IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 December 2022, accepted 14 December 2022, date of publication 16 December 2022,
date of current version 21 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3230280

RESEARCH ARTICLE

A Deep Neural Network for Cervical Cell Classification Based on Cytology Images

MING FANG[1], XIUJUAN LEI[2], (Member, IEEE), BO LIAO[3], AND FANG-XIANG WU[4,5], (Senior Member, IEEE)

[1]Division of Biomedical Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada
[2]School of Computer Science, Shaanxi Normal University, Xi'an, Shaanxi 710119, China
[3]School of Mathematics and Statistics, Hainan Normal University, Haikou, Hainan 571158, China
[4]Department of Mechanical Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada
[5]Department of Computer Science, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada

Corresponding author: Fang-Xiang Wu (faw341@usask.ca)

This work was supported in part by the Natural Science and Engineering Research Council of Canada (NSERC), in part by the China Scholarship Council (CSC), and in part by the National Natural Science Foundation of China under Grant U19A2064 and Grant 61428209.

ABSTRACT Cervical cancer is one of the most common cancers among women. Fortunately, cervical cancer is treatable if it is diagnosed timely and administered appropriately. The death rate of cervical cancer has been greatly reduced since Pap smear test was applied. However, Pap smear test is a time-consuming and error-prone process. Moreover, classifying cervical cells into different categories is clinically meaningful but also challenging in the field of cervical cancer detection. To address these concerns, computer-aided diagnosis systems with deep learning need to be designed to automatically analyze cervical cytology images. In this study, we construct a deep convolutional neural network with feature representations learned via multiple kernels with different sizes to automatically classify cervical cytology images, named DeepCELL. Firstly, we design three different basic modules of DeepCELL to capture feature information via multiple kernels with different sizes. Then, we stack several such basic modules to form the cervical cell classification model. Finally, we perform a series of experiments to evaluate the proposed method on two cervical cytology datasets: Herlev and SIPaKMeD. Our method achieves the accuracy of 95.628%, precision of 95.685%, recall of 95.647% and F-score of 95.636% on SIPaKMeD dataset, which are the highest among all competing methods. Similarly, our method also achieves satisfactory result on Herlev dataset. In summary, extensive experimental results demonstrate that our proposed method has a promising performance in cervical cell image classification.

INDEX TERMS Cell image classification, cervical cell detection, deep learning, neural networks.

I. INTRODUCTION

Cervical cancer is one of the most frequent cancers around the world. In fact, cervical cancer is the fourth most prevalent form of cancer among women [1]. Cervical cancer has a severe influence on the health and lives of women worldwide. According to data from the World Health Organization, approximately 570,000 women worldwide were diagnosed with cervical cancer, and around 311,000 women died from this cancer in 2018 [2]. The deaths continue to rise over the years, and cervical cancer has led to 342,000 deaths around the world in 2020 [3]. Moreover, cervical cancer is the primary cause of cancer deaths in 36 countries [3]. Thankfully, cervical cancer is one of the most likely forms of cancer to be successfully treated after diagnosis, as long as it can be detected at the early stage and managed in a proper manner [2]. A leading tool, i.e., Pap smear test, is a commonly adopted technique in the early diagnosis and treatment of cervical cancer. According to the research [4], the use of Pap smear test has reduced the number of deaths related to cervical cancer by 60%.

Although Pap smear test is one of the most useful techniques in cervical cancer screening, it is a manual technique that requires cytologists to identify cervical cells, which is

The associate editor coordinating the review of this manuscript and approving it for publication was Guillermo Botella Juan.

time-consuming, costly and fallible. Since manual screening is very cumbersome, it is necessary to develop automated screening systems [5]. Cytoanalyzer system is the first automated screening device for Pap smears [6]. Then, another cytology automated apparatus CYBEST [7] is designed to extract the morphological features from cell images. Those systems are not interactive. BioPEPR [8] is one of interactive image analysis systems. PAPNET is the first cytological screening system to introduce interaction in automated screening [9], and more automated systems appear such as AutoPap 300 [10].

Computer-aided diagnosis (CAD) systems can automatically analyze abnormal cells from given cervical cytology samples, which have been extensively applied in cervical cancer detection. However, there are limitations in CAD techniques. Most of them rely on the quality of cell segmentation and hand-crafted feature extraction. For example, Win et al. [11] propose a computer-assisted screening system for cervical cancer. They extract features from the regions of segmented nuclei and cytoplasm. Then they employ a random forest classifier for feature selection and implement classification combining several classifiers. Nevertheless, their method is complicated and tedious, because it involves many procedures in cell segmentation, and also feature extraction and selection. In this case, there is a huge need for more effective tools to diagnose and treat cervical cancer.

Despite years of research in the cervical screening field, accurate classification remains a huge challenge due to several issues. Firstly, many methods only consider hand-crafted features. Secondly, it is vital to extract feature patterns via multiple kernels with different sizes in image analysis tasks. However, many researchers ignore this aspect and only apply small-sized kernels in their models. To improve these issues, we propose a novel deep CNN model with three versions to directly classify cervical cytology images, named DeepCELL. Especially, we develop several different strategies to extract feature information via multiple kernels with different sizes. Firstly, three basic modules are designed and subsequently integrated to build our DeepCELL model. Specifically, we use convolutions with multiple kernels to capture feature information from neighborhoods at different levels. The first strategy is to employ 5×5, 3×3 and 1×1 convolutions to focus on feature representations with multiple sizes. The second strategy is to apply dilated convolutions to obtain feature information from distant neighbors. Considering that using a lot of convolutions heavily increases computational complexity in the first strategy, our third strategy is to replace larger convolutions with several smaller ones. That is, we replace a 5×5 convolution with two 3×3 convolutions to reduce computational cost. Then, we train our model on two Pap smear datasets. A series of experimental results demonstrate that our proposed DeepCELL outperforms some existing CNN models in terms of precision, recall and F-score, as well as accuracy on SIPaKMeD and Herlev datasets. In addition to the comparison with CNN models, we also compare our DeepCELL with some recently developed methods on both two datasets. Extensive experiments further verify the promising performance of our DeepCELL in cervical cell detection.

The main contributions of our paper are as follows:

1) Extract feature patterns via multiple kernels with different sizes from cervical cell image data. To do so, we design a new CNN model to learn feature information from different kernels for automated cervical cell image classification. Specifically, we extract feature information with multiple size kernels and apply dilated convolutions to learn feature information from distant neighbors. Moreover, we replace larger convolutions with smaller ones to reduce the number of parameters.

2) Apply our developed model to two publicly available datasets with annotated cellular images. To the best of our knowledge, most researchers use Herlev dataset for cervical cell detection in their models. According to a study [12], more than 90% of the studies in their review employ Herlev to validate the performance of proposed methods. It is insufficient to show the effectiveness of their methods when using Herlev with only 917 cropped cell samples for training and testing. Cell samples in Herlev are too fuzzy to apply well to cervical cell detection in the real world. Although some researchers also utilize private cell image data for their methods, it is tough to compare the performance of different detection methods. Thus, we employ two public cell image datasets in this study to facilitate the comparison among different methods.

3) Acquire deep features from cell image data. Unlike existing methods that rely upon cell segmentation and hand-crafted features, our method can automatically learn deep features from cell images. Hence, classification process avoids the reduction in accuracy caused by incorrect segmentation. Moreover, such an automatic classification model is less dependent on manual operation and less susceptible to human errors that occur during the diagnosis and treatment of patients.

4) Show the effectiveness of our method in cervical cell detection. We investigate the performance of our method with different evaluation metrics, including accuracy, precision, recall and F-score. Experiments show that our method produces superior performances on both Herlev and SIPaKMeD datasets. As a result, our method has a great potentiality in improving the performance of cervical cell detection.

