

# Robust whole slide image analysis for cervical cancer screening using deep learning

Shenghua Cheng<sup>1,2,7</sup>, Sib0 Liu<sup>1,2,7</sup>, Jingya Yu<sup>1,2,7</sup>, Gong Rao<sup>1,2</sup>, Yuwei Xiao<sup>1,2</sup>, Wei Han<sup>1,2</sup>, Wenjie Zhu<sup>3</sup>, Xiaohua Lv<sup>1,2</sup>, Ning Li<sup>1,2</sup>, Jing Cai<sup>4</sup>, Zehua Wang<sup>4</sup>, Xi Feng<sup>5</sup>, Fei Yang<sup>5</sup>, Xiebo Geng<sup>1,2</sup>, Jiabo Ma<sup>1,2</sup>, Xu Li<sup>1,2</sup>, Ziquan Wei<sup>1,2</sup>, Xueying Zhang<sup>1,2</sup>, Tingwei Quan<sup>1,2</sup>, Shaoqun Zeng<sup>1,2</sup>, Li Chen<sup>6</sup>, Junbo Hu<sup>3</sup> & Xiuli Liu<sup>1,2</sup>

Computer-assisted diagnosis is key for scaling up cervical cancer screening. However, current recognition algorithms perform poorly on whole slide image (WSI) analysis, fail to generalize for diverse staining and imaging, and show sub-optimal clinical-level verification. Here, we develop a progressive lesion cell recognition method combining low- and high-resolution WSIs to recommend lesion cells and a recurrent neural network-based WSI classification model to evaluate the lesion degree of WSIs. We train and validate our WSI analysis system on 3,545 patient-wise WSIs with 79,911 annotations from multiple hospitals and several imaging instruments. On multi-center independent test sets of 1,170 patient-wise WSIs, we achieve 93.5% *Specificity* and 95.1% *Sensitivity* for classifying slides, comparing favourably to the average performance of three independent cytopathologists, and obtain 88.5% true positive rate for highlighting the top 10 lesion cells on 447 positive slides. After deployment, our system recognizes a one giga-pixel WSI in about 1.5 min.

<sup>1</sup> Collaborative Innovation Center for Biomedical Engineering, Wuhan National Laboratory for Optoelectronics-Huazhong University of Science and Technology, Wuhan, Hubei, China. <sup>2</sup> Britton Chance Center and MOE Key Laboratory for Biomedical Photonics, School of Engineering Sciences, Huazhong University of Science and Technology, Wuhan, Hubei, China. <sup>3</sup> Department of Pathology, Maternal and Child Hospital of Hubei Province, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China. <sup>4</sup> Department of Obstetrics and Gynecology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. <sup>5</sup> Department of Pathology, Hubei Cancer Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China. <sup>6</sup> Department of Clinical Laboratory, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China. <sup>7</sup> These authors contributed equally: Shenghua Cheng, Sib0 Liu, Jingya Yu. ✉email: [chenliisme@126.com](mailto:chenliisme@126.com); [cqjhu@163.com](mailto:cqjhu@163.com); [xliu@mail.hust.edu.cn](mailto:xliu@mail.hust.edu.cn)

Cervical cancer is one of the most common cancers in women. In 2020, there were about 604,127 women diagnosed with cervical cancer worldwide, and 341,831 died of the disease<sup>1</sup>. Many studies show that periodic inspection can reduce the incidence and mortality of cervical cancer<sup>2–5</sup>. Traditional smear tests require doctors to read the slides under microscopes, and usually each slide has tens of thousands of cells. During the diagnosis process, the cytopathology staff need to spend a lot of time traversing all the cells and diagnosing the suspicious cells among them<sup>6</sup>. Therefore, the manual screening is very labor-intensive and experience-dependent, possibly resulting in low sensitivity and false negatives<sup>7,8</sup>. With the progresses of digital whole-slide image-scanning instruments<sup>9</sup> and computer image processing technologies<sup>10</sup>, a lot of automated lesion cell recognition methods<sup>11</sup> are developed and bring hope for accurate and efficient computer-aided cervical cancer screening.

Traditional methods, mainly based on morphological and textural characteristics<sup>12</sup>, generally consist of image segmentation, feature extraction, and cell classification. The image segmentation is used to segment the nucleus or cytoplasm through image histogram threshold, optical density measuring, and image gradient<sup>13–16</sup>. The feature extraction primarily focuses on the shape features and textural information of nuclei. Cervical cells are then classified by random forest, support vector machine, and artificial neural network<sup>17,18</sup>. The performance of such methods is highly dependent on the segmentation effect and feature engineering. Subject to the principles, traditional methods have low accuracy for distinguishing lesion cells with fuzzy classification boundaries and limited generalization for diverse cytology slides derived from staining and imaging. In order to solve this problem, some commercial systems such as BD FocalPoint Slide Profiler<sup>19</sup> and Hologic ThinPrep Imaging System<sup>20</sup> adopted a closed-loop strategy that integrates slide preparation, staining, imaging, and recognizing to ensure the accuracy and stability of the systems. In fact, they circumvented the generalization problem without solving it, which limits the wide use of the products especially in impoverished areas.

With the development of deep learning<sup>21</sup>, convolutional networks (CNNs) have been applied to the identification of cervical lesion cells. Some studies have shown that CNNs improve the effect of nucleus segmentation<sup>22,23</sup>, and others utilize image classification and object detection CNNs to directly identify lesion cells without traditional segmentation process<sup>24–29</sup>. Compared with traditional methods, CNN-based methods learn feature representations automatically and have better generalization potential. However, current CNN-based computer-assisted diagnosis algorithms are insufficient in WSI-level analysis, generalization for diverse staining and imaging, and clinical-level verification. Most of the methods mainly focus on the recognition of local lesion cells, lacking WSI-level diagnostic analysis. Even a few methods<sup>30,31</sup> analyzed the whole-slide cervical images on large-scale datasets, but they still did not solve the generalization problem in practical applications and clinical-level verification. The image volume and annotation number of existing public datasets are small, such as Herlev<sup>32</sup>, ISBI14<sup>33</sup>, ISBI15<sup>34</sup>, and CERVIX93<sup>35</sup>, and most of them are provided with image tiles instead of WSIs, which hinders the progress of WSI-level diagnostic analysis. In addition, inference speed of CNNs on giga-pixel WSIs is challenging. Consequently, it is still difficult to apply current CNN-based methods in clinical cervical cancer screening scenes.

