An Application of Transfer Learning and Ensemble Learning Techniques for Cervical Histopathology Image Classification

DAN XUE[1], XIAOMIN ZHOU[1], CHEN LI[1], YUDONG YAO[2], (Fellow, IEEE), MD MAMUNUR RAHAMAN[1], JINGHUA ZHANG[1], HAO CHEN[1], JINPENG ZHANG[1], SHOULIANG QI[1], (Member, IEEE), AND HONGZAN SUN[3]

[1]Microscopic Image and Medical Image Analysis Group, College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110819, China
[2]Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA
[3]Department of Radiology, Shengjing Hospital of China Medical University, Shenyang 110122, China

Corresponding authors: Chen Li (lichen201096@hotmail.com) and Hongzan Sun (sunhongzan@126.com)

ABSTRACT In recent years, researches are concentrating on the effectiveness of *Transfer Learning* (TL) and *Ensemble Learning* (EL) techniques in cervical histopathology image analysis. However, there have been very few investigations that have described the stages of differentiation of cervical histopathological images. Therefore, in this article, we propose an *Ensembled Transfer Learning* (ETL) framework to classify well, moderate and poorly differentiated cervical histopathological images. First of all, we have developed Inception-V3, Xception, VGG-16, and Resnet-50 based TL structures. Then, to enhance the classification performance, a weighted voting based EL strategy is introduced. After that, to evaluate the proposed algorithm, a dataset consisting of 307 images, stained by three immunohistochemistry methods (AQP, HIF, and VEGF) is considered. In the experiment, we obtain the highest overall accuracy of 97.03% and 98.61% on AQP staining images and poor differentiation of VEGF staining images, individually. Finally, an additional experiment for classifying the benign cells from the malignant ones is carried out on the Herlev dataset and obtains an overall accuracy of 98.37%.

INDEX TERMS Cervical cancer, differentiation stages, histopathology images, transfer learning, ensemble learning, classification.

I. INTRODUCTION

Cervical cancer is one of the malignant tumors, with a high incidence in women. It is also the fourth leading cause of cancer-related deaths. It poses a serious risk to women's health and can spread through direct or distant metastasis [1]. In developing countries, it is the second most prevalent malignancy after breast cancer and the third dominant cause of cancer-related deaths after breast and lung cancer [2]. Moreover, developing countries are more vulnerable to prevent cancer deaths because of a lack of awareness and adequate medical facilities, which leads to nearly 90% of cervical

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

cancer-related deaths [3]. Though the occurrence of cervical cancer in developing nations has decreased in current times, the percentage among young women grows [4]. Therefore, it is essential to make a precise diagnosis for the detection of early-stage cancer, as well as reliable techniques, which are primarily demand [5]. Even though there are some preliminary tests and non-invasive methods for detecting cancer in various organs, studies of histopathological images for detecting the cancer are inevitable. Therefore, accurate and rapid analysis of histopathological images plays a vital role in medical research [6]. A pathologist who can analyze the histological images needs to go through a rigorous training process and demands full concentration, including the time to analyze the samples. Besides, the accuracies are greatly
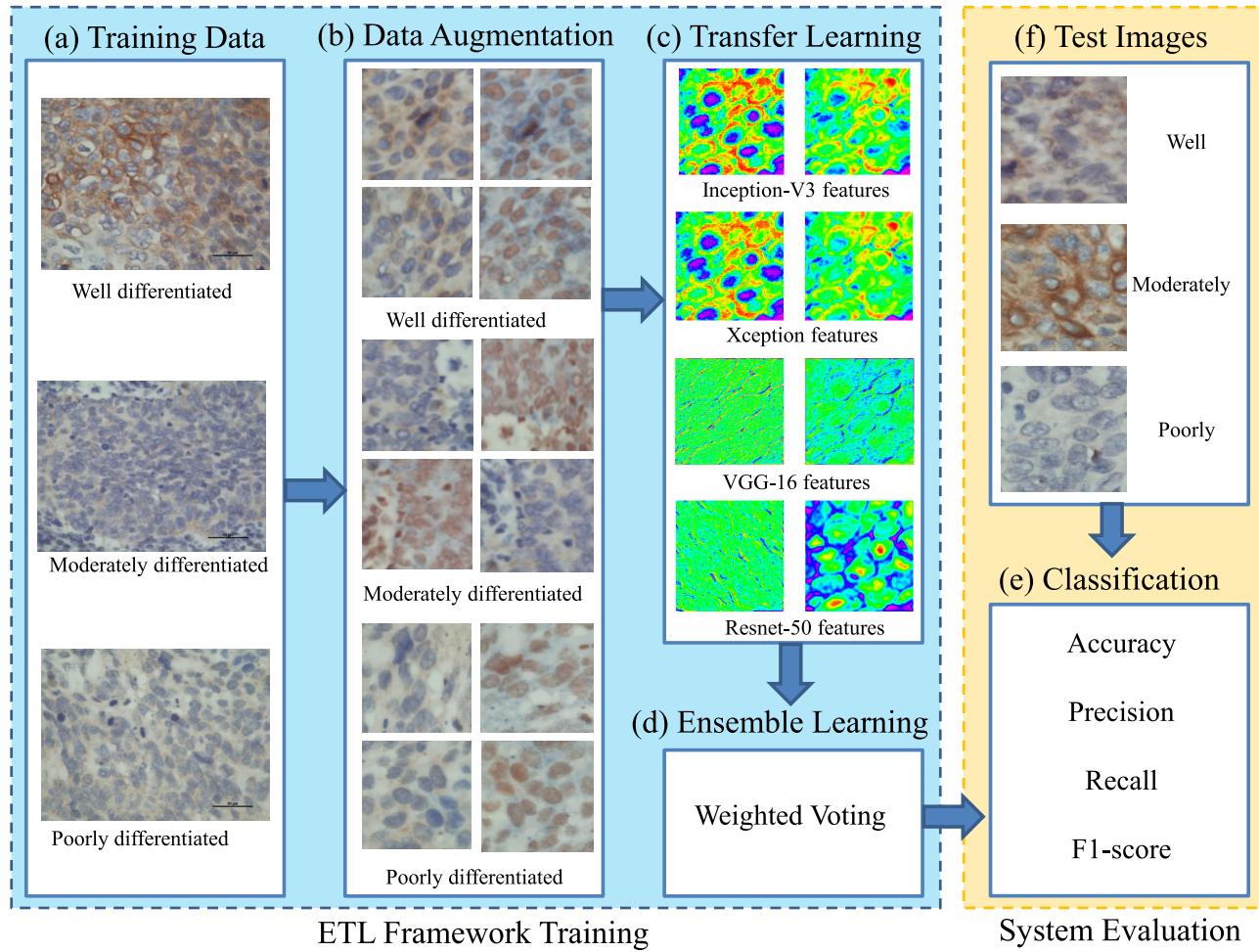
FIGURE 1. Workflow of the proposed ETL framework for the CHIC task. The blue box denotes the training process; the yellow box shows the test process.

varied from person to person. Additionally, the investigation is a very subjective and uncertain process, which brings unstable diagnostic efficiency [7]. In this regard, a *Computer Aided Diagnosis* (CAD) system can help the doctor to track the cancer of histopathological images with significant accuracy and efficiency.

As discussed earlier, the histopathological study in the field of cervical cancer is considered as the "gold standard" in clinical diagnosis [8]. However, the applications of CAD techniques for cervical cancer are still emerging and require in-depth development of research [9]. To this end, this article focuses on the *Cervical Histopathology Image Classification* (CHIC) problem by using CAD systems to solve three distinct stages of classification using an *Ensembled Transfer Learning* (ETL) framework. The workflow of the suggested ETL method is shown in Fig. 1.

In Fig. 1, the acquired microscopic images for cervical cancer are employed as training samples. Notably, the images are stained using three immunohistochemistry approaches (AQP, HIF, and VEGF). Then, due to the inadequate

dataset for deep learning training, augmentation operation is performed to enhance the classification performance. Thirdly, four *Transfer Learning* (TL) approaches are built up for the classification, including Inception-V3, Xception, VGG-16, and Resnet-50 networks. Fourthly, an *Ensemble Learning* (EL) technique, is further implemented to obtain a more accurate classification result with a weighted voting approach. Finally, the test images are utilized to evaluate the effectiveness of the proposed method, where we calculate the accuracy, precision, recall, and F1-score.

The main contributions with novelties of this study are given as follows:

(1) A state of the art approach for analyzing the cervical histopathological images is presented in this study. To the best of our knowledge, besides our previous work [10]–[13], there have been very few studies on the differentiation stage of cervical histopathology images exist.

(2) A new EL approach is introduced: In order to prevent the unicity and limitation of the classification results by a

single learner, we use four *Convolution Neural Network* (CNN) for experiments, and then adopt the weighted voting based EL method. Through multiple comparison experiments, we find that precision is an optimal choice for weighting in this work.

(3) A series of valuable data is achieved: We do the comparative experiments of the individual TL and the EL, proving the feasibility of our method.

The structure of the paper is as follows: Sec. II introduces the related work about the CHIC. Sec. III gives an overview of the proposed EL method, including TL using the Xception, Inception-V3, VGG-16 and Resnet-50 networks, and the proposed weighted voting based approach. Sec. IV elucidates the experimental results, including the evaluation and analysis of the proposed method. Sec. V concludes this paper and discusses the future work.

II. RELATED WORK

A. RELATED WORK IN MEDICAL FIELDS

In [14], an overview on the staging diagnosis of uterine tumors on MRI images is described. The paper of [15] finds that the discrete compactness is a reproducible parameter for a computer assisted quantification of the invasion front pattern and it may be a phenotypic feature of cervical cancer cells. In [16], a research on feature quantification and abnormal detection of cervical cells is proposed, which realizes the detection and identification of individual cancer cells. First, normal and abnormal cells are distinguished by the differences in the shapes of cytoplasm and nuclei. After that, affinity propagation algorithm is implemented for further analysis on the detected abnormal cells. Finally, the grading accuracy of abnormal cells is 76.47%. In [17], a method for automatic detection of images of cervical cancer cells captured from thin liquid based cytology slides is proposed, a two-level cascade classifier is developed to obtain the dimensional features in morphology and texture. The system classifies 20000 cells, where the final recognition rate is 95.642%, and the false negative rate and false positive rate are 1.44%, respectively. In [18], a machine learning model with a cross-validation algorithm for the cervical cytology image is created. In clinical practice, the examinations for diagnosing cervical cancer include medical imaging examinations, tumor biomarker detection, and HPV DNA testing. The diagnosis of cervical cancer depends on the analysis of pathologists by observing pap smear or biopsy samples, which is an expensive method. This involves examining the abnormal cell level of samples and determining the spread of abnormal cells [19]. In [20], cancer is classified according to the spread of abnormal cells, which is a tedious, subjective, and time-consuming process.

