IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 24, 2019, accepted May 21, 2019, date of publication May 27, 2019, date of current version June 12, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2919390

Fine-Grained Classification of Cervical Cells Using Morphological and Appearance Based Convolutional Neural Networks

HAOMING LIN[1,2,3], YUYANG HU[1,2,3], SIPING CHEN[1,2,3], JIANHUA YAO[4], AND LING ZHANG[5]

[1]School of Medicine, Shenzhen University, Shenzhen 518060, China
[2]National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Shenzhen 518060, China
[3]Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Shenzhen 518060, China
[4]Tencent Holdings Limited, Shenzhen 518057, China
[5]Nvidia Corporation, Bethesda, MD 20814, USA

Corresponding author: Ling Zhang (zhangling0722@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 81601510 and Grant 81501545, in part by the National Natural Science Foundation of Guangdong Province under Grant 2016A030310047, in part by the Medical Science and Technology Research Foundation of Guangdong Province under Grant A2019107, and in part by the New teacher natural science research project of Shenzhen University under Grant 2018011.

ABSTRACT Fine-grained classification of cervical cells into different abnormality levels is of great clinical importance but remains very challenging. Contrary to the traditional classification methods that rely on hand-crafted or engineered features, convolution neural network (CNN) can classify cervical cells based on automatically learned deep features. However, CNN in previous studies does not involve cell morphological information, and it is unknown whether morphological features can be directly modeled by CNN to classify cervical cells. This paper presents a CNN-based method that combines cell image appearance with cell morphology for classification of cervical cells in Pap smear. The training of cervical cell dataset consists of adaptively re-sampled image patches coarsely centered on the nuclei. Several CNN models (AlexNet, GoogLeNet, ResNet, and DenseNet) pre-trained on ImageNet dataset were fine-tuned on the cervical dataset for comparison. The proposed method is evaluated on the Herlev cervical dataset by five-fold cross-validation at patient-level splitting. The results show that by adding cytoplasm and nucleus masks as raw morphological information into appearance-based CNN learning, higher classification accuracies can be achieved in general. Among the four CNN models, GoogLeNet fed with both morphological and appearance information obtains the highest classification accuracies of 94.5%, 71.3%, and 64.5%, for two-class (abnormal versus normal), four-class (''The Bethesda System''), and seven-class (''World Health Organization classification system'') classification tasks, respectively. Our method demonstrates that combining cervical cell morphology with appearance information can provide improved classification performance. Although the initial results are promising, deep learning-based fine-grained cervical cell classification remains a very challenging task for a high-precision diagnosis.

INDEX TERMS Fine-grained classification, cell morphology, deep learning, Pap smear.

I. INTRODUCTION

Cervical cancer is one of the most common lethal malignant disease among woman [1]. The greatest factor for cervical cancer is the infection with some types of human papilloma virus (HPV) which may lead to dysplastic changes in cells before development of cervical cancer [2]. These dysplastic changes of cells typically develop over a prolonged process and refer to a wide spectrum of abnormality. Pap smear, one of the most popular screening tests for prevention and early detection of cervical cancer, has been extensively used in developed countries and credited with reducing the mortality rate of cervical cancer significantly [3]. However, population-wide screening is still not widely available in developing countries [3], partly due to the tedious and complexity nature of manually screening of the abnormal cells from a cervical cytology specimen [4]. Such diagnosis is also subject to error even for experienced doctors [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Junxiu Liu.

To address these concerns, automation-assisted reading systems have been developed to improve efficiency and increase availability over the past few decades. These automation-assisted reading systems are based on automated image analysis techniques [4]–[6], which automatically select potentially abnormal cells from a given cervical cytology slide for further review and fine-grained classification by the cytoscreener or cytopathologist. According to the World Health Organization classification system, premalignant dysplastic changes of cells can include four stages, which are mild, moderate, severe dysplasia and carcinoma in situ (CIS) [7]. The lesions are generally no more than manifestations of HPV infection for the mild stage, but the risk of progression to cancer is relatively high for the more severe stages if not detected and treated [8]. Early staging of dysplastic changes is important for preventing the developments of precancerous cells. It is known that such a task is highly challenging and subjective. A misclassification may either cause unnecessary biopsy (e.g., classify mild as CIS) or treatment delay (e.g., classify CIS as moderate). Therefore, fine-grained classification of cervical cells is highly desired in real clinical diagnosis practice. However, almost all previous studies of cervical cell classification focus on classification of cervical cells into abnormal and normal groups, which is useful for screening, but not enough for diagnosis [4]–[6], [9].

Morphological cell morphology has been widely used for computerized cell image processing and pattern recognition in biomedical applications, such as nuclei feature quantification for cancer cell cycle analysis [10], hepatocellular carcinoma feature extraction [11] and automated classification of blood cell [12]. For cervical cell application, automation-assisted reading system generally comprises three steps: cell segmentation, feature extraction/selection, and cell classification. The feature, especially morphological feature design and selection are also one of the most important factors for cervical cell classification. When dysplastic changes happen, cervical cells undergo various morphological changes which include changes in terms of size, shape, intensity and texture. Thus, feature descriptors are designed to describe these changes. In study of [13], twenty morphology-related features are extracted for cervical cell classification. Automatic method for cell nuclei detection in pap smear images based on morphological analysis is proposed in [14]. Previously, the extracted features can be grouped into handcrafted features [13], [15], [16] and engineering features [9], [17]. However, Handcrafted features are hindered by limited understanding of cervical cytology. Engineering features are derived from an unsupervised manner, and thus encode redundant information. The feature selection process potentially ignores significant clues and removes complementary information.