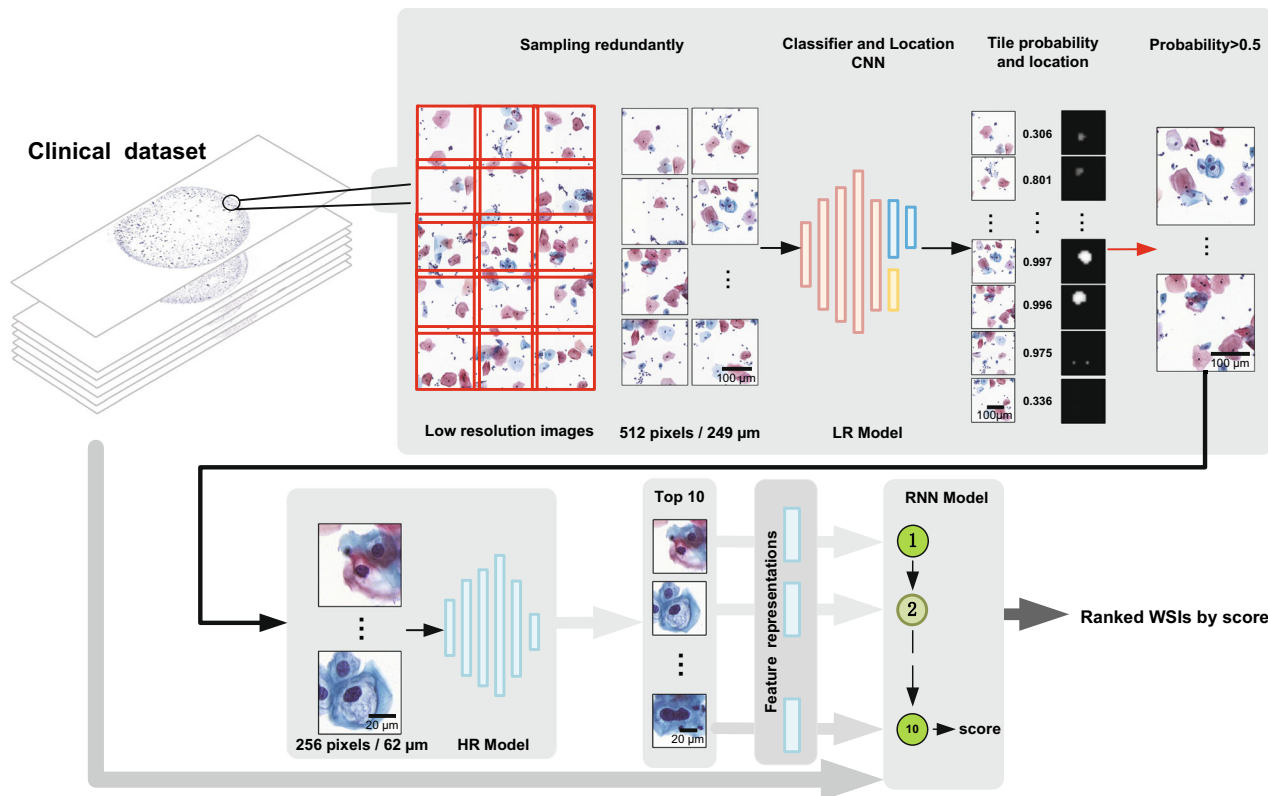
To address the above challenges in terms of WSI analysis accuracy, generalization, and speed, here we propose a computer-aided diagnosis system for cervical cancer screening based on deep learning and massive WSIs. In the cytopathologists' diagnosis process, they usually scan the slides under a low-power

microscope to find suspicious cells, and then further confirm them under a high-power microscope. Inspired by the strategy and considering the accuracy and speed, we design a progressive recognition method combining the low- and high-resolution WSIs. First, a CNN screens WSIs at low resolution (LR) to quickly locate the suspicious areas, and then these areas are further identified at high resolution (HR) by another CNN. Finally, the system recommends the 10 most suspicious lesion areas in each slide for further reviewing by cytopathologists. Besides recommending suspicious lesion areas, our system also evaluates the lesion degree of WSIs and gives a probability through developing a recurrent neural network (RNN)-based WSI classification model. The CNN image features of the top 10 areas are extracted and input to the RNN model to get the positive probability of WSIs. We integrate designed data augmentation in HSV color space, diverse data learning with group and category balancing, and hard sample mining to achieve high accuracy and good generalization of our system. We train and validate our system on patient-wise 3,545 WSIs and 79,911 annotations from five hospitals and five kinds of scanners. On multi-center independent test sets of 1,170 patient-wise WSIs, we achieve 93.5% specificity and 95.1% sensitivity for classifying slides. For the most confusing 121 WSIs of them, we achieve 50.0% specificity and 74.6% sensitivity, closely equivalent to the average level of three independent cytopathologists. The recommended top 10 lesion cells on 447 positive slides have an average true positive rate (TPR) of 88.5%. Compared with the current Hologic ThinPrep Imaging System, our system has a higher TPR of recommended cells and is more robust to staining and imaging style. When deploying the system with C++, multi-threading and TensorRt<sup>36</sup> are used to accelerate image processing and forward inference. Our system recognizes one giga-pixel WSI in about 1.5 min using one Nvidia 1080Ti GPU. This speed ensures a good user experience in clinical applications and provides the possibility of real-time augmented reality under microscopes.

In short, our work establishes a WSI-level analysis system for cytopathology screening according to cervical slide characteristics of sparsely-distributed and tiny-scale lesion cells and gives an effective demonstration of using deep learning to solve the bottleneck problems of current cervical screening methods. The extensive validation experiments demonstrate that our system can be used for effectively grading slides and recommending top-ranked lesion cells and reducing the workload of cytology screening staff. We believe our robust WSI analysis system would act as an effective cytology screening assist and help accelerate the popularization of cervical cancer screening.

## Results

**System architecture.** Our progressive recognition system consists of the LR model, HR model, and WSI classification model, as shown in Fig. 1. The LR model is designed to quickly locate suspicious lesion areas at LR. The HR model is to identify the lesion cells and recommend the top 10 lesion cells at HR. The WSI classification model uses an RNN to integrate the CNN image features of the top 10 lesion cells, and outputs the positive confidence of WSIs. The LR model and HR model are both based on ResNet50<sup>37</sup>. For the LR model, we modify the fully connected layer of original ResNet50 and add a semantic segmentation branch for generating a rough location mask (Supplementary Fig. 1). Thus, the LR model can screen WSIs and locate the suspicious lesion areas. The semantic segmentation branch is constructed with residual blocks of dilated convolutions. The LR model accepts an image tile of  $512 \times 512$  pixels ( $0.486 \mu\text{m}/\text{pixel}$ ) as input and outputs a lesion probability and a location heatmap (Supplementary Fig. 2). Afterwards, for the areas with a



**Fig. 1 The proposed cervical cancer aided screening system.** Our system consists of WSI redundant division, LR model, HR model, and RNN model. The LR model takes a divided image tile of  $512 \times 512$  pixels ( $0.486 \mu\text{m}/\text{pixel}$ ) as input and outputs a lesion probability and a location heatmap to identify and locate the suspicious lesion areas on WSIs. The HR model takes an image tile of  $256 \times 256$  ( $0.243 \mu\text{m}/\text{pixel}$ ) cropped according to the location heatmap as input and outputs a new lesion probability. The RNN model integrates the CNN image features of the recommended top 10 lesion cells and outputs positive probabilities of WSIs. The clinical dataset images in this figure were created by us.