B. RELEVANT CLASSICAL WORKS IN THE CAD FIELD

In [21], CAD techniques for tissue slice image analysis are summarized. The paper of [22] carries out automatic nuclei detection, segmentation, and feature calculation of the cervical images. In [23], the functions of the envisioned CAD system for colposcopy are described, and modular and suggests open system design for image analysis by providing a framework and foundation for the CAD system of cervical cancer. The paper of [24] proposes a prototype automated CAD system for the diagnosis of CIN using ultra-large virtual slides of cervical cancer, the segmentation achieves an average accuracy more than 94.25%. The work of [25] reports on methodologies and outcomes of a study aiming at developing robust tool to evaluate and classify histology images of cervical cancer, by using the histology images acquired from the pathology laboratories in an Indonesian hospital to classify cervical biopsy images based on four well known discriminatory features. The paper of [26] uses k-means clustering, Gabor wavelet transform, graph cutting, color segmentation algorithms, cellular morphological methods, and binary tree algorithms are used to classify epithelial cells and stromal cells in the histopathological images of cervical cancer. At the same time, a computer aided decision support system tool is presented to help pathologists in their examination of cervical cancer biopsies. The main aim of the proposed system is to identify abnormalities and quantify cancer grading in a systematic and repeatable manner. The result shows that the specificity of k-means, Gabor wavelet, and Hybrid Graph cut and colour segmentation methods in CAD is 80%, 87%, and 97%, respectively. The paper of [27] develops an automated CIN grade classification of vertical segmented epithelium regions, and explores a Particle Swarm Optimization (PSO) and Receiver Operating Characteristic curve (ROC) for CIN classification showing exact grade labeling accuracy as high as 90%. In [28], an automated approach for detecting cervical cancer is proposed. The *Gray Level Co-occurrence Matrix* (GLCM) is used to divide the image into ten vertical images to extract texture features. Then, using k-means clustering and Marker control watershed algorithm to segment the image. Finally, based on the texture and lesion area features, the *Support Vector Machine* (SVM) method is used to recognize cervical cancer, and 90% accuracy is obtained. In our previous work [10], a TL framework for Inception-V3 network is proposed. First, the images are augmented by using data rotation and mirroring. Then, a TL method based on the Inception-V3 network is constructed to extract deep learning features. Finally, the extracted features are designed for the final classification. In the result, an average accuracy of 77.3% is achieved. In another work [11], we suggest a multilayer hidden conditional random fields (MHCRFs) to classify well, moderate and poorly differentiation stages of cervical cancer, and an accuracy of 88% is obtained on a practical histopathological image dataset with more than 100 AQP stained samples. Meanwhile, in [12], a novel MHCRFs based cervical histopathology image classification model is proposed to classify well, moderate, and poorly differentiated stages of cervical cancer using a weakly supervised learning strategy. In [29], we utilize graph and unsupervised learning methods in a tissue structure clustering task, and divide the histopathological images of cervical cancer into sparse areas

and viscose areas to predict the risk of the tissues. For more information, please refer to our previous survey papers [30].

C. RELEVANT DEEP LEARNING WORKS IN THE CAD FIELD

In recent years, deep learning approaches show a robust development trend in the cervical cancer image classification field. For instance, the paper of [31] introduces a method for the diagnosis of histopathological images of cervical cancer using SVM and *Artificial Neural Network* (ANN). In [32], a superpixel and CNN based segmentation method for cervical cancer cells is proposed with an accuracy of 94.50%. In [33], the author introduces a computerized technique to distinguish normal and abnormal cervical cells using deep ANN and a learning vector quantization algorithm. As a development of [33], an ANN algorithm to extract new features of cervical cells is proposed, providing a classification method for cervical smear examination using the ANN, and comparing it with *k*-means and Bayesian classifiers. In our recent work [13], we propose an ETL framework to classify well, moderately and poorly differentiated cervical histopathology images based on VGG-16 and Inception-V3 networks followed by an early fusion approach, using a practical dataset with 100 VEGF stained cervical histopathology images, an average accuracy of 80% is finally achieved. Meanwhile, in [34], we introduce ensemble learners of multiple deep CNNs for pulmonary nodule classification using CT images obtained from the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) database, and achieve prediction accuracy of 84.0%. Based on the methods mentioned above, we carry out a series of contrast experiments in Sec. IV-F, Table. 6.

Over the last few years, CAD techniques for cervical histopathological images are focusing on practicing classical and machine learning feature extraction approaches for segmentation and pathologic abnormality screening. However, there has been a minimal contribution to the differentiation of cervical histopathological images. At the same time, our ETL approach has shown an average performance on pulmonary CT and histopathological image analysis. Hence, based on these research points, we develop a new ETL method in this paper.

III. METHOD

A. TRANSFER LEARNING

ANN is one of the primary tools employed in machine learning. It consists of input and output layers, as well as (in most cases) hidden layers consisting of units that transform the input into something that the output layer can use. The input data enters the neural network and comes to the hidden layer. After the activation function, we get the final output. The deep ANN usually refers to an ANN with more than two hidden layers [35].

Transfer Learning (TL) is a method that uses ANNs pre-trained on a large annotated image database (such as ImageNet) to complete various tasks. TL focuses on storing

TABLE 1. The parameter settings for four TL networks.

Networks \ Parameters	Initial input size	Initial learning rate	Batch-size	Epoch	Optimizer
Inception-V3	$299 \times 299 \times 3$	0.0001	64	100	Adam
Xception	$299 \times 299 \times 3$	0.0001	64	100	Adam
VGG-16	$224 \times 224 \times 3$	0.001	64	100	Adam
Resnet-50	$224 \times 224 \times 3$	0.001	64	100	Adam

knowledge gained in solving a problem and applying it to different but related problems. It essentially uses additional data so that ANNs can decode by using the features of past experience training, after that the ANNs can have better generalization ability [36]. With TL technology, we can directly use pre-trained deep learning models that are trained through a large number of readily available datasets. Then, find out the layer that can be reused. Finally, we can use the output of these layers as the input to train a network with fewer parameters and smaller scales. This small-scale network only needs to understand the internal relationships of specific problems, and learns the patterns contained in the data through the pre-trained models [37].

In this paper, we have compared the VGG series, Inception series, and Resnet series. The final selection is based on the comprehensive classification performance and number of parameters. We finally apply Inception-V3 [38], Xception [39], VGG-16 [40] and Resnet-50 [41] networks for the TL process, where the parameters are pre-trained on the ImageNet dataset [42]. For these four networks, the settings of the hyperparameters are shown in Table. 1. Among them, the learning rate uses decay learning rate, and the decay_steps is 5, the decay_rate is 0.1. Fig. 2 shows an example of the feature maps extracted by these four TL networks, where the TL method can obtain some representative information from the images.

B. ENSEMBLE LEARNING

Ensemble learning (EL) is the strategic generation and combination of multiple models, such as classifiers or experts, to solve particular computational intelligence problems [43]. EL is primarily used to improve the performance (e.g., classification, prediction, and function approximation) of a model, or reduce the likelihood of an unfortunate selection of a poor one. Other applications of EL include assigning confidence to the decision made by the model, selecting optimal (or near optimal) features, data fusion, incremental learning, nonstationary learning, and error-correcting [44].

The combination of learners may provide benefits in terms of statistics, computation and representation. In this paper, the EL adopts the weighted voting based method. Its integration includes T base learners $\{h_1, h_2, h_3, \ldots, h_T\}$, where the output of h_i on example x is $h_i(x)$. Particularly, in this paper, we select four TL methods, so $T = 4$. The learner h_i predicts a label from the set of category labels $\{c_1, c_2, \ldots, c_N\}$, the predicted output of h_i on example x is represented as

(a) Inception-V3 feature maps.

(b) Xception feature maps.

(c) VGG-16 feature maps.

(d) Resnet-50 feature maps.
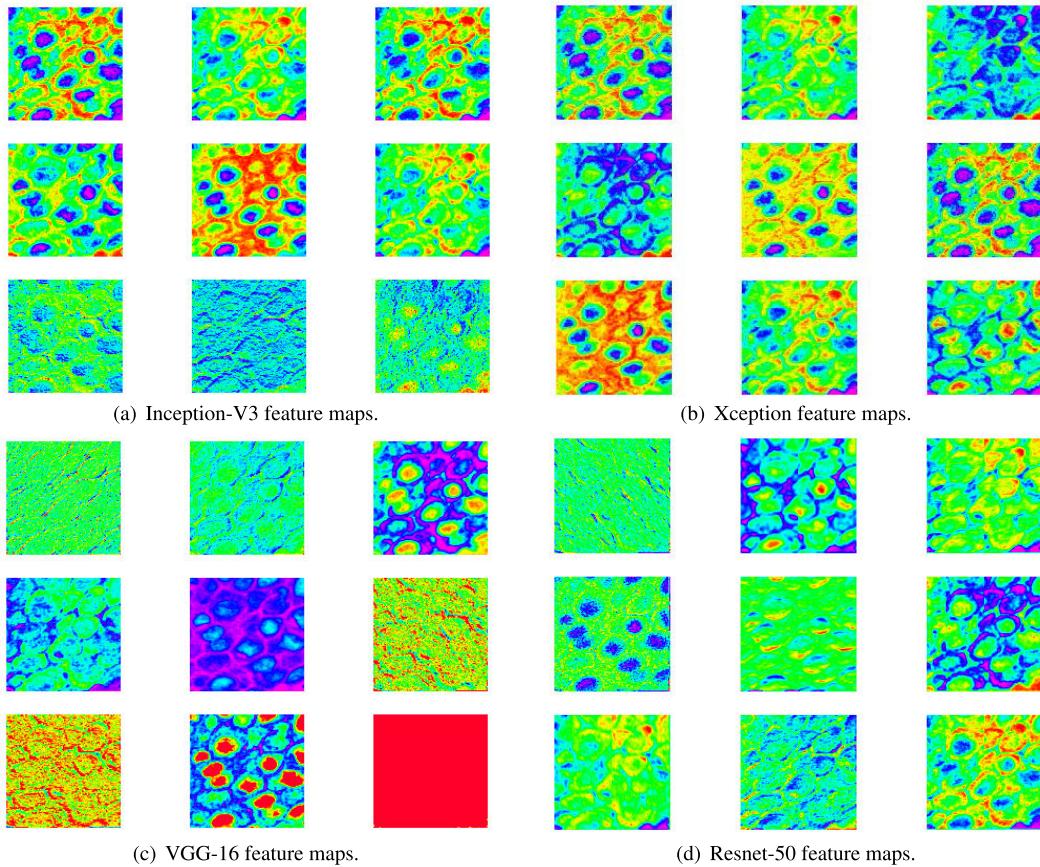
FIGURE 2. An example of the feature maps by different TL approaches. All the feature maps are extracted from the third convolutional layer of each network, respectively.

a N dimensional vector $(h_i^1(x); h_i^2(x); \ldots; h_i^N(x))$, where $h_i^N(x)$ is the output of h_i on the category labels [45]. Lastly, the whole EL process is defined refer to (1).

$$H(x) = c_{arg_j max \sum_{i=1}^{T} w_i h_i^j(x)}. \tag{1}$$

where $H(x)$ is the final EL result, w_i is the weight of h_i, usually, $w_i \geq 0$ and $\sum_{i=1}^{T} w_i = 1$. Furthermore, we pretest four evaluation indexes as the candidates of the weight, including classification accuracy, recall, precision, and F1-score. Finally, as the recall to be the weight achieves the best classification result, we select it as the weight in this paper

IV. EXPERIMENTS AND ANALYSIS

A. IMAGE DATASET

To test the effectiveness of the proposed ETL method in this paper, a practical histopathology image dataset of cervical cancer tissue sections is applied. The detailed information about this dataset is as follows.