The rest of our paper is organized as follows. Section 2 shows the related work of deep learning in cervical cell analysis. Section 3 describes the cervical cytology image data sources used in this study, and presents our proposed automated cervical cell classification method with multiple size kernels. In Section 4 and 5, the experimental results are analyzed and discussed. Finally, Section 6 summarizes this study and draws the conclusion.

II. RELATED WORK

This section introduces some convolutional neural networks (CNNs) and applications of deep learning in cervical cell segmentation and classification.

In recent years, deep learning techniques have gained extensive attention and been widely developed for computer vision tasks, such as image classification [13], segmentation [14] and object detection [15]. As a promising model of deep learning, CNNs have been successfully utilized in various tasks. Also, methods of accelerating inference stage are important for CNNs. For example, [16] develops a hardware-efficient dataflow based on CNN to reduce computations. Reference [17] proposes a low-power and low-cost NCS2 cluster to accelerate the inference stage. Reference [18] constructs a low-cost real-time hardware accelerator and decreases extra internal memory usage. Unlike traditional methods that depend heavily on hand-crafted features, CNNs can directly classify raw image data and automatically learn features from original data. Many research has proven that CNN models have excellent performance in image analysis, such as AlexNet [19], VGGNet [20], ResNet [21], GoogleNet [22], Inception-v3 [23], and DenseNet [24].

VGGNet [20] is proposed by Simonyan and Zisserman from the University of Oxford. Its main idea is to implement small 3×3 convolutional kernels and build the deeper networks. VGGNet demonstrates that an increase in the depth of a neural network can yield better performance. However, the accuracy would get saturated and degrade rapidly with the increasing number of network layers. Consequently, He et al. [21] develops a novel CNN model, called ResNet. They introduce a special residual block for CNN, which greatly improve the network performance. In residual block, they add a skip connection, also known as identity mapping. A lot of residual blocks are stacked into ResNet. Both VGGNet and ResNet can gain great performance, while they have a large number of parameters and demand high computational costs.

Different from the above network models, GoogleNet [22], i.e., Inception-v1 has a low computational cost. The major highlight of GoogleNet is the inception module that combines convolutions of multiple kernels with different sizes to extract features from data. GoogleNet is the main combination of multiple inception modules. Then, several variants are created to improve Inception-v1, such as Inception-v2 and Inception-v3 [23]. They employ small convolution kernels to replace large ones, and change traditional kernels into asymmetric kernels. In this way, the deeper networks can be obtained. DenseNet [24] directly connects all previous layers to a current layer to gain more feature information. DenseNet reuses features and greatly reduces the number of parameters, which ensure that it has an excellent performance in computer vision tasks.

Recently, CNN models have greatly improved performance in various medical imaging tasks, such as lung disease classification and lymph node detection based on CT images [25], breast cancer prediction on ultrasound images [26], and brain tissue segmentation from MRI images [27]. Moreover, a growing number of studies have shown that CNN models have great potential in cell image classification, such as HEp-2 cells [28], red blood cells [29] and white blood cells [30]. Thus, CNN can be regarded as an important tool for analyzing cervical cell images [31], [32], thereby assisting doctors in the fight against cervical cancer to a certain extent.

Huang et al. [33] develop a generative adversarial network called Cell-GAN for cervical cell segmentation. Firstly, Cell-GAN is trained to get a probability distribution of cell morphology, then a single cell image is generated for each cell. Finally, the contour of the generated cell is treated as the segmentation line, and next the input image is cropped. A study [34] proposes a lightweight feature attention network model to segment the nucleus and cytoplasm regions in cervical cell images. This model can acquire rich features of input images based on a lightweight feature encoder module, and employ the feature layer attention module to capture the channel dependency of features. Manna et al. [35] use three classic CNN models to form an ensemble model so that several individual models can be considered in the predictions. The ensemble method has a great performance compared to some state-of-the-art methods on two benchmark cervical cytology image datasets. However, this method is more complicated than the individual base models. It integrates several base models to make the final decision, so it is more likely to get superior performance to the individual base models. A study [36] shows an internet of health things-driven framework to classify whether cervical cell is normal or abnormal. They fuse transfer learning into this framework with some pretrained CNN models, which are utilized to extract features from cervical cell images. Then, feature classification is performed to predict cell category with some classifiers. In [37], a hybrid deep feature fusion approach is proposed based on deep learning for cervical cell classification. In this approach, different models including ResNet50, VGG16, VGG 19 and XceptionNet are employed to extract features and then the feature fusion approach is utilized. Zhang et al. [38] employ a CNN pretrained on large-scale image datasets and then fine-tune it for cervical cell classification. In their work, a public Herlev and a private HEMLBC datasets are employed to assess the proposed approach, and good results are shown on both datasets. In [31], a lightweight CNN with knowledge distillation is designed to obtain an efficient model with lower computational costs for the classification task of cervical cells. The experimental results demonstrate that the proposed method achieves great performance under limited resources. Reference [32] develops an image segmentation approach with multiple phases for overlapping cervical cells. A CNN is used to recognize cell locations and get probabilistic image maps in the first phase. The second phase mainly presents multi-layer random walker graph-based region growing to

separate the cytoplasm. The final segmentation is obtained using the Hungarian algorithm in the third phase.

III. MATERIALS AND METHODS

A. DATASETS

Our DeepCELL method with three strategies is evaluated on two publicly available cervical cytology image datasets: Herlev and SIPaKMeD.

1) HERLEV DATASET

Herlev [39] is acquired from the Herlev University Hospital (Denmark) with a digital camera and microscope, which contains 917 images of single cells. Among them, 675 cells are abnormal, and 242 cells are normal. Each cell sample is examined by two experts, and a cell sample is discarded if there is any disagreement. Herlev Pap smear cell images are classified by experienced doctors and cyto-technicians into seven classes, including (1) mild squamous non-keratinizing dysplasia, (2) moderate dysplasia, (3) severe dysplasia, and (4) carcinoma in situ, (5) superficial squamous epithelia, (6) intermediate squamous epithelia, (7) columnar epithelia. These seven classes can be further categorized into two groups: abnormal and normal. The first four classes are abnormal cells, while the rest of classes are normal cells, as shown in Table 1. We would like to determine whether a cell sample is normal or abnormal, which is a binary classification problem.

2) SIPaKMeD DATASET

SIPaKMeD [40] is collected with an optical microscope and a camera, which includes 4049 images of single cells. SIPaKMeD are categorized by skilled cytopathologists based on the cellular structure and morphology into five classes, including (1) Dyskeratotic, (2) Koilocytotic, (3) Metaplastic, (4) Parabasal, (5) Superficial/Intermediate. The first two classes are abnormal cells, the third class is benign cells, and the remaining classes are considered as normal cells, as shown in Table 2.

B. DATA PREPROCESSING

Each image dataset is randomly split into three parts: 60% for training, 20% for validation and 20% for testing. Herlev dataset consists of 917 single-cell images, so 549, 184 and the remaining 184 single-cell images are available for training, validation and testing, respectively, in each experiment. SIPaKMeD dataset contains 4049 cellular images. Likewise, we obtain 2426, 811 and the remaining 812 single-cell images for training, validation and testing, respectively, in each experiment. Although the original images in a Pap smear dataset have varying sizes, in this study all images are resized to 224 $\times$ 224 pixels and then are input into our DeepCELL.

