IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 24, 2019, accepted August 5, 2019, date of publication August 19, 2019, date of current version September 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2936017

Segmentation and Classification of Cervical Cells Using Deep Learning

KURNIANINGSIH[1], (Senior Member, IEEE), KHALID HAMED S. ALLEHAIBI[2], LUKITO EDI NUGROHO[3], (Member, IEEE), WIDYAWAN[3], LUTFAN LAZUARDI[4], ANTON SATRIA PRABUWONO[5], (Senior Member, IEEE), AND TEDDY MANTORO[6], (Senior Member, IEEE)

[1]Department of Electrical Engineering, Politeknik Negeri Semarang, Semarang 50275, Indonesia
[2]Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[3]Department of Electrical and Information Engineering, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia
[4]Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia
[5]Faculty of Computing and Information Technology Rabigh, King Abdulaziz University, Rabigh 21911, Saudi Arabia
[6]Faculty of Engineering and Technology, Sampoerna University, Jakarta 12780, Indonesia

Corresponding author: Kurnianingsih (kurnianingsih@polines.ac.id)

This work was supported in part by the Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia, in part by the Universitas Gadjah Mada, Indonesia, and in part by the Sampoerna University, Indonesia.

ABSTRACT Cervical cancer is the fourth most prevalent disease in women. Accurate and timely cancer detection can save lives. Automatic and reliable cervical cancer detection methods can be devised through the accurate segmentation and classification of Pap smear cell images. This paper presents an approach to whole cervical cell segmentation using a mask regional convolutional neural network (Mask R-CNN) and classifies this using a smaller Visual Geometry Group-like Network (VGG-like Net). ResNet10 is used to make full use of spatial information and prior knowledge as the backbone of the Mask R-CNN. We evaluate our proposed method on the Herlev Pap Smear dataset. In the segmentation phase, when Mask R-CNN is applied on the whole cell, it outperforms the previous segmentation method in precision (0.92 ± 0.06), recall (0.91 ± 0.05) and ZSI (0.91 ± 0.04). In the classification phase, VGG-like Net is applied on the whole segmented cell and yields a sensitivity score of more than 96% with low standard deviation ($\pm2.8\%$) for the binary classification problem and yields a higher result of more than 95% with low standard deviation (maximum 4.2% in accuracy measurement) for the 7-class problem in terms of sensitivity, specificity, accuracy, h-mean, and F1 score.

INDEX TERMS Mask R-CNN, VGG-like Net, cell segmentation, cell classification, pap smear.

I. INTRODUCTION

Cancer is a life-threatening disease and has become a major burden worldwide. Global cancer data reveals that cervical cancer is the fourth most prevalent disease among females, with an approximately 90% fatality rate in underdeveloped and developing nations due to the absence of public knowledge of its causes and impacts [1]. Fortunately, this lethal disease can be detected by the regular Pap smear testing of the cervical cells. The cell samples which are collected at the outer opening of the cervix are placed on a glass tube and stained by a pathologist for examination under a microscope to determine if there are any defects/abnormalities that indicate a pre-cancerous phase [2].

The associate editor coordinating the review of this article and approving it for publication was Shovan Barma.

Manual cell screening often results in large variations in the quality of specimens, such as the uneven distribution of the cellular material that can lead to dense clumps which light cannot penetrate whereas other parts of the specimen may have many overlapping cells which hinders an accurate interpretation. Moreover, a manual visual examination is time consuming and the analysis and classification of hundreds or thousands of cells can be inaccurate due to human error. When cell examination for abnormality is carried out by a computer, the cell must be scanned at high resolution to reliably extract the features. Due to size and shape variations of normal and abnormal cells, accurate cell segmentation and classification is crucial to differentiate between normal and abnormal cells.

Several research studies have been undertaken to develop an automated screening system through the image analysis

method [3]–[6]. Such systems automatically classify normal and abnormal cervical cells. However, an automatic two-level cascade classification system proposed in [7] produced both a false negative rate and false positive rate of 1.44%.

The aim of this paper is to develop a better system for the automatic detection of cancer cells using a deep learning approach on Pap smear images. Deep learning techniques can be used to identify patterns in complex big data starting with preprocessing the data, training the model and testing it [8]. The primary contributions of this paper are as follows.

(1) As far as we know, this work is the first to implement Mask R-CNN and the transfer learning technique to segment the whole cervical cell.

(2) As far as we are aware, this work is the first to implement a VGG-like Net in which whole cervical cells are classified.

We evaluate the accuracy, sensitivity, specificity, and Zijdhenbos similarity index (ZSI) of the models.

The rest of this paper is structured as follows. Section 2 overviews the related works on cervical cell segmentation and classification; Section 3 discusses the materials and describes the methods used to segment and classify cervical cells. The experiment analysis and evaluation of segmentation and classification is given in Section 4 and a discussion is presented in Section 5. Finally, this study concludes in Section 6.

II. RELATED WORKS

Research on the automated screening of Pap smears has moved from cytology to histology over recent years. The combination of information from a multitude of computerized histology and cytology documents was used on the Brazilian Cervical Cancer Information System (SISCOLO) for sensitivities above 90% [9]. However, cytology testing continues to be used in most countries because of its affordability and efficiency in identifying cervical cancer in routine testing.

In almost all imaging system analysis, image segmentation is an important and demanding task. It is difficult for individuals to precisely analyze the segmentation of all parts of cervical cells (nuclei and cytoplasm) in Pap smears. Poor cell segmentation can lead to poor analysis results. Accurate and automatic computer-assisted segmentation on the whole cervical cell is necessary for cervical cancer screening and diagnosis. A set of 50 images was screened for the segmentation of cervical cells using mean-shift and median filtering, and for the further processing of the segmentation result using morphological operators [10].

Three SVM-based approaches (standard SVM, SVM combined with RFE algorithm, and SVM combined with the PCA algorithm) are used to classify the cervical cancer dataset from the repository of University of California at Irvine [11]. Nucleus and cytoplasm segmentation and classification using multi-class SVM classifiers such as polynomial SVMs, quadratic SVMs, Gaussian RBF SVMs, and linear SVMs resulted in 95% accuracy [12]. SVMs were also used to separate the nucleus from the cervical smear model with 95.134% precision for adaptive segmentation based on the GVF Snake model [13]. In order to improve the classification performance, the artifacts were removed from the cytology images in the Bethesda System dataset using an SVM, resulting in a true classification of normal and abnormal cells of 85.19% and 88.1% respectively [14]. Using ultra-large cervical histological digital images, a combination of SVMs and the block-based segmentation technique utilizes robust texture feature vectors to enhance classification efficiency for cervical intraepithelial neoplasia (CIN) diagnosis [15].

A segmentation method is applied to separate the cell nuclei from its cytoplasm and then classifies them using the K-Nearest Neighbor (KNN), which resulted in an 84.3% classification accuracy with no validation and 82.9% classification accuracy with 5-fold cross validation [16], [17]. A KNN method is also used to classify normal and cancerous cells on microscopic biopsy images after the segmentation process using k-means [18].

A clustering technique using fuzzy C-means (FCM) was used to segment Pap smear images [19]. One of the drawbacks of FCM clustering is that it fails to detect all the valid clusters in a colour image segmentation. William *et al.* [20] presented a Trainable Weka Segmentation classifier for cell segmentation and an enhanced fuzzy C-means algorithm to classify cervical cells.

Deep learning has achieved enormous success in many applications, including cancer research. Deep learning was used to segment abnormal cells from conventional Pap smear digital images [21], [22]. Song *et al.* [23] proposed cervical cytoplasm and nuclei segmentation using superpixels and convolutional neural networks (CNNs). The automatic segmentation of cervical nuclei using Mask R-CNN in combination with the local fully connected conditional random field (LFCCRF) is presented by Liu *et al.* [24].

Several research studies on segmenting and classifying the nucleus have been overviewed in this section. However, it might not be possible to classify cervical cells with only nucleus data. The segmentation of the whole cell is therefore more suitable [25]–[28]. Each cell is then classified using specific classifiers after the segmentation step. Su *et al.* [7] created a two-level cascade classifier to automatically detect cervical cancer cells from thin liquid-based cytology slides. The neural MLP feedforward network of Levenberg - Marquardt was used to classify the cervical images of 100 patients [29]. Classification of cervical cell images is done with deep learning [30], [31]. The performance of this type of classification, however, is not very high [32].

In this study, the whole Pap smear cell is segmented and classified using deep learning. The evaluation was carried out on the Herlev Pap smear dataset [2], [33]. Mask R-CNN was used in the segmentation process. A cell image is segmented into cell (a combination of nucleus and cytoplasm) and background. Mask R-CNN, an extension of Faster R-CNN, is a well-known method for tackling the issue of instance segmentation by predicting a segmentation mask

TABLE 1. Distribution of 7-classes of HERLEV Pap smear dataset.

Class	Cell Type	Cell Count	Category
1	Superficial squamous epithelial	74	Normal
2	Intermediate squamous epithelial	70	Normal
3	Columnar epithelial	98	Normal
4	Mild squamous non-keratinizing dysplasia	182	Abnormal
5	Moderate squamous non-keratinizing dysplasia	146	Abnormal
6	Severe squamous non-keratinizing dysplasia	197	Abnormal
7	Squamous cell carcinoma in situ intermediate	150	Abnormal

pixel-to-pixel for each region of interest (RoI). Mask R-CNN implementation is simple and requires only a small computational overhead, therefore quick experimentation is possible. The segmented whole cell (nucleus and cytoplasm) regions from the segmentation step are classified into a 2-class problem (normal and abnormal) and a 7-class problem (superficial squamous, intermediate squamous, columnar, mild dysplasia, moderate squamous, severe dysplasia, carcinoma in situ) using a smaller Visual Geometry Group-like Network (VGG-like Net).

III. METHOD

A. DATASET

The Herlev Pap smear dataset collected by Herlev University Hospital (Denmark) and the Technical University of Denmark [2], [33] was used to evaluate the proposed framework. The dataset consists of 917 images, categorized manually by qualified cytotechnicians and physicians into 7 classes as outlined in Table 1.