In the past few years, deep convolutional neural networks (CNN) have proven to be great success in many computer vision tasks when training on large-scale annotated datasets (i.e. ImageNet) [18]. In contrast to classical machine learning methods that use a series of handcrafted features, CNNs automatically learn multi-level features from the training data set. As the development of more powerful hardware with higher computing power (i.e., Graphics Processing Units, GPUs), CNN architecture has become more and more deep and complicated. A variety of CNN models have been introduced in the literatures, such as LeNet [19], AlexNet [20], GoogLeNet [21], ResNet [22], DenseNet [23] and their variants and so on. The original LeNet only consists of 5 layers while the ResNet has already surpassed 100 layers, even reach to more than 1000 layers. In addition to increase depth directly, the GoogLeNet introduces an inception module, which concatenates feature-maps produced by filters of different sizes, to make the network wider and deeper. ResNets have achieved state-of-the-art performance on many challenging image recognition, localization, and detection tasks, such as ImageNet object detection. Large amounts of labeled data are crucial to the performance of CNN. However, the labeled data is limited for cervical cells images because high quality annotation is costly and challenging even if for experts. Fortunately, transfer learning [24] is an effective method to address this problem. CNNs have already significantly improved performance in various medical imaging analysis applications [25]–[29]. But it is still unclear which network or what is the best network depth and width for cervical cells classification given limited training data.

Besides being directly used as a classifier, CNNs can be used as feature selectors. When training with large-scale data, low-to-high-level features of data can be obtained from the shallow convolutional layer to the deep convolutional layer of CNN. Learned features extracted from pre-trained model can be combined with existing handcrafted features, such as local binary pattern (LBP), Histogram of Oriented Gradient (HOG), and then fed to other classifier (e.g. support vector machine, SVM) [30]. In study of [31], cell dynamic morphology is classified by CNN to represent different cell physiological states. However, it is still unknown whether morphological features can be directly modeled by CNN to classify cervical cells. Although CNN has recently been used to directly classify cervical cells in a recent study [32], several problems need more investigation: 1) it only involve raw RGB information; 2) it only evaluates classification performance of 2-class (normal vs. abnormal) task, while 4-class or 7-class are more challenging and more desirable; 3) it only evaluates AlexNet which may not represent the capability of the state-of-art deep classification models; 4) it's worth pointing out that the previous method with five-fold cross-validation (CV) does not guarantee patient-level separation on the Herlev dataset, which does not meet the real clinical setting.

In this paper, we present a CNN-based method that combines cell image appearance with cell morphology for classification of cervical cells in Pap smear. In our approach, cell morphology was directly represented by cytoplasm and nucleus binary masks, which were then combined with raw RGB-channels of the cell image to form a five-channel image, on which training data was sampled from a square image patch coarsely centered on the nucleus (left part

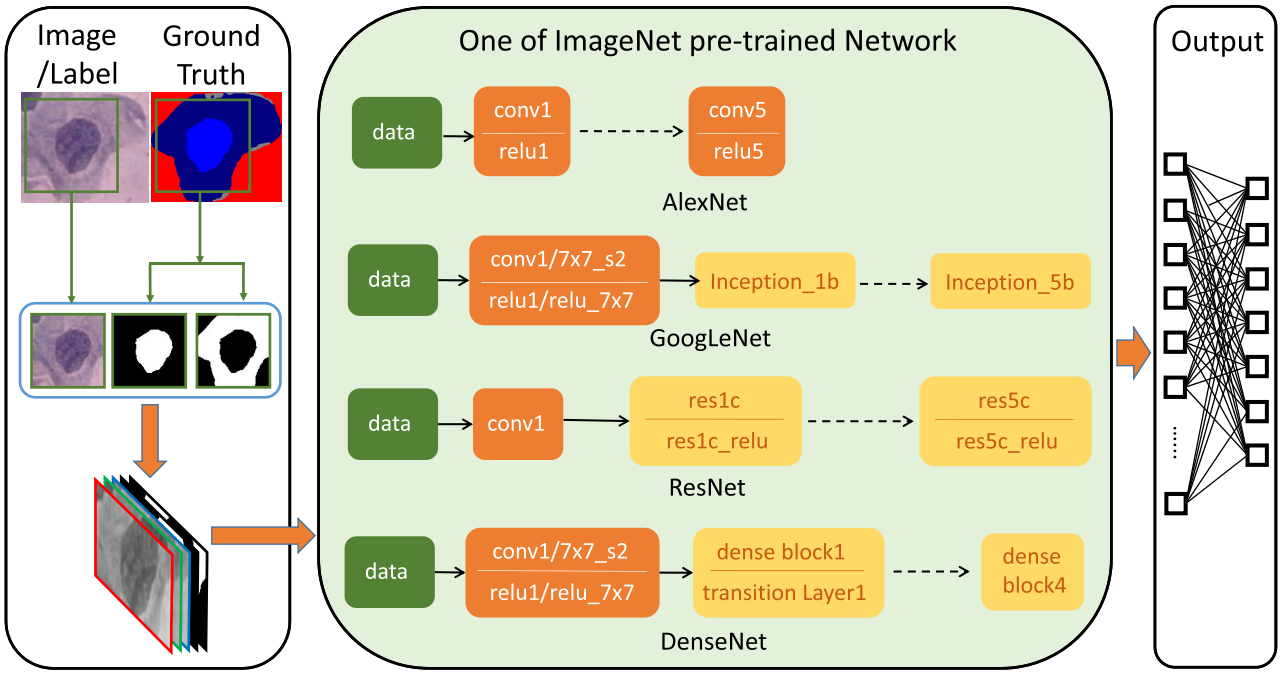


FIGURE 1. The overall flow-chart of our CNN framework for 7-class classification problem.

in Fig.1). Then this dataset was fed into CNNs for classification of cervical cell. Different CNN models (AlexNet, GoogLeNet, ResNet and DenseNet) pre-trained on ImageNet dataset were fine-tuned for 2-class (abnormal vs. normal) and 7-class ("World Health Organization classification system") cervical cell classification based on deep hierarchical features. Note that the 4-class ("The Bethesda System (TBS)") result is derived from 7-class by combining some of the subcategories.