C. DeepCELL

Our DeepCELL is to analyze cervical cell images by learning feature patterns via multiple kernels with different sizes. Our DeepCELL focuses on receptive fields with multiple size kernels. The workflow of our DeepCELL method is shown in Figure 1, where orange, yellow and green represent 5×5, 3×3 and 1×1 convolutions, respectively. First of all, cervical cell images are taken as the input to our DeepCELL. The structure is inspired by [41]. Here we adopt three strategies to learn feature patterns with multiple size kernels from cervical cell images. For the first strategy DeepCELL-v1, our method has the ability to extract feature information via multiple kernels with different sizes. For the second strategy DeepCELL-v2, we utilize dilated convolutional layers to replace conventional convolutional layers. Dilated convolutions can extract feature information from distant neighbors [42]. At the same time, the dilated convolutional layers can also improve computational efficiency of our DeepCELL, because they do not increase the number of parameters while obtaining feature information from distant neighbors [43]. In addition, the more convolution kernels with different sizes, the greater the number of parameters and computational cost in DeepCELL-v1. Thus, the third strategy DeepCELL-v3 is used to solve this problem. We replace one 5×5 convolution with two 3×3 convolutions based on kernel factorization. In this way, we can get the same receptive field yet with fewer parameters, and also increase the depth of network. Several such basic structures are stacked to construct our classification model. Finally, the output layer yields whether cervical cells are abnormal or normal for binary classification, or it produces more specific categories for multi-class problem.

Table 3 is the specific description of the architecture of DeepCELL. Figure 2 presents the basic module in DeepCELL. In Figure 2(a), yellow, green and orange represent 3×3, 1×1 and dilated convolutions, respectively. Moreover, one 5×5 or dilated convolution can be replaced with two 3×3 convolutions, which are represented as red color. In Figure 2(b), orange, yellow and green represent 5×5, 3×3 and 1×1 convolutions, respectively.

We utilize several branches in our DeepCELL model to learn feature information by putting multiple convolutions with different kernel sizes in parallel. We set our DeepCELL to five phases, and follow several principles in [41] to determine the number of layers in each phase. The first phase consists of one layer including 5×5, 3×3 and 1×1 convolutions. Then, two layers are utilized in the second phase, including 5×5, 3×3 and 1×1 convolutions in the first layer and identity is added in the second layer. The third and fourth phase have six and twelve layers respectively. In these two phases, their first layers both contain 5×5, 3×3 and 1×1 convolutions and identity is added for the rest of their layers. The structure of the last phase is similar to that of the first phase. In the first phase, we only use one layer, because this phase runs with large-resolution images, which is time-consuming. Similarly, the images from the second phase still have large resolution, so we employ two layers in this phase. There are more channels in the last phase, thus we only utilize one layer to save parameters. In addition, we arrange most of the layers into other phases.

TABLE 1. Herlev Pap smear dataset.

Class No.	Category	Description of Cell Type	Cell Count	Subtotal
1	Abnormal	Mild squamous non-keratinizing dysplasia	182	
2	Abnormal	Moderate squamous non-keratinizing dysplasia	146	
3	Abnormal	Severe squamous non-keratinizing dysplasia	197	
4	Abnormal	Squamous cell carcinoma in situ intermediate	150	675 Abnormal
5	Normal	Superficial squamous epithelial	74	
6	Normal	Intermediate squamous epithelial	70	
7	Normal	Columnar epithelial	98	242 Normal
Total			917	

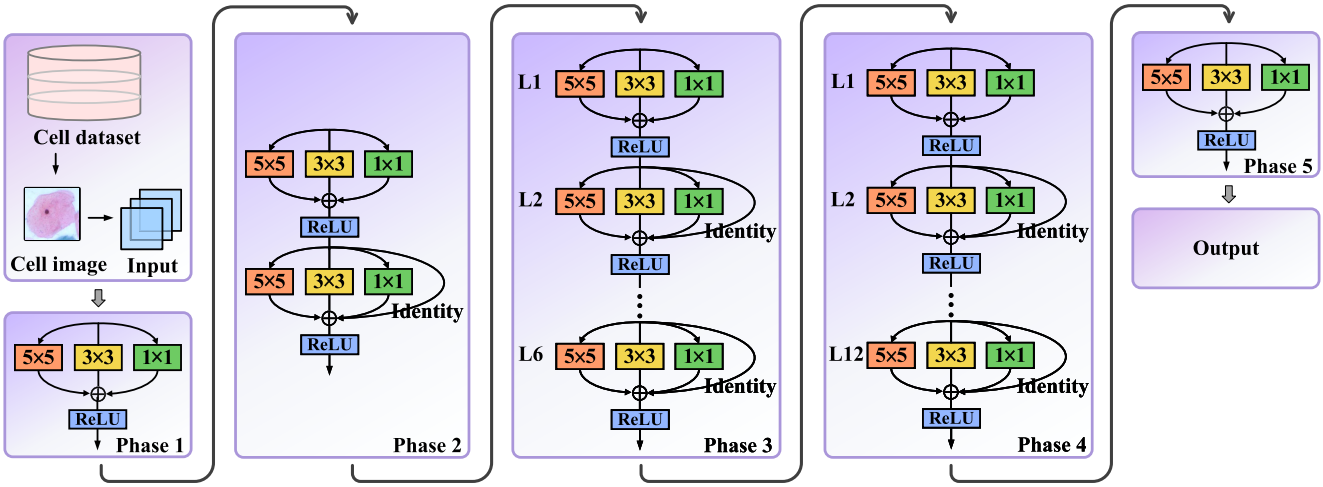
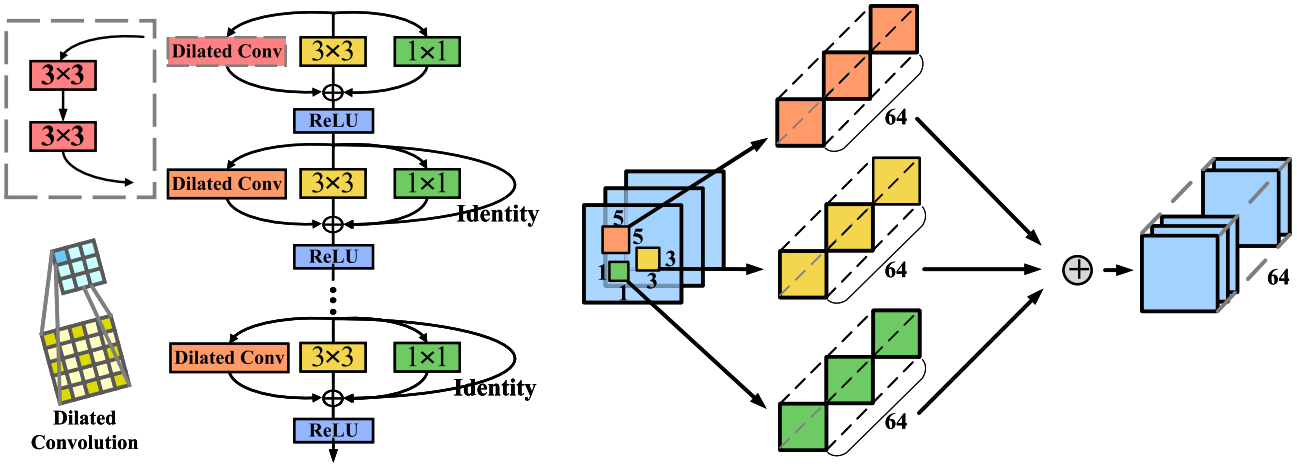


FIGURE 1. Workflow of DeepCELL. Orange, yellow and green represent 5 × 5, 3 × 3 and 1 × 1 convolutions respectively.



(a) The basic module of DeepCELL-v2

(b) The details of basic module in DeepCELL

FIGURE 2. Basic module in DeepCELL. (a) Yellow, green and orange represent 3 × 3, 1 × 1 and dilated convolutions respectively. (b) Orange, yellow and green represent 5 × 5, 3 × 3 and 1 × 1 convolutions respectively.