B. PROPOSED METHOD

The objective of this work is to develop a method to segment whole cervical cells, both single and overlapping, from conventional Pap smear images, and then classify them to identify normal and abnormal cells. The proposed method comprises two steps. The first stage partitions the cell regions using Mask R-CNN segmentation. The second stage defines the whole cell area (nucleus and cytoplasm) by classifying the segments from the initial stage. The classification in the second phase includes a training and testing phase as shown in Figure 1. We employ Mask R-CNN in the proposed segmentation process and use ResNet10 to fully utilize the spatial information and prior knowledge as the backbone of the Mask R-CNN. The primary concept of Mask R-CNN is to segment and build pixel masks for each image item automatically. We employ a smaller VGG-like Net to classify the segmentation results, which is inspired by the family of VGG networks.

In the segmentation training stage as shown in Figure 1(a), transfer learning is applied on Mask R-CNN weights trained

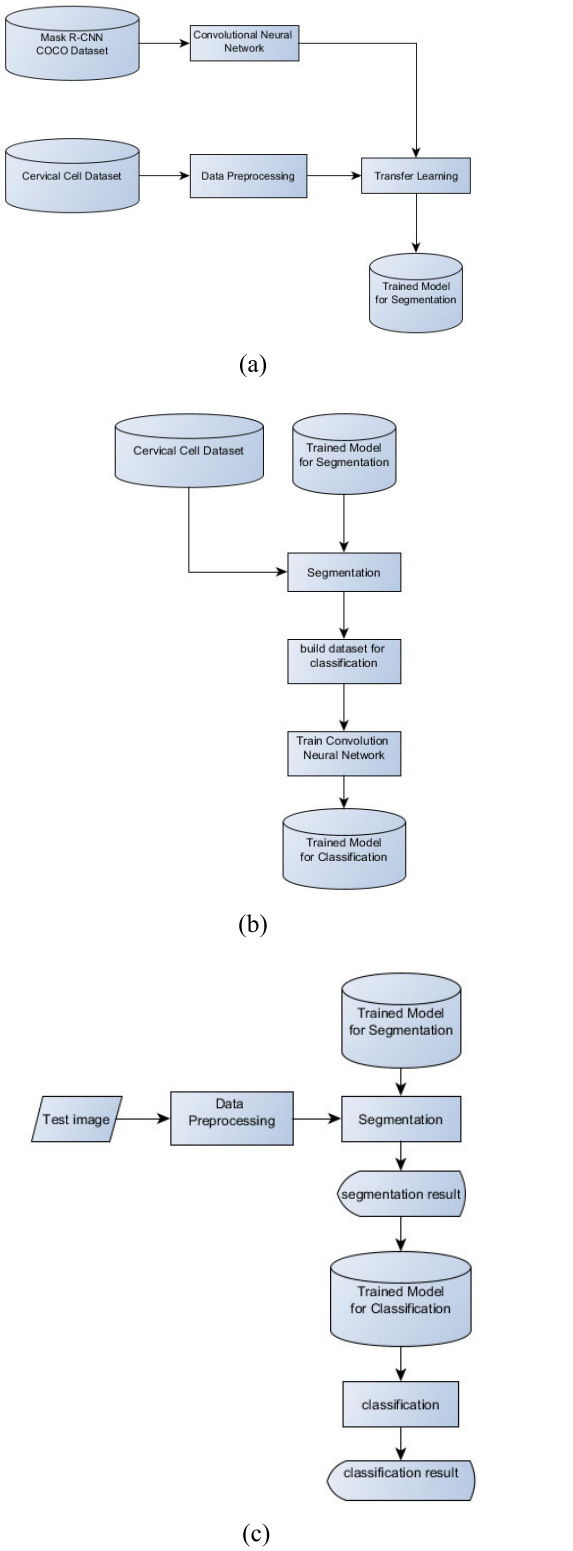

FIGURE 1. Proposed method for the automatic detection of cervical cells.

using the COCO dataset. The COCO dataset has 2,500,000 labeled instances in 328,000 images and contains 91 common object categories with 82 of these having more than 5,000 labeled instances [34]. In this proposed method, the purpose

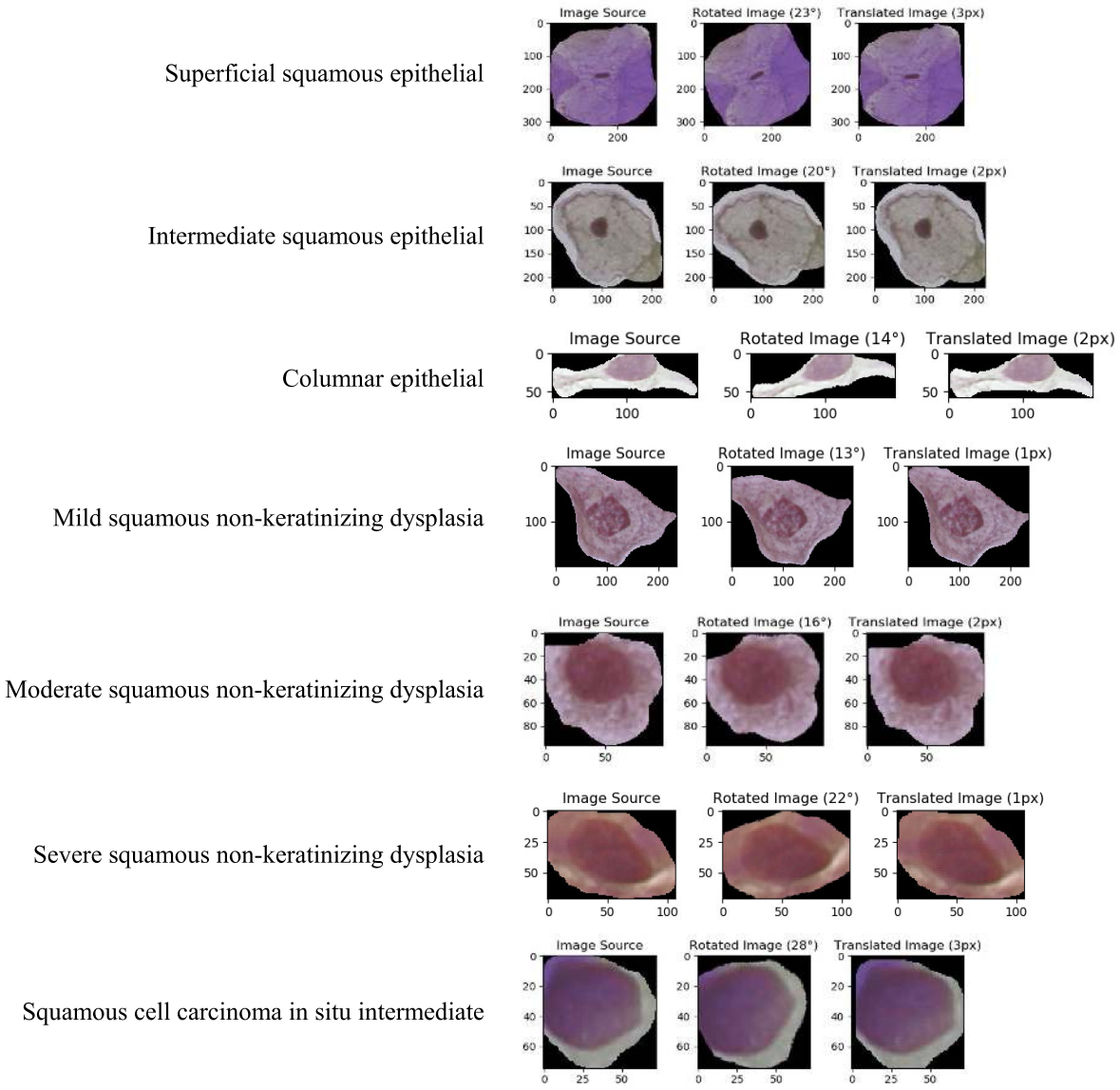

FIGURE 2. **Image rotation and translation on the Herlev dataset.**

of segmentation is to isolate the cervical cell area from its surroundings. The segmented area of the cervical cell covers both nuclei and cytoplasm. The cytoplasm can influence how a cervical cell is classified.

The results of the segmentation are then applied on the original image dataset before being handed over to the classification training algorithm, as illustrated in Figure 1(b). The input image (image source) for classification is the cervical cell (colored black), as shown in Figure 2. In the classification stage, we employ the VGG-like network, which is a more compact version of the VGG network for faster training.

Figure 1(c) illustrates the testing process during classification. The cervical cell images are segmented using Mask R-CNN to isolate the cervical cells and then they are processed in the trained VGG-like network. Based on the final score for each class (the 2 or 7 classification problem), the system determines in which class the cervical image belongs.

C. DATA PREPROCESSING

The training phase in segmentation and classification has a different preprocessing scheme. In the segmentation stage, preprocessing begins by separating the image data of the cervical cell from its mask. In the case of the Herlev dataset that we use, the original image and mask data are still mixed in one folder which corresponds to the cancer class name. This collection of images is read based on the file name pattern and is then separated into only two types of images, namely the original image of the cervical cell and the mask. When the preprocessing application finds that what is being read is a mask image, the image will be converted into a binary image, that is, white for pixels which are a part of the

cervical cells (a combination of cell nuclei and cytoplasm) and black for the other pixels. The original image of the cervical cell and its binary mask image is then resized into 200 pixels with a length that is proportionally adjusted based on the new width. The two groups of images are ready for further processing, namely network training using Mask R-CNN.

In the classification section, the application will read all the images in the Herlev dataset based on the cancer class. The images used at the classification stage are only the cervical cell regions. By involving the image's mask, before the image of the cervical cell is copied and grouped according to the cancer class, the binary mask image will be applied to the original image so that a new image consists of only two parts, namely the cell part of the cervix and the background (colored black). This new image is then resized into 200 pixels for its width and is proportional in length. The image that has been resized is then copied into a specific folder according to the classification case that we want to train, namely two folders for binary classification cases and seven folders for 7 class classification cases. The dataset is then ready to be trained by the VGG-like network.