The Herlev dataset consisting of 917 cervical cell images is used to test our method. We carefully split the cells to perform five-fold cross-validation at patient-level – this means that all the cells from the same patient will be assigned to training set or validation set alone. Experimental results demonstrated that by adding cytoplasm and nucleus masks as raw morphological information into conventional appearance-based CNN learning, higher classification accuracies can be achieved in general. Among the four CNN models, GoogLeNet fed with both morphological and appearance information obtains the highest classification accuracies of 94.5% for 2-class classification task, 64.5% for 7-class classification task and 71.3% for 4-class classification task.

Our main contributions are summarized as follows: 1) the combination of raw cytoplasm and nucleus binary masks and RGB appearance was proposed for CNN-based cervical cell classification. 2) State-of-the-art CNN models were fine-tuned to evaluate and compare the performances of cervical cell classification at patient-level cell splitting. 3) Besides distinguishing normal and abnormal cervical cells, the performances of 7-class and 4-class fine-grained classification of cervical cell were also investigated.

II. METHODS

In this study, the cervical cell images which concatenate cytoplasm/nucleus binary masks and raw RGB channels were fed into CNNs, and both the 2-class and 7-class classification performances of state-of-the-art CNN models were evaluated and analyzed. The details are described as below.

A. DATA PREPROCESSING

1) IMAGE PATCH AND CELL MORPHOLOGY EXTRACTION

As mentioned in TBS rules, cervical cells can be categorized into four classes: normal, Low grade Squamous Intraepithelial Lesion (LSIL), High grade Squamous Intraepithelial Lesion (HSIL) and Carcinoma-in-situ (CIS) [33]. Different stages of cervical cytology abnormalities are associated with different nucleus characteristics. Therefore, nucleus features in themselves already include substantial discriminative information. Since the main topic of this work is CNN classification, we follow the strategy in [32] to extract training samples. Specifically, image patches of size $m \times m$ centered on the nucleus centroid and included a certain amount of cytoplasm were extracted. This strategy allows for embedding not only the nucleus scale/size information, but also the contextual clues in the extracted patches. Although there are methods for automated extraction of nucleus, we only focus on the classification task in this paper. We directly use the centroid of ground truth mask of nucleus to extract the image patches,

and the corresponding morphology of nucleus and cytoplasm can be obtained directly from the ground truth mask.

2) DATA AUGMENTATION

Data augmentation is critical to improve the accuracy of CNNs and reducing overfitting. Because cervical cells are rotationally invariant, each cell image is performed N_r rotations with an angle step of θ degree. Zero padding is also used to avoid region that lies outside of the image boundary. Considering that detecting the centroid of the nucleus may be inaccurate in practice, the centroid of each nucleus was randomly translated (by up to d pixels) N_t times to obtain N_t coarse nucleus centers. Accordingly, N_t patches of size $m \times m$ centered at these locations are extracted. These patches not only simulate inaccurate nucleus center detection, but also increase the amount of image samples for CNNs. Other data augmentation approaches such as scale and color transformations are not utilized, because 1) the concern that the abnormality may be changed by changing the cell intensity/staining, e.g., a moderate large nucleus with dark staining tends to be abnormal but can be normal if with light staining; 2) adding color transformation may improve (or not) the accuracy on this dataset, but may result in lower robustness on new data.

Note that the distribution of different types of cells in Herlev dataset is imbalance, so classifiers have a tendency to exhibit bias towards the majority classes. For example, the amount of abnormal cell images is larger than that of normal cell images in Herlev dataset. In order to balance the proportions of positive and negative samples, we apply a higher sampling proportion to the normal patches. For 7-class task, similar sampling methods are utilized for balancing.

B. CONVOLUTIONAL NEURAL NETWORK ARCHITECTURES

CNN is a deep learning model in which multiple stages of convolution, non-linearity and pooling layers are stacked, followed by more convolutional and fully-connected layers. In our experiments, we mainly explore four convolutional neural network models (AlexNet, GoogLeNet, ResNet and DenseNet) which are shown in the green part in fig.1. The input of CNNs is image patch with five-channels which includes two channels of binary masks of the cervical nucleus and cytoplasm and three-channels of raw RGB image (left part in fig.1). To demonstrate the additive value of cell morphological features, raw RGB image is used as the only input of CNNs for performance comparison. The output layer of CNNs comprises of several neurons each corresponding to one class. In our case, there are 2 and 7 neurons in the output layer for 2-class and 7-class classification tasks, respectively. The backpropagation algorithm is used to minimize the classification error on the training dataset for optimization of weight parameters in CNNs.

AlexNet [20]: AlexNet is the winner of ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 and has received extensive attention in computer vision. ImageNet dataset consists of 1.2 million 256×256 images belong to 1000 categories. AlexNet contains five convolution layers, three pooling layers, and three full-connected layers. AlexNet achieves 15.3% top-5 classification error.

GoogLeNet [21]: GoogLeNet is more complex and deeper than AlexNet and is the winner of ImageNet ILSVRC 2014. GoogLeNet introduces a new module named ''inception'', which concatenates filters of different sizes and dimensions into a single new filter. Overall, GoogLeNet has two convolution layers, two pooling layers, and nine ''inception'' layers. Each ''inception'' layer consists of six convolution layers and one pooling layer. GoogLeNet obtains 6.67% top-5 classification error on ImageNet dataset challenge.

ResNet [22]: ResNet is about 20 times deeper than AlexNet. ResNet utilizes shortcut connections to jump over some layers to avoid the problem of vanishing gradient. ResNet wins the ImageNet ILSVRC 2015, and have achieved impressive, record-breaking performance on many challenging image recognition, localization, and detection tasks [22].

DenseNet [23]: Similar but different from ResNets, direct connections from any layer to all subsequent layers are introduced in DenseNets, which encourages feature reuse throughout the network. Moreover, the DenseNets can achieve state-of-the-art performances with substantially fewer parameters and less computation than ResNet.