1) DeepCELL-v1

A DeepCELL-v1 basic module consists of 5×5 convolutions, 3×3 convolutions and 1×1 convolutions, as well as identity mapping. Our DeepCELL integrates convolutions of different size kernels (5×5, 3×3, 1×1) to capture different feature representations, inspired by the inception modules [22]. The modules with different size kernels can capture various spatial information. The basic module of DeepCELL-v1 is shown in Figure 1 and Figure 2(b).

2) DeepCELL-v2

A DeepCELL-v2 basic module mainly contains dilated convolutions, 3×3 convolutions and 1×1 convolutions, as well as identity mapping. Dilated convolution, also known as Atrous

TABLE 2. SIPaKMeD Pap smear dataset.

Class No.	Category	Specific Category	Cell Count	
1	Abnormal	Dyskeratotic	813	
2	Abnormal	Koilocytotic	825	1638 Abnormal
3	Benign	Metaplastic	793	793 Benign
4	Normal	Parabasal	787	
5	Normal	Superficial/Intermediate	831	1618 Normal
Total			4049	

TABLE 3. Specific description of the architecture of DeepCELL.

Phase	Input Size	Output Size	Input Channel	Output Channel
1	224×224	112×112	3	64
2	112×112	56×56	64	64
3	56×56	28×28	64	128
4	28×28	14×14	128	256
5	14×14	7×7	256	1024

convolution, comes from the wavelet decomposition [44]. Dilated convolution has recently been utilized for a variety of tasks. In DeepCELL-v2, dilated convolutions are applied to obtain the feature information from distant neighbors. The basic structure of DeepCELL-v2 is illustrated in the following Figure 2(a).

3) DeepCELL-v3

In order to reduce the number of parameters in DeepCELL-v1, we replace one 5×5 convolution with two 3×3 convolutions to construct the basic module of DeepCELL-v3, which is inspired by [23]. The replacement of large-size convolution into small-size convolution can increase the depth of CNN while reducing the number of parameters and computational complexity. We know that the receptive field of one 5×5 convolution is the same as the receptive field of two consecutive 3×3 convolutions, while the number of parameters of the latter is less than that of the former. The basic module of DeepCELL-v3 is shown with the dotted box in Figure 2(a).

The information flow of DeepCELL module can be represented as $y = x + f_1(x) + f_2(x) + f_3(x)$, where $f_1(x)$ expresses a 5×5 convolution or dilated convolution or two consecutive 3×3 convolutions, $f_2(x)$ is a 3×3 convolution, and $f_3(x)$ is a 1×1 convolution. Specifically, we adopt the convolutions with multiple kernel sizes to learn the comprehensive features, then merge these extracted features, and next pass them to the next layer. Integrating convolutions with different kernel sizes can acquire rich information representations. Considering that 5×5, 3×3, and 1×1 convolutions have different kernel sizes, we use the padding operation to extend the image in order to ensure that convolutions of different kernel sizes can also obtain feature maps with the same sizes. After that, we can add these feature maps together, and then transfer the output to the next layer. Our DeepCELL model is mainly built by a stack of several such basic modules and a classification layer.

IV. EXPERIMENTAL RESULTS

A. EVALUATION METRICS

To investigate the performance of our proposed DeepCELL model, we use several evaluation metrics, i.e., accuracy, precision, and recall, as well as F-score. The definitions of these evaluation metrics are shown below.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

where TP (true positive) represents the number of positive samples that are correctly classified. FP (false positive) is the number of samples in negative class that are misidentified as positive samples. TN (True negative) denotes the number of negative samples that are correctly detected. FN (false negative) is the number of positive samples that are incorrectly classified as negative samples.

Accuracy indicates the proportion of samples that are correctly identified. Precision measures the classification potential of correctly recognizing positive samples from all the predicted positive samples. Recall measures the classification potential of correctly detecting positive samples from all the true samples. F-score is the harmonic mean of precision and recall, which assesses the trade-off between precision and recall. Generally speaking, the higher the values of precision, recall and F-score are, the better the performance of the method is.

B. PERFORMANCE COMPARISON

To demonstrate the effectiveness of our proposed cervical cell classification method DeepCELL, we independently train our DeepCELL with three versions of strategies on two cervical cell image datasets. In addition, we compare its performance with state-of-the-art models, including VGG-16, GoogleNet, DenseNet-121, Inception-v3 and ResNet-18. All of the experiments are performed using Python installed on a 64-bit Windows 10 operating system with Intel Core i7 CPU. The GPU is the NVIDIA GeForce RTX 2070.

The experimental settings for our DeepCELL are as follows. The stochastic gradient descent (SGD) with momentum is utilized to train our method for 100 epochs. The initial learning rate, weight decay and momentum are set to 0.01, 1e-4, and 0.9 respectively. Our DeepCELL trains and tests 917 cell images from Herlev dataset and 4049 cell images from SIPaKMeD dataset. The input images of our DeepCELL are 224×224 pixels. Pytorch framework is adopted to design our proposed method for cervical cell image classification. The robustness of our DeepCELL method is assessed by repeating the experiments ten times to average the results.

Table 4 displays the performance of our DeepCELL in comparison with existing methods on Herlev dataset

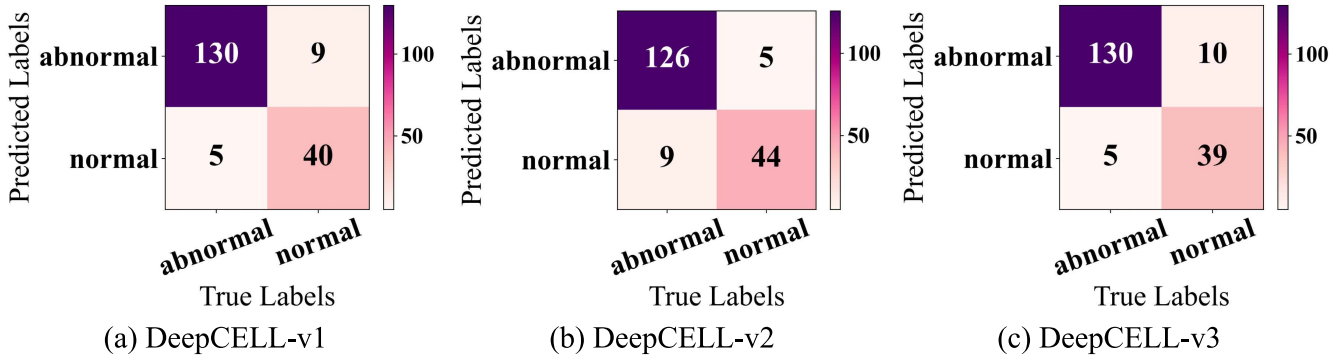


FIGURE 3. Confusion matrix of our method with three strategies on Herlev dataset.

TABLE 4. Performance of cervical cell classification methods on Herlev dataset.

Models	Categories	Precision (%)	Recall (%)	F-score (%)
VGG-16	Abnormal	89.130	91.111	90.110
	Normal	73.913	69.388	71.579
GoogleNet	Abnormal	89.362	93.333	91.304
	Normal	79.070	69.388	73.913
DenseNet-121	Abnormal	84.967	96.296	90.278
	Normal	83.871	53.061	65.000
Inception-v3	Abnormal	85.161	**97.778**	91.034
	Normal	**89.655**	53.061	66.667
ResNet-18	Abnormal	89.231	85.926	87.547
	Normal	64.815	71.429	67.961
DeepCELL-v1	Abnormal	93.525	96.296	**94.891**
	Normal	88.889	81.633	85.106
DeepCELL-v2	Abnormal	**96.183**	93.333	94.737
	Normal	83.019	**89.796**	**86.275**
DeepCELL-v3	Abnormal	92.857	96.296	94.545
	Normal	88.636	79.592	83.871

considering cell categories (abnormal and normal). From Table 4, one can see that the best F-score is 94.891% for abnormal cells and 86.275% for normal cells, both of which are obtained from our proposed DeepCELL. On the other hand, ResNet-18 gets an F-score of 87.547% for abnormal cells and 67.961% for normal cells. Moreover, F-score is 91.034% for abnormal cells while F-score is 66.667% for normal cells in Inception-v3. There is an apparent difference in the ability to classify abnormal cells and normal cells for Inception-v3. We can see that all models have more ability to identify abnormal cells compared with the ability to identify normal cells. Unlike these benchmark models, there is a small difference in the ability to detect different categories of cervical cells in our DeepCELL. In addition to our model, the benchmark model has relatively obvious differences in the classification ability between abnormal and normal cells. In other words, our DeepCELL can get superior performance no matter which categories of cervical cells are. Hence, our DeepCELL can get more satisfactory classification results than other models.

