



# Deep Learning-based Ensemble Approach for Conventional Pap Smear Image Classification

Paisit Khanarsa<sup>1</sup> and Satanat Kitsiranuwat<sup>2</sup>

## ABSTRACT

Cervical cancer screening allows the early signs of precancerous abnormalities in the cervix before they develop into invasive cancer. The Pap Smear is a widely used screening for early detection and prevention of cervical cancer. In many remote areas, the number of cytologists available to interpret pap smear screening tests is insufficient. This lack of personnel makes the test interpretation very time-consuming. To address this, deep learning techniques have been employed to detect cervical cancer cells and support cytologists. Therefore, an integrative approach with deep learning models and the ensemble techniques such as the maximum occurrence and the maximum probability score of cervical cells was proposed. The multi-cell assessment of the Pap smear slide allowed aggregate predictions of single cervical cell images using the proposed method. The classification results between pre-trained deep learning models and the proposed method were compared. In the experimental results, the proposed method can achieve an accuracy score of more than 97%, while the best pre-trained deep learning model can attain an accuracy score of more than 85%. Hence, the proposed method may have the potential to assist physicians or cytologists in the classification of cervical cell types for Pap Smear images.

## Article information:

**Keywords:** Cervical Cancer, Conventional Pap Smear, Deep Learning Models, Ensemble Techniques, Image Classification

## Article history:

Received: November 1, 2023

Revised: February 2, 2024

Accepted: February 8, 2024

Published: February 10, 2024

(Online)

**DOI:** 10.37936/ecti-cit.2024181.254621

## 1. INTRODUCTION

Cervical cancer is one of the most prevalent cancers among women globally, in which cells develop abnormally in the cervix [1, 2]. According to the World Health Organization (WHO), the global burden of cervical cancer remains considerable with an estimated 604,000 women diagnosed with the disease and approximately 342,000 fatalities from the disease recorded in 2020. More than 90% of new cervical cancer cases and deaths occurred in low- and middle-income countries due to a lack of screening and treatment resources [3]. Cervical cytopathology techniques such as conventional and liquid-based cytology are the most widely used screening methods to identify cervical cancer because of their cost-effectiveness [4, 5].

According to Thailand's Department of Medical Services under the Ministry of Public Health, approximately 5,320 new cervical cancer cases were reported in 2020 [6]. Phaliwong et al. [7] determined the prevalence of abnormal cervical cytology in screening methods such as conventional and liquid-based cytology.

Their findings showed that the prevalence of abnormal Pap Smears from conventional and liquid-based cytology was 4.8% and 5.7%, respectively. Conventional cytology remains the primary issue with specimen adequacy, while liquid-based cytology serves as an alternative method for improving specimen adequacy. However, some hospitals in Thailand continue to utilize conventional cytology screening because of its comparatively lower cost [7]. Fortunately, cervical cancer has a long-term precancerous stage, allowing for timely detection and treatment through cytology screening; moreover, the disease can be treated effectively when it is detected early. The diagnosis of cervical cancer based on cytology samples is time-consuming and subject to human error [8]. Computer-assisted analysis such as deep learning models can be employed to detect cervical cancer cells in order to prevent the issue.

<sup>1</sup> The author is with the Institute of Field Robotics, King Mongkut's University of Technology Thonburi, Bangkok, 10140, Thailand., E-mail: paisit.kha@mail.kmutt.ac.th

<sup>2</sup> The author is the Department of Computer Science, Faculty of Science and Technology, Thammasat University, Pathum Thani, 12120, Thailand., E-mail: satanat@tu.ac.th

<sup>2</sup> Corresponding author: satanat@tu.ac.th

In 2020, Khamparia *et al.* [9] proposed a novel Internet of health things (IoHT)-driven deep learning framework for the detection and classification of cervical cancer in Pap smear images using transfer learning. The framework integrated pre-trained convolutional neural network (CNN) models with various conventional machine learning techniques. Feature extraction was performed using CNN models such as InceptionV3, VGG19, SqueezeNet, and ResNet50 for classifying normal and abnormal cervical cells. The proposed framework achieved a high classification rate of 97.89% using the ResNet50 model in combination with the random forest classifier.

In 2021, Chandran *et al.* [10] focused on the diagnosis of cervical cancer using ensemble deep-learning networks based on colposcopy images. The study proposed two deep learning architectures, VGG19 and Colposcopy Ensemble Network (CYENET), for automatically detecting cervical cancer. The CYENET model achieved high scores of sensitivity, specificity, and kappa with more than 92%, 96%, and 88%, respectively. The accuracy score of the CYENET model was 92%, which improved to 19% when compared to the VGG19 model. These findings indicated that the CYENET model had the potential for accurately and efficiently detecting cervical cancer. Moreover, Rahaman *et al.* [2] introduced Deep-Cervix, a deep learning-based framework for precisely categorizing cervical cells, to enhance the manual cervical cancer screening procedure in 2021. They suggested the hybrid deep feature fusion (HDFE) method based on deep learning techniques to detect cervical cells. The classification results showed that the accuracy scores were 99.85%, 99.38%, and 99.14% for 2-class, 3-class, and 5-class when using the SIPaKMeD dataset, while the accuracy scores were 98.32% for 2-class and 90.32% for 7-class when using the Herlev dataset.

In 2022, Alquran *et al.* [11] proposed a novel system based on a combination of machine learning and deep learning approaches to effectively classify cervical cancer on Pap Smear images. Automated features were extracted using ResNet101. Then, the SVM model was employed to distinguish among levels of normal cases. The results showed that the novel system achieved high performance with an accuracy score of approximately 92%.