D. DATA AUGMENTATION

The aim of applying data augmentation is to increase the generalizability of the model which can increase the dataset size and classification accuracy while preventing overfitting [35]. In this study, data augmentation is used both in the segmentation training phase and the classification training phase. We used several geometric transformation methods on the Herlev dataset for data augmentation, i.e. top-down translation, left-right translation, horizontal reflection, vertical reflection and rotation. For each training data image, the application will select randomly what kind of geometric transformations will be applied to the image.

Figure 2 shows the augmented data results for classification using 30-degree rotations and 5 pixels of translation applied on the Herlev dataset.

E. SEGMENTATION

There are three primary goals of object detection [36] i.e., given an input image to obtain 1) a list of bounding boxes for each object in the image, 2) a class label associated with each bounding box and 3) the confidence score associated with each bounding box and class label. Instance segmentation takes object detection a step further. Instead of predicting a bounding box for each object in an image, we now want to predict a mask for each object, giving us a pixel-wise segmentation of the object rather than a coarse, perhaps even unreliable bounding box.

Instance segmentation algorithms attempt to partition the image into meaningful parts and associate every pixel in an input image with a class label (e.g., person, road, car, bus) [37]. While object detection produces a bounding box, instance segmentation produces a pixel-wise mask for each individual object. However, instance segmentation does not require every pixel in an image to be associated with a label. Instance segmentation can be solved using two steps, i.e., performing object detection to draw bounding boxes around each instance of a class and then performing semantic segmentation on each of the bounding boxes [37].

The Mask R-CNN algorithm was first introduced by He *et al.* [38]. Mask R-CNN is based on the previous object detection work of R-CNN, Fast R-CNN, and Faster R-CNN by Girshick et al. The first R-CNN paper, Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, was published in 2014 by Girshick *et al.* [39]. In the first step, we input an image to the R-CNN algorithm. We then run a region proposal algorithm such as Selective Search (or equivalent). The Selective Search algorithm takes the place of sliding windows and image pyramids, intelligently examining the input image at various scales and locations, thereby dramatically reducing the total number of proposed ROIs that will be sent to the network for classification. We can thus think of Selective Search as a smart sliding window and image pyramid algorithm.

Once we have our proposed locations, we crop each of them individually from the input image and apply transfer learning via feature extraction. R-CNN utilizes feature extraction to enable a downstream classifier to learn more discriminating patterns from these CNN features. The fourth and final step is to train a series of SVMs on top of these extracted features for each class.

The problem with the original R-CNN approach is that it is still incredibly slow. Furthermore, we are not actually learning to localize via deep neural network, instead, we are leaving the localization to the Selective Search algorithm. R-CNN only classifies the ROI once it has been determined as ''interesting'' and ''worth examining'' by the region proposal algorithm, which is Selective Search.

Similar to the original R-CNN, the Fast R-CNN algorithm [40] still utilizes Selective Search to obtain region proposals, but a novel contribution, Region of Interest (ROI) Pooling, is made. In this new approach, Fast R-CNN applies the CNN to all the input images and extracts a feature map from it. ROI Pooling works by extracting a fixed-size window from the feature map and then passing it into a set of fully-connected layers to obtain the output label for the ROI.

The network of the Fast R-CNN comprises the following phases: (1) use an image and its bounding box as the inputs; (2) extract the feature map; (3) obtain the ROI feature vector by applying ROI Pooling; (4) for each region proposal, calculate the bounding box location and the class label prediction using two fully connected layers.

Due to dependency in the Selective Search (or equivalent) for the region proposal algorithm, although the network is now end-to-end trainable, the inference time performance (i.e. at prediction) dramatically declines. Ren et al. collaborated with Ren *et al.* [41] to create an additional component to create the R-CNN architecture, a Region Proposal Network (RPN). As the name suggests, the goal of the RPN is to remove the requirement of running Selective Search prior to

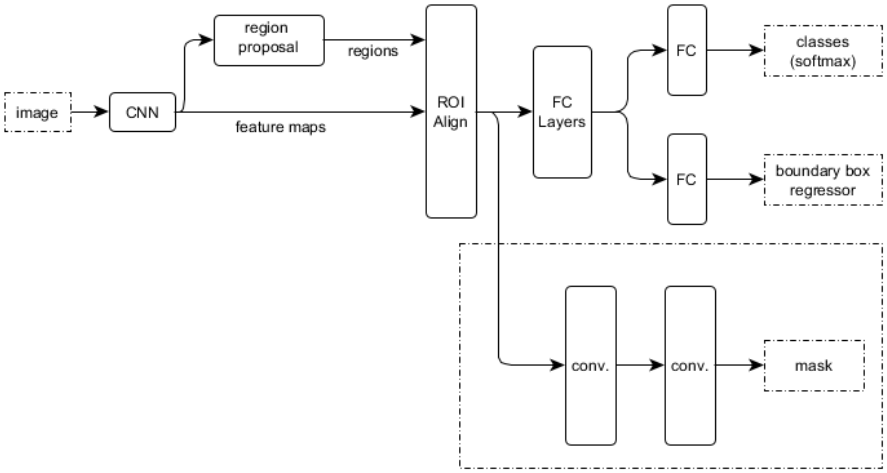

FIGURE 3. Mask R-CNN architecture.

inference and instead, makes the region proposal directly into the R-CNN architecture.

An input image is presented to the network and its features are extracted via the pre-trained CNN (i.e., the base network). These features, in parallel, are sent to two different components of the Faster R-CNN architecture. The first component, the RPN, is used to determine where in an image a potential object could be. At this point, we do not know what the object is, just that there is potentially an object at a certain location in the image. The proposed bounding box ROIs are based on the ROI Pooling module of the network along with the features extracted in the previous step. ROI Pooling is used to extract fixed-size windows of features which are then passed into two fully connected layers (one for the class labels and one for the bounding box coordinates) to obtain our final localizations. In essence, we now place anchors spaced uniformly across the whole image at varying scales and aspect ratios. The RPN will then examine these anchors and output a set of proposals as to where it is possible an object exists. In this Faster R-CNN, the complete object detection pipeline which takes place inside the network is: (1) region proposal; (2) feature extraction; (3) computing the bounding box coordinates of the objects; and (4) providing class labels for each bounding box.

The Mask R-CNN approach builds on the Faster R-CNN and makes two significant contributions: (1) it replaces the ROI Pooling module with a more accurate ROI Align module; and (2) it adds a branch for each ROI, as shown in Figure 3. This additional branch is responsible for predicting the actual mask of an object/class. The masking branch splits off from the ROI Align module prior to our FC layers and then consists of two CONV layers responsible for creating the mask predictions themselves. The Mask R-CNN output has three kinds of prediction, i.e. class/label prediction, bounding box prediction and mask prediction. Mask R-CNN can leverage different architectures such as ResNet, VGG, SqueezeNet, and MobileNet as their backend/backbone, making it possible to decrease the size of the model produced by the segmentation training stage, making it feasible for deployment on a mobile device and potentially increase frame per second (FPS) throughput as well. In our study, the Mask-RCNN backbone is applied as a ResNet-based Feature Pyramid Network (FPN) with the refined extraction layers of features and the reduced subsequent extraction layers of features according to all the cervical cell images.

Mask R-CNN utilizes a region proposal network (RPN) to generate image regions that possibly contain an object. Each region is ranked based on its "objectness score" (i.e. the probability of an object being present in a specified area) and then the top N most probable object regions are maintained.

The value 2000 was used as the N-value in the original Faster R-CNN [41]. In practice, a much lower N, such as $N = 10$, 100, 200 and 300 can be used to obtain reasonable results. In this paper, we use the same N value as He *et al.* [38], which is 300. Each of these 300 ROIs passes through three separate network sections to predict the label, the bounding box and the image mask itself.

F. CLASSIFICATION

We employ a VGG-like network as the basis of our deep learning training for the classification stage. The idea of VGGNet introduced by Simonyan and Zisserman [42] is to improve the recognition performance by increasing the depth of the CNN. The network has deep architectures from 11 to 19 weight layers and only uses small filters (3×3 convolution layer filters). The deeper the network used, the larger the number of filters learned by each convolution layer. To reduce the volume size, max pooling layers (2×2) are applied every time the number of convolutional filters doubles. Another characteristic of VGGNet is that there are several fully connected layers at the end of the network before the last layer, which uses the softmax activation function as a classifier. This framework achieves state-of-the-art results which

TABLE 2. VGG-like net architecture.

	Input	Conv1	Conv2	Conv3	Conv4	Conv5	Fc6	Fc7
Filter size	-	3x3	3x3	3x3	3x3	3x3	-	-
Channels/shape	200x200x3	32	64	64	128	128	1024	2 or 7
Activation	-	ReLU	ReLU	ReLU	ReLU	ReLU	ReLU	Softmax
Batch Normalization	-	True	True	True	True	True	True	-
Max Pooling	-	3x3	-	2x2	-	2x2	-	-
Dropout	-	0.25	-	0.25	-	0.25	0.5	-

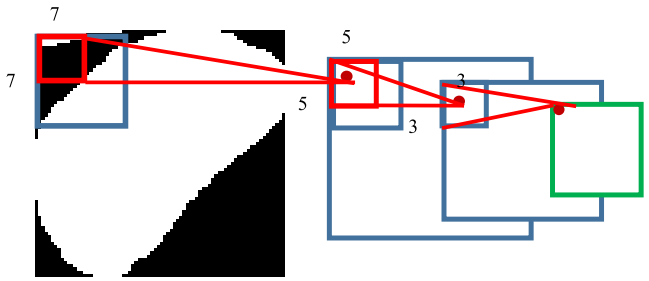

FIGURE 4. Sample of VGGNet applied in this study.

are equivalent to the results obtained by GoogLeNet [43] on ILSVRC 2014 classification without outside training data.