Figure 3 depicts the confusion matrix of our method with three strategies on Herlev dataset. As shown in Figure 3, some cervical cells are easier to be detected whereas others are relatively difficult. For example, our DeepCELL-v1 correctly detects 130 out of 135 cells, and 40 among 49 cells. In addition, our DeepCELL-v2 correctly predicts 126 from 135 cells, and 44 out of 49 cells. We can see that our DeepCELL can recognize abnormal cells better than normal cells, but it performs well in classifying abnormal and normal cells.

Table 5 shows the performance of different methods for classifying each of the five cell categories on SIPaKMeD dataset, considering cell categories (Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, Superficial/Intermediate). As it can be seen in Table 5, most of the existing models lead to F-score of above 80% on five categories of cervical cells, except for Inception-v3 that acquires F-score of less than 80% on Koilocytotic and Metaplastic. Specifically, F-score values of Dyskeratotic, Parabasal and Superficial/Intermediate exceed 90% in VGG-16, GoogleNet and DenseNet-121, which are higher than those of Koilocytotic and Metaplastic. Different from these models, F-score of Superficial/Intermediate is 91.765% for Inception-v3, which is better than those of other four categories. ResNet-18 has worse performance on Koilocytotic compared with the other four categories, which only achieves an F-score of 84.571%. Our DeepCELL-v1 and DeepCELL-v2 both yield F-score of 100% on one type of normal cell, i.e., Parabasal category. Clearly, our DeepCELL-v1 and DeepCELL-v2 reach perfect performance on this type of cell. Among all other types of cells, our method with three strategies obtains F-score values of more than 90%. However, the performances of other competing methods are relatively lower than our proposed method, and some of their F-score values are just more than 80% in five cell categories.

Figure 4 illustrates the confusion matrix of our method with three strategies on SIPaKMeD dataset. Certain categories of cervical cells are likely to be classified correctly, yet some other cell categories have fewer chances to be recognized precisely. As shown in Figure 4, our

TABLE 5. Performance of different methods for five categories on SIPaKMeD dataset.

Models	Categories	Precision (%)	Recall (%)	F-score (%)
VGG-16	Dyskeratotic	95.513	91.411	93.417
	Koilocytotic	88.816	81.818	85.174
	Metaplastic	81.667	92.453	86.726
	Parabasal	98.039	94.937	96.463
	Superficial/Intermediate	95.906	98.204	**97.041**
GoogleNet	Dyskeratotic	95.597	93.252	94.410
	Koilocytotic	87.742	82.424	85.000
	Metaplastic	88.344	90.566	89.441
	Parabasal	98.649	92.405	95.425
	Superficial/Intermediate	87.701	98.204	92.655
DenseNet-121	Dyskeratotic	95.706	95.706	95.706
	Koilocytotic	82.184	86.667	84.366
	Metaplastic	93.197	86.164	89.542
	Parabasal	99.342	95.570	97.419
	Superficial/Intermediate	91.477	96.407	93.878
Inception-v3	Dyskeratotic	89.308	87.117	88.199
	Koilocytotic	78.431	72.727	75.472
	Metaplastic	80.921	77.358	79.100
	Parabasal	81.714	90.506	85.886
	Superficial/Intermediate	90.173	93.413	91.765
ResNet-18	Dyskeratotic	97.315	88.957	92.949
	Koilocytotic	80.000	89.697	84.571
	Metaplastic	91.558	88.679	90.096
	Parabasal	94.410	96.203	95.298
	Superficial/Intermediate	**96.319**	94.012	95.152
DeepCELL-v1	Dyskeratotic	95.783	97.546	96.657
	Koilocytotic	91.515	**91.515**	91.515
	Metaplastic	96.026	91.195	93.548
	Parabasal	**100**	**100**	**100**
	Superficial/Intermediate	94.186	97.006	95.575
DeepCELL-v2	Dyskeratotic	**97.561**	**98.160**	**97.859**
	Koilocytotic	**96.732**	89.697	**93.082**
	Metaplastic	**97.436**	95.597	**96.508**
	Parabasal	**100**	**100**	**100**
	Superficial/Intermediate	91.160	**98.802**	94.828
DeepCELL-v3	Dyskeratotic	96.319	96.319	96.319
	Koilocytotic	95.395	87.879	91.483
	Metaplastic	92.216	**96.855**	94.479
	Parabasal	**100**	99.367	99.683
	Superficial/Intermediate	94.220	97.605	95.882

DeepCELL-v1 can identify 158 from 158 Parabasal cells, which is perfect. It also accurately detects 159 out of 163 cells, 151 out of 165 cells, 145 out of 159 cells, and 162 out of 167 cells on Dyskeratotic, Koilocytotic, Metaplastic and Superficial/Intermediate categories, respectively. Similarly, our DeepCELL-v2 can detect 158 from 158 Parabasal cells. That is to say, Parabasal cells are detected with the highest performance. Our DeepCELL-v2 successfully predicts 160 among 163 Dyskeratotic cells, 148 among 165 Koilocytotic cells, 152 among 159 Metaplastic cells, and 165 among 167 Superficial/Intermediate cells. It is obvious that our designed model has more potential to classify Dyskeratotic and Parabasal cells compared to other three cervical cell categories. However, our DeepCELL model can generally identify all five cervical cell categories well.

The performance of our method and other competing methods on two Pap smear datasets is shown in Table 6, which is obtained by averaging the results of ten experiments.

Figures 5 and 6 present the evaluation metrics histograms for each method to present the performance more directly. As shown in Figure 5 and Table 6, our DeepCELL outperforms other methods in all evaluation metrics on Herlev

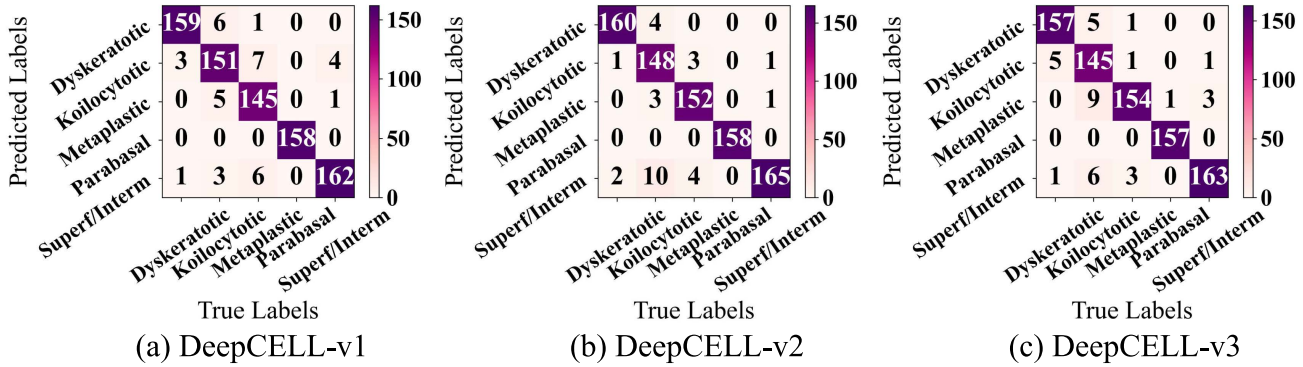


FIGURE 4. Confusion matrix of our method with three strategies on SIPaKMeD dataset.