C. TRANSFER LEARNING

Transfer learning refers to the fine-tuning of deep learning models that are pre-trained on other large-scale image datasets. Due to limited cervical image data in this study, for each CNN architecture, pre-trained models trained on ImageNet dataset are used as the basis of our network, where the weights of the first convolution layer and last several task-specific full-connection layers are randomly initialized, and other network layers are transferred to the same locations of our model. All of these layers in our models are jointly trained on our cervical cell dataset. Note that only the first convolution layer and the last several task-specific full-connection layers are trained from scratch.

In testing, the random-view aggregation and multiple crop testing are used following the approach in [32].

III. EXPERIMENTS AND RESULTS

A. DATA SET

The utilized cervical cell data is publicly available (`http://mdelab.aegean.gr/downloads`), which is collected at the Herlev University Hospital by a digital camera and microscope [13]. The image resolution is 0.201 μm per pixel. The specimens are prepared via conventional Pap smear and Pap staining. There are 917 images in the Herlev dataset, where each image contains one cervical cell with its segmentation of nucleus and cytoplasm and the class label. In order to maximize the certainty of diagnosis, cervical images in Herlev dataset were diagnosed by two cytotechnicians and a doctor and were categorized into seven classes. These seven classes further belong to two categories:

TABLE 1. The 917 cells (242 normal and 675 abnormal) from Herlev dataset.

Category	Class	Cell type	Num.
Normal	1	Superficial squamous epithelial	74
Normal	2	Intermediate squamous epithelial	70
Normal	3	Columnar epithelial	98
Abnormal	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

normal (class 1-3) and abnormal (class 4-7), as showed in Table 1. For each cell image in the Herlev dataset, rotations and translations (up to 10 pixels) are performed to yields a relatively balanced data distribution. After augmentation, each class has roughly 12000 images. The RGB image patch size is set to $m = 128$ pixels to cover some cytoplasm region for most cells, and to contain most of the nucleus region for the largest one. Then segmentation masks of the nuclei and cytoplasm with the same size and location as RGB image patch are extracted. These image patches and masks are then up-sampled to a size of $256 \times 256 \times 3$ and $256 \times 256 \times 2$ pixels via nearest interpolation, in order to facilitate the transfer of pre-trained CNN model. The image patches and masks are concatenated to obtain five-channel dataset with a size of $256 \times 256 \times 5$.

B. NETWORK ARCHITECTURES AND IMPLEMENTATION

In this study, two different inputs (i.e., raw RGB-channel dataset and five-channel dataset) are fed into four different CNN models, i.e., AlexNet, GoogLeNet, ResNet-50 and DenseNet-121, and the classification performances for different tasks (2-class and 7-class problems) are compared. Note that there are deeper architectures for ResNet and DenseNet (e.g. ResNet-152, DenseNet-161). However, we found that ResNet-50 and DenseNet-121 have better performances than their deeper versions on our dataset. Here, the base CNN models (denoted as AlexNet-B, GoogLeNet-B, ResNet-B, and DenseNet-B) are pre-trained on the ImageNet classification dataset. AlexNet-B contains three fully connection layer (*fc*6-*fc*8), and the number of neurons in last fully connection layer is determined by the number of output class. As shown in [32], reducing the number of neurons of *fc*6 and *fc*7 layer will tend to have slightly higher accuracy in this cervical dataset, and the neuron number of the *fc*6 and *fc*7 layer in our AlexNet model (denote as AlexNet-T) was set to be 1024-256. The AlexNet-T and the AlexNet-B share the same network, except for the first convolution layer and the three fully connection layers with weights of random initialized Gaussian distribution. For the other networks, weights of the first convolution layer and the last fully connection layer in our models (denote as GoogLeNet-T, ResNet-T and DenseNet-T) are initialized with random Gaussian distribution, and the other initial weights are copied from the same location of GoogLeNet-B, ResNet-B, and DenseNet-B, respectively. For each CNN model, we report classification results with different inputs for 2-class and 7-class classification tasks: CNN-3C (three-channel RGB dataset as input), CNN-2C (two-channel dataset of nucleus and cytoplasm binary mask as input), CNN-5C (five-channel dataset as input). Note that only CNN-3C and CNN-5C are used for 7-class classification. Therefore, throughout this paper, we refer to a total of twenty models. All these models are implemented on Caffe platform [34], using two Nvidia GeForce GTX 1080 Ti GPUs with a total memory of 22 GB.

C. TRAINING AND TESTING PROTOCOLS

From each 256×256 training image patch and its mirrored version, a 227×227 sub-patch is randomly cropped for AlexNet-T, while a 224×224 sub-patch is randomly cropped for the other networks in this study. Stochastic Gradient Descent (SGD) is utilized to train the model for 30 epochs. The mini-batch sizes of training are 256, 32, 20 and 12 for AlexNet-T, GoogLeNet-T, ResNet-T and DenseNet-T respectively. The base learning rates are 0.01, 0.005, 0.01 and 0.01 for AlexNetT, GoogLeNet-T, ResNet-T and DenseNet-T, respectively, and are decreased by a factor of 10 at every tenth epoch. Weight decay and momentum are set to be 0.0005 and 0.9 for AlexNet-T, and 0.0002 and 0.9 for the other networks.

D. EVALUATION METHODS

Most previous methods using cross validation on the Herlev dataset are random split, including study of [32]. When using random split, cells from the same patient may be split into both training and testing set. While in real clinical practice, all the cells from a testing patient are unseen to the training set. Therefore, in this study, five-fold cross-validation is performed on patient-level. In each of the 5 iterations, 4 of 5 folds are used as the training set and the other one as validation set. We carefully ensure that cells of the same patient can only be in the training set or the validation set. Note that data augmentation is performed after the training/validation spitting of cell population. Final performances of models are obtained by averaging the results from 5 validation sets. The performance evaluation metrics include sensitivity (*Sens*), specificity (*Spec*), accuracy (*Acc*) and area under ROC curve (*AUC*), where *Sens* indicates the proportion of correctly identified abnormal cells, *Spec* is the proportion of correctly identified normal cells, and *Acc* is the global percentage of correctly identified classified cell. The confusion matrix is used to show the classification performance of 7-class problem. The average accuracy of classification of cervical cells is calculated by averaging the values on the main diagonal of confusion matrix.