Data source: Two practical medical doctors from Shengjing Hospital of China Medical University provide image samples and give image-level labels;

Staining method: Immunohistochemical (IHC) Staining, including AQP, HIF, and VEGF approaches;

Magnification: $400\times$;

Microscope: Nikon (Japan);

Acquisition software: NIS-Elements F 3.2;

Image size: 1280×960 pixels;

Image format: *.png;

Image bits per pixel depth: $3 \times 8 = 24$.

Image category and morphological characteristics: There are 307 images in the dataset, where 33 are well differentiated, 35 are moderate differentiation, and 28 are poor differentiation of AQP staining; 39 are well differentiation, 38 are moderately differentiation, and 34 are poor differentiation of HIF staining; 29 are well differentiation, 33 are moderately differentiation, and 38 are poor differentiation of VEGF staining.

- Well differentiation: The tumor cells are closer to normal cells, cell heteromorphism is relatively small, cell size and morphology are similar.
- Moderate differentiation: Most cancer cells are concentrated in moderate differentiation. The characteristic is between well differentiation and poor differentiation of cervical cancer cells.
- Poor differentiation: The cell structure is not visible, and the parts are disordered.

Among them, the well differentiated tumor cells are the least malignant, the poorly differentiated tumor cells have the highest degree of malignancy, and the moderately

(a) AQP staining.

(b) HIF staining.

(c) VEGF staining.
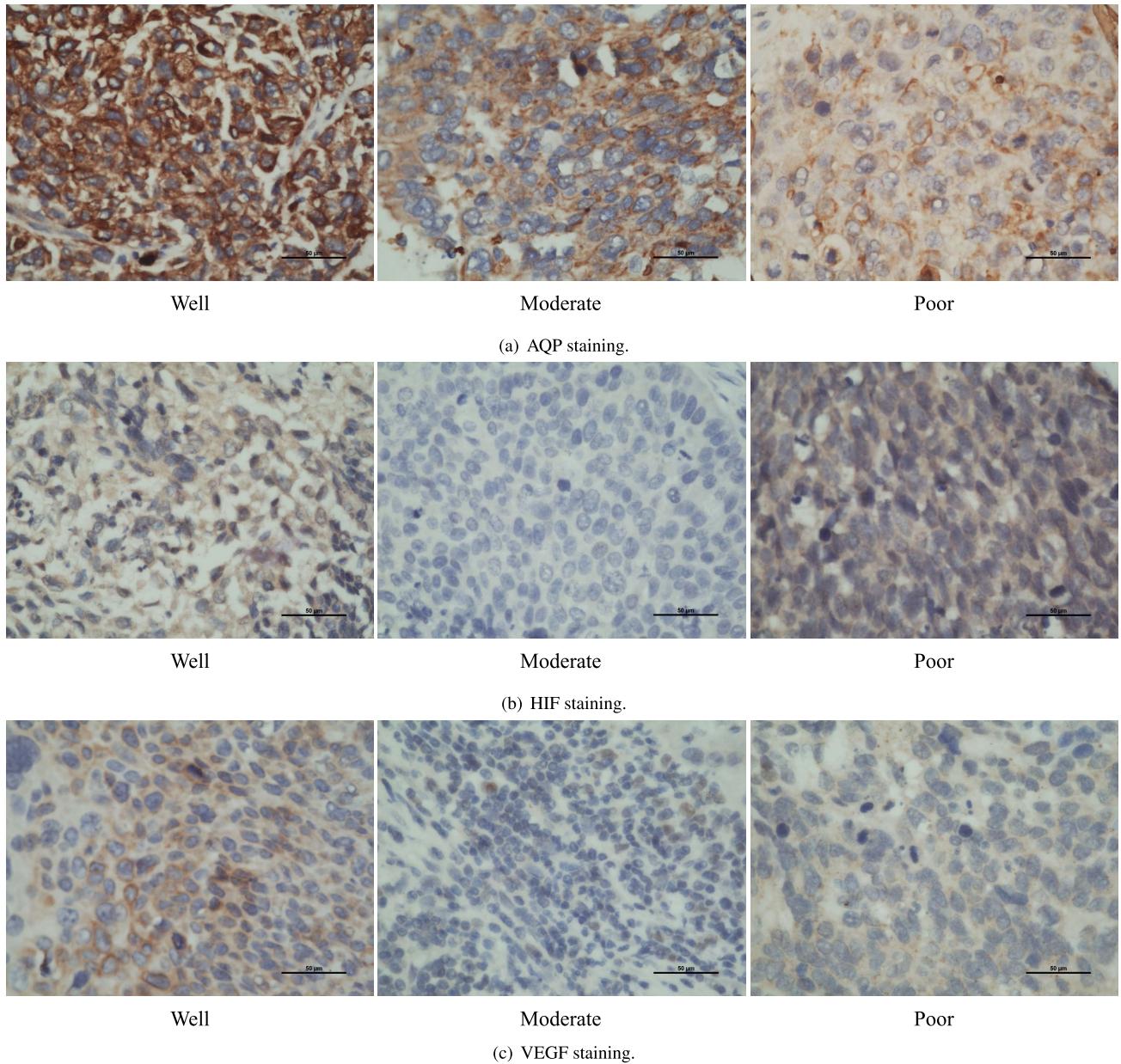
FIGURE 3. An example of the cervical IHC image dataset. Each row shows the images of an IHC staining method.

differentiated tumor cells are moderately malignant [46]. An example of this dataset is shown in Fig. 3.

B. DATA AUGMENTATION

Data augmentation adds value to the underlying data by transforming the information inside the dataset. Since the total number of sample datasets is too small (only 307), the neural network trained on the small dataset is prone to over-fitting. So, we use the data augmentation technique to enhance the original data set. The increasing training set can improve the generalization ability of the neural networks, as well as help the neural networks to learn some features with scale,

rotation, and color invariance, thus improving the prediction performance of the classifier.

We use data rotation and mirroring to augment our images. For each sample picture x_i, $i = 1, 2, \ldots, n$, n is the total number of pictures in a sample set X, we first divide it into 9 equal-sized sub-images $z_{(i,j)}$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, 9$. Then we use the mirror edge padding to fill in the sub-images of equal length and width, and obtain the image $z'_{(i,j)}$. For each $z'_{(i,j)}$, we apply two data augmentation operations: In the first operation, we rotate each image into 0°, 90°, 180° and 270°; In the second operation we do the horizontal flipping, vertical flipping, and channel flipping. So that each sub-image $z'_{(i,j)}$ can generate 16 images, where the image labels are the same

TABLE 2. The experimental data setting of the IHC cervical histopathology image dataset. The first row indicates the staining methods. The second row indicates three differentiation stages. The third to the last rows indicate training, validation and testing datasets for each stage and staining method, respectively.

Dataset	AQP			HIF			VEGF		
	Well	Moderate	Poor	Well	Moderate	Poor	Well	Moderate	Poor
Training	3802	4032	3226	4493	4378	3917	3341	3802	4378
Validation	475	504	403	562	547	490	418	475	547
Test	475	504	403	561	547	489	417	475	547

TABLE 3. Evaluation metrics.

Assessments	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F1-score	$\frac{2TP}{2TP+FP+FN}$

as the original image x_i. Hence, each sample image x_i is augmented to 144 images. Finally, the size of the data set is increased from 307 to 44208 after data augmentation. Our images have a small field of view, and most of them are cancer regions with a single stage of differentiation when they are prepared. Therefore, we make the patch label inherit the image-level label.

C. EXPERIMENTAL SETTING

Histopathological images are usually noisy. However, the proposed technique is a deep learning based approach, which provides feature detection and extraction automatically [47]. Therefore, the proposed method does not need a noise reduction step. In the experiment, the augmented 44208 images are applied to examine the proposed EL method. The training set, validation set, and test set are divided according to the ratio of 8 : 1 : 1 as shown in Table. 2. In order to make full use of the dataset, we randomly select 80% of the data set as the training set, and then randomly select 50% from the remainder of the dataset as the validation set, and the rest of the data as the test set.

D. EXPERIMENTAL EVALUATION

The performance of the classifiers is evaluated using accuracy, precision, recall, and F1-score metrics. The accuracy is the ratio of the number of samples correctly classified by the classifier to the total number of samples. The precision reflects the proportion of the positive samples that are determined by the classifier to be positive samples. The recall rate reflects the positive case of correct judgment accounting for the proportion of the total positive samples and the F1-score is an indicator that comprehensively considers the accuracy and the recall [48]. Table. 3 describes these performance metrics. In this paper, the samples of the categories studied at this time are positive samples, and the samples of other categories are negative samples.

As shown in Table. 3, TP is the True Positive (positive sample is predicted to be positive), TN is the True Negative (negative sample is predicted to be negative), FP is the False Positive (negative sample is predicted to be positive), and FN is the False Negative (positive sample is predicted to be negative).