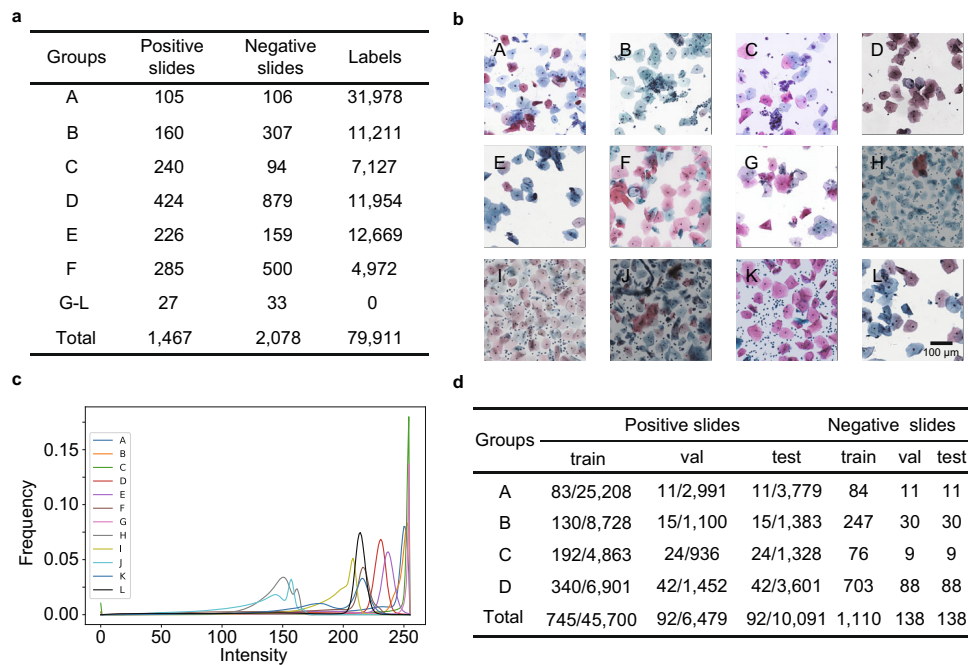
probability higher than 0.5 predicted by the LR model, we perform some morphological operations on the corresponding location heatmap to generate the location mask. A cropped image tile of  $256 \times 256$  ( $0.243 \mu\text{m}/\text{pixel}$ ) according to the location mask is input to the HR model and a new lesion probability is obtained. Finally, all identified lesion cells in WSIs are sorted by lesion probabilities, and the top 10 typical lesion cells are recommended for cytopathologist reviewing. Further, the RNN model integrates the CNN image features of the recommended top 10 lesion cells to classify WSIs. For each lesion cell tile, 2,048-dimensional features are extracted by the HR model. Then the total  $10 \times 2,048$ -dimensional features are input to the RNN model, and positive probabilities of WSIs are output.

**Multi-center WSI datasets.** To assess the robustness and clinical applicability of our system, we collected 12 groups of datasets from five hospitals and five kinds of imaging instruments (see “Dataset sources” in “Methods”), which are referred to as groups A–L (Fig. 2a). These 12 datasets include 1,467 (41.4%) positive WSIs and 2,073 (58.6%) negative WSIs with 79,911 annotated lesion cells by a consensus of three cytopathologists. Each WSI represents one unique patient. The 12 datasets show completely different image styles of staining and imaging characteristics (Fig. 2b) and we quantified the difference in their numerical distributions (Fig. 2c). Groups A–D are used for training our system. Groups E–L are treated as a completely independent test set to evaluate the generalization of our system. Groups A–D are randomly divided into training set, validation set, and test set with a slide-wise ratio of 8:1:1 (Fig. 2d). WSIs of all groups are scanned under  $\times 20$  or  $\times 40$  magnification microscopes. We

uniformly interpolate them to  $0.243 \mu\text{m}/\text{pixel}$  in data preprocessing, since the different resolutions of various imaging instruments.

In order to verify the effect of our recognition system in practical applications, we invited three cytopathologists to evaluate the prediction results of 1,170 slides in groups E and F. Groups E and F are independent from the training data with new styles and thus are suitable for clinical-level experiments. We performed the below assessments: slide level accuracy, tile level accuracy, and true positive rate of recommended top 10 lesion cells. All the skilled cytopathologists for annotation have been trained and certified by the Chinese Society for Colposcopy and Cervical Pathology.

**Assessment at the slide level.** To assess the effectiveness of our system at the slide level, we compared the RNN classifier and cytopathologists in classifying WSIs on the independent groups E and F of 1,170 slides. Figure 3a shows the ROC (receiver operating characteristic) curve of our system for classifying positive and negative slides, achieving 93.5% specificity and 95.1% sensitivity with 0.979 AUC (the area under ROC). The most confusing 121 slides of the slides were classified by the RNN and the three cytopathologists. Each red dot in Fig. 3b refers to 1-specificity and sensitivity of a cytopathologist’s interpretation result. Our system achieves 50.0% specificity and 74.6% sensitivity with 0.647 AUC, which is comparable with the average level of cytopathologists. In addition, our system processes one giga-pixel WSI in about 1.5 min after deploying on a single GPU card, much faster than manual slide reading time.



**Fig. 2 Overview of multi-center WSI datasets.** **a** The collected 12 groups of WSI datasets from five hospitals and five kinds of scanners. There are in total 3,545 WSIs with 79,911 annotated lesion cells by a consensus of three cytopathologists. **b** Cervical image instances of the 12 groups, showing diverse image styles of staining and imaging characteristics. **c** The numerical distributions of the 12 groups in the value channel of HSV space (hue, saturation, value), further confirming the difference of image styles. **d** The division of training set, validation set, and test set with a slide-wise ratio of 8:1:1 on groups A–D, which are used for training our system.

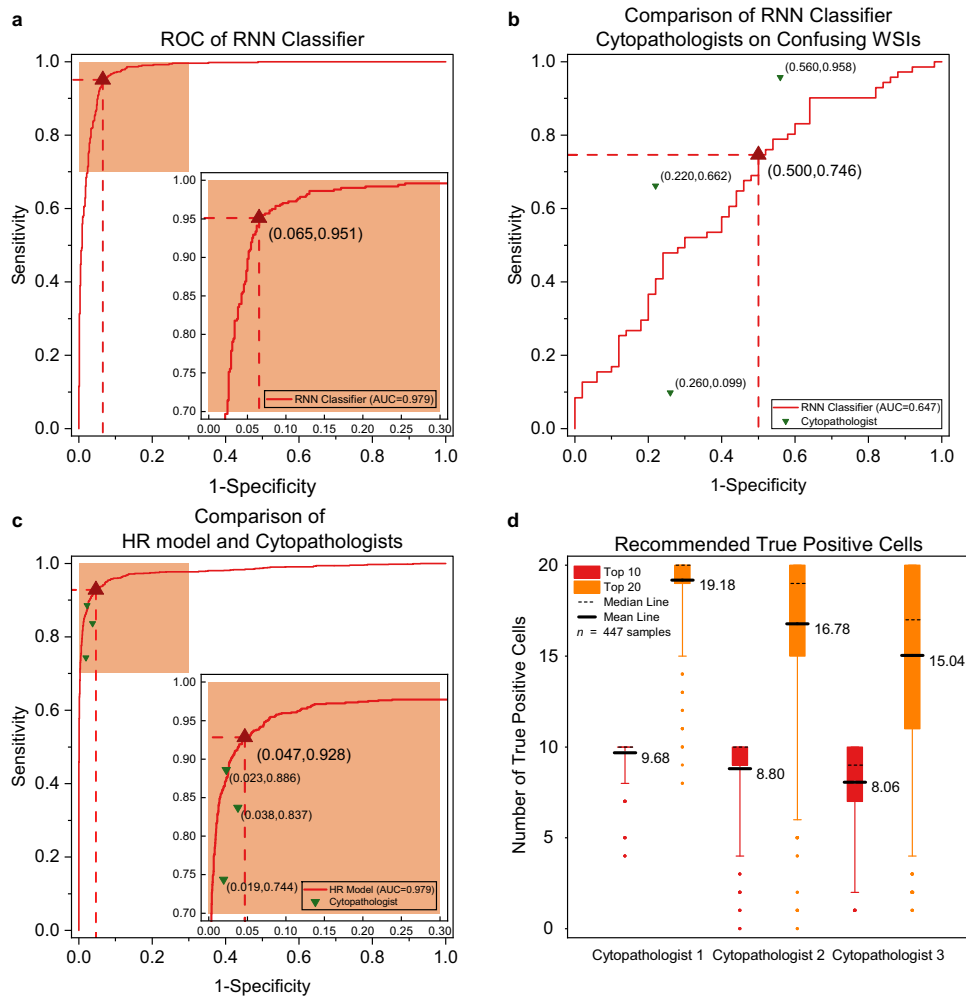
To test the performance of our system in the case of HPV testing first or in combination with cytology, we further analyzed the cervical cell slides from 395 HPV-positive patients (cytology positive 169, negative 226). We achieved 81.9% specificity and 79.3% sensitivity with 0.890 AUC (Supplementary Fig. 3). The main reason for the decrease is that many of these HPV-positive and cytology-negative samples are accompanied by bacterial infections, which may increase the classification difficulty. Our system was designed for the general population of women, and there was no special training for HPV-positive slides. We conducted the trial with the read-made networks without retraining. Although the preliminary results are interesting, the classification effect is not enough. If the HPV-positive samples are trained purposefully in the future, the classification effect is expected to be improved.

