

Cervical Cancer Image Processing with Convolutional Neural Network for Detection

Aulia Arif Iskandar, S.T., M.T.
Department of Biomedical Engineering
(Swiss German University)
Tangerang, Indonesia
aulia.arif@sgu.ac.id

Elnora Listianto Lie
Department of Biomedical Engineering
(Swiss German University)
Tangerang, Indonesia
elnoralistianto@gmail.com

Kholis Abdurachim Audah, Ph.D.
Department of Biomedical Engineering
(Swiss German University)
Tangerang, Indonesia
kholis.audah@sgu.ac.id

dr. Rose Khasana Dewi
Faculty of Medicine
(Brawijaya University)
Malang, Indonesia
rosenade.dr@gmail.com

Abstract—The diagnostic method for detecting cervical cancer using Pap smear can be laborious and time-consuming. Therefore, research on computer-aided diagnosis is essential. The purpose of this study is to aid the distinguishing of Pap smear images from various categories of cervical cells by creating an alternative image processing and classification method. This is so that in the future, the burden on pathologists to manually analyze many Pap smear images can be reduced. The developed method will be able to help in the detection of abnormality or cancer. The processing methods include Gaussian filtering, Otsu thresholding, Canny edge detection, and Convolutional Neural Network. The analytical methods utilized were accuracy and loss curves, and the evaluation measures of accuracy, precision, recall, and F1 measure. The most optimal trained model had an accuracy, precision, recall, and F1 measure of 93.26%, 92.55%, 91.52%, and 91.84% respectively. It was concluded that the image processing and classification method could be used to distinguish multi-cell Pap smear images. Even with some limitations, it has the potential to improve single-cell analysis and also aid in classification. In the future, this method may be used in the medical field to help diagnose cervical cancer in Indonesia.

Keywords—cervical cancer, computer-aided diagnosis, convolutional neural network, image processing, pap smear

I. INTRODUCTION

Cancer is a large family of dangerous diseases that occurs when abnormal cells grow out of control, invade, and spread (metastasize) to other parts of the body. As recorded in 2018, cancer affected public health around the world with an estimated 9.6 million deaths from cancer [1]. Based on the World Health Organization (WHO), one of the most common types of cancer is cervical cancer, which is the fourth most commonly diagnosed and fourth most common cause of death. In 2020, an estimated 604,000 cases of cervical cancer occurred worldwide, killing approximately 342,000 people [2], which increased since 2018, when WHO reported approximately 570,000 cases and 311,000 deaths [3]. In Indonesia, cervical cancer is the second most common cancer in women after breast cancer, with 23.4 cases and a mortality rate of 13.9 per 100,000 population [4].

Pap smear, short for Papanicolaou smear or Papanicolaou test, is a diagnostic test for cervical cancer and is considered one of the most effective and successful cancer screening tests in decades [5]. Cells from the cervix will be extracted with an Ayre Spatula or a cytobrush, fixed with 95% alcohol, stained with hematoxylin and eosin, and observed under a light microscope by a pathologist. High accuracy in testing and

interpreting these samples is essential to avoid fatal misdiagnosis, for example, if a benign tumor is diagnosed as malignant [6]. In addition, a pathologist can be given one or two microscope slides from one patient, and with one microscope slide, the entire area is covered with clusters of cells in the low and high-power fields. The analysis requires a lot of time and effort from the pathologist to diagnose whether cells are abnormal, normal, or cancerous. Apart from that, analysis can also be subjective due to human error. For this reason, the development of a method for computer-aided diagnosis (CAD) is required to help pathologists. Images of the stained sample can be taken under a microscope for analysis using a computer. In Indonesia, screening for early detection of cervical cancer with a Pap smear is still low, even if it is said to be cheaper, very effective, and efficient [7].

In light of current research, Win *et al.* [8] proposed an image processing detection system for cervical cancer for Pap smear images with the ordered steps of image enhancement, segmentation, feature extraction, feature selection, and finally classification. Nandanwar *et al.* [9] discovered that the use of the Otsu method in Pap smear images of cervical cancer can enable accurate segmentation. Agustyawati *et al.* [10] built an application design for cervical cancer detection, using images from acetic acid visual inspection method (IVA tests), with the utilization of a convolutional neural network (CNN) and the pre-processing steps for feature extraction using 8-bin histogram, thresholding, Gaussian filtering, and Canny edge detection.