E. EXPERIMENTAL RESULTS

1) EVALUATION OF THE TL AND ETL RESULT

We use accuracy, precision, recall, and F1-score to evaluate the TL and the proposed ETL method. Especially, we carry out 9-fold cross-validation to obtain an overall balanced evaluation on the IHC dataset (each staining method 3-folds), and the results are shown in Table. 4. The results are analyzed in three aspects: differentiation stage (well, moderate, and poor differentiation), staining method (AQP, HIF, and VEGF staining), and experimental method (TL and ETL method).

For the first aspect: Differentiation stage (well, moderately, and poor differentiation). It can be seen from Table. 4 that the poor differentiation stage achieves the best classification accuracy on all TL methods and the ETL method. Furthermore, this stage also achieves the best classification results on precision and F1-score among others. For recall, the moderate differentiation stage achieves the best classification results on all TL methods and the ETL method except the VGG-16 TL approach. As for the VGG-16 TL method, the poor differentiation stage achieves the best classification results. Moreover, all evaluation indexes of the proposed ETL method have higher values than those of the individual TL method. On the whole, the poor differentiation stage achieves the best classification results among the three differentiation stages, and the proposed ETL method has the best classification performance.

For the second aspect: The staining method (AQP, HIF, and VEGF staining). It is observed from Table. 4 that for accuracy, the VEGF staining method obtains the best classification results among all TL methods and the ETL method except Resnet-50 network. As for the Resnet-50 network, the HIF staining technique achieves the best classification results. For precision, the VEGF staining algorithm achieves the best classification results compared to other TL methods and the ETL method. Furthermore, the VEGF staining method also achieves the best classification results on F1-score of all. For recall, the AQP staining technique achieves the best classification results on Inception-V3 and Xception TL methods and the ETL method. As for VGG-16 TL approach, the VEGF staining algorithm achieves the best classification results, and the HIF staining technique achieves the best classification results for ResNet-50. Moreover, the proposed ETL method has a higher value than those of the individual TL method. Overall, the VEGF staining system achieves the best distribution results among the three staining approaches.

TABLE 4. The TL and ETL results. The third to the last columns in the first row denote four evaluation indexes, respectively. The second to the last rows in the first column denote TL and ETL methods, respectively. The second to the last rows in the second column denote three staining methods, respectively. The second to the last rows in the third column denote three differentiation stages, respectively. (In [%].)

Result			Accuracy	Precision	Recall	F1-score
Inception-V3 TL	AQP	Well	81.84	73.24	74.32	73.77
		Moderate	87.55	82.81	**83.13**	74.19
		Poor	85.60	75.89	74.19	75.03
	HIF	Well	84.10	79.46	73.80	76.52
		Moderate	86.10	78.07	82.63	80.28
		Poor	88.35	80.48	81.80	81.14
	VEGF	Well	83.32	69.71	75.06	72.29
		Moderate	83.04	74.63	73.68	74.15
		Poor	**88.88**	**87.14**	83.00	**85.02**
Xception TL	AQP	Well	87.63	86.19	76.21	80.89
		Moderate	89.80	82.35	**91.67**	86.76
		Poor	90.74	84.29	83.87	84.08
	HIF	Well	87.54	87.87	74.87	80.85
		Moderate	86.35	74.74	90.86	82.01
		Poor	91.67	89.21	82.82	85.90
	VEGF	Well	89.65	84.01	79.38	81.63
		Moderate	89.02	80.08	88.84	84.23
		Poor	**92.56**	**92.47**	87.57	**89.95**
VGG-16 TL	AQP	Well	95.44	93.83	92.84	93.33
		Moderate	96.45	94.00	96.43	95.20
		Poor	96.67	95.19	93.30	94.24
	HIF	Well	95.87	95.08	93.05	94.05
		Moderate	95.93	93.35	94.88	94.11
		Poor	96.68	94.31	94.89	94.60
	VEGF	Well	96.11	94.13	92.33	93.22
		Moderate	95.55	92.37	94.32	93.33
		Poor	**98.05**	**97.61**	**97.26**	**97.44**
Resnet-50 TL	AQP	Well	95.95	96.45	91.58	93.95
		Moderate	96.96	93.92	**98.02**	95.92
		Poor	96.96	94.57	95.04	94.80
	HIF	Well	96.43	97.37	92.34	94.78
		Moderate	96.68	92.88	97.81	95.28
		Poor	**98.25**	97.14	97.14	97.14
	VEGF	Well	96.04	94.33	91.85	93.07
		Moderate	95.34	91.46	94.74	93.07
		Poor	98.19	**98.15**	97.07	**97.61**
Our ETL method in this paper	AQP	Well	97.61	98.46	94.53	96.46
		Moderate	98.19	95.62	**99.60**	97.57
		Poor	98.26	97.26	96.77	97.01
	HIF	Well	97.18	99.05	92.87	95.86
		Moderate	97.00	92.65	99.09	95.76
		Poor	98.43	97.74	97.14	97.44
	VEGF	Well	97.78	97.77	94.48	96.10
		Moderate	97.50	93.81	98.95	96.31
		Poor	**98.61**	**99.25**	97.07	**98.15**

TABLE 5. The overall accuracy of De-novo trained CNNs, TL methods and ETL method. The first row shows three staining methods. (In [%].)

	AQP	HIF	VEGF
Inception-V3	73.81	75.77	73.97
Xception	79.59	78.77	81.31
VGG-16	88.28	90.17	91.38
Resnet-50	90.88	91.86	90.69
Inception-V3 TL	77.50	79.27	77.62
Xception TL	84.08	82.78	85.62
VGG-16 TL	94.28	94.24	94.86
Resnet-50 TL	94.93	95.68	94.79
Our ETL method in this paper	**97.03**	**96.31**	**96.94**

For the third aspect: Experimental method (De-novo trained CNNs, TL and ETL method). Table. 5 exhibits the overall performance of the proposed techniques. It can be seen from the table that the accuracy of each TL algorithm is about 1% to 3% higher than the method of training models from scratch (De-novo trained CNNs). At the same time, each transfer training process saves about 4 hours than the De-novo trained CNNs. This understanding recommends training CNNs using TL. Finally, it is observed that the advanced ETL technique has higher correctness than any of the single TL methods, which proves that the sophisticated ETL approach has the best classification performance.

2) VISUALIZED ANALYSIS

In order to show the classification performance of the proposed model more intuitively, we exhibit confusion matrix and loss curve for a visualized analysis. As shown in Fig. 4 and Fig. 5, the ETL method has low probability of error classification, and the ETL method has higher stability. The accuracy and loss curve of training and validation sets are shown in Fig. 6. It is clear to see from the curve that the VGG-16 and Resnet-50 networks are more stable.

Moreover, the histogram of the individual TL and EL with their accuracy is provided. As shown in Fig. 7, after EL, the accuracy of well, moderate, and poorly differentiated is increased. For well differentiation classification results, the accuracy of AQP staining is improved by 1.66%, HIF staining is increased by 0.75%, and VEGF staining is increased by 1.74% compared with the highest accuracy of single TL classification. For moderate differentiation, the accuracy of AQP staining is risen by 1.23%, HIF staining by 0.32% and VEGF staining by 2.16% compared with TL. Similarly, for the poor differentiation, AQP staining, HIF staining and VEGF staining improved by 1.3%, 0.18% and 0.42%, respectively. It is worth noting that compared with the highest accuracy of single TL classification, the overall accuracy of AQP staining increased by 2.10%, HIF staining by 0.63%, and VEGF staining by 2.15%. Above all, the effectiveness of the EL strategy is proved.
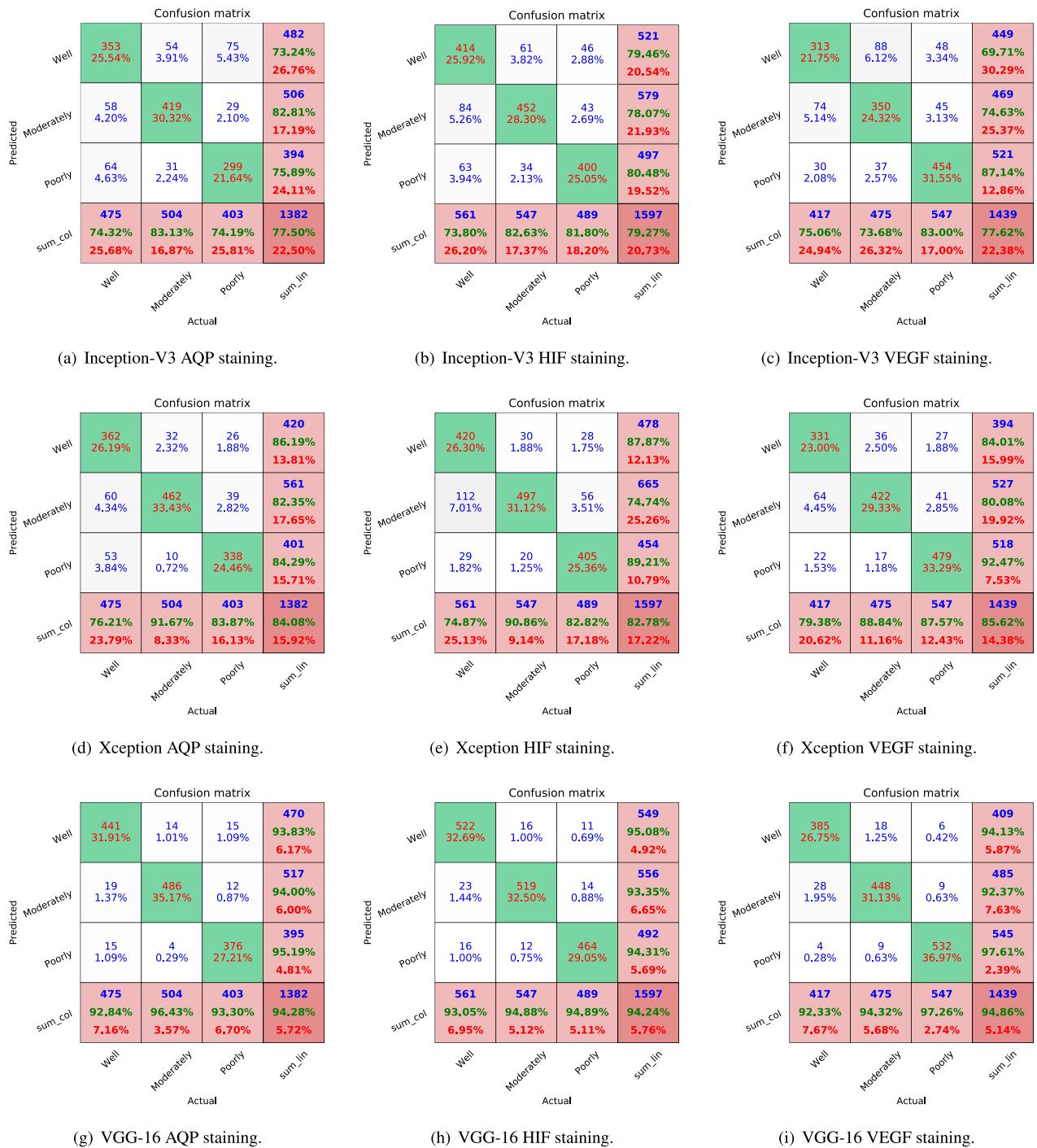
FIGURE 4. The confusion matrix of TL and ETL methods using different staining methods. (a)-(c) are confusion matrix of three staining methods of Inception-V3 network, (d)-(f) are confusion matrix of three staining methods of Xception network, (g)-(i) are confusion matrix of three staining methods of VGG-16 network.