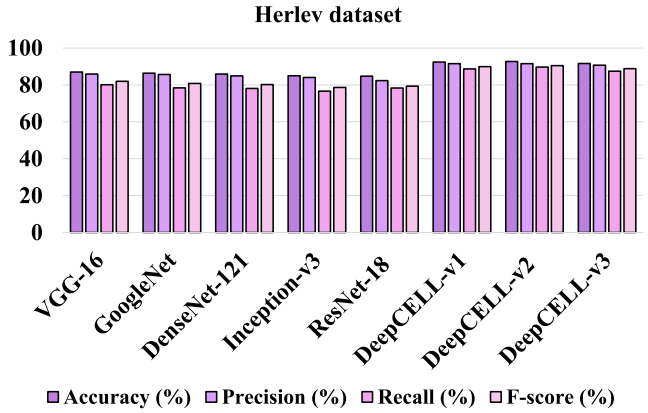


FIGURE 5. Overall performance comparison of different methods on Herlev dataset.

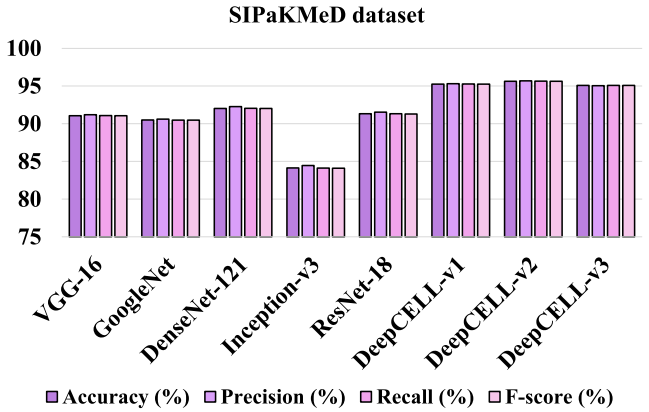


FIGURE 6. Overall performance comparison of different methods on SIPaKMeD dataset.

dataset after running models ten times and then taking the average. Specifically, the performance of VGG-16 is the best among the benchmark methods. Accuracy, precision, recall and F-score of VGG-16 are 86.957%, 85.923%, 80.125% and 81.954% respectively. Our DeepCELL achieves the highest results including an accuracy of 92.717%, a precision of 91.565%, a recall of 89.707%, and a F-score of 90.475% respectively. The best accuracy of 92.717% is acquired by our DeepCELL-v2, which is almost 6% higher than the next-best VGG-16 and approximately 8% higher than ResNet-18. Similarly, our method achieves an F-score of 90.475%, which is better than the second highest F-score of 81.954%. In addition, although there is a slight difference in the results of our model with three versions, they perform better than those competing methods. Obviously, the performance of our method is relatively stable for binary classification of abnormal and normal cervical cells.

As shown in Figure 6 and Table 6, our DeepCELL is superior to those competing methods in all evaluation metrics on SIPaKMeD dataset based on the average of ten experiments. As we can see, DenseNet-121 achieves the accuracy of 92.032%, precision of 92.264%, recall of 92.036% and F-score of 92.025%, which is the most remarkable result compared to other methods. Among those metrics, our DeepCELL produces the highest accuracy of 95.628%, which is around 3.6% higher than the second best DenseNet-121 and about 11.5% higher than Inception-v3. Similarly, our DeepCELL yields the great performance with the F-score of 95.636%, which is 3.611% higher than the second best model. Also, three versions of our DeepCELL are not very different for their performance in each evaluation metric even though the results of DeepCELL-v1 and DeepCELL-v3 are slightly lower than that of DeepCELL-v2. This clearly illustrates that our DeepCELL is stable enough for multi-classification of cervical cells.

Through the above analysis, we can find that our proposed DeepCELL method has the best performance on both Herlev and SIPaKMeD datasets comparing with other methods. In particular, our DeepCELL achieves the highest classification accuracy and F-score, all of which reach more than 90% (highlighted in bold in Table 6). The results of our DeepCELL far exceed other competing methods in all evaluation metrics.

To further show the cell classification ability of our proposed method, we compare it with some existing methods on Herlev and SIPaKMeD datasets in terms of accuracy, precision, recall and F-score, respectively. Table 7 displays the comparison of classification performance among different methods on Herlev dataset. It can be concluded from

TABLE 6. Performance of cervical cell classification methods on two Pap smear datasets.

Datasets	Models	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
Herlev	VGG-16	86.957	85.923	80.125	81.954
	GoogleNet	86.359	85.697	78.418	80.802
	DenseNet-121	85.924	84.891	78.122	80.225
	Inception-v3	84.946	84.084	76.610	78.617
	ResNet-18	84.783	82.385	78.319	79.368
	DeepCELL-v1	92.391	**91.565**	88.704	89.923
	DeepCELL-v2	**92.717**	91.545	**89.707**	**90.475**
	DeepCELL-v3	91.630	90.660	87.471	88.871
SIPaKMeD	VGG-16	91.071	91.186	91.082	91.054
	GoogleNet	90.493	90.603	90.486	90.476
	DenseNet-121	92.032	92.264	92.036	92.025
	Inception-v3	84.138	84.464	84.118	84.097
	ResNet-18	91.318	91.539	91.327	91.280
	DeepCELL-v1	95.259	95.305	95.278	95.262
	DeepCELL-v2	**95.628**	**95.685**	**95.647**	**95.636**
	DeepCELL-v3	95.074	95.051	95.085	95.086

Table 7 that the proposed DeepCELL has better classification performance than other methods in general. In particular, the proposed DeepCELL-v2 outperforms other methods in both accuracy and F-score. Specifically, DeepCELL-v2 obtains an improvement of 6.25%, 3.81% and 3.41% respectively compared with others in F-score. The method of Liu et al. [45] reaches the lowest F-score of 84.23% among all competing methods. The method of Bhatt et al. [46] exhibits the lowest accuracy of 78.142% compared to other methods. Our DeepCELL-v2 achieves an improvement of 4.69% and 14.58% in accuracy than the methods proposed by Liu et al. [45] and Bhatt et al. [46], respectively. Our DeepCELL gains the higher value in precision, which means that it correctly identifies more positive cells from all the predicted ones. Besides, our DeepCELL-v1 achieves the second highest accuracy with only 0.326% lower than DeepCELL-v2. Our DeepCELL-v2 gets better performance in terms of precision and F-score, even though it produces only a slightly higher value in accuracy compared to RF (Random Forest) [47]. It is clear that our proposed DeepCELL improves the performance of cervical cell recognition.

Table 8 presents the comparison of the proposed method with existing studies applied on SIPaKMeD dataset. The proposed DeepCELL-v2 gains the best performance with an accuracy and F-score of 95.628% and 95.636%, respectively. The method proposed by Win et al. [11] in 2020 yields the accuracy of 94.09%. On the contrary, our DeepCELL-v2 achieves the highest accuracy of 95.628%, which is 1.54% higher than the method of Win et al. [11]. The higher accuracy suggests that the model effectively detects the total cells. Our DeepCELL-v2 is better than the second best method [35] proposed in 2021 with an improvement of less than 1.0% in terms of all metrics. However, this ensemble method is more complex than ours. The method [35] adopts an ensemble learning technique, which combines several base models to make the final decision thus outperforming than the individual base models. Additionally, although our proposed DeepCELL-v1 is slightly lower than DeepCELL-v2 among all metrics, it also obtains the decent classification performance.

TABLE 7. Comparison of classification performance among different methods on Herlev dataset.