High accuracy has been achieved with the methods listed. However, different approaches and different datasets can change results. Changing different parameters can improve accuracy, and replacing one step with another may lead to better results. Thus, to improve and further assist diagnosis of cervical cancer through Pap smear images, finding new methods or alternatives needs to be done. Based on the study by Agustyawati *et al.* [10], using colour histogram, thresholding, Gaussian filtering, Canny edge detection was previously performed for cervical cancer detection from IVA test images instead of Pap smear. In this study, the usage of several steps from the mentioned methodology was applied to create another possible image processing method that can assist pathologists in Pap smear analysis. Gaussian filtering, Otsu thresholding, Canny edge detection, and CNN was incorporated to develop an alternative method for processing and classifying Pap smear images, which has the potential to improve cervical cancer detection in the future.

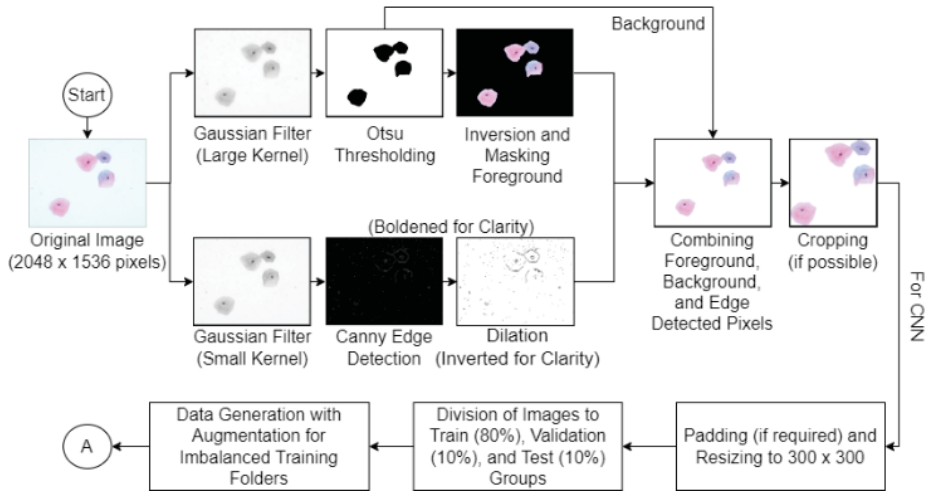


Fig. 1 Image Processing Steps Executed in Visual Studio Code

This method aimed to remove the unnecessary background of Pap smear images because noisy backgrounds have been previously proven to affect CNN performance negatively [11], and possibly reduce image size. Furthermore, the action to highlight or make prominent the features of the cells can give easier classification, and help in cropping or analysis of single-cell images, which can be achieved in the future.

II. METHODOLOGY

A. Materials and Equipment

There were two datasets utilized in this study, namely the SIPaKMeD and Mendelely Liquid-Based Cytology (LBC) datasets. The first Pap smear dataset that was used is the publicly available cervical cytology dataset researched by Plissiti *et al.* [12], namely the SIPaKMeD dataset. The distribution of the images is shown in Figure 2. The dataset contains both single-cell images (isolated cells) and multi-cell images (cell clusters) obtained from Pap smear slides. The images are grouped into five categories, namely dyskeratotic, koilocytotic, metaplastic, parabasal, and superficial-intermediate.

SIPaKMeD Dataset Image Distribution

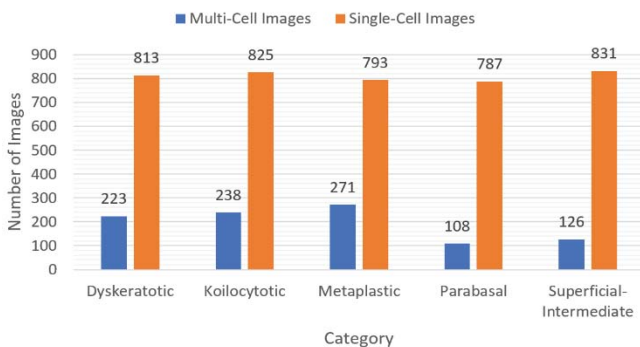


Fig. 2 SIPaKMeD Dataset Image Distribution for Each Category

The second Pap smear dataset that was used is the Mendelely LBC dataset. This is another publicly available benchmark dataset provided and researched by Hussain *et al.* [13]. The distribution of the images is shown in Figure 3. The dataset is licensed under CC BY 4.0. Both datasets were combined for this research to gain classes or categories that align with The Bethesda System, a standard for classifying cervical cell abnormalities [14].