For our proposed method, we use a more compact version of VGGNet. We design a network architecture with a total of 7 layers and with a convolution filter channel value that is smaller (a maximum of 128 channels) than the original version of VGGNet. We do this to save computing costs and speed up the calculation process.

Unlike the original VGGNet which uses 3 fully connected layers before the softmax layer, we only use 1 fully connected layer. In each convolution layer that uses the max pooling layer, a dropout layer is added with a ratio of 0.25 while the dropout ratio of the only fully connected layer (beside the softmax layer) is 0.5. In addition, there is a batch normalization at each network layer except in the input and softmax layers. The proposed VGG-like Net architecture is shown in Table 2.

A given sample in this study using VGG-like Net, shown in Figure 4, has a stack of convolutional layers with small filters (3×3) and a (7×7) receptive field of the input image as a result of segmentation.

G. PERFORMANCE MEASURES

We implemented the algorithm in Python and performed all of the experiments using NVidia K80s 12GB, Linux operating system, 4 virtual CPUs and 61 GB memory. We trained the Mask-RCNN using ResNet101 as a backbone architecture for 40 epochs using a learning momentum of 0.9, a learning rate of 0.001, and weights decayed by 0.0001.

In this study, there are two kinds of performance measurements, i.e. segmentation and classification performance. We summarize the performance of our segmentation using precision, recall, a Zijdenbos similarity index (ZSI) and specificity, whereas the performance of the classification is evaluated using F1 score, accuracy, sensitivity, specificity, and h-mean. The prediction results obtained from the confusion matrix include:

- True Positives (TP): The number of pixels correctly identified as a mask (white pixels).
- True Negatives (TN): The number of pixels correctly identified as not part of a mask (black pixels).
- False Positives (FP): The number of pixels incorrectly identified as a mask.
- False Negatives (FN): The number of pixels incorrectly identified as not part of a mask.

$$precision = \frac{TP}{TP + Fp} \tag{1}$$

$$Recall, Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$ZSI = \frac{2TP}{2TP + FP + FN} \tag{3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$F1\ score = \frac{2TP}{2TP + FP + FN} \tag{6}$$

Precision denotes a classifiers' exactness measure, whereas recall denotes a classifiers' completeness measure. Using both recall and precision, the F1 score is used to evaluate the detection results. An excellent performance for both recall and precision is preferred over an exceptionally good performance in one aspect and a bad performance in the other. According to Zijdenbos *et al.* [44], if ZSI is greater than 0.7, it shows the detected segmentation boundary is extremely well matched with the ground truth.

Accuracy refers to a classifier being correctly categorized in a two-class issue, i.e., normal or abnormal, whereas in the seven-class problem, accuracy refers to a classifier being correctly classified as carcinoma insitu, mild dysplasia, moderate dysplasia, columnar, intermediate squamous, superficial squamous, or severe dysplasia. Sensitivity denotes that a classifier correctly classifies abnormal data as abnormal (true positive). Specificity denotes that a classifier correctly classifies normal data as normal (true negative).

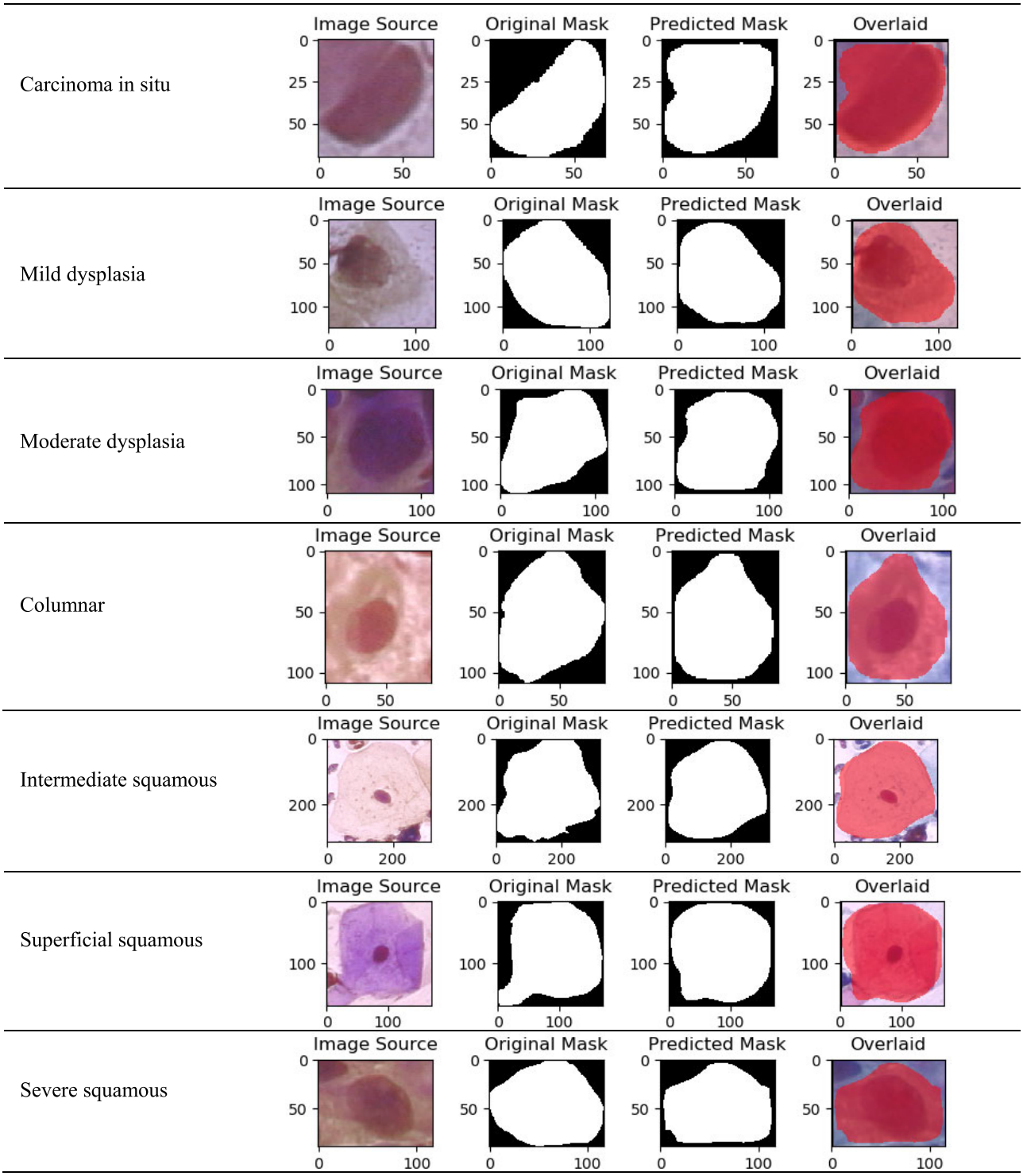

FIGURE 5. Sample of cervical cell segmentation results using Mask R-CNN.

IV. RESULTS

A. CERVICAL CELL SEGMENTATION

The objective of cervical segmentation is to divide a cell into two areas, i.e., the whole cell which consists of the cytoplasm and nucleus, and the background. Sample images from the Herlev data set at every phase in the Mask-RCNN segmentation are shown in Figure 6. The original images were masked with a white color for cytoplasm and nuclei, and black for the background.

As seen in Figure 5, the image source column is the cervical cell images taken from the Herlev dataset as is, without any image processing. The original mask is converted from the

TABLE 3. Cell segmentation performance results using mask R-CNN.

Cell Type	Precision	Recall	ZSI	Specificity
Carcinoma in situ	0.93 ± 0.06	0.91 ± 0.05	0.92 ± 0.04	0.81 ± 0.12
Mild dysplasia	0.92 ± 0.06	0.92 ± 0.04	0.92 ± 0.03	0.83 ± 0.10
Moderate dysplasia	0.91 ± 0.08	0.91 ± 0.05	0.91 ± 0.05	0.81 ± 0.11
Columnar	0.86 ± 0.10	0.89 ± 0.06	0.87 ± 0.07	0.76 ± 0.14
Intermediate squamous	0.97 ± 0.02	0.92 ± 0.03	0.95 ± 0.02	0.93 ± 0.05
Superficial squamous	0.96 ± 0.03	0.92 ± 0.06	0.94 ± 0.03	0.91 ± 0.06
Severe dysplasia	0.91 ± 0.07	0.91 ± 0.04	0.91 ± 0.04	0.78 ± 0.12
Average	**0.92 ± 0.06**	**0.91 ± 0.05**	**0.91 ± 0.04**	**0.83 ± 0.10**

FIGURE 6. Training and validation graph for the binary classification of cervical cells.

ground truth mask (from color to a binary image) provided in the original Herlev dataset. This converted mask will be used to train the Mask R-CNN and to measure the quality of our network. The predicted mask is the binary mask generated by the trained Mask-RCNN while the overlaid image shows the area from the image source which is predicted as the cell area and will be fed to the VGG-like network.

As shown in Table 3, our proposed segmentation using Mask R-CNN produces high average performance, i.e. 92% precision, 91% recall and 91% ZSI for all cell types with low standard deviation. Only the normal columnar type produces a performance result below 90%.

B. CERVICAL CELL CLASSIFICATION

We implemented two classification scenarios i.e. 2-class and 7-class classification problems. Figure 6 illustrates the training process of our VGG-like network during its 250 epochs for the binary classification problem in one of its folds. The training accuracy is quite stable while the validation accuracy sometimes drops in the middle of a full epoch. Therefore, we train the network in hundreds of epochs without an early stopping mechanism to reduce overfitting.