Method	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
RF	92.040	80.000	95.500	87.066
Bhatt et al.	78.142	79.268	**95.588**	86.666
Liu et al.	88.030	-	-	84.230
DeepCELL-v1	92.391	**91.565**	88.704	89.923
DeepCELL-v2	**92.717**	91.545	89.707	**90.475**
DeepCELL-v3	91.630	90.660	87.471	88.871

TABLE 8. Comparison of the proposed method with existing methods on SIPaKMeD dataset.

Method	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
Ensemble	95.430	95.340	95.380	95.360
Win et al.	94.090	-	-	-
DeepCELL-v1	95.259	95.305	95.278	95.262
DeepCELL-v2	**95.628**	**95.685**	**95.647**	**95.636**
DeepCELL-v3	95.074	95.051	95.085	95.086

TABLE 9. Ablation study of the proposed method on Herlev dataset.

Method	Run Time (s)	Params Size (MB)	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
[41]	1438	29.87	89.423	86.767	85.801	86.211
DeepCELL([41] + 5 × 5 conv)	2927	148.93	92.587	91.730	88.994	90.174
DeepCELL(replace with 3 × 3 conv)	3334	147.75	91.497	90.048	87.737	88.770

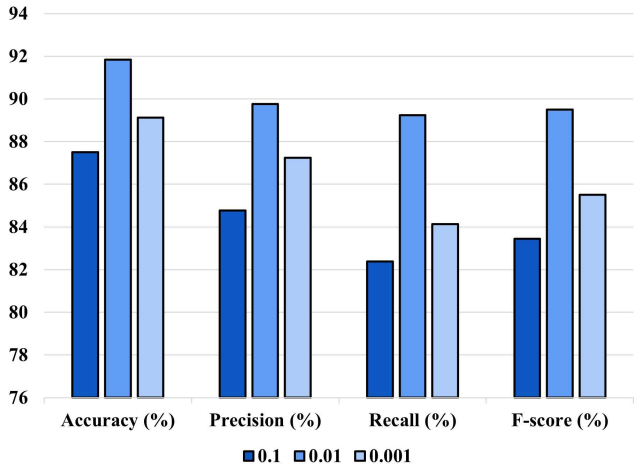


FIGURE 7. Comparison of different initial learning rates for our method.

V. DISCUSSION

A. PARAMETER SELECTION

In this subsection, we compare different initial learning rates using our models on the Herlev dataset in Figure 7. It can be seen that our model has the best classification performance in terms of accuracy, precision, recall and F-score when the initial learning rate is set to 0.01. Therefore, we select 0.01 as the initial learning rate in this study.

B. ABLATION STUDY

To further show the performance of our method, we apply five-fold cross validation strategy to train our method on the Herlev dataset. In five-fold cross validation, four of five folds are used as training set, and the remaining one as validation set. This process is repeated for five times, and the final performance of the methods are obtained by averaging results from five validation sets.

The structure of our models is inspired by [41], which mainly consists of 1 × 1 and 3 × 3 convolutions. To learn features with multiple size kernels from cervical cell images, we add 5×5 convolutions based on the model from [41] in our model to capture different feature representations. It means that our model mainly consists of 5 × 5, 3 × 3 and 1 × 1 convolutions. Specifically, we compare our method with the method from [41] with different convolutional layers using five-fold cross validation on the Herlev dataset. As shown in Table 9, our method achieves better performance than the method [41], which can demonstrate that adding extra 5 × 5 convolutions is effective in cervical cell classification.

In order to reduce the number of parameters in our method, we also adopt another strategy, which is to use two 3 × 3 convolutions to replace one 5 × 5 convolution. From Table 9, we can find that the parameter size is reduced after replacing one 5 × 5 convolution with two 3 × 3 convolutions in the proposed DeepCELL method. Overall, our models have higher computational costs than the model from [41], but they obtain better performance in accuracy, precision, recall and F-score. This can also illustrate the effectiveness of our DeepCELL on cervical cell classification task.

C. LIMITATIONS OF THE PROPOSED METHOD

The proposed method has good performance in cervical cell classification, but also has some limitations. Firstly, the proposed method has only been applied to cervical cell image classification in this study and has not been implemented in other tasks. In addition, our method has more abilities to classify abnormal cells than normal cells on the Herlev dataset. Hence, the ability of our method to classify normal cells needs to be improved.

VI. CONCLUSION

More and more researchers have recently tended to design highly efficient deep neural networks with great accuracy for cervical cell detection. Therefore, we develop an end-to-end automated cervical cell detection model with three different versions, which can extract deep feature information via multiple size kernels from cell images. Then, we assess the performance of our designed model on two cervical cell image datasets, Herlev and SIPaKMeD. Experimental performance of our DeepCELL is better than that of other competing CNN models for cervical cell detection on both datasets, including accuracy, precision, recall and F-score. Moreover, our proposed method achieves a certain improvement in cervical cell detection compared with other existing methods applied to two benchmark cervical cytology datasets, which illustrates that our DeepCELL is robust and comparable to the state-of-the-art approaches. The superiority of our proposed method is further illustrated through the comparison of cervical cell classification performance among different methods on two datasets. It has evidenced that the exploration of features via kernels with multiple sizes contributes to improve the detection ability of cervical cells to a certain extent. The purpose of our study is to provide a segmentation-free and highly accurate classification model for cervical cells. We hope that the construction of our DeepCELL can assist physicians and cytologists in cervical cancer screening to a certain extent.

REFERENCES

[1] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA, Cancer J. Clinicians*, vol. 61, no. 2, pp. 69–90, 2011.

[2] *World Health Organization*. Accessed: Nov. 2, 2021. [Online]. Available: https://www.who.int/health-topics/cervical-cancer

[3] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021.

[4] M. Kuko and M. Pourhomayoun, "Single and clustered cervical cell classification with ensemble and deep learning methods," *Inf. Syst. Frontiers*, vol. 22, no. 5, pp. 1039–1051, Oct. 2020.

[5] 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.

[6] W. E. Tolles and R. C. Bostrom, "Automatic screening of cytological smears for cancer: The instrumentation," *Ann. New York Acad. Sci. USA*, vol. 63, no. 6, pp. 1211–1218, Mar. 1956.

[7] S. Watanabe, "An automated apparatus for cancer prescreening: CYBEST," *Comput. Graph. Image Process.*, vol. 3, no. 4, pp. 350–358, Dec. 1974.

[8] D. J. Zahniser, P. S. Oud, M. C. Raaijmakers, G. P. Vooys, and R. T. Van De Walle, "BioPEPR: A system for the automatic prescreening of cervical smears.," *J. Histochem. Cytochem.*, vol. 27, no. 1, pp. 635–641, Jan. 1979.

[9] L. J. Mango, "Computer-assisted cervical cancer screening using neural networks," *Cancer Lett.*, vol. 77, nos. 2–3, pp. 155–162, Mar. 1994.

[10] S. J. Bernstein, L. Sanchez-Ramos, and B. Ndubisi, "Liquid-based cervical cytologic smear study and conventional papanicolaou smears: A meta-analysis of prospective studies comparing cytologic diagnosis and sample adequacy," *Amer. J. Obstetrics Gynecology*, vol. 185, no. 2, pp. 308–317, Aug. 2001.

[11] K. P. Win, Y. Kitjaidure, K. Hamamoto, and T. M. Aung, "Computer-assisted screening for cervical cancer using digital image processing of pap smear images," *Appl. Sci.*, vol. 10, no. 5, p. 1800, Mar. 2020.

[12] M. M. Rahaman, C. Li, X. Wu, Y. Yao, Z. Hu, T. Jiang, X. Li, and S. Qi, "A survey for cervical cytopathology image analysis using deep learning," *IEEE Access*, vol. 8, pp. 61687–61710, 2020.