The images were grouped into normal, abnormal, and cancer. The “normal” group included superficial-intermediate and parabasal images. Since images under negative intraepithelial lesion or malignancy (NILM) may include abnormality, it was excluded. The “abnormal” group included metaplastic, a category for benign cells, and koilocytotic, features of abnormal cells that can be found in low or high-grade squamous intraepithelial lesions (LSIL or HSIL). Since most research has utilized the SIPaKMeD dataset, LSIL and HSIL were currently excluded from this research. The last group of “cancer” includes squamous cell carcinoma (SCC). Cells undergoing dyskeratosis have not been specifically listed in Bethesda terminology because of the few consensus definitions. It may be used to elaborate on the morphological feature of cells, however, cytologists should not use the term in an interpretive way in reports [14]. Thus, images under dyskeratotic were not utilized in this research.

The image processing and segmentation scripts in this research were developed and implemented on a local PC that has Visual Studio Code. Meanwhile, the CNN scripts were implemented in Kaggle, an online platform commonly used for data science, with the provided GPU accelerator utilized. Both were run with Python as the programming language, in Windows 10 and a 64-bit operating system.

Mendelely LBC Dataset Image Distribution

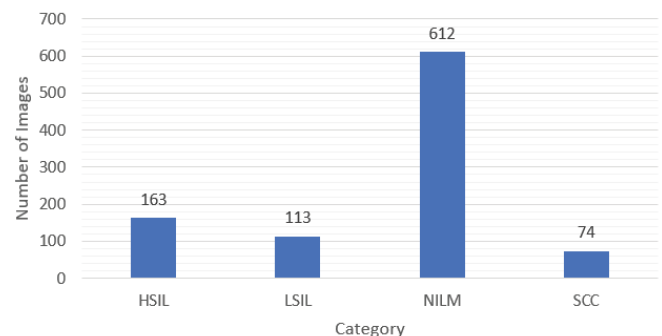


Fig. 3 Mendelely LBC Dataset Image Distribution for Each Category

B. Processing and Classification Method

The processing steps before classification with CNN are shown in Figure 1. The steps include filtering with Gaussian filter of a large kernel before Otsu thresholding, while the original image was also blurred with Gaussian filter of a smaller kernel. The processed image from the Otsu

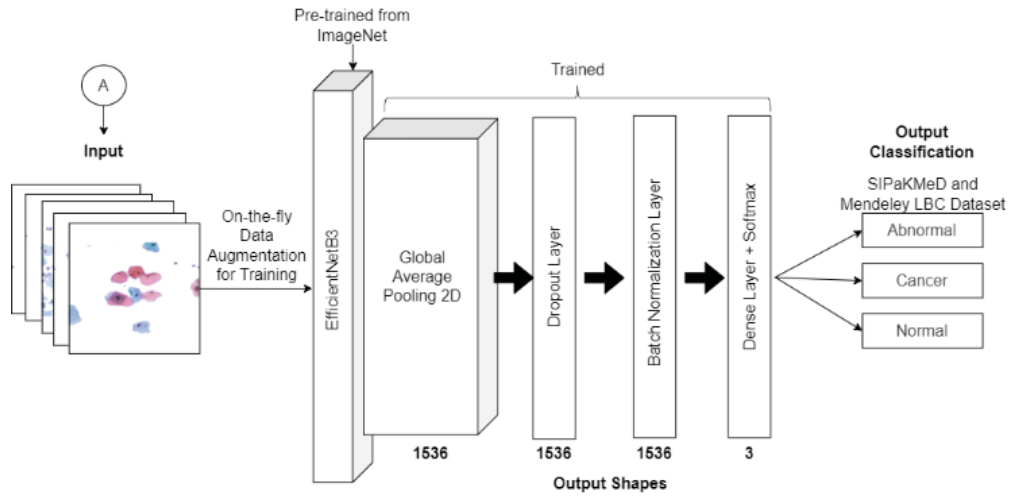


Fig. 4 Image Classification Steps with CNN in Kaggle

thresholding was inverted so that a mask can be created. The mask captured only the area of the cells (the nucleus and cytoplasm) as the foreground so that masking of the original image can be done to remove the background. The result of this process was combined with the original image processed with the smaller kernel of the Gaussian filter before Canny edge detection and dilation were applied. The combined processed images were then cropped if the mask did not occupy the whole image to possibly reduce size. These processed images were kept separately while also being processed further to a different folder for classification with CNN. The images were padded if the ratio of the image is not 1:1, and all the processed images were resized to 300 x 300 pixels, the standard image input for using EfficientNetB3 in Keras. Existing research has utilized lower resolutions for classifying Pap smear multi-cell images using CNN. A higher resolution would be best, but due to the limitations in this current study, the standard image input size was currently used. The images were then grouped into three groups: train (80%), validation (10%), and test (10%) groups. Images split into the training group were augmented to generate more data if the image numbers of the classes were very imbalanced. The image processing steps were also developed with OpenCV [15].