**Analysis of false-positive and false-negative slides.** From the frequency histogram of slide classification scores of groups E and F (Fig. 4), our system produces 0.8% false-negative slides (the slide score threshold value is 0.5), all of which were confirmed as ASC-US slides (atypical squamous cells of undetermined significance). As we know that cervical cytology ASC-US slides and part hard negative slides are confusable; thus, it is acceptable to misjudge a small number of ASC-US slides. Meanwhile, our system produces 26.3% false-positive slides, which is in line with the original intention of cervical cytology computer-aided diagnosis. These false positives will be further reviewed by cytopathologists. Further, our system achieves 49.3% specificity while retaining 100% sensitivity, indicating that 49.3% negative slides in the test groups E–F can be excluded. For different test groups, this ratio value may vary since that it depends on the lowest score of positive slides. The results indicate that our WSI analysis system can be applied for prescreening part completely normal slides and reducing the workload of cytotechnologists.

**Assessment at the tile level.** To evaluate the difference between our system and three cytopathologists at the tile level, we randomly selected 1,018 positive tiles and 3,047 negative tiles with a size of  $256 \times 256$  ( $0.243 \mu\text{m}/\text{pixel}$ ) from the test data of groups E and F. As shown in Fig. 3c, the ROC curve describes the performance of our system, and each red dot represents 1-specificity and sensitivity of a cytopathologist's classification result. Our system achieves 95.3% specificity and 92.8% sensitivity with 0.979 AUC better than the average level of cytopathologists.

**Assessment on the recommended top 10 lesion cells.** Three cytopathologists evaluated the recommended top 10 lesion cells on 447 positive slides in groups E and F. The average true positive rates of top 10 recommended cells evaluated separately by three cytopathologists are 96.8%, 88.0%, and 80.6% (Fig. 3d), with an average value of 88.5%, and the average true positive rate of top 20 recommended cells is 85.0%. For some atypical positive slides with a few lesion cells, such as ASC-US, the recommended true positive cell number may be very low as shown in the box plot (Fig. 3d). In this case, our system employs voting the evaluated results of three cytopathologists, and the final result shows we do not miss any positive slide, i.e., positive slides at least have one true lesion cell in the recommended top 10 or top 20 cells. Figure 4 shows the recommended cells of slides with different classification scores. For high-risk slides, our system recommended typical lesion cells such as koilocytotic cells or hyperchromatic cells with large nucleus and irregular nuclear membrane. For medium-risk slides, some suspicious cells with slightly large or deep-stained nucleus were recommended. No typical lesion cells were recommended on low-risk slides. Our system can accurately recommend a few top-ranked lesion cells, allowing cytopathologists to focus on these suspicious areas.

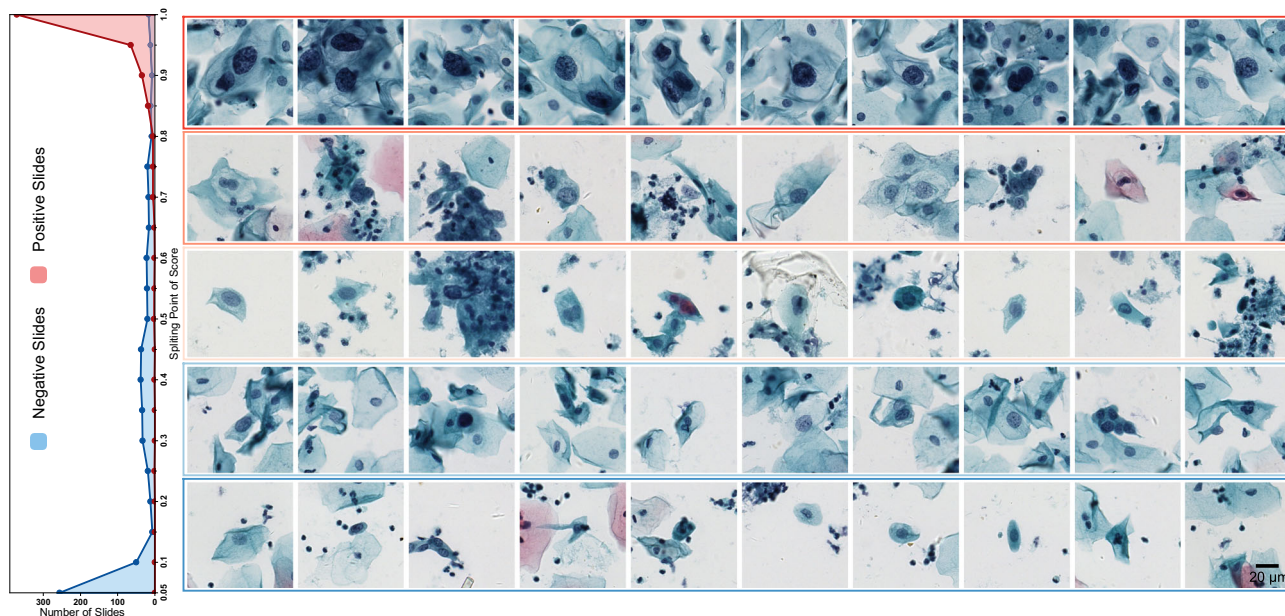
**Comparison of our system and Hologic ThinPrep Imaging System on recommending top lesion cells.** To further verify the