According to TBS, cervical cell in Herlev dataset can be categorized into four classes: normal (class 1-3), LSIL (class 4), HSIL (class 5-6) and CIS (class 7). Therefore, we calculate

TABLE 2. Performance comparison of different models for 2-class classification task. 2C, 3C and 5C are corresponding two-channel dataset (binary masks of nucleus and cytoplasm), three-channel dataset (Raw RGB data) and five-channel dataset (combining raw RGB data with binary masks of nucleus and cytoplasm) as network input, respectively. Bold indicates the highest value in each column.

Model	AUC	$Acc(\%)$	$Sens(\%)$	$Spec(\%)$
AlexNet-2C	0.946 ± 0.022	88.8 ± 2.3	95.4 ± 3.0	79.7 ± 4.8
AlexNet-3C	0.962 ± 0.008	89.7 ± 1.8	94.6 ± 4.2	83.0 ± 4.3
AlexNet-5C	0.964 ± 0.016	91.5 ± 2.8	96.5 ± 2.9	84.7 ± 4.1
GoogLeNet-2C	0.947 ± 0.022	89.0 ± 2.0	95.0 ± 2.5	80.8 ± 4.3
GoogLeNet-3C	0.979 ± 0.005	93.6 ± 1.1	96.2 ± 2.6	90.1 ± 2.5
GoogLeNet-5C	0.984 ± 0.012	94.5 ± 2.8	97.4 ± 2.7	90.4 ± 3.1
ResNet-2C	0.950 ± 0.022	88.9 ± 2.8	96.8 ± 1.9	78.2 ± 5.2
ResNet-3C	0.978 ± 0.018	92.3 ± 2.6	94.8 ± 3.3	89.1 ± 6.3
ResNet-5C	0.979 ± 0.011	92.1 ± 2.0	97.3 ± 2.8	85.2 ± 4.3
DenseNet-2C	0.934 ± 0.016	86.8 ± 1.3	95.1 ± 3.6	75.5 ± 4.6
DenseNet-3C	0.970 ± 0.013	92.6 ± 2.0	96.6 ± 2.5	87.1 ± 2.9
DenseNet-5C	0.980 ± 0.009	93.3 ± 2.0	95.6 ± 2.8	90.0 ± 3.6

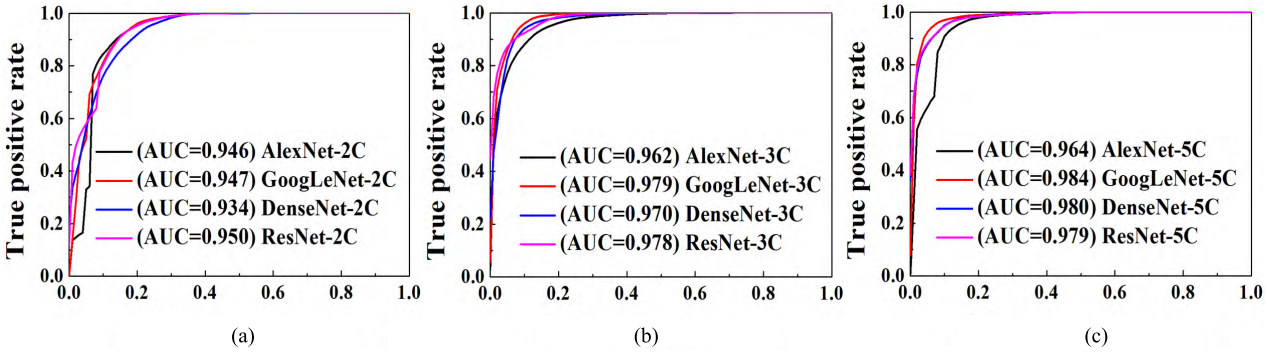


FIGURE 2. ROC curve comparison of different models for 2-class classification task. (a)–(c) are corresponding to CNN-2C model, CNN-3C model and CNN-5C model, respectively.

a 4-class classification result based on the 7-class's result in order to align with TBS.

E. RESULTS

Table 2 and Fig. 2 show the classification performances (*Sens*, *Spec*, *Acc* and *AUC*) of CNN-5C model (AlexNet-5C, GoogLeNet-5C, ResNet-5C and DenseNet-5C) in compared with CNN-3C model (AlexNet-3C, GoogLeNet-3C, ResNet-3C and DenseNet-3C) and CNN-2C model (AlexNet-2C, GoogLeNet-2C,ResNet-2C and DenseNet-2C) for 2-class classification task. It can be seen that each of models with five-channel dataset as input outperforms its corresponding three-channel-input and two-channel-input model in *AUC* metrics. Among them, the model GoogLeNet-5C obtains the best performance. The mean values of *Sens*, *Spec*, *Acc* and *AUC* for the model GoogLeNet-5C are 97.4%, 90.4%, 94.5% and 0.984, respectively. Some classification examples can be seen in fig. 3, where images misclassified by GoogLeNet-3C can be correctly classified by GoogLeNet-5C. GoogLeNet-5C also obtains the highest accuracy (67.0%).