The objective of this research is to classify the conventional Pap Smear images into five categories: Superficial-Intermediate, Parabasal, Koilocytotic, Metaplastic, and Dyskeratotic. The cervical single-cell images were extracted from a cervical multi-cell image or Pap Smear slide to enhance the classification performance when using pre-trained deep-learning models. The pre-trained deep learning models such as VGG19, ResNet152V2, NASNet-Large, and EfficientNetB7 were employed for cervical cell classification. However, these classification

results were reported for cervical single-cell, not cervical multi-cell. Therefore, ensemble techniques such as the maximum occurrence of cervical cells and the maximum probability score of cervical cells were proposed to classify the category of a multi-cell image based on the classification predictions of all extracted single-cell images. After that, the performance results of pre-trained deep learning models when using with and without ensemble techniques were evaluated and compared to the efficiency of cervical multi-cell image classifications. The remaining parts of the paper are organized as follows. Section II describes the proposed techniques such as the maximum occurrence and the maximum probability score of cervical cells. The results of comparing the classification performance among multi-cell classification without ensemble techniques, with the maximum occurrence of the cervical cells technique, and the maximum probability score of the cervical cells technique are presented in Section III. Section IV describes the efficiency and the limitation of the ensemble techniques. The last section is the conclusion and the future work.

## 2. MATERIAL AND METHOD

The overview of this work is illustrated in Fig. 1. First, the conventional Pap Smear images were collected from a public dataset. Then, the data preprocessing was performed to extract cervical single-cell images. After that, the proposed ensemble techniques such as maximum occurrence of cervical cells and maximum probability score of cervical cells were operated to refer to the classification prediction of a cervical multi-cell image based on the classification prediction of all cervical single-cell images. Finally, the classification results were compared between using pre-trained deep learning models with and without ensemble techniques.

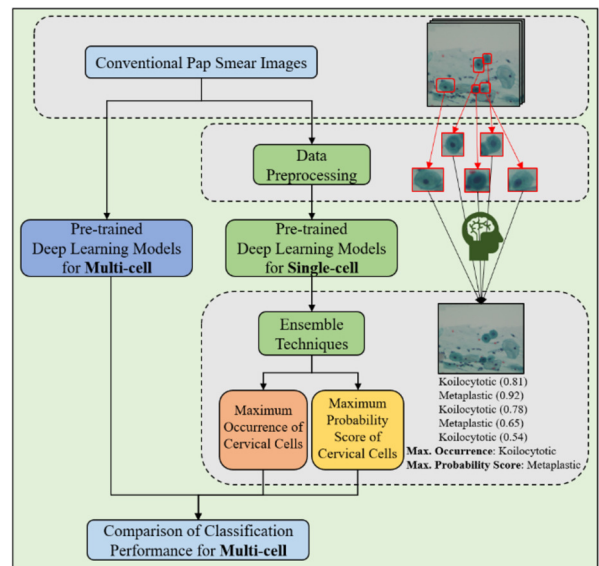


Fig.1: The Overview of this Work.

## 2.1 Dataset and Data Preparation

The conventional Pap Smear images were retrieved from the SiPakMed [12] database, which is publicly available via <https://www.cs.uoi.gr/~marina/sipakmed.html>.

The SiPakMed database consists of 966 cervical cell images. The cervical cells were classified by cytopathologists into five classes. They are 126 Superficial-Intermediate cells, 108 Parabasal cells, 238 Koilocytotic cells, 271 Metaplastic cells, and 223 Dyskeratotic cells. In particular, the category of normal cervical cells contains Superficial-Intermediate and Parabasal cells. On the other hand, Koilocytotic and Dyskeratotic are categorized as abnormal cells. Moreover, the Metaplastic cells are recognized as benign cells that develop in most cervical precancerous and cancerous diseases.

To prepare groups of cervical multi-cell images, all 966 cervical cell images were randomly split into three groups (64% for the training set, 16% for the validating set, and 20% for the test set) according to the proportional number in each class. E.g., in Dyskeratotic cells, there are 142 cells for the training set, 36 cells for the validating set, and 45 cells for the test set. In Parabasal cells, there are 69 cells for the training set, 17 cells for the validating set, and 22 cells for the test set.

For data pre-processing, single-cell images were manually cropped from the multi-cell image by expert observations. Each cervical single-cell image contains the boundary of the cytoplasm and nucleus to represent the cervical cell type. All cervical single-cell images were retrieved from the SiPakMed database [12]. There are 831, 787, 825, 793, and 813 single-cell images for Superficial-Intermediate, Parabasal, Koilocytotic, Metaplastic, and Dyskeratotic, respectively.

## 2.2 Pre-trained Deep Learning Models

The pre-trained deep learning models were employed for predicting cervical cell types of multi-cell and single-cell images. Transfer learning was utilized to enhance a classifier by leveraging knowledge from a pre-trained CNN. Transfer learning can be utilized for the feature extractor and the fine-tuning of the CNN [13, 14]. Deep learning has significantly transformed the fields of computer vision and image recognition.

In this study, pre-trained deep learning models such as VGG19, ResNet152V2, NASNetLarge, and EfficientNetB7 were exploited for cervical cell classifications. VGG19 model utilizes a deep architecture with convolutional and fully connected layers to capture intricate visual features [15]. ResNet152V2 addresses the vanishing gradient problem using residual connections to train deep networks effectively [16]. NASNetLarge employs neural architecture search techniques to automatically discover optimized network architectures [17]. EfficientNetB7 achieves a

balance between accuracy and computational efficiency through compound scaling [18].