Table 4 shows that our proposed method for the binary classification problem (normal and abnormal) achieves high performance results with low standard deviation in all metrics for 250 epochs, i.e. 96.5% F1 score, 98.1% accuracy, 96.7% sensitivity, 98.6% specificity, and 97.7% h-mean. The confusion matrices on the testing dataset and on all the datasets confirm this claim, as shown in Table 5 and Table 6, respectively. From all the datasets, as shown in Table 6, only one instance of abnormal cells was misclassified as normal and three instances of normal cells were misclassified as abnormal.

Table 7 details the confusion matrix of the testing dataset (20% of all datasets) in 7-class classification. The classification report as shown in Table 8 achieves 94% F1 score of the micro average, 94% F1 score of the macro average, and 95% weighted average.

Table 9 details the confusion matrix of all the datasets (917 data) in 7-class classification. The classification report, as shown in Table 10, achieves 99% F1 score of the micro average, 99% F1 score of the macro average, and 99% weighted average.

Figure 7 shows that the 7-class problem network training suffers the same issue as that of the binary classification problem, namely the validation accuracy sometimes drops significantly in the middle of its full epoch training.

Table 11 shows that the same proposed VGG-like network can also address the 7-classification problem without suffering much loss to the binary classification. The average results of classification performance yield a high accuracy of 95.9%,

TABLE 4. Two-class classification performance (normal and abnormal) using 30 degree rotation and 5 pixel translation.

Epoch	Fold	F1 score	Accuracy	Sensitivity	Specificity	h-mean
250	1	92.9%	96.1%	95%	96.6%	95.8%
	2	94.8%	97.3%	95%	98%	96.5%
	3	97.7%	98.8%	97.1%	99.4%	98.2%
	4	99.7%	98.8%	97.5%	99.3%	98.4%
	5	99.2%	99.6%	98.8%	99.9%	99.3%
	Average	**96.5% ± 2.5%**	**98.1% ± 1.4%**	**96.7% ± 1.6%**	**98.6% ± 1.3%**	**97.7% ± 1.4%**
400	1	86.4%	92.4%	92.1%	92.4%	92.3%
	2	97.7%	98.8%	98.3%	99%	98.7%
	3	98.8%	99.3%	98.3%	99.7%	99%
	4	97.8%	98.8%	99.2%	98.7%	98.9%
	5	99%	99.5%	98.3%	99.9%	99.1%
	Average	**95.9% ± 5.3%**	**97.8% ± 4%**	**97.3% ± 2.8%**	**97.9% ± 3.1%**	**97.6% ± 2.9%**

TABLE 5. Confusion matrix of testing dataset in binary classification.

		Predicted	
		Abnormal	Normal
Actual	Abnormal	135	0
	Normal	3	45

TABLE 6. Confusion matrix of all datasets in binary classification.

		Predicted	
		Abnormal	Normal
Actual	Abnormal	674	1
	Normal	3	45

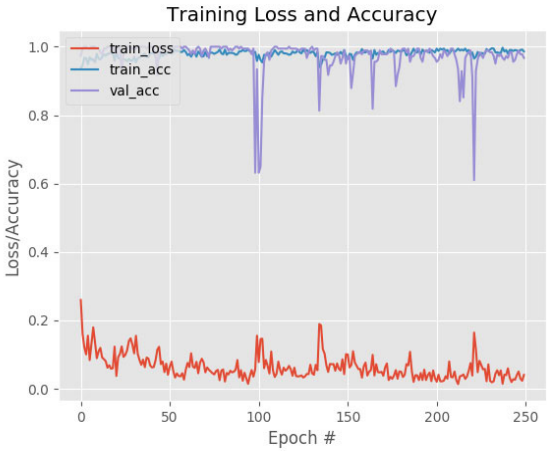

FIGURE 7. Training and validation graph for 7-class classification of cervical cells.

high sensitivity of 96.2%, high specificity of 99.3%, and high h-mean of 97.7%. The results detailed in Table 9 show the confusion matrix of all images in the Herlev dataset. The most misclassified cell type is moderate dysplastic with 3 instances (wrongly) predicted as mild/light dysplastic which is still in the same category as abnormal cells.

V. DISCUSSION

This section provides an in-depth discussion of the segmentation and classification results applied on the Herlev dataset obtained by the proposed method. Previous approaches applied to the problem of cervical cell detection in the Herlev dataset are compared to the results obtained from our research. We employed Mask-RCNN which builds on the Faster R-CNN as a promising approach that uses pixel-level prior information to acquire better semantic features so that it can efficiently detect and localize the whole cervical boundary regions with high accuracy while simultaneously generating a high-quality segmentation mask for each instance. It overcomes the difficulties that are widespread in whole cervical cell images. More importantly, the Mask R-CNN approach is conceptually simple to implement because it does not need complex pre-processing steps since feature selection is conducted by the Mask R-CNN algorithm. Mask R-CNN has a small computational overhead that enables a rapid system for training.

Most of the existing segmentation algorithms in the Herlev dataset focus on nuclei segmentation [23], [31], [32], and scant research focuses on whole cell segmentation which involves both nuclei and cytoplasm. In Table 12, the performance of several segmentation methods and our method are compared. Our method on whole cell segmentation using Mask R-CNN achieves more than 91% with low standard deviation. Specifically, the values for precision, recall, and ZSI are 0.92±0.06, 0.91±0.05, 0.91±0.04, respectively. Compared to previous research on whole cell segmentation using FCM proposed by Chankong *et al.* [25] and HCM proposed by Bezdek [45] as shown in Table 12, in terms of precision average, they both achieve 0.95±0.08, while Mask R-CNN achieves 0.92±0.06, a slight difference of 0.03 in precision average. In terms of recall average and ZSI average, Mask R-CNN achieves satisfactory performance with 0.91±0.05 and 0.91±0.04, respectively,

TABLE 7. Confusion matrix of the testing dataset (181 data) in 7-class classification.

		Predicted						
		Carcinoma in situ	Mild dysp	Moderate dysp	Columnar	Inter squamous	Sup squamous	Severe dysp
Actual	Carcinoma in situ	27	1	0	0	0	0	0
	Mild dysp	0	36	0	0	0	0	0
	Moderate dysp	0	3	26	0	0	0	0
	Columnar	0	0	1	18	0	0	0
	Inter squamous	0	0	0	0	13	1	0
	Sup squamous	0	0	0	0	0	14	0
	Severe dyspl	0	0	0	0	0	0	37

TABLE 8. Classification report of testing dataset in 7-class classification.

	Precision	Recall	F1 score	Support
Carcinoma in situ	1.00	0.90	0.95	30
Mild dysp	0.86	1.00	0.92	36
Moderate dysp	0.96	0.90	0.93	29
Columnar	1.00	0.95	0.97	19
Inter squamous	1.00	0.93	0.96	14
Sup squamous	0.82	1.00	0.90	14
Severe dyspl	1.00	0.95	0.97	39
Micro avg	**0.94**	**0.94**	**0.94**	**181**
Macro avg	**0.95**	**0.95**	**0.94**	**181**
Weighted avg	**0.95**	**0.94**	**0.95**	**181**

TABLE 9. Confusion matrix of all datasets (917 data) in the 7-class problem.

		Predicted						
		Carcinoma in situ	Mild dysp	Moderate dysp	Columnar	Inter squamous	Sup squamous	Severe dysp
Actual	Carcinoma in situ	146	2	0	0	0	0	0
	Mild dysp	0	182	0	0	0	0	0
	Moderate dysp	0	3	143	0	0	0	0
	Columnar	0	0	1	97	0	0	0
	Inter squamous	0	0	0	0	69	1	0
	Sup squamous	0	0	0	0	0	74	0
	Severe dyspl	0	0	0	0	0	0	195

whereas Chankong *et al.* [25] achieves 0.80±0.12 for recall average and 0.86±0.08 ZSI average, Bezdek [45] achieves 0.79±0.13 recall average and 0.85±0.09 for ZSI average. Our study results are higher compared to the Watershed method [46].

The approach used by Chankong *et al.* [25] applied feature extraction from the nucleus and cytoplasm in each image, whereas feature extraction in our work is conducted by deep CNN that will simplify the pre-processing steps. Chankong's approach also involves manual selection to choose the best threshold that gives the minimum error both when applying the median filter and the FCM result to build the mask of an object. Our approach using Mask R-CNN inserts an additional branch to predict automatically the actual mask of an object so Mask R-CNN is fast to train. Chankong's approach using FCM involves the manual selection of threshold and produces a slight difference of 0.03 for precision which is higher compared to the approach using Mask R-CNN. A higher result for precision indicates that the approach is able to detect more pixels which are identified correctly as part of a mask. On the other side, lower recall shows that the approach detects more pixels which are identified correctly as not part of a mask. Chankong's and Bezdek's approach shows a higher difference of 0.15 and 0.16 between precision and recall, respectively. Our approach using Mask R-CNN has a very slight difference of only 0.01 between precision and recall which is almost a balance between both of them as well as a ZSI measure which achieves 0.91, which is the same

TABLE 10. Classification report of all datasets in the 7-class problem.

	Precision	Recall	F1 score	Support
Carcinoma in situ	1.00	0.97	0.99	150
Mild dysp	0.96	1.00	0.98	182
Moderate dysp	0.99	0.98	0.99	146
Columnar	1.00	0.99	0.99	98
Inter squamous	1.00	0.99	0.99	70
Sup squamous	0.96	1.00	0.98	74
Severe dyspl	1.00	0.99	0.99	197
Micro avg	**0.99**	**0.99**	**0.99**	**917**
Macro avg	**0.99**	**0.99**	**0.99**	**917**
Weighted avg	**0.99**	**0.99**	**0.99**	**917**

TABLE 11. Performance results for 7-classification problem using 30 degree rotation and pixel translation.

Fold	Accuracy	Sensitivity	Specificity	h-mean
1	89.3%	90.2%	98.2%	94%
2	94.3%	94.7%	99%	96.8%
3	97.6%	97.8%	99.6%	98.7%
4	99.3%	99.4%	99.9%	99.6%
5	98.8%	98.8%	99.8%	99.3%
Average	**95.9% ± 4.2%**	**96.2% ± 3.8%**	**99.3% ± 0.7%**	**97.7% ± 2.3%**

result as recall. The value of ZSI in our study is greater than 0.9 which shows that the detected segmentation boundary is extremely well matched with the ground truth.