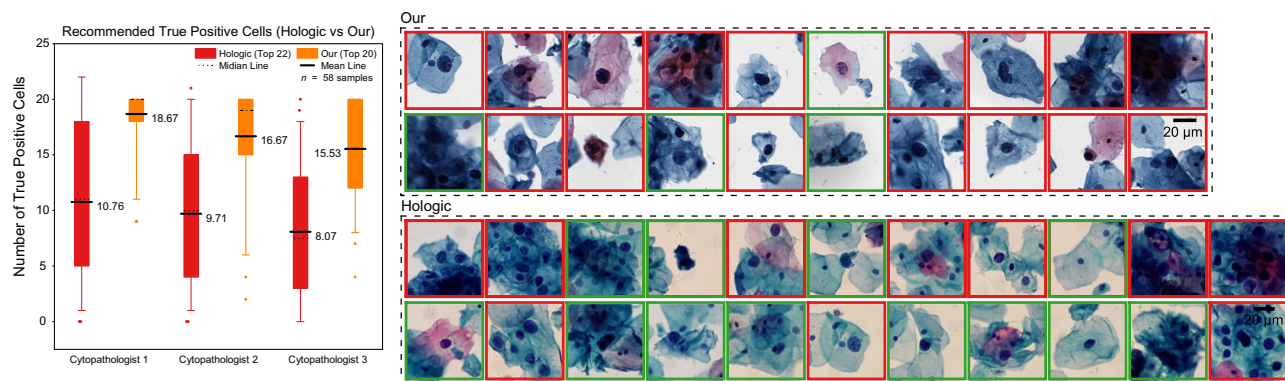
**Fig. 3 Clinical-level experimental results.** **a** The ROC of the RNN model for classifying the slides of groups E and F ( $n = 1,170$ ). **b** Comparison of the RNN model and three cytopathologists at the slide level ( $n = 121$ ). The most confusing 121 slides of the 1,170 slides were classified by the RNN model and the three cytopathologists. Each green triangle refers to 1-specificity and sensitivity of a cytopathologist’s result. **c** Comparison of the HR model and the three cytopathologists at the tile level. Randomly selected 1018 positive test tiles and 3,047 negative test tiles with a size of  $256 \times 256$  ( $0.243 \mu\text{m}/\text{pixel}$ ) from groups E and F ( $n = 4,065$  tiles) were simultaneously classified by the HR model and three cytopathologists. **d** The average true positive numbers of the recommended top 10 and top 20 lesion cells on positive slides in groups E and F ( $n = 447$ ), evaluated by three cytopathologists separately. The boxes indicate the upper and lower quartile values, the whiskers indicate the 95% and 5% quantile values, the middle bold solid line and dotted lines indicate the mean and median values, and the scatter dots indicates outliers.

effectiveness of our aided screening system, we compared it with the Hologic ThinPrep Imaging System (referred to as TIS). The test data are cervical cytology samples from 58 positive patients in Maternal and Child Hospital of Hubei Province equipped with TIS. First, 58 glass slides were prepared from the 58 samples, stained, imaged, and identified by TIS in the hospital. Twenty-two suspicious fields of view were recommended by TIS on each slide. Then, we used another instrument (Shenzhen Shengqiang Technology Ltd with  $0.180 \mu\text{m}/\text{pixel}$  under  $\times 40$  magnification) to scan the 58 glass slides, and we used our system to recommend the top 20 suspicious cell regions (about  $60 \times 60 \mu\text{m}^2$ , far smaller than TIS’ fields of view) for each slide. We asked three cytopathologists to evaluate the results recommended by TIS and our system at the same time. The statistical results in Fig. 5 show that the true positive rate of our system is higher than that of TIS. Notably, TIS can only work under the closed-loop strategy of preparation, staining, imaging, and recognition, while our system is robust to staining and imaging of various sources.

**Importance of designed data enhancement, hard sample mining, and diverse data learning.** We conducted a set of ablation experiments to demonstrate the importance of designed data enhancement in HSV color space, hard sample mining, and diverse data learning with group and category balancing (see “Methods”). We used the three learning strategies to train a series of control high-resolution models step by step, and gave the classification accuracies on the test sets of groups A–F. Notably, the ratio of positive and negative tiles in the test set is 1:1. The ablation experimental model configs and results are provided in Fig. 6a. To evaluate model generalization, we treated groups E–F as the independent test data and showed the ROC curves of these control models on groups E–F in Fig. 6b. According to the results, with the designed data enhancement and hard sample mining, performance of the enhanced and mined models on groups E–F made great progress with the AUC value increase of 0.138 and 0.072. The results indicate that our designed data enhancement and hard sample mining strategies are effective for improving model generalization and accuracy. Further, as more groups of



**Fig. 4** The recommended top 10 lesion cells of WSIs with different classification scores. The left subgraph is the frequency histogram of slide scores from 0 to 1 with an interval of 0.05 in groups E and F ( $n = 1,170$ ). The right subgraph is the recommended top 10 lesion cells of slides with different scores.



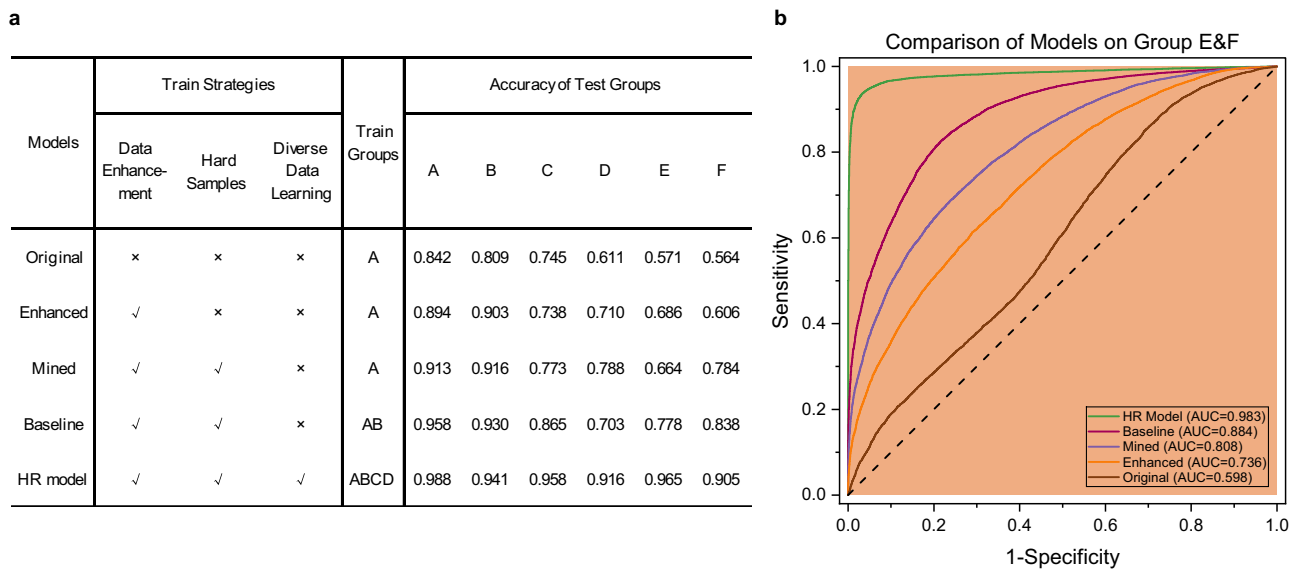
**Fig. 5** Comparison of our system and Hologic TIS on recommending top lesion cells. The histogram on the left shows the true positive cell number of our system's top 20 cell regions (about  $60 \times 60 \mu\text{m}^2$ ) and TIS' top 22 fields of view (far greater than  $60 \times 60 \mu\text{m}^2$ ) on 58 positive slides ( $n = 58$ ). The boxes indicate the upper and lower quartile values, the whiskers indicate the 95% and 5% quantile values, the middle bold solid line and dotted lines indicate the mean and median values, and the scatter dots indicate outliers. The subgraph on the right is recommended cells of one positive slide. Notably, the most suspicious cells were cropped from TIS' top 22 fields of view according to cytopathologists' evaluation results. The cells with the red border line were evaluated as true positive cells by cytopathologists.

training datasets were used, the AUC values of the mined, baseline, and HR models increased gradually from 0.808 to 0.983. The results indicate that the diverse data learning of multiple groups with different styles is important for model generalization.