Moreover, the parameters of the maximum number of epochs, optimizer, and learning rate for all deep learning models were set to 100, Adam, and 0.001, respectively. However, the number of epochs varied for each model to prevent unnecessary training epochs. Then, the training will stop if the validation loss does not improve. Moreover, the learning rate (0.001) is the default value for Adam in Keras. Furthermore, the global average pooling was assigned to represent the classifier layer for ResNet152V2, the NASNetLarge, and the EfficientNetB7 models. On the other hand, the VGG19 model consists of two fully connected layers.

## 2.3 Ensemble Techniques

To aggregate the classification prediction of single-cell images into the classification prediction of multi-cell images, ensemble techniques such as the maximum occurrence of cervical cells and the maximum probability score of cervical cells were proposed, as follows.

### 2.3.1 Maximum Occurrence of Cervical Cells

The maximum occurrence of cervical cells is defined by the most frequently predicted cervical cell types of all predicted single-cell images. The classification prediction of the cervical multi-cell image  $i^{th}$  can be computed based on the maximum occurrence of single-cell images or  $MO_i$ , as shown in (1):

$$MO_i = mode([SC_1, SC_2, \dots, SC_n]), \quad (1)$$

where  $SC_j$  is the classification prediction of cervical single-cell image  $j^{th}$  such that  $1 \leq j \leq n$  and  $n$  is the total number of extracted single-cell images from multi-cell image  $i^{th}$ . In statistical analysis, the  $mode()$  function was employed to calculate the maximum predicted occurrence of cervical cells. Moreover, if the numbers of predicted classes were equal,  $MO_i$  was assigned in the following order: Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-intermediate cells. This ordering depended on the level of the cervical cells (i.e. normal, benign, or abnormal cells).

### 2.3.2 Maximum Probability Score of Cervical Cells

The maximum probability score of cervical cells is defined by the highest predicted score of cervical single-cell images. The classification prediction of cervical multi-cell image  $i^{th}$  can be calculated based on the maximum probability score of single-cell images or  $MP_i$ , as shown in (2):

$$MP_i = SC[\argmax(S_1, S_2, \dots, S_n)], \quad (2)$$

where  $S_j$  is the predicted score of cervical single-cell image  $j^{th}$  such that  $1 \leq j \leq n$  and  $n$  is the total

number of extracted single-cell images from multi-cell image  $i^{th}$ . Then, the *argmax* function was utilized to find the index of the single-cell image that gives the maximum predicted score. After that, the output of the *argmax* function was assigned as the argument for  $SC = [SC_1, SC_2, \dots, SC_n]$ . More precisely, the  $SC_k$  is the classification prediction of cervical single-cell image  $k^{th}$ .

E.g., there are five single-cell images in which classification results show three Koilocytotic cells with predicted scores of 0.81, 0.78, and 0.54, respectively, and two Metaplastic cells with predicted scores of 0.92 and 0.65, respectively, as shown in Fig. 1. Then, the maximum occurrence of cervical cells is Koilocytotic with 3 of 5 images. Therefore, the classification prediction for the multi-cell image is Koilocytotic cell when using the maximum occurrence of cervical cells technique. On the other hand, the classification prediction for the multi-cell image is Metaplastic when using the maximum probability score of cervical cells technique. Because the Metaplastic cell has a maximum probability score of 0.92.

## 2.4 Evaluating Performance

Performance scores such as precision, recall, F1-score, and accuracy scores were employed in evaluating the classification results. Precision refers to the proportion of correctly classified instances within the total number of instances predicted as positive. Recall, on the other hand, measures the proportion of correctly classified instances out of all the actual positive instances. The F1-score can be calculated by the harmonic mean of precision and recall scores. Accuracy represents the overall correctness of the model's predictions by calculating the proportion of correctly classified instances out of the total number of instances.

To represent crucial region-based color visualization, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed in the final convolutional feature map for deep-learning models. The Grad-CAM technique utilizes a label's gradient information to identify essential regions for classifying images [19]. The areas of essential features for categorizing images are represented by dark tones in a Grad-CAM output.

## 3. RESULTS

The performance results such as multi-cell classification without ensemble techniques, single-cell classification without ensemble techniques, multi-cell classification with the maximum occurrence of cervical cells technique, and multi-cell classification with the maximum probability score of cervical cells technique are demonstrated in this section.

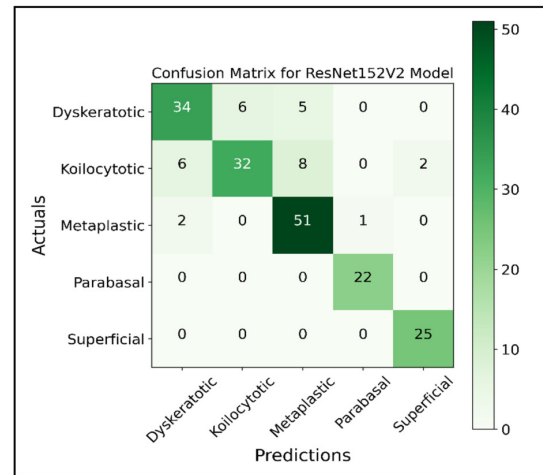
### 3.1 Multi-cell Classification without Ensemble Techniques

For cervical multi-cell classification, the performance results of pre-trained deep learning models and the confusion matrix of the best predicted-model are presented in Table 1 and Fig. 2, respectively.

**Table 1:** Performance for Predicting Multi-cell.

Deep Learning Models	Performance Scores			
	Precision	Recall	F1	Accuracy
VGG19	0.82	0.81	0.81	0.81
ResNet152V2	0.85	0.85	0.84	0.85
NASNet Large	0.82	0.77	0.77	0.77
EfficientNetB7	0.82	0.82	0.82	0.82

Table 1 shows that the performance scores such as precision, recall, F1, and accuracy of predicting cervical multi-cell images are more than 0.81, 0.84, 0.77, and 0.82 for VGG19, ResNet152V2, NASNet-Large, and EfficientNetB7 models, respectively. The ResNet152V2 model outperformed the other models in predicting multi-cell images, with all performance scores above 84%. Because this was the best prediction model, the confusion matrix for ResNet152V2 is illustrated in Fig. 2.