Devi *et al.* [47] demonstrate that their segmentation method using NGCS achieves higher average precision and recall compared to Mask R-CNN. Devi's approach has more complex segmentation steps, consisting of 6 layers and each layer has a different algorithm. Devi's approach and our approach are not significantly different, with only a slight difference of 0.03 in precision and 0.04 in recall.

Table 13 and Table 14 compare the performance results of previous classifiers and our method in terms of sensitivity, specificity, accuracy, h-mean, and F1 score for the 2-class problem and 7-class problem, respectively. A higher result for the precision of segmentation will lead to a higher sensitivity of the classification result, whereas a higher result for recall of segmentation will lead to the higher specificity of the classification results. Most of the existing classification algorithms for both the 2-class problem and 7-class problem result in an accuracy of above 90%, except for KNN and Bayesian, which when applied on the nucleus, yield an accuracy of below 90% for the 2-class problem, while the SVM classifier with watershed segmentation achieved a lower performance with an accuracy below 80%.

The classification performance of our method on the 2-class problem is: 96.7% sensitivity, 98.6% specificity, 98.1% accuracy, 97.7% h-mean and 96.5% F1 score. Similarly, the classification performance of our method on the 7-class problem is: 96.2% sensitivity, 99.3% specificity, 95.9% accuracy, 97.7% h-mean and 99% F1 score. The results show that our method, when applied on the whole cell, achieves a higher accuracy of 95.9% and the best specificity of 99.3% for the 7-class problem, and a specificity of 98.6% for the 2-class problem, compared to the method presented by Chankong *et al.* [25].

Whole cell segmentation is a more difficult problem than nucleus segmentation. Accurate whole cell segmentation is paramount to achieving high accuracy in classification performance. The advantages of applying Mask R-CNN as our segmentation method compared to the other aforementioned methods are: (1) it is simple, flexible, and fast to train and does not need complex algorithms or parameter tuning; (2) it selects the features automatically; (3) it is conceptually simple and does not need complex pre-processing steps; and (4) it is flexible and can leverage different architectures such as ResNet, VGG, SqueezeNet, and MobileNet as its backbone. The advantages of applying VGG-like Net are: (1) our network is deep enough to obtain high accuracy; (2) it is faster for training; (3) it is possible to decrease the size of the model produced by the segmentation training stage; and (4) it is feasible for deployment on a mobile device and potentially increase frame per second (FPS) throughput as

TABLE 12. Performance comparison of segmentation method on herlev dataset.

Author	Method	Coverage	Cell Type	Precision	Recall	ZSI
Genctav et al. [32]	Multi scale hierarchical segmentation algorithm	Nuclei	Carcinoma in situ	0.89 ± 0.15	0.90 ± 0.08	0.92 ± 0.17
			Mild dysplasia	0.88 ± 0.17	0.86 ± 0.16	0.96 ± 0.16
			Moderate dysplasia	0.91 ± 0.10	0.86 ± 0.14	0.97 ± 0.07
			Columnar	0.85 ± 0.15	0.77 ± 0.18	0.98 ± 0.05
			Intermediate squamous	0.79 ± 0.29	0.73 ± 0.31	0.98 ± 0.12
			Superficial squamous	0.69 ± 0.37	0.63 ± 0.37	0.98 ± 0.12
			Severe dysplasia	0.90 ± 0.12	0.89 ± 0.11	0.95 ± 0.13
			Average	**0.88 ± 0.15**	**0.93 ± 0.15**	**0.89 ± 0.15**
Liu et al. [24]	Mask R-CNN + LFC-CRF	Nuclei	Carcinoma in situ	0.96 ± 0.05	0.96 ± 0.10	0.95 ± 0.09
			Mild dysplasia	0.96 ± 0.04	0.98 ± 0.07	0.97 ± 0.07
			Moderate dysplasia	0.96 ± 0.04	0.97 ± 0.08	0.96 ± 0.08
			Columnar	0.93 ± 0.10	0.94 ± 0.17	0.92 ± 0.16
			Intermediate squamous	0.95 ± 0.05	0.99 ± 0.02	0.97 ± 0.02
			Superficial squamous	0.96 ± 0.06	0.97 ± 0.12	0.95 ± 0.16
			Severe dysplasia	0.97 ± 0.04	0.95 ± 0.12	0.95 ± 0.12
			Average	**0.96 ± 0.05**	**0.96 ± 0.11**	**0.95 ± 0.10**
Li et al. [26]	Radiating Gradient Vector Flow (RGVF)	Nuclei	Carcinoma in situ	0.84 ± 0.18	0.88 ± 0.11	0.86 ± 0.24
			Mild dysplasia	0.92 ± 0.13	0.90 ± 0.16	0.96 ± 0.08
			Moderate dysplasia	0.89 ± 0.15	0.87 ± 0.17	0.94 ± 0.13
			Columnar	0.83 ± 0.16	0.76 ± 0.20	0.97 ± 0.08
			Intermediate squamous	0.95 ± 0.03	0.92 ± 0.06	0.98 ± 0.02
			Superficial squamous	0.92 ± 0.12	0.88 ± 0.14	0.98 ± 0.02
			Severe dysplasia	0.88 ± 0.15	0.90 ± 0.13	0.90 ± 0.19
			Average	**0.83 ± 0.20**	**0.96 ± 0.13**	**0.87 ± 0.19**
Zhang et al. [30]	Fully Convolutional Networks and Graph (FCN-G)	Nuclei	**Average**	**-**	**-**	**0.92 ± 0.09**
Chankong et al. [25]	Fuzzy C-Means (FCM)	Whole cell	Carcinoma in situ	0.93 ± 0.08	0.81 ± 0.11	0.86 ± 0.06
			Mild dysplasia	0.96 ± 0.07	0.78 ± 0.11	0.85 ± 0.07
			Moderate dysplasia	0.95 ± 0.07	0.77 ± 0.12	0.85 ± 0.08
			Columnar	0.95 ± 0.07	0.72 ± 0.15	0.81 ± 0.10
			Intermediate squamous	0.97 ± 0.12	0.87 ± 0.12	0.91 ± 0.13
			Superficial squamous	0.99 ± 0.04	0.84 ± 0.13	0.90 ± 0.08
			Severe dysplasia	0.93 ± 0.08	0.80 ± 0.11	0.85 ± 0.07
			Average	**0.95 ± 0.08**	**0.80 ± 0.12**	**0.86 ± 0.08**
Bezdek [45]	Hard C-Means (HCM)	Whole cell	Carcinoma in situ	0.92 ± 0.10	0.81 ± 0.11	0.85 ± 0.07
			Mild dysplasia	0.95 ± 0.08	0.78 ± 0.12	0.85 ± 0.07
			Moderate dysplasia	0.93 ± 0.09	0.77 ± 0.12	0.84 ± 0.08
			Columnar	0.93 ± 0.09	0.73 ± 0.16	0.80 ± 0.10
			Intermediate squamous	0.98 ± 0.09	0.83 ± 0.16	0.88 ± 0.13
			Superficial squamous	0.99 ± 0.04	0.82 ± 0.14	0.89 ± 0.09
			Severe dysplasia	0.91 ± 0.10	0.80 ± 0.11	0.85 ± 0.08
			Average	**0.95 ± 0.08**	**0.79 ± 0.13**	**0.85 ± 0.09**
Soille et al. [46]	Watershed	Whole cell	Carcinoma in situ	0.63 ± 0.17	0.90 ± 0.13	0.72 ± 0.11
			Mild dysplasia	0.74 ± 0.29	0.85 ± 0.14	0.74 ± 0.20
			Moderate dysplasia	0.66 ± 0.27	0.86 ± 0.14	0.70 ± 0.17
			Columnar	0.77 ± 0.22	0.76 ± 0.17	0.73 ± 0.13
			Intermediate squamous	0.96 ± 0.12	0.87 ± 0.14	0.91 ± 0.13
			Superficial squamous	0.96 ± 0.05	0.88 ± 0.10	0.91 ± 0.06
			Severe dysplasia	0.58 ± 0.23	0.89 ± 0.16	0.66 ± 0.16
			Average	**0.76 ± 0.19**	**0.86 ± 0.14**	**0.77 ± 0.14**
Devi et al. [47]	Neutrosophic Graph Cut-based Segmentation (NGCS)	Whole cell	**Average**	**0.95 ± 0.11**	**0.96 ± 0.06**	**-**
Our study	Mask R-CNN	Whole cell	Carcinoma in situ	0.93 ± 0.06	0.91 ± 0.05	0.92 ± 0.04
			Mild dysplasia	0.92 ± 0.06	0.92 ± 0.04	0.92 ± 0.03
			Moderate dysplasia	0.91 ± 0.08	0.91 ± 0.05	0.91 ± 0.05
			Columnar	0.86 ± 0.10	0.89 ± 0.06	0.87 ± 0.07
			Intermediate squamous	0.97 ± 0.02	0.92 ± 0.03	0.95 ± 0.02
			Superficial squamous	0.96 ± 0.03	0.92 ± 0.06	0.94 ± 0.03
			Severe dysplasia	0.91 ± 0.07	0.91 ± 0.04	0.91 ± 0.04
			Average	**0.92 ± 0.06**	**0.91 ± 0.05**	**0.91 ± 0.04**

TABLE 13. Performance comparison of classification method on herlev dataset for 2-class problem.