Table 3 shows the classification performance of CNN-5C models and CNN-3C models for 7-class problem. CNN-5C models provide higher classification accuracy than CNN-3C models except for DenseNet. Among these models, GoogLeNet-5C obtains the highest accuracy (64.5%). Fig. 4 shows the confusion matrix for the model of GoogLeNet-5C,

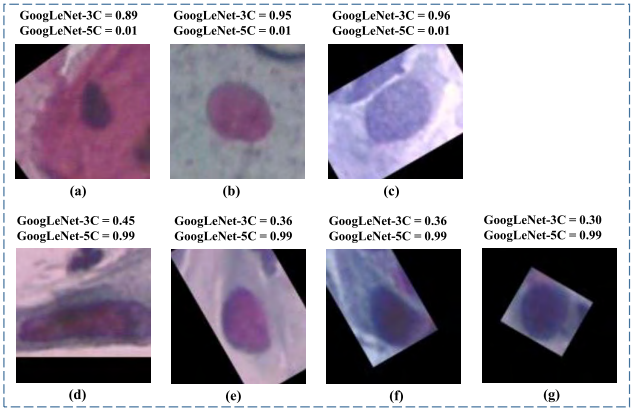


FIGURE 3. Examples of classified cervical cells by GoogLeNet-3C and GoogLeNet-5C. The ground truth labels of cells in the first row are Normal [(a) - (g) are superficial, intermediate and columnar] and the second row are Abnormal [(d) - (g) are mild dysplasia, moderate dysplasia, severe dysplasia and carcinoma]. Score = 1 corresponds a 100% probability of representing an abnormal cell.

with the average accuracy (averaging the values on main diagonal of confusion matrix) of 64.8%, which surpasses the previous non-deep-learning result of 61.1% [13]. Table 4 and Fig. 5 show the 4-class classification results obtained from 7-class classification results. The model of GoogLeNet-5C also achieves highest classification accuracy of 71.3%.

TABLE 3. Accuracy comparison of different models for 7-class classification. Bold indicates the highest value in each column.

Model	$Acc(\%)$
Benchmark [13]	61.1 ± 25.1
AlexNet-3C	57.8 ± 4.4
AlexNet-5C	60.8 ± 4.0
GoogLeNet-3C	62.5 ± 3.1
GoogLeNet-5C	64.5 ± 4.2
ResNet-3C	60.8 ± 3.7
ResNet-5C	63.7 ± 3.8
DenseNet-3C	63.9 ± 2.0
DenseNet-5C	61.0 ± 3.7

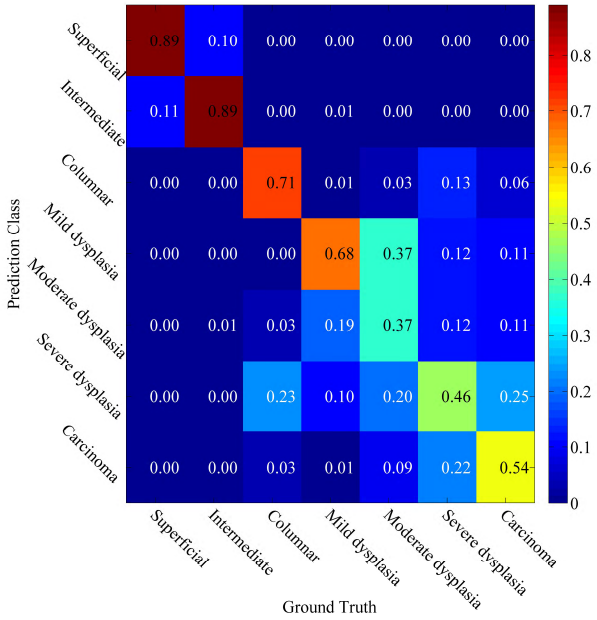


FIGURE 4. Confusion matrix of GoogLeNet-5C for 7-class classification task.

TABLE 4. Accuracy comparison of different models for 4-class classification. Bold indicates the highest value in each column.

Model	$Acc(\%)$
AlexNet-3C	66.5 ± 4.3
AlexNet-5C	68.3 ± 3.7
GoogLeNet-3C	69.7 ± 3.9
GoogLeNet-5C	71.3 ± 2.7
ResNet-3C	67.0 ± 4.6
ResNet-5C	70.3 ± 4.7
DenseNet-3C	71.1 ± 2.6
DenseNet-5C	67.6 ± 2.1

IV. DISCUSSION

Fine-grained cervical cell classification is highly desired in pathologistsâĂŹ daily diagnosis practice, which has long been ignored by most of previous (automated) studies. Our work raises the question that whether the state-of-the-art deep learning could push forward this field. We investigate the ability of deep learning in utilizing cell (raw) morphological and appearance features for classification. For the 2-class classification problem, networks only using morphological information or appearance information as input,

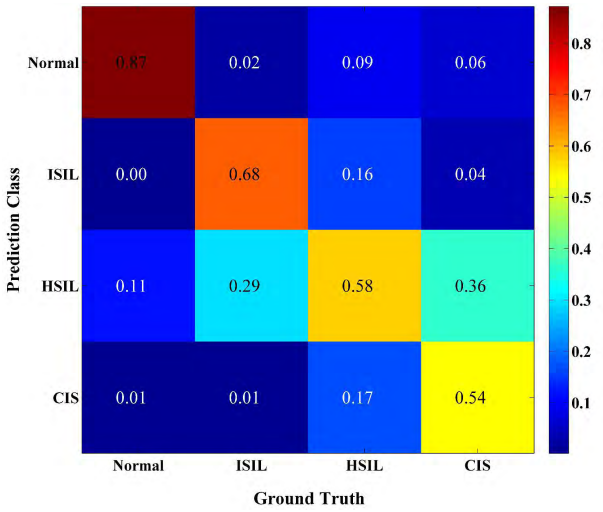


FIGURE 5. Confusion matrix of GoogLeNet-5C for 4-class classification task.

are compared with networks using both morphological and appearance information as input. As shown in Table 2, the classification performances of morphology-based CNNs are slightly lower than that of appearance-based CNNs, and classification performances of appearance-morphology-based CNNs are obviously higher, which indicates that morphology and appearance provide complementary information to each other, thereby improving the classification performance.

Among CNNs-2C model, CNNs-3C model and CNNs-5C model, the GooLeNet has the best performance, and the GoogLeNet-5C model has the highest classification accuracy, sensitivity, and specificity (94.5%, 97.4%, and 90.4%). The GoogLeNet also has the highest classification accuracy of 64.5% and 71.3% for the 7-class and 4-class problem, respectively. The networks deeper than GoogLeNet, such as ResNet and DenseNet, do not provide higher performance on our task. This may due to the relatively small number of cells in Herlev cervical cell dataset used in this study, which may lead training on such complicate network to be overfit. For the 2-class problem, the best previous study obtained 96.8% classification accuracy by optimizing the features derived from the manually segmented cytoplasm and nucleus [35]. Deep learning method has also used on this Herlev data with an accuracy of 98.3% [32]. The performances are not directly comparable since different data splitting methods are used, i.e., we split data at patient-level, while previous studies are random split. Also note that our method does not use any feature engineering, only raw RGB image and segmentation masks are used for CNN learning.