The method was then continued by training the CNN model with the images grouped for training (Figure 4). The CNN was built with EfficientNetB3 architecture using Keras [16], pre-trained with the ImageNet dataset, and frozen during training, followed by a Global Average Pooling 2D, a Dropout layer, a Batch Normalization layer, and a dense layer with softmax function for the classification layers. EfficientNetB3 was used in this study because it has been researched to gain higher accuracies with a smaller number of parameters when compared to other architectures [17], and has been proven to obtain the highest accuracy for classifying multi-cell Pap smear images [18]. The training images were loaded with on-the-fly data augmentation while the validation and test images were not. The final output classification included three classes as mentioned previously: abnormal, cancer, and normal.

The Adam optimizer, a batch size of 32, a learning rate of $1e^{-3}$, and a categorical cross-entropy loss function were utilized with 30 epochs. The random transformations for data augmentation include using a random rotation from the range -90 to 90 degrees, either a horizontal or vertical flip, random brightness changes in the range of 0.8 to 1.2, a random zoom

scaling in the range of 0.5 to 1.5, a random channel shift in the range of 20, and with the fill mode of “nearest.” These parameters were set after trials and were based on references from other existing sources of similar research [18]–[21].

C. Analytical Method

The method of analyzing the results was first done from the accuracy and loss curves of the trained model. The accuracy is the degree of correctness of the model’s predictions during training, and the loss is the calculated error. The next results that were analyzed were gained from the prediction of the testing data to produce confusion matrices. From the confusion matrices, the evaluation measures of accuracy, precision, sensitivity or recall, and the F1-score or F-measure were calculated.

Accuracy is known as the degree of how close the measurement is to the true value. Precision is the degree how which, under the same conditions, the same results can be obtained again after repetition, which shows the quality of the measure and how the model returns relevant results. Sensitivity, or recall, is how often input data is classified as a positive class correctly. Lastly, the F1 score, or F-score or F1 measure, is the measurement of the model’s accuracy, which can be calculated from the harmonic mean of precision and recall. F1 score is usually used as an overall score of how well the model performs, 0.0 would be the worst outcome and 1.0 the best [22].

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

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

$$Sensitivity/Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F - measure = \frac{2 * TP}{2 * TP + FP + FN} \quad (4)$$

III. RESULTS AND DISCUSSIONS

A. Image Processing Results

Gaussian filtering is one of the most common methods used in pre-processing to enhance an image's quality by the removal of noises through smoothing or blurring the image [22]. Gaussian filtering was used to eliminate unnecessary pixels to aid the next steps in the image processing flow. A large kernel was used before the Otsu thresholding method so that the area captured included only the essential pixels of, mainly, the cells (foreground). Gaussian filtering helps smoothen the image for the next image processing step, and even though Gaussian filtering blurs the image to a great degree it can be utilized to generate a mask for segmentation [23], which was executed in this study. The small kernel was used before Canny edge detection to help the capturing of the edge-detected lines to not include noises.