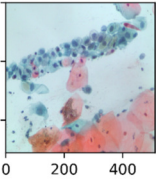
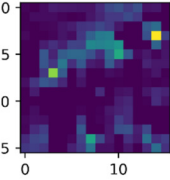
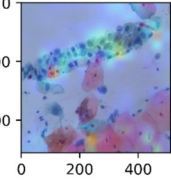
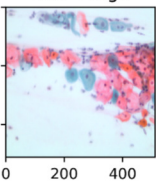
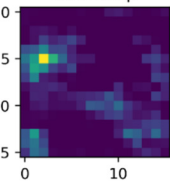
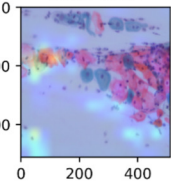
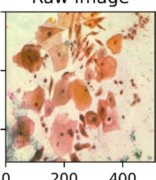
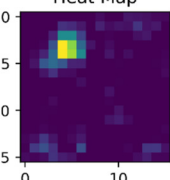
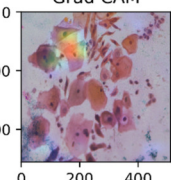


**Fig.2:** The confusion matrix for ResNet152V2 to predict multi-cell.

Fig. 2 shows that the ResNet152V2 model has an ability to classify the positive instances which are actually correct for Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate cells with the precision scores of 0.81 (34/42), 0.84 (32/38), 0.80 (51/64), 0.96 (22/23), and 0.93 (25/27), respectively. Moreover, the ResNet152V2 model has the ability to find all the positive samples from actual positives, such as Parabasal and Superficial-Intermediate cells. On the other hand, the



**Table 2:** Grad-CAM Visualization for Predicting Multi-cell using the ResNet152V2 Model.

Image Name	Grad-CAM Visualization			Predicted Class (Predicted Score)	Actual Class
2_129.bmp				Dyskeratotic (0.5936)	Koilocytotic
3_145.bmp				Dyskeratotic (0.5161)	Metaplastic
4_150.bmp				Koilocytotic (0.4351)	Dyskeratotic

ResNet152V2 model can find all the positive samples from actual positives for Dyskeratotic, Koilocytotic, and Metaplastic cells with recall scores of 0.76 (34/45), 0.67 (32/48), and 0.94 (51/54), respectively. The results indicate that the prediction of cervical multi-cell classes such as Dyskeratotic, Koilocytotic, and Metaplastic cells is characterized by confusion. Then, the cervical multi-cell images that were classified into false-positive groups (Image names: 2\_129.bmp, 3\_145.bmp, and 4\_150.bmp) are shown in Table 2.

Table 2 contains four columns which are image name, Grad-CAM visualization, predicted class, and actual class. The Grad-CAM visualization is composed of a raw image of cervical cells, a heat map which highlight the important areas for the model's decision, and Grad-CAM which depict the process of superimposing the heat map onto the raw image. The actual class represents the ground truth class label of the cervical multi-cell image, while the predicted class refers to the class label that assigned by the ResNet152V2 model. Regions of warmer colours (red or yellow) represent significant areas that the ResNet152V2 model for the classification of multi-cell images.

### 3.2 Single-cell Classification without Ensemble Techniques

This part presents the classification results of cervical single-cell images, which were manually cropped from multi-cell images by expert observation. For cervical single-cell classification, the performance results

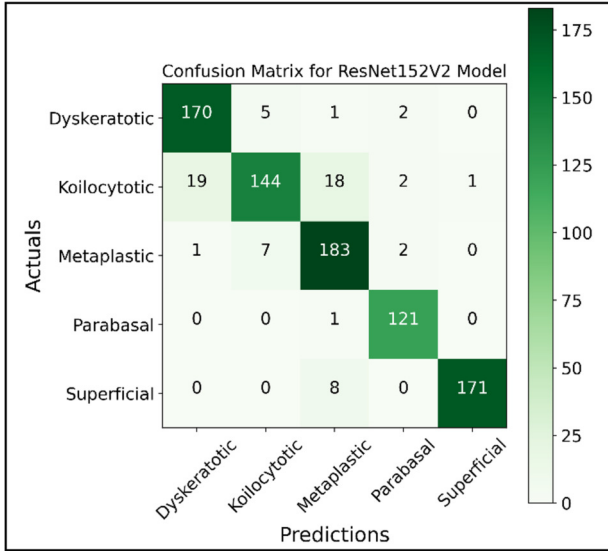
of pre-trained deep learning models and the confusion matrix of the best-predicted model are presented in Table 3 and Fig. 3, respectively.

**Table 3:** Performance for Predicting Single-cell.

Deep Learning Models	Performance Scores			
	Precision	Recall	F1	Accuracy
VGG19	0.87	0.85	0.84	0.85
ResNet152V2	0.92	0.92	0.92	0.92
NASNet Large	0.86	0.86	0.86	0.86
EfficientNetB7	0.87	0.86	0.87	0.86

Table 3 shows that the performance scores such as precision, recall, F1, and accuracy of predicting cervical single-cell images are more than 0.84, 0.92, 0.86, and 0.86 for VGG19, ResNet152V2, NASNet-Large, and EfficientNetB7 models, respectively. Interestingly, the classification performance results of the predicting single-cell are better than the predicting multi-cell for VGG19, ResNet152V2, NASNet-Large, and EfficientNetB7 models. Moreover, the ResNet152V2 model outperforms predicting cervical single-cell images with more than 92% of performance scores. Therefore, the confusion matrix for the best-predicted model, the ResNet152V2 model, is illustrated in Fig. 3.