The Herlev dataset has 7-class cervical cells, specifically containing four categories of cervical abnormalities, and three categories of normal cells. There is only one previous study (i.e., [13]) that reports the confusion matrix of fine-grained classification result on this dataset. In our study,

GoogLeNet-5C model obtains an average accuracy of 64.8% for 7-class problem, which is 3.7% higher than that of 61.1% in [13]. As shown in the confusion matrix in Fig. 4, some cells are easier to be classified, while some are harder; superficial and intermediate cells are classified with the highest accuracy. Some columnar cells are wrongly classified as severe dysplasia cells because severe dysplasia cells have similar characteristics in appearance and morphology with columnar cells (e.g., dark nuclei and small-sized cytoplasm); This is indeed a difficult point in the identification of cervical cells, but it may be possible to improve this difficulty by adding nuclear size characteristics. The fine-grained classification of abnormal cells (i.e., mild dysplasia, moderate dysplasia, severe dysplasia and carcinoma) remains very challenging. In general, such a task is hard even for cyto-pathologists but highly desirable and with significantly clinical value; The most difficult case is moderate abnormal cell with correct classification rate of only 37%.

Similar results can also be observed in Fig. 5 (i.e., 4-class TBS). Only about 1% of Normal and LSIL cervical cells are misclassified as cancer cells, which is potentially meaningful for reducing unnecessary biopsy. The accuracy rate of cancer cell classification is 54%. Improving of which is the direction of future research effort.

The segmentation of nucleus and cytoplasm are pre-required for applying our method. This is directly obtained from the ground truth segmentation in this paper. Note that screening of abnormal of cells in practice using the proposed method requires automated segmentation of the nucleus and cytoplasm. If the segmentation results are not reliable, it may affect the classification performances. But our goal here is fine-grained classification of the cervical cells, which is even a very difficult task for experienced doctors. In order to improve the classification accuracy first, we do not consider the problem of automatic segmentation. The task of automatic segmentation may be achieved by using fully convolutional networks (e.g., U-Net) [36] or specifically designed algorithm [37] for semantic segmentation of cervical cells. The effects of automated segmentation on classification performance still need to be analyzed in future study. Nevertheless, the current results demonstrate that fine-grained classification of cervical cells into different abnormal levels remains to be very challenging even with accurate cell segmentation available.

V. CONCLUSION

This paper proposes an appearance and morphology based convolutional neural network method for cervical cell fine-grained classification. Unlike the previous CNN-based method which only uses raw image data as network input, our method combines the raw image data with segmentation masks of the nucleus and cytoplasm as network input. Our method consists of extracting cell image/mask patches coarsely centered on the nucleus, transferring features from another pre-trained model into a new model for fine-tuning on the cervical cell image dataset, and forming the final network output. State-of-the-art CNN networks including AlexNet, GoogLeNet, ResNet and DenseNet are trained for performance comparison. The results show that the combination of raw RGB data with segmentation masks of nuclei and cytoplasm as network input can provide higher performance in the fine-grained classification of cervical cells. Although the initial results are promising, deep learning based fine-grained cervical cell classification remains a very challenging task for high precision diagnosis. Moreover, the effects of automated segmentation of nucleus and cytoplasm on classification performance still need to be analyzed in future study.

ACKNOWLEDGMENT

This work was performed when J. Yao and L. Zhang were in the National Institutes of Health.

REFERENCES

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

[2] A. Gadducci, C. Barsotti, S. Cosio, L. Domenici, and A. R. Genazzani, "Smoking habit, immune suppression, oral contraceptive use, and hormone replacement therapy use and cervical carcinogenesis: A review of the literature," *Gynecol. Endocrinol.*, vol. 27, no. 8, pp. 597–604, 2011.

[3] D. Saslow, D. Solomon, H. W. Lawson, M. Killackey, S. L. Kulasingam, J M. Cain, F. A. R. Garcia, A. T. Moriarty, A. G. Waxman, D. C. Wilbur, N. Wentzensen, L. S. Downs, M. Spitzer, A. B. Moscicki, E. L. Franco, M. H. Stoler, M. Schiffman, P. E. Castle, E. R. Myers, "American cancer society, American society for colposcopy and cervical pathology, and American society for clinical Pathology screening guidelines for the prevention and early detection of cervical cancer," *Cancer J. Clinicians*, vol. 62, no. 3, pp. 147–172, 2012.

[4] G. G. Birdsong, "Automated screening of cervical cytology specimens," *Hum. Pathol.*, vol. 27, no. 5, pp. 468–481, 1996.

[5] L. Zhang, H. Kong, C. T. Chin, S. Liu, X. Fan, T. Wang, and S. Chen, "Automation-assisted cervical cancer screening in manual liquid-based cytology with hematoxylin and eosin staining," *Cytometry A*, vol. 85, no. 3, pp. 214–230, 2014.

[6] E. Bengtsson and P. Malm, "Screening for cervical cancer using automated analysis of PAP-smears," *Comput. Math. Methods Med.*, vol. 2014, Mar. 2014, Art. no. 842037.

[7] R. M. DeMay, *Practical Principles of Cytopathology*. Chicago, IL, USA: American Society Clinical Pathology Press, 2007.

[8] T. P. Canavan and N. R. Doshi, "Cervical cancer," *Amer. Family Phys.*, vol. 61, no. 5, pp. 1369–1376, 2000.

[9] L. Nanni, A. Lumini, and S. Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis," *Artif. Intell. Med.*, vol. 49, no. 2, pp. 117–125, 2010.

