health Polytechnic, in 2017. She works as a civil servant expert midwife at the Kalimanah Public Health Center in Purbalingga Regency, Central Java, Indonesia. She can be contacted at email: tika.micu@gmail.com.

Futiat Diana Kartika received her medical doctor degree from the Faculty of Medicine, Universitas Islam Indonesia, Yogyakarta, Indonesia. She is currently serving as a general practitioner and civil servant doctor at a primary health care facility in Purbalingga Regency, Central Java, Indonesia. She can be contacted at email: futiatdk@gmail.com.