Fig. 3 shows that the ResNet152V2 model has an ability to classify the positive instances which are actually correct for Dyskeratotic, Koilocytotic,



**Fig.3:** The confusion matrix for ResNet152V2 to predict single-cell.

Metaplastic, Parabasal, and Superficial-Intermediate cells with the precision scores of 0.89 (170/190), 0.92 (144/156), 0.87 (183/211), 0.95 (121/127), and 0.99 (171/172), respectively. On the other hand, the ResNet152V2 model can find all the positive samples from actual positives for Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate cells with recall scores of 0.96 (170/178), 0.78 (144/184), 0.95 (183/193), 0.99 (121/122), and

0.96 (171/179), respectively.

In Addition, the cervical single-cell images that extracted from cervical multi-cell images (Image names: 2\_129.bmp, 3\_145.bmp, and 4\_150.bmp) are shown in Table 4. The 2\_129.bmp, 3\_145.bmp, and 4\_150.bmp images were selected to compare with Grad-CAM visualization for multi-cell classifications. The significant areas that the ResNet152V2 model used to classify single-cell images are represented as regions of warmer colours (red or yellow).

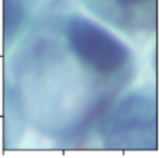
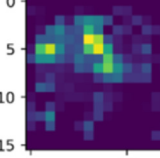
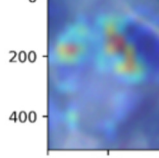
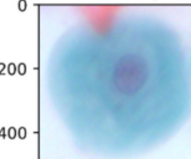
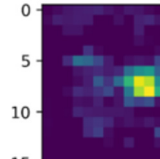
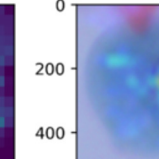
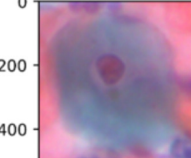
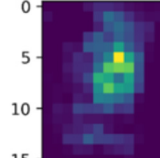
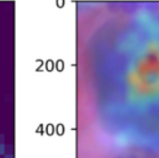
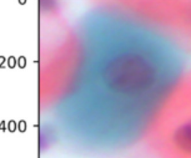
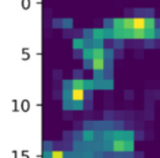
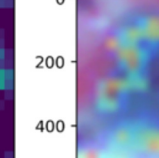
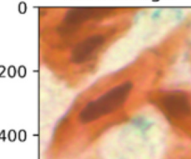
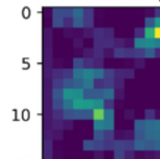
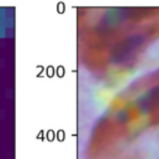
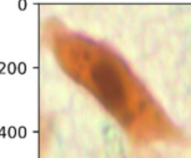
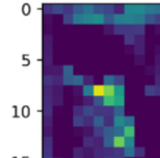
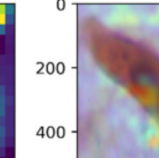
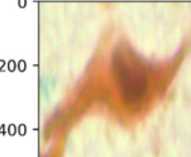
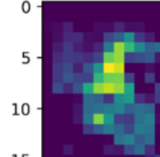
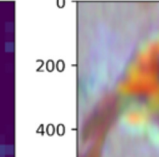
### 3.3 Multi-cell Classification using the Maximum Occurrence of Cervical Cells

The maximum occurrence of cervical cells was proposed to aggregate the cervical single-cell classifications for identifying the predictions of cervical multi-cell classifications. The performance results of deep learning models for predicting cervical multi-cell based on the maximum occurrence of cervical cells and the confusion matrix of the best-predicted model are described in Table 5 and Fig. 4, respectively.

Table 5 shows that the performance scores such as precision, recall, F1, and accuracy of predicting cervical multi-cell images based on the maximum occurrence technique are more than 0.88, 0.95, 0.89, and 0.90 for VGG19, ResNet152V2, NASNetLarge, and EfficientNetB7 models, respectively. Remarkably, the classification performance results of predicting multi-cell with the maximum occurrence of cervical cells are better than those of predicting multi-cell for

**Table 4:** Grad-CAM Visualization for Predicting Single-cell using the ResNet152V2 Model.

Image Name	Grad-CAM Visualization			Predicted Class (Predicted Score)	Actual Class
2_129.bmp	Raw Image	Heat Map	Grad CAM	Metaplastic (0.6404)	Koilocytotic
	Raw Image	Heat Map	Grad CAM	Metaplastic (0.3677)	Koilocytotic
	Raw Image	Heat Map	Grad CAM	Metaplastic (0.4151)	Koilocytotic

Image Name	Grad-CAM Visualization			Predicted Class (Predicted Score)	Actual Class
	  			Metaplastic (0.7022)	Koilocytotic
3_145.bmp	  			Metaplastic (0.9544)	Metaplastic
	  			Metaplastic (0.9784)	Metaplastic
	  			Parabasal (0.7943)	Metaplastic
4_150.bmp	  			Koilocytotic (0.8391)	Dyskeratotic
	  			Koilocytotic (0.5522)	Dyskeratotic
	  			Dyskeratotic (0.8393)	Dyskeratotic

**Table 5:** Performance for Predicting Multi-cell when using the Maximum Occurrence.

Deep Learning Models	Performance Scores			
	Precision	Recall	F1	Accuracy
VGG19	0.90	0.89	0.88	0.89
ResNet152V2	0.95	0.95	0.95	0.95
NASNetLarge	0.89	0.89	0.89	0.89
EfficientNetB7	0.90	0.90	0.90	0.90

VGG19, ResNet152V2, NASNetLarge, and EfficientNetB7 models. Then, the ResNet152V2 model outperforms predicting cervical multi-cell images with more than 95% of performance scores when using the maximum occurrence technique. Therefore, the confusion matrix for the best-predicted model, the ResNet152V2 model, is illustrated in Fig. 4.