Finally, Fig. 8 exhibits the output of our recommended algorithms. It can be seen from the figure that the correctly predicted images have full information about the stage of differentiation, so they are easy to distinguish. The wrongly predicted images contain little information about the stages of differentiation of tumor cells. Furthermore, in the data augmentation process, we use image padding, which may add redundant information. These reasons can disturb the computer to classify the differentiation stages of cervical cancer histopathology microscopic images.

F. COMPARISON WITH PREVIOUS WORKS

In our previous work, different methods are introduced to classify well, moderate and poorly differentiated cervical

(a) Resnet-50 AQP staining.

(b) Resnet-50 HIF staining.

(c) Resnet-50 VEGF staining.

(d) Ensemble AQP staining.

(e) Ensemble HIF staining.

(f) Ensemble VEGF staining.
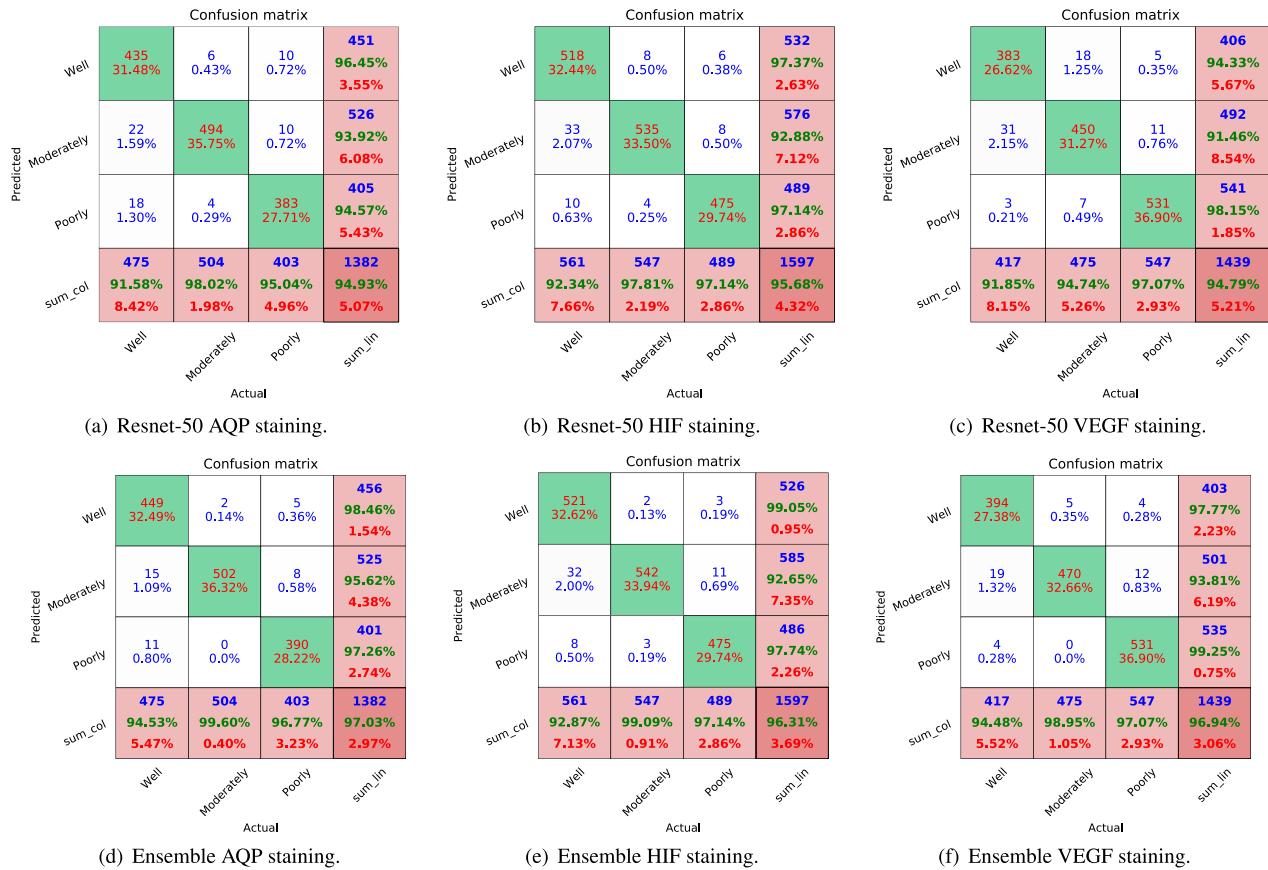
FIGURE 5. The confusion matrix of TL and ETL methods using different staining methods. (a)-(c) are confusion matrix of three staining methods of Resnet-50 network, (d)-(f) are confusion matrix of three staining methods of EL method.

histopathological images. In order to show the performance of the proposed ETL method in this paper, we compare it with the TL method based on Inception-V3, the ETL method based on VGG-16 and Inception-V3, the method based on multi-layer hidden conditional random field and the methods combining classical features with SVM, ANN, and Random Forests (RF) classifiers respectively. The classical features include Scale-invariant Feature Transform (SIFT), GLCM, Histogram of Oriented Gradient (HOG), and color histogram features. In Table. 6, all the comparison methods and results mentioned above are summarized.

As shown in Table. 6, the accuracy of the single TL method [10] is about 30% higher than the average accuracy of the classical method. The accuracy of the ETL method [13] is 2.7% higher than the single TL method. However, our proposed method achieves the highest accuracy of 98.61% among all the methods, and the accuracy is 10.61% higher than the second [11], [12], showing the effectiveness of our ETL method in this paper.

G. COMPUTATIONAL TIME

Finally, we describe the computational time of our proposed ETL method. In our experiment, we use a workstation with Intel® Core™ i7-7700 CPU with 3.60 GHz, 32 GB RAM, and GeForce GTX 1080 8GB. The training time for each

staining method is about 8.5 hours, and the test time of each staining method is about 5 seconds. Although the training time takes more time, the test time costs only a few seconds, showing the feasibility of our proposed ETL method.

H. ADDITIONAL EXPERIMENT

1) IMAGE DATASET

In order to evaluate the generalization ability of the proposed ETL method, a publicly available Herlev dataset [49] is employed, which is a cytopathological image set for cervical cells. This dataset includes 917 single-cell images, where 242 images belong to a benign class and 675 images belong to a malignant class. According to experimental setting of the existing work, the training, validation, and test sets are divided according to the ratio of 3: 1: 1. So, the number of benign cells in these three datasets is 144, 49, and 49 respectively, while the number of malignant cells in these three datasets is 405, 135, and 135 respectively. To solve the problem of the unbalanced dataset, the benign cell images in the training set and the validation set are augmented. First, each original image is rotated by 180° and flipped by a mirror. Then, the images of benign cells with the same number of malignant cells are randomly selected as the training and validation sets. An example of the Herlve dataset is shown in Fig. 9.
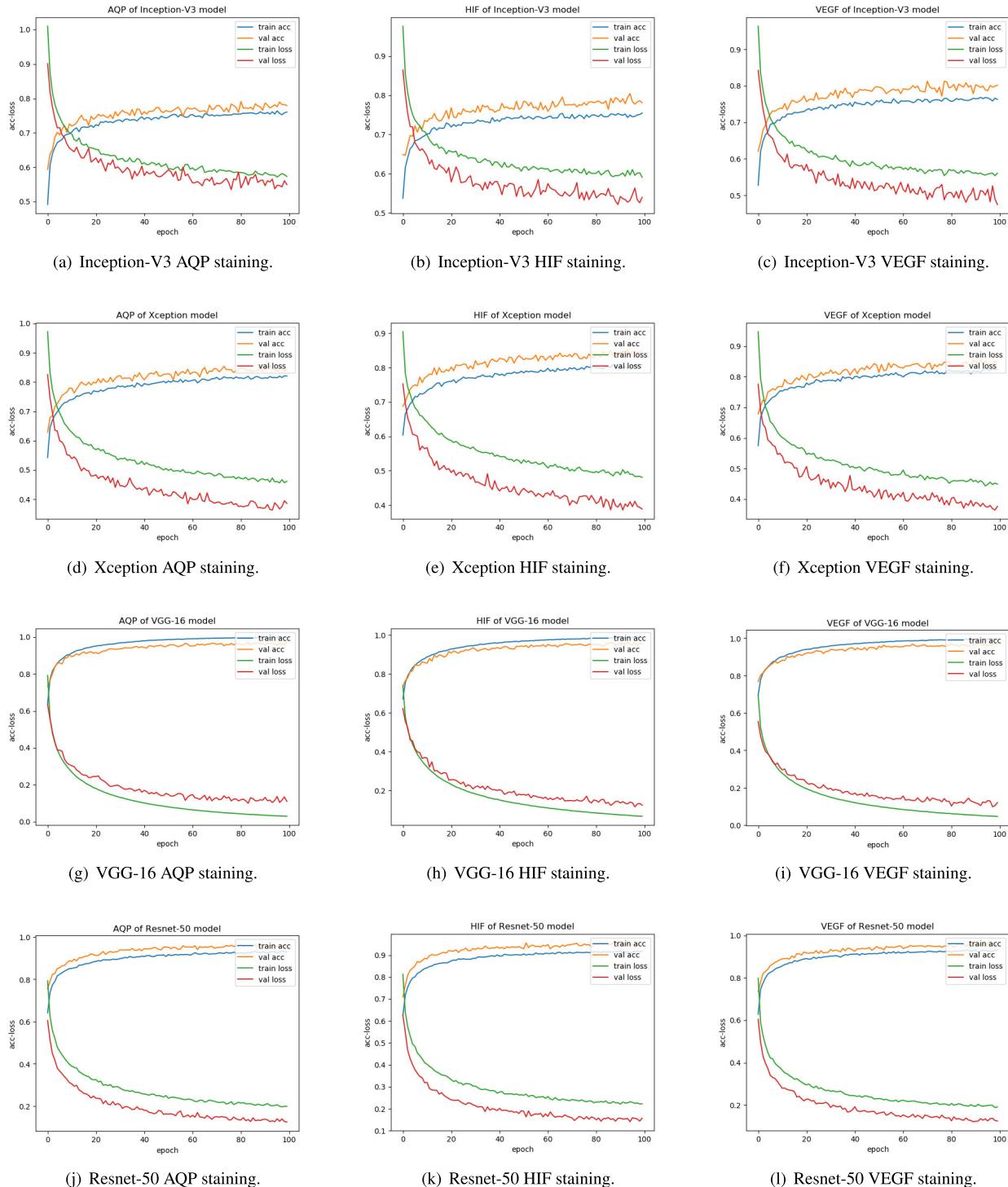
FIGURE 6. The accuracy and loss curve of TL methods. (a)-(c) are curves of three staining methods of Inception-V3 network, (d)-(f) are curves of three staining methods of Xception network, (g)-(i) are curves of three staining methods of VGG-16 network, (j)-(l) are curves of three staining methods of EL method.