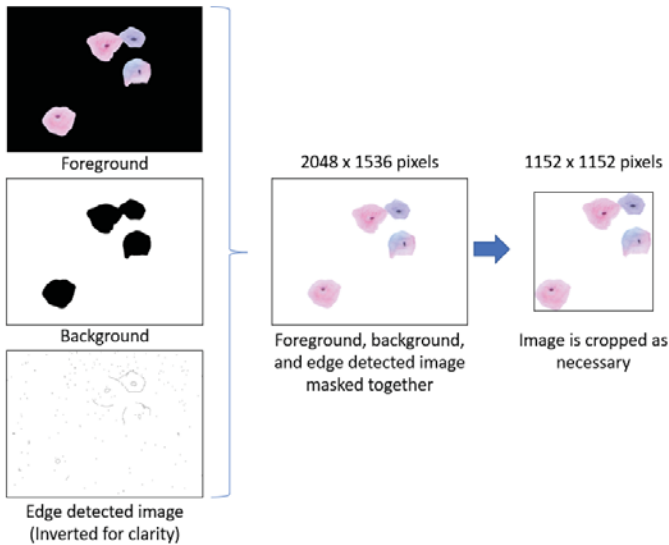


Fig. 5 Final Processed Image Example

Otsu thresholding was then applied to the image processed with the large kernel so that the segmented area can have smoother boundaries. Thresholding is a technique to create a binary image from a grayscale image by separating the pixels of the image into two classes, and this separation is based on which class has the largest between-class variance [24]. Since the area desired to be captured and segmented is only that of the cells, and the Otsu threshold captured the cells as the background due to their darker intensities, the inversion technique was implemented after Otsu thresholding so that the area of the cells was labeled as foreground. This is implemented for the next step, which is masking the original image's cell pixels to only be limited to the area of the foreground obtained by the Otsu threshold. The mask for the foreground needed to be labeled white with pixel intensity 255 after Otsu thresholding. This is why the inversion technique from OpenCV was also applied after the Otsu method.

The final steps included masking together all the processed images (Figure 5). The foreground masked from the original image, the white background of the Otsu processed image, and the Canny edge detected image were combined to form the final processed image. This final processed image contained only the necessary pixel values of the foreground, which included the nucleus and cytoplasm, while the background was removed and eliminated to the same pixel intensity of 255. Due to this, if the region of interest for the cells was

smaller than the original image size, the area of interest may be reduced. Cropping was done, if possible, to reduce the image file size, and as mentioned it is only if none of the areas captured by the Otsu thresholding process will be eliminated.

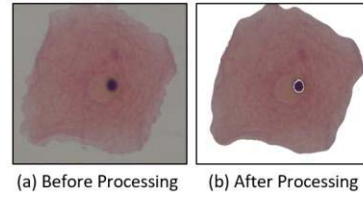


Fig. 6 Image Processing Effect Focusing on Single "Normal" Cell Categorized as Superficial-Intermediate (Borders were added for clarity)

The cells can be segmented from the background and the nucleus' borders were highlighted and extracted (Figure 6). The determination of the abnormality or normality of cells that can lead to cancer detection depends on the nucleus and cytoplasm. For instance, the nucleus to cytoplasm ratio, the shape, size, and structure of the nucleus can be an indicator.

The image processing steps are also important in which an abnormal cell under the category of koilocytotic can have its unique features highlighted (Figure 7). Koilocytotic cells may have binuclear characteristics and a perinuclear halo as some of their important indicators. A perinuclear halo displays a steep change in color in the cytoplasm from lighter to darker colors, and this can be caught by the Canny edge detection method. The binuclear characteristics can also be extracted by the edge detection method.

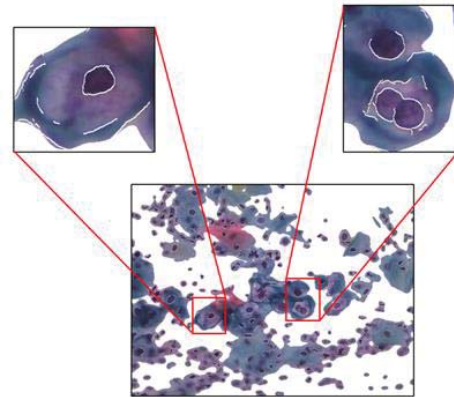


Fig. 7 Koilocytotic Cells with Perinuclear Halo and Binuclear Characteristics Highlighted (Borders were added for clarity)

B. CNN Training for Two Experimental Models

Three trials were run for two experimental models. The first experimental model involved the combined dataset images to be randomly divided by a Python script, following the fixed ratio of training, validation, and testing. The second experimental model only differs in that the testing data was hand-picked by a pathologist. These images were those that are labeled representative with clear cell characteristics of that category to be classified. Since there were no Pap smear images from the Mendeley LBC dataset under the category of SCC that was deemed representative enough based on the manual observation of the pathologist, no images were chosen for this category. Moreover, the test images were also input to be predicted manually. Accuracy and loss curves were obtained for both experimental models.