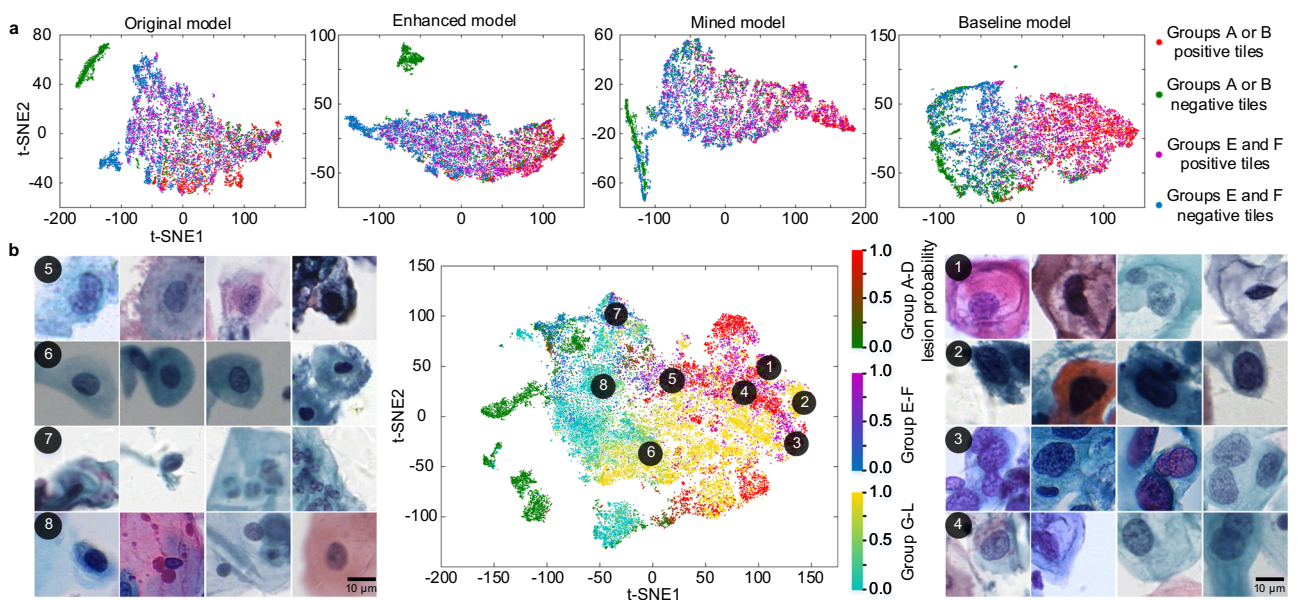
#### Generalized and rich feature representations of our models.

We analyzed the alignment of the features of high-resolution models between different groups of data by feature visualization. The dimension-reduced features of the original, enhanced, mined, and baseline models by t-SNE<sup>38</sup> on groups A, B, E, and F are shown in Fig. 7a. For the models, groups E and F are independent test data. From the original model to the baseline model, features of positive and negative tiles are gradually separated. Further, the features are gradually aligned between groups A-B and groups E-F. The results indicate that the designed data enhancement, hard sample mining, and diverse data learning strategies improve the discrimination and alignment of features on unseen groups E-F.

We further analyzed the feature representations of the HR model on groups A–L in Fig. 7b. The tiles with high and low lesion probabilities from the 12 groups are clustered and well separated. Tiles corresponding to the far-right points are the typical lesion cells, including koilocytotic cells and hyperchromatic cells with a large nucleus and an irregular nuclear membrane. These lesion cells with different staining and imaging characteristics are clustered together and share similar features. Normal cells from different groups are clustered on the left points. At the junction regions are the suspicious cells with about 0.5 lesion probabilities. The suspicious cells generally contain a slightly large nucleus or deep-stained nucleus, but the degree is not enough. In addition, artifacts from staining and imaging may cause the suspicious cells. The results indicate that the learned features represent cervical lesion cell morphology well and the features are aligned between datasets with different staining and imaging characteristics. This is the key reason why our system has good generalization for unseen datasets of new styles.



**Fig. 6 Importance of designed data enhancement, hard sample mining, and diverse data learning.** **a** The ablation experimental results about the designed three learning strategies of data enhancement, hard sample mining, and diverse data learning on the HR model. We used these learning strategies to train a series of control models step by step, and gave the classification accuracies on the test sets of groups A-F. **b** AUC-ROC comparison of the original, enhanced, mined, baseline, and HR models on groups E and F.



**Fig. 7 Generalized and rich feature representations of our models.** **a** Gradually aligned features between different groups of datasets. The last 2,048-dimensional features of the original, enhanced, and mined models were reduced to two dimensions using t-SNE on the test sets of groups A, E, and F and were plotted in the first three subgraphs respectively. The dimension-reduced features of the baseline model on the test sets of groups A, B, E, and F were plotted in the fourth subgraphs. In all, 1,000 positive test tiles and 1,000 negative test tiles in each group were randomly selected for the visualization. **b** The distribution of dimension-reduced features of the HR model on the 12 groups test sets and the interpretation of the feature representations by examining the corresponding cervical image tiles. In all, 2,000 test tiles in each group were randomly selected for the feature visualization. The different color bars refer to different groups and the gradient colors within a single color bar refer to the lesion probabilities of tiles.

**Discussion**

In this paper, we propose a clinical-level-aided diagnosis system for cervical cancer screening based on deep learning and massive WSIs. Compared with the existing methods, our system has the following key advantages: (a) WSI-level analysis system rather than tile-level evaluation; (b) the integrated strategy of low- and high-resolution combination, data augmentation, diverse data learning, and hard sample mining for achieving high accuracy, good generalization, and fast speed of our system; (c) the

human-computer comparison verification at both tile and WSI levels to prove the effectiveness of our system; (d) the practical deployment with C++, processing a giga-pixel WSI in as short as about 1.5 min with one GPU.

The mean reported positive rates of cervical cytology screening are less than 10%<sup>39,40</sup>, and that of the physical examination population is even lower. Therefore, if some negative samples can be distinguished, it should be a great aid to cytotechnology. In our study, the distribution of slide positive scores (0.8% false-negative slides) indicates

that our system has potential to exclude a considerable number of negative slides and reduce a lot of cytotechnologists' workloads. The 10/20 most suspicious areas on each slide with relatively high 88.5%/85.0% TPRs are given, which would help free doctors from the task of traversing and searching for suspicious targets and concentrate on the task of identifying recommended suspicious cells. For underdeveloped areas lacking cytopathology stuff, our system has important clinical and social significance for accelerating the popularization of cervical cancer screening.