2) PERFORMANCE COMPARISON BETWEEN TL AND ETL METHODS

The TL and ETL methods are used for binary classification on the Herlev dataset [49]. Their 3-fold cross-validation results are shown in Table. 7.

From Table. 7 we can find that among the single TL methods, the classification performance of the Resnet-50 network is the best, and the highest value of all evaluation indicators is obtained. Among them, accuracy is 95.65%, which is about 2.5% to 5.5% higher than other networks. Precision
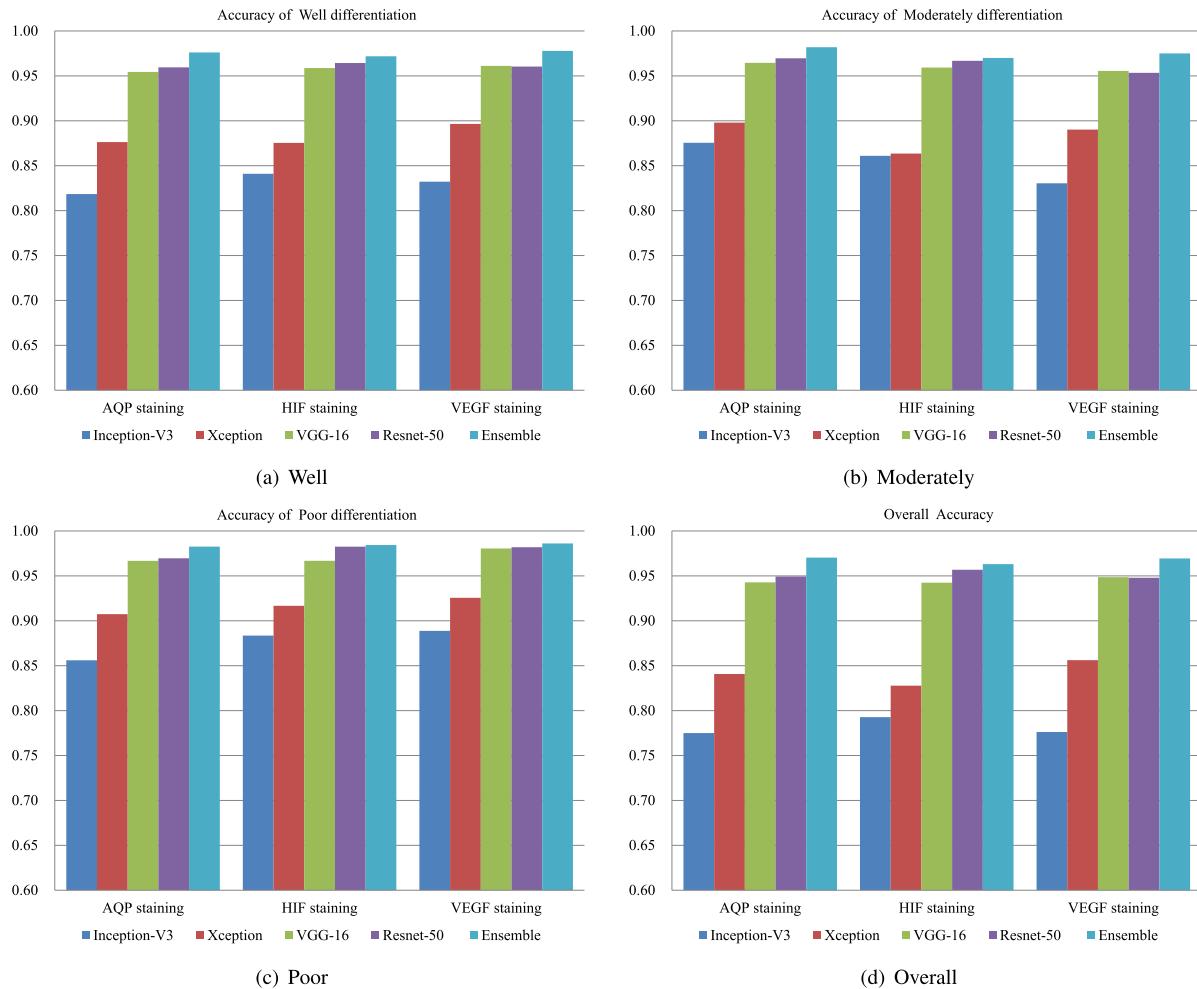
FIGURE 7. A comparison between individual TL and EL classification performance on three differentiation stages.
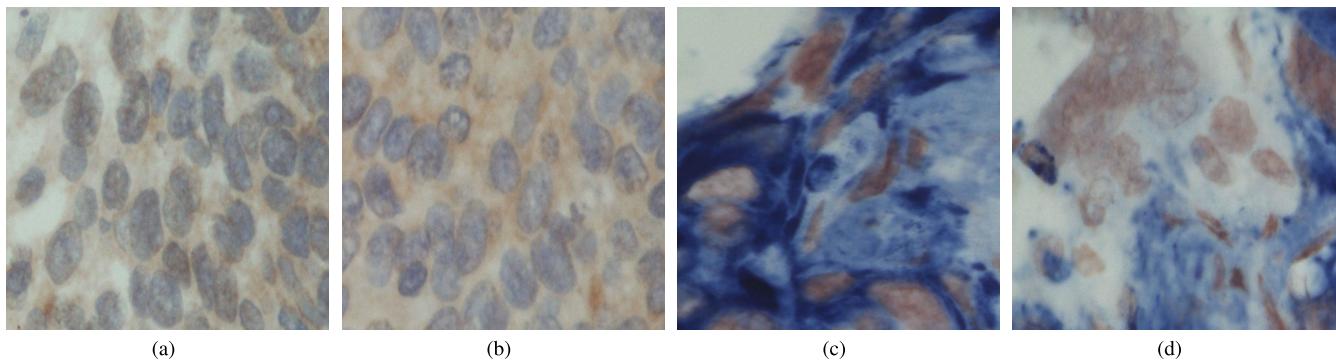
FIGURE 8. An example of the classification result. (a) and (b) are correctly classified images. (c) and (d) are wrongly classified images.

is 96.35%, which is about 1.5% to 2.5% higher than other networks. Recall is 97.78%, which is about 3% to 5% higher than other networks. F1-score is 97.06%, which is about 1.8% to 3.8% higher than other networks. However, it is worth noting that the ETL method performs better than all the single TL methods, obtaining an accuracy of 98.37%, a precision of 98.53%, a recall of 99.26%, and a F1-score of 98.89%. It shows that the ETL method proposed in this paper can
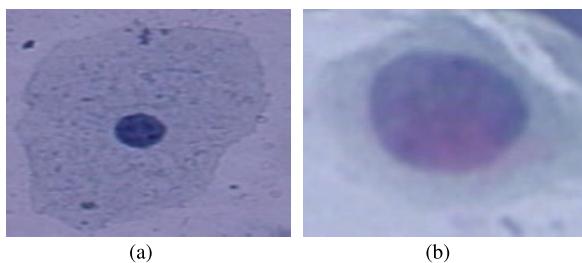
improve the performance of classification on Herlev dataset effectively.

3) PERFORMANCE COMPARISON BETWEEN ETL AND OTHER EXISTING METHODS

In order to evaluate the effectiveness of the proposed method, a comparative analysis with existing work on Herlev dataset is presented in Table. 8. For more details, please refer to our

TABLE 6. A comparison of ETL method with other existing methods. (In [%].)

| Method	Accuracy
Color histogram features & RBF-SVM	34.62
Color histogram features & Linear-SVM	62.82
Color histogram features & ANN (6 layers)	71.30
Color histogram features & RF (2^11 trees)	63.57
SIFT features & RBF-SVM	34.62
SIFT features & Linear-SVM	45.16
SIFT features & ANN (6 layers)	50.28
SIFT features & RF (2^11 trees)	47.61
GLCM features & RBF-SVM	34.62
GLCM features & Linear-SVM	42.55
GLCM features & ANN (6 layers)	59.54
GLCM features & RF (2^11 trees)	57.76
HOG features & RBF-SVM	45.33
HOG features & Linear-SVM	40.14
HOG features & ANN (6 layers)	46.65
HOG features & RF (2^11 trees)	45.73
TL framework based on Inception-V3 [10]	77.30
ETL framework based on Inception-V3 and VGG-16 [13]	80.00
Multilayer hidden conditional random fields [11], [12]	88.00
Our ETL method in this paper	**98.61**

FIGURE 9. An example of cell images on Herlev dataset. (a) is benign cell, (b) is malignant cell.

TABLE 7. TL and ETL classification results on Herlev dataset. (In [%].)

	Accuracy	Precision	Recall	F1-score
Inception-V3 TL	90.22	94.12	94.81	94.46
Xception TL	91.85	94.85	95.56	95.20
VGG-16 TL	92.93	93.99	92.59	93.28
Resnet-50 TL	95.65	96.35	97.78	97.06
ETL	**98.37**	**98.53**	**99.26**	**98.89**

previous survey paper about "Cervical Cytopathology Image Analysis" [50].

From the comparison Table. 8, we can find that our proposed ETL method has a certain degree of competition in accuracy and precision compared with the existing methods. These results validate the robustness of our proposed

TABLE 8. ETL comparative analysis with existing methods on Herlev dataset. Adaptive Neuro Fuzzy Inference System (ANFIS), Least Square Support Vector Machine (LSSVM). (In [%].)