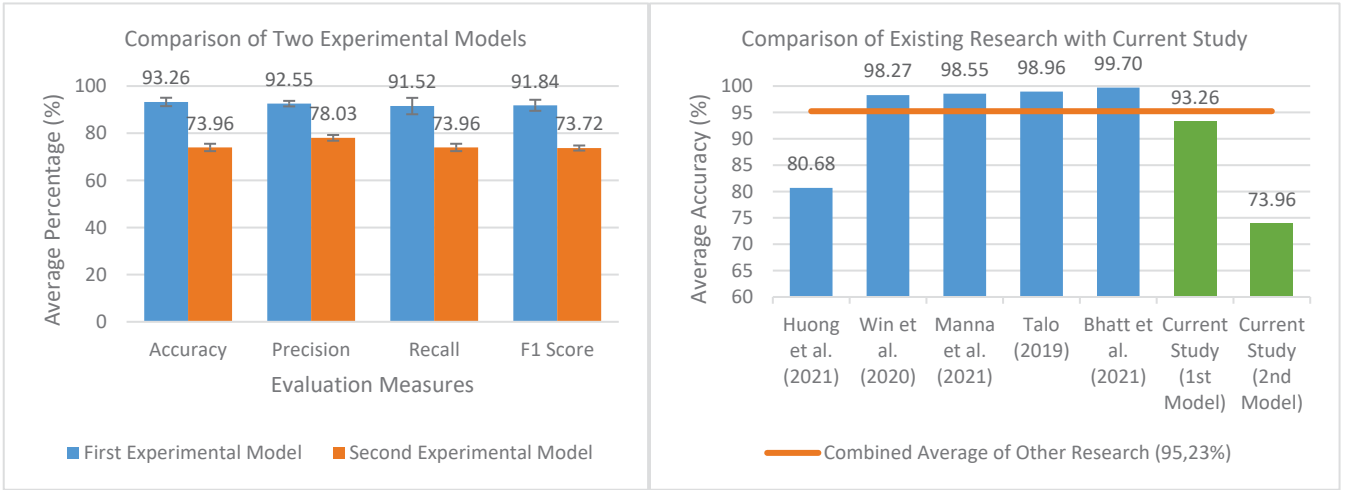


Fig. 8 Overall Comparison of Results for Two Experimental Models (Left) and Accuracy Comparison of Current Study with Existing Literature (Right)

The first experimental model showed a good-fit model, while the second experimental model did not show the same results (Figure 9). A model is defined as a good fit model when the training and validation curves for both accuracy and loss do not have much generalization gap. This means that the model learned and generalized well, showing no indication of overfitting. If a model is overfitting, the training and validation accuracies over epochs will have a larger difference, the validation accuracy will also be much lower wherein the model could not predict unseen data besides the training data well. Aside from the generalization gap for both accuracy curves, if the loss curve of the validation data is much higher than the training loss curve, that is also an indication of overfitting [25].

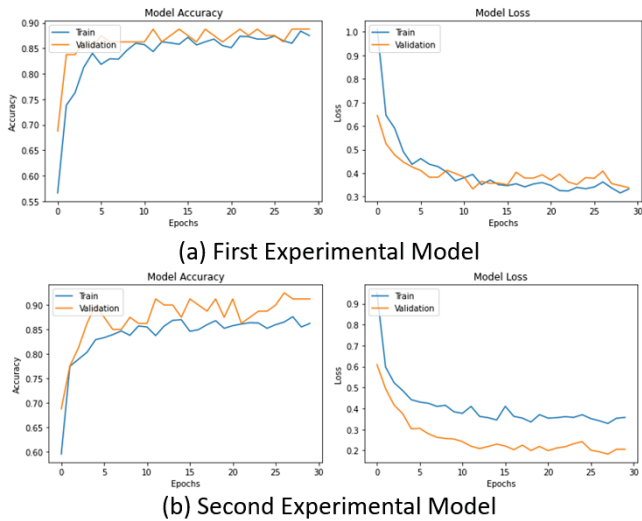


Fig. 9 Accuracy and Loss Curves of the First (a) and Second (b) Experimental Model

For the second experimental model (Figure 9b), the division of the images was not optimal since the validation dataset was shown to be too easy for the model to predict. Since the validation dataset affects the final model indirectly where it tunes by updating higher-level parameters during training [26], the final model was shown to not be trained well enough. Thus, the second model trained is not in its best-trained form to predict and classify the test images for Pap smear.

The confusion matrices for three trials were obtained from two experimental models, and the average evaluation

measures were calculated from these confusion matrices (Figure 10) for accuracy, precision, sensitivity, and F1 score. There were more correct predictions in the first experimental model than in the second, as well as more testing data predicted for the calculation of the evaluation measures. The test images used for each class were different in values due to the imbalanced dataset utilized.