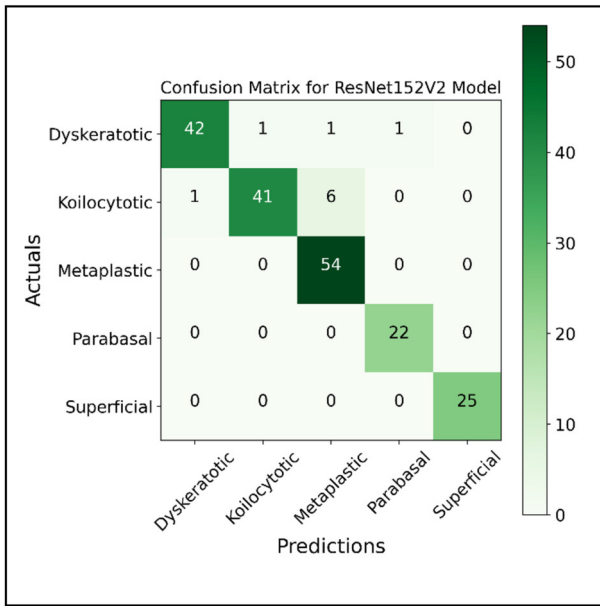
**Fig.4:** The confusion matrix for ResNet152V2 to predict multi-cell using the maximum occurrence.

Fig. 4 shows that the ResNet152V2 model has the ability to find all the positive samples from actual positives, such as Metaplastic, Parabasal, and Superficial-Intermediate cells. On the other hand, the ResNet152V2 model can detect all positive classes from actual positives for Dyskeratotic and Koilocytotic with recall scores of 0.93 (42/45) and 0.85 (41/48) respectively. Moreover, the ResNet152V2 model has the ability to classify the positive instances which are actually correct for Dyskeratotic, Koilocytotic, Metaplastic, and Parabasal cells with precision scores of 0.98 (42/43), 0.98 (41/42), 0.89 (54/61), and 0.96 (22/23), respectively. Furthermore, the ResNet152V2 model has the ability to classify

the positive instances which are actually correct for Superficial-Intermediate.

### 3.4 Multi-cell Classification using the Maximum Probability Score of Cervical Cells

The maximum probability score of predicting cervical cells technique was proposed to aggregate the cervical single-cell classifications for identifying the predictions of cervical multi-cell classifications. The performance results of deep learning models for predicting cervical multi-cells based on the maximum probability score of cervical cells and the confusion matrix of the best-predicted model are described in Table 6 and Fig. 5, respectively.

**Table 6:** Performance for Predicting Multi-cell when using the Maximum Probability Score.

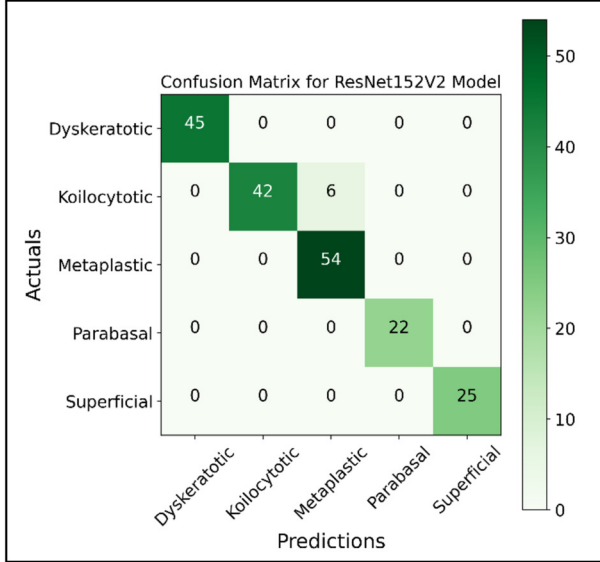
Deep Learning Models	Performance Scores			
	Precision	Recall	F1	Accuracy
VGG19	0.92	0.90	0.90	0.90
ResNet152V2	0.97	0.97	0.97	0.97
NASNetLarge	0.91	0.91	0.90	0.91
EfficientNetB7	0.93	0.92	0.92	0.92

Table 6 shows that the performance scores such as precision, recall, F1, and accuracy of predicting cervical multi-cell images based on the maximum probability score technique are more than 0.90, 0.97, 0.90, and 0.92 for VGG19, ResNet152V2, NASNetLarge, and EfficientNetB7 models, respectively. Remarkably, the classification performance results of predicting multi-cells with the maximum probability of cervical cells are better than those of predicting multi-cell for VGG19, ResNet152V2, NASNetLarge, and EfficientNetB7 models. Interestingly, the classification performance results for predicting multi-cells using the maximum probability score dominate when compared with the maximum occurrence of cervical cells. As a result, the ResNet152V2 model based on the maximum probability score technique outperforms with more than 97% of performance scores. Therefore, the confusion matrix for the best-predicted model, the ResNet152V2 model, is illustrated in Fig. 5.