[13] X. Cao, J. Yao, Z. Xu, and D. Meng, "Hyperspectral image classification with convolutional neural network and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4604–4616, Jul. 2020.

[14] Q. Xu, Y. Zeng, W. Tang, W. Peng, T. Xia, Z. Li, F. Teng, W. Li, and J. Guo, "Multi-task joint learning model for segmenting and classifying tongue images using a deep neural network," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 9, pp. 2481–2489, Sep. 2020.

[15] T. Wollmann and K. Rohr, "Deep consensus network: Aggregating predictions to improve object detection in microscopy images," *Med. Image Anal.*, vol. 70, May 2021, Art. no. 102019.

[16] S. Lee, S. Joo, H. K. Ahn, and S.-O. Jung, "CNN acceleration with hardware-efficient dataflow for super-resolution," *IEEE Access*, vol. 8, pp. 187754–187765, 2020.

[17] J. Mas, T. Panadero, G. Botella, A. A. D. Barrio, and C. García, "CNN inference acceleration using low-power devices for human monitoring and security scenarios," *Comput. Electr. Eng.*, vol. 88, Dec. 2020, Art. no. 106859.

[18] J. Kim, J.-K. Kang, and Y. Kim, "A low-cost fully integer-based CNN accelerator on FPGA for real-time traffic sign recognition," *IEEE Access*, vol. 10, pp. 84626–84634, 2022.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 25, Dec. 2012, pp. 1097–1105.

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[24] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[25] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[26] M. Byra, K. Dobruch-Sobczak, Z. Klimonda, H. Piotrzkowska-Wroblewska, and J. Litniewski, "Early prediction of response to neoadjuvant chemotherapy in breast cancer sonography using Siamese convolutional neural networks," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 3, pp. 797–805, Mar. 2021.

[27] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. De Vries, M. J. N. L. Benders, and I. Isgum, "Automatic segmentation of MR brain images with a convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1252–1261, May 2016.

[28] Z. Gao, L. Wang, L. Zhou, and J. Zhang, "HEp-2 cell image classification with deep convolutional neural networks," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 2, pp. 416–428, Mar. 2017.

[29] M. Xu, D. P. Papageorgiou, S. Z. Abidi, M. Dao, H. Zhao, and G. E. Karniadakis, "A deep convolutional neural network for classification of red blood cells in sickle cell anemia," *PLOS Comput. Biol.*, vol. 13, no. 10, Oct. 2017, Art. no. e1005746.

[30] P. Huang, J. Wang, J. Zhang, Y. Shen, C. Liu, W. Song, S. Wu, Y. Zuo, Z. Lu, and D. Li, "Attention-aware residual network based manifold learning for white blood cells classification," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 4, pp. 1206–1214, Apr. 2021.

[31] W. Chen, L. Gao, X. Li, and W. Shen, "Lightweight convolutional neural network with knowledge distillation for cervical cells classification," *Biomed. Signal Process. Control*, vol. 71, Jan. 2022, Art. no. 103177.

[32] T. L. Mahyari and R. M. Dansereau, "Multi-layer random Walker image segmentation for overlapped cervical cells using probabilistic deep learning methods," *IET Image Process.*, vol. 16, no. 11, pp. 2959–2972, Sep. 2022.

[33] J. Huang, G. Yang, B. Li, Y. He, and Y. Liang, "Segmentation of cervical cell images based on generative adversarial networks," *IEEE Access*, vol. 9, pp. 115415–115428, 2021.

[34] Y. Zhao, C. Fu, S. Xu, L. Cao, and H.-F. Ma, "LFANet: Lightweight feature attention network for abnormal cell segmentation in cervical cytology images," *Comput. Biol. Med.*, vol. 145, Jun. 2022, Art. no. 105500.

[35] 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, Jul. 2021.

[36] A. Khamparia, D. Gupta, V. H. C. De Albuquerque, A. K. Sangaiah, and R. H. Jhaveri, "Internet of health things-driven deep learning system for detection and classification of cervical cells using transfer learning," *J. Supercomput.*, vol. 76, no. 11, pp. 8590–8608, 2020.

[37] M. M. Rahaman, C. Li, Y. Yao, F. Kulwa, X. Wu, X. Li, and Q. Wang, "DeepCervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104649.

[38] 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 Informat.*, vol. 21, no. 6, pp. 1633–1643, Nov. 2017.

[39] J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, "Pap-smear benchmark data for pattern classification," in *Proc. Nature Inspired Smart Inf. Syst.*, 2005, pp. 1–9.

[40] 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 *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3144–3148.

[41] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "RepVGG: Making VGG-style ConvNets great again," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13733–13742.

[42] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.

[43] S. C. Kosaraju, J. Hao, H. M. Koh, and M. Kang, "Deep-Hipo: Multi-scale receptive field deep learning for histopathological image analysis," *Methods*, vol. 179, pp. 3–13, Jul. 2020.

[44] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," in *Wavelets*. Berlin, Germany: Springer, 1990, pp. 286–297.

[45] D. Liu, S. Wang, D. Huang, G. Deng, F. Zeng, and H. Chen, "Medical image classification using spatial adjacent histogram based on adaptive local binary patterns," *Comput. Biol. Med.*, vol. 72, pp. 185–200, May 2016.
[46] 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, Feb. 2021.
[47] F. B. Akyol and O. Altun, "Detection of cervix cancer from pap-smear images," *Sakarya Univ. J. Comput. Inf. Sci.*, vol. 3, no. 2, pp. 99–111, 2020.

MING FANG received the B.Sc. degree from the North University of China, Taiyuan, China, in 2016, and the M.Sc. degree in computer application technology from Shaanxi Normal University, Xi'an, China, in 2019. She is currently pursuing the Ph.D. degree with the Division of Biomedical Engineering, University of Saskatchewan, Saskatoon, Canada. Her research interests include bioinformatics, medical image analytics, machine/deep learning, and big biomedical data analytics.

XIUJUAN LEI (Member, IEEE) received the M.S. and Ph.D. degrees from Northwestern Polytechnical University, Xi'an, China, in 2001 and 2005, respectively. She was a Visiting Scholar at The State University of New York at Buffalo, Buffalo, NY, USA, from 2009 to 2010, Peking University, from 2013 to 2014, and Georgia State University, Atlanta, GA, USA, from 2019 to 2020. She is currently a Professor with the School of Computer Science, Shaanxi Normal University, Xi'an. Her research interests include bioinformatics, intelligent computing, data mining, and artificial intelligence.

BO LIAO received the Ph.D. degree in computational mathematics from the Dalian University of Technology, Dalian, China, in 2004. He was a Postdoctoral Fellow with the University of Chinese Academy of Sciences, Beijing, China, from 2004 to 2006. He is currently working as a Professor and the Dean of the School of Mathematics and Statistics, Hainan Normal University, Haikou, China. He has authored more than 130 papers in international conferences and journals. His research interests include data mining, machine learning, big data processing, and bioinformatics.

FANG-XIANG WU (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in applied mathematics from the Dalian University of Technology, in 1990 and 1993, respectively, the first Ph.D. degree in control theory and its applications from Northwestern Polytechnical University, in 1998, and the second Ph.D. degree in bioinformatics and computational biology from the University of Saskatchewan (U of S), in August 2004. From September 2004 to August 2005, he worked as a Postdoctoral Fellow with the Biomedical Research Center, Laval University. He is currently a Full Professor with the Department of Computer Science, Division of Biomedical Engineering, and the Department of Mechanical Engineering, U of S. His research interests include artificial intelligence, machine/deep learning, computational biology, health informatics, medical image analytics, and complex network analytics. He is serving as an editorial board member for several international journals, a guest editor for numerous international journals, and the program committee chair or a member for many international conferences.