Method	Coverage	Segmentation	Classifier	Sensitivity	Specificity	Accuracy	h-mean	F1 score
Jantzen et al. [16]	Nucleus	Commercial Software Package CHAMP	Benchmark	98.8%	79.3%	98.6%	88.0%	NA
Zhang et al. [30]	Nucleus	-	Deep Convolutional Network	98.2%	98.3%	98.3%	98.3%	98.8%
Bora et al. [48]	Nucleus	Maximally Stable Extremal Region (MSER)	Ensemble Classifier	98.96%	89.67%	96.51%	NA	93.13%
		Bayesian	Bayesian	97.78%	60.42%	87.98%	NA	72.50%
		SVM	SVM	97.93%	87.50%	95.20%	NA	90.52%
		KNN	KNN	80.49%	81.82%	89.39%	NA	90.00%
Paul et al. [49]	Nucleus	K-Means Cluster	Minimum Distance	89.29%	100%	92.37%	NA	NA
			KNN	97.62%	100%	98.31%	NA	NA
Devi et al. [47]	Nucleus + cytoplasm	-	Neutrosophic Graph Cut-based Segmentation (NGCS)	98.52%	99.42%	99.42%	NA	NA
Marinakis et al. [50]	Nucleus + cytoplasm	Commercial software package CHAMP	Nearest neighbour	NA	NA	92.8%	NA	NA
William et al. [20]	Nucleus, cytoplasm, background, debris	Trainable Weka Segmentation	Enhanced Fuzzy C-Means	99.28%	97.47%	98.88%	NA	NA
Chankong et al. [25]	Whole cell	FCM	ANN	99.85%	96.53%	99.27%	NA	NA
			SVM	95.11%	96.53%	95.36%	NA	NA
		HCM	ANN	99.26%	92.36%	98.05%	NA	NA
			SVM	97.33%	95.14%	96.95%	NA	NA
		Watershed	ANN	98.96%	88.89%	97.19%	NA	NA
			SVM	95.70%	88.89%	94.51%	NA	NA
Our study	**Whole cell**	**Mask R-CNN**	**Deep CNN**	**96.7%**	**98.6%**	**98.1%**	**97.7%**	**96.5%**

TABLE 14. Performance comparison of classification method on herlev dataset for 7-class problem.

Method	Coverage	Segmentation	Classifier	Sensitivity	Specificity	Accuracy	h-mean	F1 score
Marinakis et al. [50]	Nucleus + cytoplasm	Commercial software package CHAMP	Nearest Neighbour	NA	NA	92.8%	NA	NA
Chankong et al. [25]	Whole cell	FCM	ANN	98.96%	96.69%	93.78%	NA	NA
			SVM	94.22%	92.56%	85.39%	NA	NA
		HCM	ANN	98.07%	87.60%	88.88%	NA	NA
			SVM	95.70%	91.74%	83.53%	NA	NA
		Watershed	ANN	96.59%	83.47%	85.39%	NA	NA
			SVM	88.74%	85.54%	76.12%	NA	NA
Our study	**Whole cell**	**Mask R-CNN**	**Deep CNN**	**96.2%**	**99.3%**	**95.9%**	**97.7%**	**99%**

well. In general, the results show that our work using Mask R-CNN and a deep CNN classifier with a smaller VGGNet is effective.

VI. CONCLUSION AND FUTURE WORKS

This work proposes a method of cervical cell segmentation and classification. The Herlev Pap smear dataset was used for testing. First, we employed the Mask R-CNN segmentation algorithm to partition the cell regions. Second, by classifying the segments detected from the first phase with a smaller Visual Geometry Group Network, we identified the whole cell areas. To fully utilize the spatial information and prior knowledge in Mask R-CNN, we use RestNet10 as the network backbone. In this paper, we use two types of performance measures, i.e., segmentation and classification performance. We summarize the performance of our segmentation using precision, recall, ZSI and specificity, whereas the classification performance is evaluated using F1 score, accuracy, sensitivity, specificity and h-mean.

Our proposed segmentation using Mask R-CNN produces the best average performance, i.e. 0.92±0.06 precision, 0.91±0.05 recall and 0.91±0.04 ZSI and 0.83±0.10

specificity for all cell types with a low standard deviation. Only the normal columnar type produces a lower performance result of below 0.90.

We implemented two classification scenarios i.e. 2-class and 7-class classification problems. Our proposed method for the binary classification problem (normal and abnormal) yields high performance results with a low standard deviation for all metrics for 250 epochs, i.e. 96.5% F1 score, 98.1% accuracy, 96.7% sensitivity, 98.6% specificity, and 97.7% h-mean, whereas the classification performance for the 7-class problem yields a high accuracy of 95.9%, high sensitivity of 96.2%, high specificity of 99.3%, and high h-mean of 97.7%.

The advantage of our method is that we do not need complex pre-processing steps since feature selection is conducted by the Mask R-CNN algorithm. The limitation of our work is the need for higher processing power compared to the other methods. Future study should focus on the use of a deeper network to improve the performance results.

ACKNOWLEDGMENT

The authors would like to thank for the Herlev Pap smear dataset collected by Herlev University Hospital (Denmark) and the Technical University of Denmark.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, 2018.

[2] E. Martin, "Pap-smear classification," M.S. thesis, Dept. Automat., Tech. Univ. Denmark, Lyngby, Denmark, 2003.

[3] D. L. Rosenthal, "Computerized scanning devices for pap smear screening: Current status and critical review," *Clinics Lab. Med.*, vol. 17, no. 2, pp. 263–284, Jun. 1997.

[4] E. Bengtsson and P. Malm, "Screening for cervical cancer using automated analysis of PAP-smears," *Comput. Math. Methods Med.*, vol. 2014, pp. 1–12, Mar. 2014.

[5] M. E. Plissiti, C. Nikou, and A. Charchanti, "Automated detection of cell nuclei in Pap smear images using morphological reconstruction and clustering," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 2, pp. 233–241, Mar. 2011.

[6] Y.-F. Chen, P. C. Huang, K. C. Lin, H. H. Lin, L. E. Wang, C. C. Cheng, T. P. Chen, Y. K. Chan, and J. Y. Chiang, "Semi-automatic segmentation and classification of pap smear cells," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 1, pp. 94–108, Jan. 2014.

[7] J. Su, X. Xu, Y. He, and J. Song, "Automatic detection of cervical cancer cells by a two-level cascade classification system," *Anal. Cellular Pathol.*, vol. 2016, pp. 1–11, Apr. 2016.

[8] B. Jan, H. Farman, M. Khan, M. Imran, I. U. Islam, A. Ahmad, S. Ali, and G. Jeon, "Deep learning in big data Analytics: A comparative study," *Comput. Electr. Eng.*, vol. 75, pp. 275–287, May 2019.

[9] R. F. Costa, A. Longatto-Filho, C. Pinheiro, L. C. Zeferino, and J. H. Fregnani, "Historical analysis of the Brazilian cervical cancer screening program from 2006 to 2013: A time for reflection," *PLoS One*, vol. 10, no. 9, Sep. 2015, Art. no. e0138945.

[10] C. Bergmeir, M. G. Silvente, J. E. López-Cuervo, and J. M. Benítez, "Segmentation of cervical cell images using mean-shift filtering and morphological operators," *Proc. SPIE, Med. Imag., Image Process.*, vol. 7623, Mar. 2010, Art. no. 76234C. doi: 10.1117/12.845587.

[11] W. Wu and H. Zhou, "Data-driven diagnosis of cervical cancer with support vector machine-based approaches," *IEEE Access*, vol. 5, pp. 25189–25195, 2017.

[12] D. Kashyap, A. Somani, J. Shekhar, A. Bhan, M. K. Dutta, R. Burget, and K. Riha, "Cervical cancer detection and classification using independent level sets and multi SVMs," in *Proc. 39th TSP*, Jun. 2016, pp. 523–528.

[13] J. W. Zhang, S. S. Zhang, G. H. Yang, D. C. Huang, L. Zhu, and D. F. Gao, "Adaptive segmentation of cervical smear image based on GVF Snake model," in *Proc. ICMLC*, Jul. 2013, pp. 890–895.

[14] R. R. Kumar, V. A. Kumar, P. N. S. Kumar, S. Sudhamony, and R. Ravindrakumar, "Detection and removal of artifacts in cervical cytology images using Support Vector Machine," in *Proc. ITiME*, Dec. 2012, pp. 717–721.

[15] Y. Wang, D. Crookes, O. S. Eldin, S. Wang, P. Hamilton, and J. Diamond, "Assisted diagnosis of cervical intraepithelial neoplasia (CIN)," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 1, pp. 112–121, Feb. 2009.

[16] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, "Pap-smear benchmark data for pattern classification," in *Proc. NiSIS*, Oct. 2005, pp. 1–9.

[17] M. Sharma, S. Kumar Singh, P. Agrawal, and V. Madaan, "Classification of clinical dataset of cervical cancer using KNN," *Indian J. Sci. Technol*, vol. 9, no. 28, pp. 1–5, Jul. 2016.

[18] R. Kumar, R. Srivastava, and S. Srivastava, "Detection and classification of cancer from microscopic biopsy images using clinically significant and biologically interpretable features," *J. Med. Eng.*, vol. 2015, pp. 1–14, Apr. 2015.

[19] J. Talukdar, C. Kr Nath, and P. H. Talukdar, "Fuzzy clustering based image segmentation of Pap smear images of cervical cancer cell using FCM Algorithm," *Int. J. Eng. Innov. Technol.*, vol. 3, no. 1, pp. 460–462, Jul. 2013.

[20] W. William, A. Ware, A. H. Basaza-Ejiri, and J. Obungoloch, "Cervical cancer classification from Pap-smears using an enhanced fuzzy C-means algorithm," *Informat. Med. Unlocked*, vol. 14, pp. 23–33, Feb. 2019.

[21] P. Liang, G. Sun, and S. and Wei, "Application of deep learning algorithm in cervical cancer MRI image segmentation based on wireless sensor," *J. Med. Syst.*, vol. 43, no. 156, pp. 1–7, Jun. 2019.

[22] F. H. D. Araújo, R. R. V. Silva, D. M. Ushizima, M. T. Rezende, C. M. Carneiro, A. G. C. Bianchi, and F. N. S. Medeiros, "Deep learning for cell image segmentation and ranking," *Computerized Med. Imag. Graph.*, vol. 72, pp. 13–21, Mar. 2019.