Fig. 5 shows that the ResNet152V2 model has the ability to find all the positive samples from actual positives, such as Dyskeratotic, Metaplastic, Parabasal, and Superficial-Intermediate cells. On the other hand, the ResNet152V2 model can detect all positive classes for Koilocytotic with a recall value of 0.88 (42/48). Moreover, the ResNet152V2 model can identify the positive predictions that are actually correct for Dyskeratotic, Koilocytotic, Parabasal, and Superficial-Intermediate cells. Furthermore, the



ResNet152V2 model has the ability to identify the positive predictions that are actually correct for Metaplastic with a precision value of 0.90 (54/60). As a consequence, the ResNet152V2 model based on the maximum probability technique has issues predicting the class label of Metaplastic cells, whereas the ground truth class labels correspond to Koilocytotic cells.



**Fig.5:** The confusion matrix for ResNet152V2 to predict multi-cell using the maximum probability score.

#### 4. DISCUSSION

This study proposed the cervical multi-cell classification technique through the analysis of cervical single-cell data. It is based on the integration of deep learning models and ensemble techniques, such as the maximum occurrence of cervical cells and the maximum probability score of cervical cells. This approach classifies multi-cell Pap Smear images into five classes, Superficial-Intermediate, Parabasal, Koilocytotic, Metaplastic, and Dyskeratotic cells. The experimental results showed that the ResNet152V2 model outperforms for both cervical multi-cell and single-cell classifications.

Moreover, the performance scores such as precision, recall, F1, and accuracy obtained from predicting cervical single-cell images surpass those achieved by predicting cervical multi-cell images for all VGG19, ResNet152V2, NASNetLarge, and EfficientNetB7 models, as shown in Table 1 and Table 3. For the best-predicted model (the ResNet152V2), the accuracy scores for predicting cervical multi-cell and single-cell are 0.85 and 0.92, respectively.

To provide support for the assumption regarding the presence of obscuring substances, the Grad-CAM visualization for predicting cervical multi-cell and single-cell using the ResNet152V2 model are dis-

played in Table 2 and Table 4, respectively. In the experimental results, the highlight regions for multi-cell predictions were incorrectly identified for cervical classifications while most of the focus areas of single-cell predictions were correctly diagnosed as cervical cell types. The results of Grad-CAM visualization for single-cell predictions showed that the highlight regions were located around nuclei and cytoplasm. It is possible that deep learning models endeavour to recognize various factors such as locations of the nucleus in the cytoplasm, density, shapes, and stains of cervical cells for classification. The findings are consistent with other studies [20]. Several previous studies proposed multiple crucial factors such as sizes of nucleus, length of perimeters/diameters/brightness of nuclei and cytoplasm, and locations of nucleus in the cytoplasm for promising cervical classification outcomes [20, 21]. Therefore, the extracted multi-cell images into single-cell images could potentially serve as an alternative approach for cervical cell classifications. Subsequently, ensemble methods were proposed to gather the single-cell predictions into the multi-cell predictions.

Furthermore, the classification results of the ensemble techniques such as the maximum occurrence numbers and the maximum probability score of cervical cells were compared. The results showed that the ResNet152V2 model outperforms for both ensemble techniques. The performance scores obtained from deep learning models based on the maximum probability score technique surpass those achieved by deep learning models based on the maximum occurrence technique, as shown in Table 5 and Table 6. For the best-predicted model (the ResNet152V2), the accuracy score for predicting cervical cells based on the maximum occurrence of cervical cells and the maximum probability score techniques are 0.95 and 0.97, respectively. It is possible that the low quality of the images, the overlapping nuclei/cytoplasm of cervical cells, or the presence of obscuring substances contribute to the observed limitations when extracting multi-cell images into single-cell images. Such situations can potentially result in mistaken single-cell image predictions. Consequently, the classification results obtained from the maximum occurrence technique are comparatively lower than the results from the maximum probability technique. Therefore, the classification results for cervical cells based on ensemble techniques are influenced by the characteristics of extracting single-cell images.

In addition, the ResNet152V2 model based on the maximum probability score of cervical cells yields the best classification results for multi-cell images with more than 97% of accuracy. The classification result of image 4.150.bmp showed the example of correct and incorrect predictions when using the maximum probability score and the maximum occurrence of cervical cells, respectively. Therefore, the proposed

method could have the potential to support doctors or specialist nurses in classifying cervical cell types of Pap Smear images. However, the confusion matrix for the ResNet152V2 model based on the maximum probability score technique demonstrates that there are six predictions of Metaplastic cells, whereas the ground truth labels correspond to the Koilocytotic cells, as shown in Fig. 5. This may have been due to the overlapping cytoplasm of cervical cells or the presence of obscuring substances, as shown in Table 4 (Image name: 2\_129.bmp). Therefore, the cervical cells images classification is subject to the influence of image quality. In future studies, image processing techniques would be implemented in order to enhance the quality of cervical single-cell images and improve the classification performance.

Furthermore, the performance of the proposed technique was compared to the state-of-the-art models such as VGG19, ResNet152V2, NASNetLarge, and EfficientNetB7. The best performance scores among these state-of-the-art models (ResNet152V2) achieved an accuracy of 82%, as shown in Table 1. Moreover, the proposed technique, which integrated the maximum probability score, outperformed these models, reaching an accuracy of 97% (Table 6). This difference in performance may be attributed to the presence of obscuring substances such as mucous or blood in the conventional Pap Smear slides or multi-cell images, leading to a relatively low accuracy score for predicting cervical multi-cell images. However, it is noteworthy that the model's performance can be improved by applying multi-cell classification through single-cell analysis.