[10] F. Li, X. Zhou, J. Ma, and S. T. C. Wong, "Multiple nuclei tracking using integer programming for quantitative cancer cell cycle analysis," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 96–105, Jan. 2010.

[11] P.-W. Huang and Y.-H. Lai, "Effective segmentation and classification for HCC biopsy images," *Pattern Recognit.*, vol. 43, pp. 1550–1563, Apr. 2010.

[12] N. Theera-Umpon and S. Dhompongsa, "Morphological Granulometric features of nucleus in automatic bone marrow white blood cell classification," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 3, pp. 353–359, May 2007.

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

[14] 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. 2010.

[15] 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.

[16] Y. Marinakis, M. Marinaki, and G. Dounias, "Particle swarm optimization for pap-smear diagnosis," *Expert Syst. Appl.*, vol. 35, no. 4, pp. 1645–1656, 2008.

[17] Y. Guo, G. Zhao, and M. Pietikäinen, "Discriminative features for texture description," *Pattern Recognit.*, vol. 45, no. 10, pp. 3834–3843, 2012.

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[19] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[21] 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.*, Jun. 2015, pp. 1–9.

[22] 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.

[23] 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.

[24] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3320–3328.

[25] H.-C. Shin *et al.*, "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] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-Ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3462–3471.

[27] J. De Fauw *et al.*, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature Med.*, vol. 24, no. 9, pp. 1342–1350, 2018.

[28] K. Yan, X. Wang, L. Lu, L. Zhang, A. P. Harrison, M. Bagheri, and R. M. Summers, "Deep lesion graphs in the wild: Relationship learning and organization of significant radiology image findings in a diverse large-scale lesion database," in *Proc. IEEE CVPR*, Jun. 2018, pp. 9261–9270.

[29] L. Zhang, L. Lu, R. M. Summers, E. Kebebew, and J. Yao, "Convolutional invasion and expansion networks for tumor growth prediction," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 638–648, Feb. 2018.

[30] L. Nanni, S. Ghidoni, and S. Brahnam, " Ensemble of convolutional neural networks for bioimage classification," *Appl. Comput. Inform.*, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2210832718301388

[31] H. Li, F. Pang, Y. Shi, and Z. Liu, "Cell dynamic morphology classification using deep convolutional neural networks," *Cytometry A* , vol. 93, no. 6, pp. 628–638, 2018.

[32] 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.

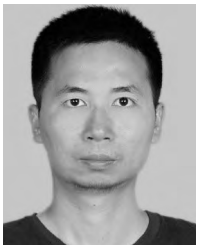
[33] R. Nayar and D. C. Wilbur, "The pap test and bethesda 2014," *Acta Cytol.*, vol. 59, no. 2, pp. 121–132, 2015.

[34] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.

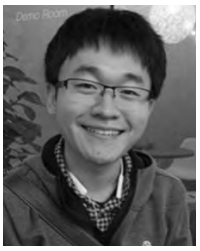
[35] Y. Marinakis, G. Dounias, and J. Jantzen, "Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification," *Comput. Biol. Med.*, vol. 39, no. 1, pp. 69–78, 2009.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Springer, vol. 9351, Nov. 2015, pp. 234–241.

[37] L. Zhang, M. Sonka, L. Lu, R. M. Summers, and J. Yao, "Combining fully convolutional networks and graph-based approach for automated segmentation of cervical cell nuclei," in *Proc. IEEE 14th Int. Symp. Biomed. Imag.*, Apr. 2017, pp. 406–409.



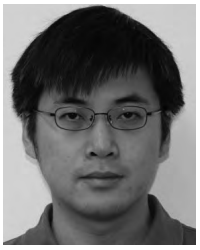
HAOMING LIN received the Ph.D. degree in biomedical engineering from Zhejiang University, China, in 2014. From 2015 to 2017, he was a Postdoctoral Researcher for ultrasound imaging with Shenzhen University, Shenzhen, China, where he has been an Assistant Professor with the School of Biomedical Engineering, since 2017. His research interests include medical ultrasound image, ultrasound elastography, and medical image computing and analysis.



YUYANG HU received the B.S. degree in biomedical engineering from Shenzhen University, Shenzhen. in 2015, where he is currently pursuing the M.S. degree in biomedical engineering. His current research interests include development of MAT-MI techniques and ultrasound imaging techniques.

SIPING CHEN received the Ph.D. degree in biomedical engineering from Xi'an Jiaotong University, Shanxi, China, in 1987. After a Postdoctoral Fellowship at Zhejiang University, Zhejiang, China, he joined Shenzhen Anke High-tech Co., Ltd., as the Chief Technology Officer, in 1989. Having served at Anke for 16 years, he joined Shenzhen University as the Vice President and founded the Biomedical Engineering Branch at Shenzhen University, in 2005. His main research interests include biomedical ultrasound imaging, medical instrumentation, tissue elasticity imaging, multimodal ultrasound imaging, and image processing.

JIANHUA YAO received the Ph.D. degree in computer science from Johns Hopkins University, in 2002. He was an Associate Scientist and the Manager with the Department of Radiology and Imaging Sciences, NIH, where he managed the Clinical Image Processing Services Laboratory. He was with the Computer-Aided Detection (CAD) and Image Biomarker Laboratory, NIH. He is currently an Expert Fellow with AI Lab, Tencent Holdings Ltd., China. His research interests include machine learning, clinical image processing, deformable model, nonrigid registration, CAD, CT colonography, and tumor growth modeling.



LING ZHANG received the Ph.D. degree in biomedical engineering from Zhejiang University, China, in 2013. From 2013 to 2016, he was a Postdoctoral Researcher with the Iowa Institute for Biomedical Imaging, IA, USA. From 2016 to 2018, he was a Visiting Fellow with the Radiology and Imaging Sciences, National Institutes of Health, MD, USA. Since 2018, he has been a Research Scientist with Nvidia Corporation, MD, USA. His research interests include medical image computing and analysis, machine learning, and image processing.