Method	Accuracy	Precision	Recall	F1-score
TL based on AlexNet & Decision tree [51]	99.30	-	-	-
TL based on AlexNet & SVM [52]	99.19	99.51	99.50	-
Hybrid classifier (SVM & ANFIS) [53]	99.10	-	-	-
Hybrid ensemble technique (composed of 15 different classifiers) [54]	98.57	-	-	-
TL based on CNN [55]	98.30	-	-	-
TL based on CNN & SVM [56]	95.10	-	-	-
AlexNet & classifiers (LSSVM/SoftMax Regression) [57]	94.61	-	-	-
VGG16 and ResNet [58]	86.56	85.94	79.04	82.16
Our ETL method in this paper	98.37	98.53	99.26	98.89

algorithm, which is not only applicable for cervical histopathological image analysis, but also suitable for cervical cytopathological images.

V. LIMITATIONS AND FUTURE WORK

In this paper, weighted voting based EL is proposed to classify cervical histopathological images. Especially, three cervical cancer differentiation stages are classified, where the highest overall accuracy of 97.03% is achieved on AQP staining method. Meanwhile, the highest accuracy of 98.61% is achieved on poorly differentiated of VEGF staining. However, this method presents some deficiencies worthy to be pointed out. Firstly, we use four base learners, it requires a very sophisticated computer for its implementation and presents very high computational cost, and it adds to the workload as well. Secondly, there is no patient-level label for our current IHC dataset, so this paper does not do basic studies based on patients. At the moment, we are collecting more data and working in this direction. In the future, we plan to develop more efficient and simple systems to do the classification of cervical histopathology images.

REFERENCES

[1] D. M. Parkin, F. Bray, J. Ferlay, and P. Pisani, "Global cancer statistics, 2002," *CA, Cancer J. Clinicians*, vol. 55, no. 2, pp. 74–108, Mar. 2005.

[2] C. Li, D. Xue, Z. Hu, H. Chen, Y. Yao, Y. Zhang, M. Li, Q. Wang, and N. Xu, "A survey for breast histopathology image analysis using classical and deep neural networks," in *Information Technology in Biomedicine* (Advances in Intelligent Systems and Computing), vol. 1011, E. Pietka, P. Badura, J. Kawa, and W. Wieclawek, Eds. Berlin, Germany: Springer, 2019, pp. 222–233.

[3] J. Zhang, J. Lu, D. Hong, L. Wu, W. Xie, and X. Peng, "A correction algorithm for cone-beam computed tomography dose calculations in patients with cervical cancer," *Chin. J. Med. Imag. Technol.*, vol. 33, no. 1, pp. 114–118, 2017.

[4] S. E. Waggoner, "Cervical cancer," *Lancet*, vol. 361, no. 9376, pp. 2217–2225, 2003.

[5] N. Crossley, C. Tipton, T. Meier, M. Sudhoff, and J. Kharofa, "The value of hybrid interstitial tandem and ring applicators for organ at risk dose reduction in small volume cervical cancer," *Brachytherapy*, vol. 17, no. 4, p. S111, Jul. 2018.

[6] C. Ngelange, N. Muñoz, F. X. Bosch, M. R. Festin, J. Deacon, M. V. Jacobs, M. Santamaria, C. J. L. M. Meijer, and J. M. M. Walboomers, "Causes of cervical cancer in the philippines: A case-control study," *JNCI, J. Nat. Cancer Inst.*, vol. 90, no. 1, pp. 43–49, Jan. 1998.

[7] J. Eluf-Neto, M. Booth, N. Muñoz, F. Bosch, C. Meijer, and J. Walboomers, "Human papillomavirus and invasive cervical cancer in Brazil," *Brit. J. Cancer*, vol. 69, no. 1, pp. 114–119, Jan. 1994.

[8] J. A. A. Jothi and V. M. A. Rajam, "A survey on automated cancer diagnosis from histopathology images," *Artif. Intell. Rev.*, vol. 48, no. 1, pp. 31–81, Jun. 2017.

[9] G. L. Wied, P. H. Bartels, G. F. Bahr, and D. G. Oldfield, "Taxonomic intra-cellular analytic system (TIICAS) for cell identification," *Acta Cytologica*, vol. 12, no. 3, p. 180, 1968.

[10] C. Li, D. Xue, X. Zhou, J. Zhang, H. Zhang, Y. Yao, F. Kong, L. Zhang, and H. Sun, "Transfer learning based classification of cervical cancer immunohistochemistry images," in *Proc. 3rd Int. Symp. Image Comput. Digit. Med. (ISICDM)*, 2019, pp. 102–106.

[11] C. Li, H. Chen, D. Xue, Z. Hu, L. Zhang, L. He, N. Xu, S. Qi, H. Ma, and H. Sun, "Weakly supervised cervical histopathological image classification using multilayer hidden conditional random fields," in *Information Technology in Biomedicine* (Advances in Intelligent Systems and Computing), vol. 1011, E. Pietka, P. Badura, J. Kawa, and W. Wieclawek, Eds. Berlin, Germany: Springer, 2019, pp. 209–221.

[12] C. Li, H. Chen, L. Zhang, N. Xu, D. Xue, Z. Hu, H. Ma, and H. Sun, "Cervical histopathology image classification using multilayer hidden conditional random fields and weakly supervised learning," *IEEE Access*, vol. 7, pp. 90378–90397, 2019.

[13] C. Li, D. Xue, F. Kong, Z. Hu, H. Chen, Y. Yao, H. Sun, L. Zhang, J. Zhang, T. Jiang, J. Yuan, and N. Xu, "Cervical histopathology image classification using ensembled transfer learning," in *Information Technology in Biomedicine* (Advances in Intelligent Systems and Computing), vol. 1011, E. Pietka, P. Badura, J. Kawa, and W. Wieclawek, Eds. Berlin, Germany: Springer, 2019, pp. 26–37.

[14] Q. Ji, J. Engel, and E. Craine, "Classifying cervix tissue patterns with texture analysis," *Pattern Recognit.*, vol. 33, no. 9, pp. 1561–1573, Sep. 2000.

[15] J. Einenkel, U.-D. Braumann, L.-C. Horn, N. Pannicke, J.-P. Kuska, A. Schütz, B. Hentschel, and M. Höckel, "Evaluation of the invasion front pattern of squamous cell cervical carcinoma by measuring classical and discrete compactness," *Computerized Med. Imag. Graph.*, vol. 31, no. 6, pp. 428–435, Sep. 2007.

[16] M. Zhao, L. Chen, L. Bian, J. Zhang, C. Yao, and J. Zhang, "Feature quantification and abnormal detection on cervical squamous epithelial cells," *Comput. Math. Methods Med.*, vol. 2015, no. 3, 2015, Art. no. 941680.

[17] J. Su, X. Xu, Y. He, and J. Song, "Automatic detection of cervical cancer cells by a two-level cascade classification system," *Anal. Cellular Pathol.*, vol. 2016, no. 4, 2016, Art. no. 9535027.

[18] H. Komagata, T. Ichimura, Y. Matsuta, M. Ishikawa, K. Shinoda, N. Kobayashi, and A. Sasaki, "Feature analysis of cell nuclear chromatin distribution in support of cervical cytology," *J. Med. Imag.*, vol. 4, no. 4, p. 1, Oct. 2017.

[19] J. E. Macgregor, S. M. Moss, D. M. Parkin, and N. E. Day, "A case-control study of cervical cancer screening in North East Scotland," *Brit. Mach. J.*, vol. 290, no. 6481, pp. 1543–1546, May 1985.

[20] N. Krieger, M. T. Bassett, and S. L. Gomez, "Breast and cervical cancer in 187 countries between 1980 and 2010," *Lancet*, vol. 379, no. 9824, pp. 1391–1392, Apr. 2012.

[21] S. H. Ong, X. C. Jin, Jayasooriah, and R. Sinniah, "Image analysis of tissue sections," *Comput. Biol. Med.*, vol. 26, no. 3, pp. 269–279, May 1996.

[22] M. Guillaud, D. Cox, K. Adler-Storthz, A. Malpica, G. Staerkel, J. Matisic, D. Van Niekerk, N. Poulin, M. Follen, and C. MacAulay, "Exploratory analysis of quantitative histopathology of cervical intraepithelial neoplasia: Objectivity, reproducibility, malignancy-associated changes, and human papillomavirus," *Cytometry*, vol. 60A, no. 1, pp. 81–89, 2004.

[23] H. Lange and D. G. Ferris, "Computer-aided-diagnosis (CAD) for colposcopy," *Proc. SPIE*, vol. 5747, pp. 71–84, Apr. 2005.

[24] Y. Wang, "Computer assisted diagnosis of cervical intraepithelicel neoplasia (CIN) using histological virtual slides," Queens Univ. Belfast, Belfast, U.K., Tech. Rep., 2008. [Online]. Available: https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.492486

[25] Rahmadwati, G. Naghdy, M. Ross, C. Todd, and E. Norachmawati, "Classification cervical cancer using histology images," in *Proc. 2nd Int. Conf. Comput. Eng. Appl.*, Mar. 2010, pp. 515–519.

[26] R. Rahmadwati, G. Naghdy, M. Ros, and C. Todd, "Computer aided decision support system for cervical cancer classification," *Proc. SPIE*, vol. 8499, Oct. 2012, Art. no. 849919.

[27] P. Guo, "Cervical cancer histology image feature extraction and classification," M.S. thesis, Missouri Univ. Sci. Technol., Rolla, MO, USA, 2014. [Online]. Available: https://scholarsmine.mst.edu/masters_theses/7302/

[28] L. Wei, Q. Gan, and T. Ji, "Cervical cancer histology image identification method based on texture and lesion area features," *Comput. Assist. Surg.*, vol. 22, no. 1, pp. 186–199, Oct. 2017.

[29] C. Li, Z. Hu, H. Chen, D. Xue, N. Xu, Y. Zhang, X. Li, Q. Wang, and H. Ma, "Cervical histopathology image clustering using graph based unsupervised learning," in *Proc. 11th Int. Conf. Modelling, Identificat. Control*, Lecture Notes in Computer Science, vol. 582, R. Wang, Z. Chen, W. Zhang, and Q. Zhu, Eds. Berlin, Germany: Springer, 2019, pp. 141–152.

[30] C. Li, H. Chen, X. Li, N. Xu, Z. Hu, D. Xue, S. Qi, H. Ma, L. Zhang, and H. Sun, "A review for cervical histopathology image analysis using machine vision approaches," *Artif. Intell. Rev.*, pp. 1–42, Feb. 2020, doi: 10.1007/s10462-020-09808-7.