[23] Y. Song, L. Zhang, S. Chen, D. Ni, B. Li, Y. Zhou, B. Lei, and T. Wang, "A deep learning based framework for accurate segmentation of cervical cytoplasm and nuclei," in *Proc. EMBC*, Aug. 2014, pp. 2903–2906.

[24] Y. Liu, P. Zhang, Q. Song, A. Li, P. Zhang, and Z. Gui, "Automatic segmentation of cervical nuclei based on deep learning and a conditional random field," *IEEE Access*, vol. 6, pp. 53709–53721, 2018.

[25] T. Chankong, N. Theera-Umpon, and S. Auephanwiriyakul, "Automatic cervical cell segmentation and classification in pap smears," *Comput. Methods Programs Biomed.*, vol. 113, no. 2, pp. 539–556, 2014.

[26] K. Li, Z. Lu, W. Liu, and J. Yin, "Cytoplasm and nucleus segmentation in cervical smear images using radiating GVF snake," *Pattern Recognit.*, vol. 45, no. 4, pp. 1255–1264, 2012.

[27] P. Y. Pai, C. C. Chang, and Y. K. Chan, "Nucleus and cytoplast contour detector of cervical smear image," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 154–161, Jul. 2012.

[28] M. H. Tsai, Y. K. Chan, Z. Z. Lin, S. F. Yang-Mao, and P. C. Huang, "Nucleus and cytoplast contour detector of cervical smear image," *Pattern Recognit. Lett.*, vol. 29, no. 9, pp. 1441–1453, Jul. 2008.

[29] B. Sokouti, S. Haghipour, and A. D. Tabrizi, "A framework for diagnosing cervical cancer disease based on feedforward MLP neural network and ThinPrep histopathological cell image features," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 221–232, Jan. 2014.

[30] L. Zhang, L. Lu, I. Nogues, R. M. Summers, S. Liu, and J. Yao, "DeepPap: Deep convolutional networks for cervical cell classification," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1633–1643, Nov. 2017.

[31] M. Wu, C. Yan, H. Liu, Q. Liu, and Y. Yin, "Automatic classification of cervical cancer from cytological images by using convolutional neural network," *Biosci. Rep.*, vol. 38, no. 6, pp. 1–9, Nov. 2008.

[32] A. Gençtav, S. Aksoy, and S. Önder, "Unsupervised segmentation and classification of cervical cell images," *Pattern Recognit.*, vol. 45, no. 12, pp. 4151–4168, Sep. 2012.

[33] J. Norup, "Classification of pap-smear data by transductive neuro-fuzzy methods," M. S. thesis, Dept. Automat., Tech. Univ. Denmark, Lyngby, Denmark, 2005.

[34] T. Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollar, "Microsoft COCO: Common Objects in Context," 2015, *arXiv:1405.0312*. [Online]. Available: https://arxiv.org/pdf/1405.0312

[35] L. Taylor and G. Nitschke, "Improving deep learning using generic data augmentation," Aug. 2017, *arXiv:1708.06020*. [Online]. Available: https://arxiv.org/abs/1708.06020

[36] Z. Q. Zhao, P. Zheng, S. Xu, and X. Wu, "Object detection with deep learning: A review," Apr. 2019, *arXiv:1807.05511*. [Online]. Available: https://arxiv.org/pdf/1807.05511.pdf
[37] A. Rosebrock, "Mask R-CNN and Cancer Detection," in *Deep Learning for Computer Vision With Python*. 2nd ed. Stockholm , Sweden: ImageSearch, Nov. 2018.
[38] K. He, G. Gkioxari, P. DollÃ r, and R. Girschik, "Mask R-CNN," Jan. 2018, *arXiv:1703.06870*. [Online]. Available: http://arxiv.org/abs/1703.06870
[39] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," Oct. 2014. *arXiv:1311.2524*. [Online]. Available: https://arxiv.org/pdf/1311.2524.pdf
[40] R. Girshick, "Fast R-CNN," Sep. 2015, *arXiv:1504.08083*. [Online]. Available: http://arxiv.org/abs/1504.08083
[41] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," Jan. 2016, *arXiv:1506.01497*. [Online]. Available: https://arxiv.org/abs/1506.01497
[42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Apr. 2015, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/pdf/1409.1556
[43] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," Sep. 2014, *arXiv:1409.4842*. [Online]. Available: https://arxiv.org/pdf/1409.4842
[44] A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, "Morphometric analysis of white matter lesion in MR images: Method and validation," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 716–724, Dec. 1994.
[45] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York, NY, USA: Plenum Press, 1981.
[46] P. Soille, *Morphological Image Analysis Principles and Applications*. Berlin, Germany: Springer, 2004.
[47] M. A. Devi, J. I. Sheeba, and K. S. Joseph, "Neutrosophic graph cut-based segmentation scheme for efficient cervical cancer detection," *J. King Saud Univ.-Comput. Inf. Sci.*, to be published. doi: 10.1016/j.jksuci.2018.09.014.
[48] K. Bora, M. Chowdhury, L. B. Mahanta, M. K. Kundu, and A. K. Das, "Automated classification of pap smear images to detect cervical dysplasia," *Comput. Method Programs Biomed.*, vol. 138, pp. 31–47, Jan. 2017.
[49] P. R. Paul, M. K. Bhowmik, and D. Bhattacharjee, "Automated cervical cancer detection using pap smear images," in *Proc. 4th Int. Conf. Soft Comput. Problem Solving, Adv. Intell. Syst. Comput.*, Mar. 2015, pp. 267–278.
[50] Y. Marinakis, G. Dounias, and J. Jantzen, "Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification," *Comput. Biol. Med.*, vol. 3, no. 1, pp. 69–78, Jan. 2009.

KURNIANINGSIH received the B.Eng. degree in informatics engineering from Telkom University, Indonesia, the M.Eng. degree in electrical engineering from North Sumatera University, Indonesia, and the Ph.D. degree in electrical engineering from Universitas Gadjah Mada, Indonesia. She is currently an Assistant Professor with the Department of Electrical Engineering, Politeknik Negeri Semarang, Indonesia. Her current research interests include sensor networks, machine learning, and computational intelligence. She has been an Executive Committee Member in the IEEE Region 10 (Asia-Pacific Region), since 2018, appointed as the Information Management Committee Chair. She is also a member of the IEEE Computational Intelligence Society (IEEE CIS) and the IEEE Systems, Man, and Cybernetics Society (IEEE SMCS). She was a recipient of the IEEE Region 10 Young Professionals Award in Academician, in 2018. She also serves as the Vice-Chair for the IEEE Indonesia Section.

KHALID HAMED S. ALLEHAIBI received the B.Sc. degree from King Abdulaziz University, Jeddah, Saudi Arabia, the M.Sc. degree from the University of Tulsa, OK, USA, and the Ph.D. degree from De Monfort University, U.K., in 2014, all in computer science. He was the Chairman of the Information Technology Department, Faculty of Computing and Information Technology in Rabigh. He is currently an Assistant Professor with the Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University.

LUKITO EDI NUGROHO received the M.Sc. degree from James Cook University, Australia, in 1995, and the Ph.D. degree from Monash University, Australia, in 2001. He is currently an Associate Professor with the Department of Electrical and Information Engineering, Faculty of Engineering, Universitas Gadjah Mada, Indonesia, where he was appointed as an Academic Staff Member after completing his undergraduate degree. His research interests include pervasive and mobile computing, software engineering, and applications of ICT in education. He is also a member of ACM.

WIDYAWAN received the B.Sc. degree in electrical engineering from Universitas Gadjah Mada, the M.Sc. degree from Erasmus University, The Netherlands, and the Ph.D. degree in electronic engineering from the Cork Institute of Technology, Ireland. He is currently an Assistant Professor with the Department of Electrical and Information Engineering, Universitas Gadjah Mada, Indonesia, where he has been serving as the Director of the Centre for Information Systems and Resources. He is also on the Board of Trustees, Gamatechno Indonesia. He co-founded Ubiaware, a software systems company specializing in WLAN design and location systems, and is also a Researcher with the Centre for Adaptive Wireless Systems, Ireland. His research interests include wireless sensors networks, machine learning, location technology, and ubiquitous computing to allow computing to fade quietly into the background of everyday life.

LUTFAN LAZUARDI received the Medical Doctor and Master of Public Health degrees from Universitas Gadjah Mada, and the Ph.D. degree from Innsbruck Medical University, in 2006. He is currently an Assistant Professor of public health with the Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Indonesia. His main research interest includes public health informatics. He has been actively involved in the development of a cancer registry in the Yogyakarta region under the support of the Indonesian Ministry of Health.

ANTON SATRIA PRABUWONO started his academic career at the Institute of Electronics, National Chiao Tung University, Taiwan, and the Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM), in 2006 and 2007, respectively. He joined the Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), in 2009. He then joined the Faculty of Computing and Information Technology, King Abdulaziz University, Rabigh, Saudi Arabia, in 2013. He was an Erasmus Mundus Visiting Professor with the Department of Mechanical Engineering and Mechatronics, Karlsruhe University of Applied Sciences, Germany. He is currently a Professor with the Faculty of Computing and Information Technology, King Abdulaziz University. His research interests include computer vision, intelligent robotics, and autonomous systems. He is also a Senior Member of ACM.

TEDDY MANTORO received the B.Sc., M.Sc., and Ph.D. degrees in computer science and the Ph.D. degree from the School of Computer Science, The Australian National University (ANU), Canberra, Australia. He is currently a Computer Science Professor with Sampoerna University, Jakarta, Indonesia. He has conducted intensive work in the intelligent environment that uses computational intelligence. He developed the concept and theory of context-aware computing for the intelligent environment, and as a proof of concept, he and his lab developed many prototypes that led to many awards. His research interests include information security, pervasive computing, and intelligent environment/IoT. He received five Gold, nine Silver, and eleven Bronze medals from the National and International IT Innovation Competitions, since 2009.