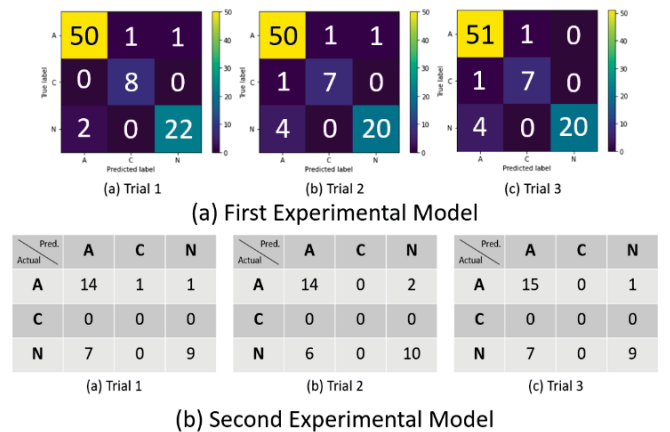


Fig. 10 Confusion Matrices of the First (a) and Second (b) Experimental Model

C. Comparison of Evaluation Measures

From Figure 8 (Left), it can be observed that the first experiment reached high evaluation parameters with the accuracy, precision, recall, and F1 score of $93.26 \pm 1.78\%$, $92.55 \pm 1.17\%$, $91.52 \pm 3.48\%$, and $91.84 \pm 2.34\%$ respectively. Meanwhile, the second experiment obtained lower evaluation parameters with the accuracy, precision, recall, and F1 score of $73.96 \pm 1.56\%$, $78.03 \pm 1.22\%$, $73.96 \pm 1.56\%$, and $73.72 \pm 1.07\%$ respectively. The lower evaluation measures from the second experimental model were due to the less optimal distribution of the images.

As shown in Figure 8 (Right), the average accuracies obtained from this study were compared to other similar research, either one which used the same datasets, or CNN, or with a classification that is close to this study's three-class classification. When compared to the average accuracy obtained from the other mentioned research (95.23%), the accuracy obtained from the two experimental models has not reached higher values. This was caused by a few existing weaknesses and limitations of the image processing method.

Reasons for errors might include that some pixels were not captured as foreground wholly by the Otsu thresholding method due to their pixels being classified as the background class with lighter pixel intensities. This can be improved by using other image processing steps such as Contrast Limited Adapted Histogram Equalization (CLAHE), which can help improve the contrast only of the cells and not with the unnecessary noises. Furthermore, some nuclei borders were not detected perfectly as well due to their intensity gradient that was not as steep. One of the ways to solve this can be by modifying the Canny edge detection method to make it more sensitive, but aside from that, there has been research that processes images in separate and different color spaces to obtain better results. Higher-level CNN techniques can also follow after these image processing steps, such as creating an ensemble or using progressive resizing with higher resolutions, since these methods have been proven to gain high evaluation measures. These are some recommendations that can be done in the future to further improve this research.

IV. CONCLUSION

The first experimental model showed a good fit model that achieved an average accuracy, precision, sensitivity, and F1 score of 93.26%, 92.55%, 91.52%, and 91.84% respectively, while the second experimental model, that could not be defined as a good fit model, achieved 73.96%, 78.03%, 73.96%, and 73.72% respectively. Thus, the image processing method developed with CNN can work in classifying multi-cell Pap smear images. Although the evaluation measures obtained have not yet reached higher values when compared to other research, especially the accuracy, the developed method has given an alternative way for Pap smear images to be analyzed and classified, which can be improved further for future research. For future study, a larger and more representative Pap smear dataset can be utilized, and more image processing methods are needed as well to overcome the method's still existing weaknesses and to open doors for single-cell image analysis and classification. Furthermore, more advanced CNN techniques can be implemented, and a higher classification for more categories in The Bethesda System, such as low-grade or high-grade squamous intraepithelial lesion (LSIL or HSIL), is recommended to bring this research to higher heights.