In recent years, WSI analysis has been widely studied in various histopathology subspecialties<sup>41–45</sup>. These algorithms generally follow the below principle: first extract classification features or confidences of local tiles, then aggregate the local information to construct WSI-level feature descriptors, and finally classify the slides. The lesions of histopathology slides are region-level and have overall background information. Cytopathology slides show sparsely-distributed and tiny-scale lesion cells, and the cells are independent and short of overall information. These characteristics of cytopathology have brought challenges to accurate WSI analysis when directly transferring the histopathology methods. In this work, we propose a WSI-level analysis system for cytopathology screening according to cervical slide characteristics, and demonstrate its effectiveness in classifying cytology slides by the extensive validation experiments.

The diversities of slide staining and imaging in different hospitals greatly limit the utility of current automated cervical cell recognition algorithms; thus, model generalization is a key factor in the practicality. Our multi-center independent test datasets include differences in slide preparation (liquid-based preparation methods: membrane-based and sedimentation), dyeing schemes (fixing, clearing, and dehydrating), imaging magnification ( $\times 20$  and  $\times 40$ ), imaging resolution (0.180–0.293  $\mu\text{m}/\text{pixel}$ ), and imaging color characteristics (Fig. 2b, c). The good results on these diverse data prove the generalization of our system and lay a foundation for the practicability in diverse data scenarios.

We consider that the basic task of the aided screening system is to reduce the workload of cytology staff by excluding low-risk slides and recommending a limited number of suspicious cells on high-risk slides for cytopathologist reviewing. The final precise diagnosis is up to cytopathology doctors, and this human–machine combination mode can reduce possible errors of artificial intelligence and ensure the accuracy of diagnosis<sup>19,20</sup>. Therefore, unlike the works of Lin et al.<sup>30</sup> and Zhu et al.<sup>46</sup>, our system focuses on distinguishing positive and negative classes instead of fine subclasses at both cell and slide levels, such as ASC-US, ASC-H (atypical squamous cells cannot exclude high-grade squamous intraepithelial lesion (HSIL)), LSIL (low-grade squamous intraepithelial lesion), HSIL, and SCC (squamous cell carcinoma)<sup>6</sup>. Moreover, the definition of the subclasses is based on cell morphology and the boundaries are often fuzzy especially at the cell level, which will produce a lot of noises in the actual manual annotations and inconsistency between different cytopathologists.

In the future, we will focus on research about AI-enhanced portable microscopy and augmented reality microscopy to further expand our system. At present, professional but expensive scanners are still required, preventing the spread of cervical cancer screening in remote and underdeveloped areas. Thus, developing portable microscope-based cervical cancer computer-aided diagnosis is necessary. In addition, developing real-time augmented reality microscopes can provide friendly human–computer interaction for AI-assisted slide screening without changing the conventional working mode of cytopathologists.

## Methods

**Dataset sources.** All 12 groups of glass slides are provided by Maternal and Child Hospital of Hubei Province (referred as H1), Tongji Hospital of Huazhong

**Table 1 Dataset sources of the 12 groups of datasets.**

Groups	Hospital	Scanner
A	H1	S1—version 1
B	H1	S1—version 2
C	H1	S3
D	H2—version 1	S2
E	H1	S2
F	H2—version 2	S2
G	H2—version 1	S1—version 2
H	H4	S4
I	H2—version 3	S4
J	H3	S4
K	H5	S5
L	H1	S2

University of Science and Technology (referred as H2), Wuhan Union Hospital of Huazhong University of Science and Technology (referred as H3), Hubei Cancer Hospital (referred as H4), and KingMed Diagnostics Ltd (referred as H5). The slide acquisition is performed in accordance with the guidelines of the Medical Ethics Committee of Tongji Medical College at Huazhong University of Science and Technology. These glass slides are scanned into WSIs by the instruments from 3DHisTech Ltd with 0.243  $\mu\text{m}/\text{pixel}$  under  $\times 20$  magnification (referred as S1), Shenzhen Shengqiang Technology Ltd with 0.180  $\mu\text{m}/\text{pixel}$  under  $\times 40$  magnification (referred as S2), Wuhan National Laboratory for Optoelectronics-Huazhong University of Science and Technology with 0.293  $\mu\text{m}/\text{pixel}$  under  $\times 20$  magnification (referred as S3), Huaiguang Intelligent Technology Ltd with 0.238  $\mu\text{m}/\text{pixel}$  under  $\times 20$  magnification (referred as S4), and Konfoong Biotech Information Ltd with 0.238  $\mu\text{m}/\text{pixel}$  under  $\times 40$  magnification (referred as S5). The details are shown in Table 1. Notably, version 1 and version 2 of the scanner S1 are different generations of instruments, and they have different imaging color characteristics. Version 1, version 2, and version 3 of the hospital H2 are different in slide preparation and staining scheme.

**Data annotating and screening.** Based on the TBS criteria<sup>6</sup>, the cervical slides are annotated by six cytopathologists using Qupath<sup>47</sup> (v0.2.0) and a home-made semi-automatic online annotation software. Considering that it is a standard operating procedure in cytopathology when the diagnoses of the two pathologists are inconsistent, ask a senior doctor to make an interpretation<sup>48,49</sup>. In this work, the final annotations were produced by a consensus of three cytopathologists. We abandoned questionable annotations. Based on the TBS criteria<sup>6</sup>, we classify all kinds of squamous epithelial cell abnormalities (including atypical squamous cells (ASC), squamous intraepithelial lesion (SIL), and SCC), and glandular epithelial cell abnormalities (including atypical endocervical/glandular cell, endocervical adenocarcinoma in situ (AIS), and adenocarcinoma) as positive labels, and classify various normal cellular elements, nonneoplastic findings such as nonneoplastic cellular variations, reactive cellular changes, and glandular cells' status post hysterectomy as negative label NILM (negative for intraepithelial lesion or malignancy). During the iterative learning of our system, we collected massive false negatives of various morphologies as negative annotations.

**Designed data enhancement.** Data enhancement is a common technique to expand the diversity of data distributions and improve the generalization ability of models. Based on the characteristics of cervical cell images, we designed a series of specific data enhancement in HSV and RGB color spaces to imitate the color distribution of various image styles derived from diverse staining and imaging. The data enhancement includes transformations on hue, saturation, brightness, contrast, flipping, and shifting, as well as adding noise such as blurring and sharpening. We determined the enhancement parameters to ensure that the recognition of cervical cells will not be affected.

**Diverse data learning.** Diverse data learning on different groups of data was employed to improve the robustness of feature representations. We first trained the LR model and HR model on groups A and B with a large number of annotations to obtain baseline models. The pre-training weights on ImageNet<sup>50</sup> were used as the initial weights. Then we incorporated groups C and D of different styles as extra training data. Based on the baseline models, we learned more robust models on mixed data of groups A–D. The key point of our diverse data learning is the group- and category-balancing strategy. Since there are multiple groups and each group has unbalanced annotations of several subtypes, we resample the sample number of all groups and categories to make them as balanced as possible.

**Hard sample mining.** Hard sample mining was adopted to improve the accuracy of our models. For training samples that fail to be successfully classified, we conducted a second round of learning on the hard samples. The proportion of hard



samples in the entire training samples affected the tendency of models. The LR model of our recognition system was designed to find all possible lesion cells quickly and coarsely, while the HR model aimed to distinguish the true lesion cells from those candidates found by the LR model. Therefore, we used a small proportion of hard samples to train the LR model to ensure a high recall rate and a large proportion to train the HR model to ensure a high precision.