## 5. CONCLUSION

An integrative deep learning-based ensemble technique was proposed to detect cervical cell types such as Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate cells for conventional Pap Smear image classifications. This study introduces a technique for multi-cell classification through single cervical cell images. The cervical single-cell images were extracted from the multi-cell image to make the classifications. After that, single-cell classification results were aggregated using the maximum occurrence and the maximum probability score techniques. Then, the classification performance results among multi-cell classification, single-cell classification, multi-cell classification with the maximum occurrence, and multi-cell classification with the maximum probability score were compared. The experimental results showed that the best-predicted model (the ResNet152V2) can achieve accuracy scores of 0.85 and 0.92 for predicting cervical multi-cell and single-cell, respectively. Consequently, single cervical cell images have the potential to be used for cervical cell classification. Moreover, the ResNet152V2 model based on the maximum prob-

ability score can achieve superior performance with 0.97 of accuracy. Hence, the proposed method can improve the performance by up to 14.12% of accuracy.

## References

- [1] T. Hanprasertpong, K. Dhanaworavibul, K. Leetanaporn and J. Hanprasertpong, "Cervical Cancer Screening in Elderly Women," *Thai Journal of Obstetrics and Gynaecology*, vol. 30, no. 6, pp. 370-375, 2022.
- [2] M. M. Rahaman *et al.*, "DeepCervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques," *Computers in Biology and Medicine*, vol. 136, p. 104649, 2021.
- [3] H. Sung *et al.*, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209-249, 2021.
- [4] D. Saslow *et al.*, "American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer," *CA: A Cancer Journal for Clinicians*, vol. 62, no. 3, pp. 147-72, 2012.
- [5] E. Davey *et al.*, "Effect of study design and quality on unsatisfactory rates, cytology classifications, and accuracy in liquid-based versus conventional cervical cytology: a systematic review," *The Lancet*, vol. 367, no. 9505, pp. 122-32, 2006.
- [6] J. Rojanamatin *et al.*, "Cancer in Thailand," *Bangkok Thailand: National Cancer Institute Ministry of Public Health*, vol. X 2016-2018, 2021.
- [7] P. Phaliwong, P. Pariyawateekul, N. Khuakoonratt, W. Sirichai, K. Bhamarapratana and K. Suwannarurk, "Cervical Cancer Detection between Conventional and Liquid Based Cervical Cytology: a 6-Year Experience in Northern Bangkok Thailand," *Asian Pacific Journal of Cancer Prevention*, vol. 19, no. 5, pp. 1331-1336, 2018.,
- [8] R. Mehrotra, S. Mishra, M. Singh and M. Singh, "The efficacy of oral brush biopsy with computer-assisted analysis in identifying precancerous and cancerous lesions," *Head & Neck Oncology*, vol. 3, no. 1, p. 39, 2011.
- [9] A. Khamparia, D. Gupta, V. H. C. de Albuquerque, A. K. Sangaiah and R. H. Jhaveri, "Internet of health things-driven deep learning system for detection and classification of cervical cells using transfer learning," *The Journal of Supercomputing*, vol. 76, no. 11, pp. 8590-8608, 2020.

- [10] V. Chandran *et al.*, “Diagnosis of Cervical Cancer based on Ensemble Deep Learning Network using Colposcopy Images,” *BioMed Research International*, vol. 2021, p. 5584004, 2021.
- [11] H. Alquran *et al.*, “Cervical Cancer Classification Using Combined Machine Learning and Deep Learning Approach,” *Computers, Materials & Continua*, vol. 72, no. 3, pp. 5117-5134, 2022.
- [12] M. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni and A. Charchanti, “Sipakmed: A New Dataset for Feature and Image Based Classification of Normal and Pathological Cervical Cells in Pap Smear Images,” *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 3144-3148, 2018.
- [13] J. Yosinski, J. Clune, Y. Bengio and H. Lipson, “How transferable are features in deep neural networks?,” *Advances in Neural Information Processing Systems (NIPS)*, vol. 27, 2014.
- [14] S. J. Pan and Q. Yang, “A Survey on Transfer Learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345-1359, 2010.
- [15] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv:1409.1556v6*, 2014.
- [16] K. He, X. Zhang, S. Ren and J. Sun, “Identity Mappings in Deep Residual Networks,” *Computer Vision – ECCV 2016*, vol. 9908, pp. 630-645, 2016.
- [17] B. Zoph, V. Vasudevan, J. Shlens and Q. V. Le, “Learning Transferable Architectures for Scalable Image Recognition,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8697-8710, 2018.
- [18] M. T. a. Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” *CoRR*, vol. abs/1905.11946, 2019. [Online]. Available: <http://arxiv.org/abs/1905.11946>.
- [19] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336-359, 2020.
- [20] C. Duangate, B. Uyyanonvara and T. Koanantakul, “A Review of Image Analysis and Pattern Classification Techniques for Automatic Pap Smear Screening Process,” *International Conference on Embedded Systems and Intelligent Technology*, 2008.
- [21] J. Jantzen and G. Dounias, “Analysis of Pap-smear Image Data,” in *Proceedings of the Nature-Inspired Smart Information Systems 2nd Annual Symposium*, 2006.



**Paisit Khanarsa** is a researcher who earned his Doctor of Philosophy in Applied Mathematics and Computational Science from Chulalongkorn University in 2020. Currently serving as a lecturer at the Institute of Field Robotics (FIBO) at King Mongkut's University of Technology Thonburi, his current research interests encompass artificial intelligence in medical fields, machine learning, data science, and time series

analysis.



**Satanat Kitsiranuwat** graduated with a doctorate in Bioinformatics and Computational Biology, from the Faculty of Science at Chulalongkorn University, Thailand. She is currently a lecturer at the Department of Computer Science, Faculty of Science and Technology, Thammasat University. Her research interest is in developing computational strategies and algorithms based on machine learning, deep learning, artificial

intelligence for drug repurposing, medical image classification, and sentiment analysis.