REFERENCES

- [1] World Health Organization, "Cancer." https://www.who.int/health-topics/cancer#tab=tab_1 (accessed Mar. 28, 2022).
- [2] H. Sung *et al.*, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA. Cancer J. Clin.*, vol. 71, no. 3, pp. 209–249, 2021.
- [3] World Health Organization, "Cervical Cancer." https://www.who.int/health-topics/cervical-cancer#tab=tab_1 (accessed Jul. 07, 2022).
- [4] Kementerian Kesehatan Republik Indonesia, "Penyakit Kanker di Indonesia Berada Pada Urutan 8 di Asia Tenggara dan Urutan 23 di Asia," *kemkes.go.id*, 2019. <http://p2p.kemkes.go.id/penyakit-kanker-di-indonesia-berada-pada-urutan-8-di-asia-tenggara-dan-urutan-23-di-asia/> (accessed Mar. 28, 2022).
- [5] W. Gray and G. Kocjan, *Diagnostic Cytopathology*, Third. Elsevier, 2010.
- [6] T. Song *et al.*, "Misdiagnosis of a small cell lung cancer resulting from inaccurate pathology," *Ann. Thorac. Surg.*, vol. 99, no. 5, pp. e125–e127, 2015.
- [7] E. P. L. Rapar, M. K. Sambuaga, and M. F. Durry, "Onkogenesis, Morfologi, dan Modalitas Deteksi Dini Karsinoma Serviks," *Med. Scope J.*, vol. 3, no. 1, pp. 47–60, 2021.
- [8] K. P. Win, Y. Kitjaiduree, K. Hamamoto, and T. Myo Aung, "Computer-assisted screening for cervical cancer using digital image processing of pap smear images," *Appl. Sci.*, vol. 10, no. 5, p. 1800, 2020.
- [9] M. P. D. Nandanwar, V. M. Wadhai, and M. A. S. Chanchlani, "Image Preprocessing Using OTSU Method On Pap Smear Images For Accurate Segmentation," *Des. Eng.*, pp. 3517–3529, 2021.
- [10] D. N. Agustyawati, H. Fauzi, and A. Pratondo, "Perancangan Aplikasi Deteksi Kanker Serviks Menggunakan Metode Convolutional Neural Network," *eProceedings Eng.*, vol. 8, no. 4, 2021.
- [11] K. Kc, Z. Yin, D. Li, and Z. Wu, "Impacts of Background Removal on Convolutional Neural Networks for Plant Disease Classification In-Situ," *Agriculture*, vol. 11, no. 9, p. 827, 2021.
- [12] M. E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, and A. Charchanti, "SIPAKMED: A new dataset for feature and image based classification of normal and pathological cervical cells in Pap smear images," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 3144–3148.
- [13] E. Hussain, L. B. Mahanta, H. Borah, and C. R. Das, "Liquid based-cytology Pap smear dataset for automated multi-class diagnosis of pre-cancerous and cervical cancer lesions," *Data Br.*, vol. 30, p. 105589, 2020.
- [14] R. Nayar and D. C. Wilbur, *The Bethesda System for Reporting Cervical Cytology*, Third. Springer, 2015.
- [15] OpenCV, "Open Source Computer Vision Library." 2015.
- [16] Y. Fu, "Image classification via fine-tuning with EfficientNet," *Keras*, 2020. https://keras.io/examples/vision/image_classification_efficientnet_fine_tuning/ (accessed May 19, 2022).
- [17] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, 2019, pp. 6105–6114.
- [18] A. R. Bhatt, A. Ganatra, and K. Kotecha, "Cervical cancer detection in pap smear whole slide images using convNet with transfer learning and progressive resizing," *PeerJ Comput. Sci.*, vol. 7, p. e348, 2021.
- [19] Y. Xiang, W. Sun, C. Pan, M. Yan, Z. Yin, and Y. Liang, "A novel automation-assisted cervical cancer reading method based on convolutional neural network," *Biocybern. Biomed. Eng.*, vol. 40, no. 2, pp. 611–623, 2020.
- [20] X. Li, Z. Xu, X. Shen, Y. Zhou, B. Xiao, and T.-Q. Li, "Detection of Cervical Cancer Cells in Whole Slide Images Using Deformable and Global Context Aware Faster RCNN-FPN," *Curr. Oncol.*, vol. 28, no. 5, pp. 3585–3601, 2021.
- [21] A. Manna, R. Kundu, D. Kaplun, A. Sinitca, and R. Sarkar, "A fuzzy rank-based ensemble of CNN models for classification of cervical cytology," *Sci. Rep.*, vol. 11, no. 1, pp. 1–18, 2021.
- [22] S. Dey, *Hands-On Image Processing with Python: Expert techniques for advanced image analysis and effective interpretation of image data*. Packt Publishing Ltd, 2018.
- [23] E. S. Gedraite and M. Hadad, "Investigation on the effect of a Gaussian Blur in image filtering and segmentation," in *Proceedings ELMAR-2011*, 2011, pp. 393–396.
- [24] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Fourth Ed. Pearson, 2018.
- [25] S. Salman and X. Liu, "Overfitting mechanism and avoidance in deep neural networks," *arXiv Prepr. arXiv1901.06566*, 2019.
- [26] T. Shah, "About Train, Validation and Test Sets in Machine Learning," *Towards Data Science*, 2017. <https://towardsdatascience.com/train-validation-and-test-sets-72cb40c9a9e7>.