[31] E. Purwanti, M. Bustomi, and R. Aldian, "Applied computing based artificial neural network for classification of cervical cancer," in *Proc. CISAK*, 2013, pp. 1–4.

[32] Y. Song, L. Zhang, S. Chen, D. Ni, B. Li, Y. Zhou, B. Lei, and T. Wang, "A deep learning based framework for accurate segmentation of cervical cytoplasm and nuclei," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 2903–2906.

[33] M. A. Devi, S. Ravi, J. Vaishnavi, and S. Punitha, "Classification of cervical cancer using artificial neural networks," *Procedia Comput. Sci.*, vol. 89, pp. 465–472, 2016.

[34] B. Zhang, S. Qi, P. Monkam, C. Li, F. Yang, Y.-D. Yao, and W. Qian, "Ensemble learners of multiple deep CNNs for pulmonary nodules classification using CT images," *IEEE Access*, vol. 7, pp. 110358–110371, 2019.

[35] M. M. Ghazi, B. Yanikoglu, and E. Aptoula, "Plant identification using deep neural networks via optimization of transfer learning parameters," *Neurocomputing*, vol. 235, pp. 228–235, Apr. 2017.

[36] T. Kamishima, M. Hamasaki, and S. Akaho, "TrBagg: A simple transfer learning method and its application to personalization in collaborative tagging," in *Proc. 9th IEEE Int. Conf. Data Mining*, Dec. 2009, pp. 219–228.

[37] F. Shoeleh and M. Asadpour, "Graph based skill acquisition and transfer learning for continuous reinforcement learning domains," *Pattern Recognit. Lett.*, vol. 87, pp. 104–116, Feb. 2017.

[38] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *Comput. Sci.*, pp. 2818–2826, Jun. 2015.

[39] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," 2016, *arXiv:1610.02357*. [Online]. Available: http://arxiv.org/abs/1610.02357

[40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Comput. Sci.*, pp. 1–14, 2014. [Online]. Available: https://arxiv.org/abs/1409.1556

[41] 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.

[42] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1063–6919.

[43] C. Qian, Y. Yu, K. Tang, Y. Jin, X. Yao, and Z.-H. Zhou, "On the effectiveness of sampling for evolutionary optimization in noisy environments," *Evol. Comput.*, vol. 26, no. 2, pp. 237–267, Jun. 2018.

[44] C. Qian, Y. Yu, and Z.-H. Zhou, "Analyzing evolutionary optimization in noisy environments," *Evol. Comput.*, vol. 26, no. 1, pp. 1–41, Mar. 2018.

[45] Z. Zhou, *Machine Learning*. Beijing, China: Tsinghua Univ. Press, 2016.

[46] Buja, L. Netter, and F. Krueger, "Netter's illustrated human pathology," Tech. Rep., 2008.

[47] E. Goceri, B. Goksel, J. B. Elder, V. K. Puduvalli, J. J. Otero, and M. N. Gurcan, "Quantitative validation of anti-PTBP1 antibody for diagnostic neuropathology use: Image analysis approach," *Int. J. Numer. Methods Biomed. Eng.*, vol. 33, no. 11, p. e2862, Nov. 2017.

[48] L. Collingwood and P. Jurka, "Creates a summary with precision, recall, and F1 scores," Tech. Rep., 2017.

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

[50] 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, no. 1, pp. 61687–61710, 2020.

[51] S. Gautam, H. K. K., N. Jith, A. K. Sao, A. Bhavsar, and A. Natarajan, "Considerations for a PAP smear image analysis system with CNN features," 2018, *arXiv:1806.09025*. [Online]. Available: http://arxiv.org/abs/1806.09025

[52] B. Taha, J. Dias, and N. Werghi, "Classification of cervical-cancer using pap-smear images: A convolutional neural network approach," in *Proc. Annu. Conf. Med. Image Understand. Anal.* Berlin, Germany: Springer, 2017, pp. 261–272.

[53] P. Sukuma and R. Gnanamurthy, "Computer aided detection of cervical cancer using pap smear images based on hybrid classifier," *Int. J. Appl. Eng. Res., Res. India Publications*, vol. 10, no. 8, pp. 21021–21032, 2015.

[54] A. Sarwar, V. Sharma, and R. Gupta, "Hybrid ensemble learning technique for screening of cervical cancer using papanicolaou smear image analysis," *Personalized Med. Universe*, vol. 4, pp. 54–62, Jul. 2015.

[55] 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.

[56] L. Nanni, S. Ghidoni, and S. Brahnam, "Handcrafted vs. Non-handcrafted features for computer vision classification," *Pattern Recognit.*, vol. 71, pp. 158–172, Nov. 2017.

[57] K. Bora, M. Chowdhury, L. B. Mahanta, M. K. Kundu, and A. K. Das, "Pap smear image classification using convolutional neural network," in *Proc. 10th Indian Conf. Comput. Vis., Graph. Image Process. (ICVGIP)*, 2016, p. 55.

[58] G. Forslid, H. Wieslander, E. Bengtsson, C. Wahlby, J.-M. Hirsch, C. R. Stark, and S. K. Sadanandan, "Deep convolutional neural networks for detecting cellular changes due to malignancy," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 82–89.

CHEN LI received the B.E. degree from the University of Science and Technology Beijing, China, in 2008, the M.Sc. degree from Northeast Normal University, China, in 2011, and the Ph.D. degree from the University of Siegen, Germany, in 2016. From 2016 to 2017, he was a Postdoctoral Researcher with the Johannes Gutenberg University of Mainz, Germany. He is currently an Associate Professor and the Head of the Research Group for Microscopic Image and Medical Image Analysis with the College of Medicine and Biological Information Engineering, Northeastern University, China. His research interests include microscopic image analysis, medical image analysis, machine learning, pattern recognition, machine vision, and multimedia retrieval. He is a Reviewer for several journals and conferences, including *Pattern Recognition*, *Future Generation Computer Systems*, *Artificial Intelligence in Medicine*, *Chemometrics and Intelligent Laboratory Systems*, IEEE Access, *Neurocomputing*, *Journal of X-Ray Science and Technology*, AAAI-20, and ITIB-2020.

YUDONG YAO (Fellow, IEEE) received the B.Eng. and M.Eng. degrees in electrical engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1982 and 1985, respectively, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, in 1988. From 1987 to 1988, he was a Visiting Student with Carleton University, Ottawa, ON, Canada. From 1989 to 2000, he was with Carleton University, Spar Aerospace Ltd., Montreal, QC, Canada, and Qualcomm Inc., San Diego, CA, USA. Since 2000, he has been with the Stevens Institute of Technology, Hoboken, NJ, USA, where he is currently a Professor and the Chair of the Department of Electrical and Computer Engineering. He holds one Chinese patent and more than 13 U.S. patents. His research interests include wireless communications, machine learning and deep learning techniques, and healthcare and medical applications. For his contributions to wireless communications systems, he was elected as a Fellow of the National Academy of Inventors, in 2015, and the Canadian Academy of Engineering, in 2017. He has served as an Associate Editor for the IEEE Communications Letters, from 2000 to 2008, and the IEEE Transactions on Vehicular Technology, from 2001 to 2006. He also served as an Editor for the IEEE Transactions on Wireless Communications, from 2001 to 2005.

DAN XUE received the B.E. degree from Shenyang Ligong University, China, in 2017, and the M.E. degree from Northeastern University, China, in 2019. She is currently an Algorithm Engineer with Neusoft Medical Company. She is also an External Researcher with the Research Group for Microscopic Image and Medical Image Analysis, Northeastern University. Her research interests include microscopic image analysis, medical image analysis, machine learning, pattern recognition, and machine vision. She is a Reviewer for IEEE Access.

MD MAMUNUR RAHAMAN received the B.Sc. degree from BRAC University, Dhaka, Bangladesh, in 2017. He is currently pursuing the master's degree with the Research Group for Microscopic Image and Medical Image Analysis, Biomedical and Information Engineering School, Northeastern University. His research interests include microscopic image analysis, medical image analysis, machine learning, pattern recognition, and machine vision.

XIAOMIN ZHOU received the B.E. degree from the University of Science and Technology Liaoning, China, in 2017. She is currently pursuing the master's degree with the Research Group for Microscopic Image and Medical Image Analysis, College of Medicine and Biological Information Engineering, Northeastern University, China. Her research interests include histopathology image analysis and deep learning.

JINGHUA ZHANG was born in 1996. He received the B.E. degree from Hefei University, China, in 2018. He is currently pursuing the master's degree with the Research Group for Microscopic Image and Medical Image Analysis, Northeastern University, China. His research interests include microscopic image segmentation and deep learning.
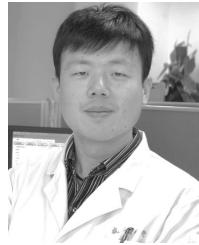
HAO CHEN received the B.E. and M.E. degrees from Northeastern University, China, in 2017 and 2019, respectively. He is currently an External Researcher with the Research Group for Microscopic Image and Medical Image Analysis, Northeastern University. His research interests include histopathology image analysis, machine learning, pattern recognition, and machine vision.

SHOULIANG QI (Member, IEEE) received the Ph.D. degree from Shanghai Jiao Tong University, in 2007. He joined the GE Global Research Center, where he was responsible for designing innovative magnetic resonance imaging (MRI) system. From 2014 to 2015, he was a Visiting Scholar with the Kempenhaeghe Epilepsy Center, The Netherlands, and the Eindhoven University of Technology. He is currently an Associate Professor with the Sino-Dutch Biomedical and Information Engineering School, Northeastern University, China. He had published more than 80 papers in peer-reviewed journals and international conferences. His research interests include conducting productive studies in intelligent medical imaging computing and modeling, machine learning, brain networks, and brain models. He received many academic awards, such as the Chinese Excellent Ph.D. Dissertation Nomination Award and the Award for Outstanding Achievement in Scientific Research from the Ministry of Education.

JINPENG ZHANG received the M.E. degree from Northeastern University, China, in 2019, where he is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering. His research interests include microscopic image analysis, medical image analysis, machine learning, pattern recognition, and machine vision.

HONGZAN SUN received the Ph.D. degree from China Medical University, China, in 2012. Since 2012, he has been an Associate Professor with the Shengjing Hospital of China Medical University. He was a visiting researcher in several hospitals in Germany and U.S. for more than one year. His research interests include medical imaging and image analysis.

• • •