**Training details.** We employed the above three training strategies to enhance the generalization ability and recognition accuracy of the LR and HR models. According to the datasets of groups A–D in Fig. 1d, we used the train set to optimize the LR and HR models, the validation set to adjust the hyper-parameters, and then tested the performance on the test set. Groups E and F were used as independent datasets for evaluating model generalization. The positive samples were cropped around the annotations of positive slides and the negative samples were randomly cropped from negative slides. The LR and HR models used Adam<sup>51</sup> as the optimizer with an initial learning rate of  $10^{-3}$ . The strategy of learning rate decay was adopted during training.

We used a simple RNN with one hidden layer of 512 units for classifying WSIs. The RNN model was trained and validated on groups A–D and evaluated on independent groups E and F. We trained three kinds of RNNs with different inputs: the HR model features of the recommended top 10, top 20, and top 30 lesion cells. For each kind, we trained two RNNs, and then integrated the total six RNNs as our WSI classifier. In the training process, we used data augmentation to improve the varieties of the input features of the RNNs, including enhancement and rearrangement of the top  $k$  lesion cell images. These strategies of model integration and data augmentation improved the robustness of the RNNs, since the limited number of slides ( $<10^4$ ). Similarly, Adam<sup>51</sup> with an initial learning rate of  $10^{-3}$  and the learning rate decay strategy were adopted during training the RNNs.

We deployed our system with C++ and utilized multi-threading to accelerate image processing. Our system processes one WSI under  $\times 20$  magnification in about 3.0 min using one Nvidia 1080Ti GPU. Further, we used TensorRT<sup>36</sup> to accelerate network model forward inference, and achieved a speed of about 1.5 min per slide. The processing speed in practice would be influenced by pixel number of slides, disk reading speed, and WSI format.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The LR model and HR model of our progressive lesion cell recognition method were initialized with the pre-training weights of ImageNet (<https://www.image-net.org>). We further provided the source codes, the C++ software and some test WSIs to facilitate the reproducibility. The original WSI and annotation data are private and are not publicly available since the protection of patients' privacy in cooperative hospitals. All data supporting the findings of this study are available on requests for non-commercial and academic purposes from the primary corresponding author (xliliu@mail.hust.edu.cn) within 10 working days. We do not require to sign a data use agreement. Source data are provided with this paper.

## Code availability

The source codes of this paper are available at <https://github.com/ShenghuaCheng/Aided-Diagnosis-System-for-Cervical-Cancer-Screening>. We also provide a C++ software and a user manual of our system with some test slides at Baidu Cloud ([https://pan.baidu.com/s/1UmQzASwvlpKLO7hbwaDc\\_A](https://pan.baidu.com/s/1UmQzASwvlpKLO7hbwaDc_A), extracting code is *cyto*) or at Google Drive (<https://drive.google.com/drive/folders/19rE9atLryLaBR8shqAlc45f8tn6z07ky?usp=sharing>, no extracting code).

Received: 30 March 2021; Accepted: 3 August 2021;

Published online: 24 September 2021

## References

- Ferlay, J. et al. *Global Cancer Observatory: Cancer Today*. International Agency for Research on Cancer. <https://gco.iarc.fr/today> (2020).
- Peto, J., Gilham, C., Fletcher, O. & Matthews, F. E. The cervical cancer epidemic that screening has prevented in the UK. *Lancet* **364**, 249–256 (2004).
- Sasieni, P., Adams, J. & Cuzick, J. Benefit of cervical screening at different ages: evidence from the UK audit of screening histories. *Br. J. Cancer* **89**, 88–93 (2003).
- Levi, F., Lucchini, F., Negri, E., Franceschi, S. & Vecchia, C. L. Cervical cancer mortality in young women in Europe: patterns and trends. *Eur. J. Cancer* **36**, 2266–2271 (2000).
- Parkin, D. M., Nguyen-Dinh, X. & Day, N. E. The impact of screening on the incidence of cervical cancer in England and Wales. *Br. J. Obstet. Gynaecol.* **92**, 150–157 (1985).
- Nayar, R. & Wilbur, D. C. *The Bethesda System for Reporting Cervical Cytology: Definitions, Criteria, and Explanatory Notes* (Springer, 2015).
- Nanda, K. et al. Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. *Ann. Intern. Med.* **132**, 810–819 (2000).
- Fahey, M. T., Irwig, L. & Macaskill, P. Meta-analysis of pap test accuracy. *Am. J. Epidemiol.* **141**, 680–689 (1995).
- Wright, A. M. et al. Digital slide imaging in cervicovaginal cytology: a pilot study. *Arch. Pathol. Lab. Med.* **137**, 618–624 (2013).
- Litjens, G. et al. A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017).
- Conceição, T., Braga, C., Rosado, L. & Vasconcelos, M. J. M. A review of computational methods for cervical cells segmentation and abnormality classification. *Int. J. Mol. Sci.* **20**, 5114 (2019).
- Cahn, R. L., Poulsen, R. S. & Toussaint, G. Segmentation of cervical cell images. *J. Histochem. Cytochem.* **25**, 681–688 (1977).
- Borst, H., Abmayr, W. & Gais, P. A thresholding method for automatic cell image segmentation. *J. Histochem. Cytochem.* **27**, 180–187 (1979).
- Chang, C. W. et al. Automatic segmentation of abnormal cell nuclei from microscopic image analysis for cervical cancer screening. In *Proceedings of the IEEE 3rd International Conference on Nano/Molecular Medicine and Engineering* 77–80 (IEEE, 2009).
- Kim, K. B., Song, D. H. & Woo, Y. W. Nucleus segmentation and recognition of uterine cervical pap-smears. In *International Workshop on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, Lecture Notes in Computer Science 153–160 (Springer, 2007).
- Chen, Y. et al. Semi-automatic segmentation and classification of pap smear cells. *IEEE J. Biomed. Health Inf.* **18**, 94–108 (2014).
- Mariarpatham, E. J. et al. Nominated texture based cervical cancer classification. *Comput. Math. Methods Med.* **2015**, 1–10 (2015).
- Huang, P. C. et al. Quantitative assessment of pap smear cells by PC-based cytopathologic image analysis system and support vector machine. In *International Conference on Medical Biometrics*, Lecture Notes in Computer Science 192–199 (Springer, 2007).
- Renshaw, A. & Elsheikh, T. M. A validation study of the Focalpoint GS imaging system for gynecologic cytology screening. *Cancer Cytopathol.* **121**, 737–738 (2013).
- Quddus, M., Neves, T., Reilly, M., Steinhoff, M. & Sung, C. Does the ThinPrep Imaging System increase the detection of high-risk HPV-positive ASC-US and AGUS? The Women and Infants Hospital experience with over 200,000 cervical cytology cases. *CytoJournal* **6**, 15 (2009).
- Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- Zhang, L. et al. Combining fully convolutional networks and graph-based approach for automated segmentation of cervical cell nuclei. In *Proceedings of the IEEE 14th International Symposium on Biomedical Imaging* 406–409 (IEEE 2017).
- Song, Y. et al. A deep learning based framework for accurate segmentation of cervical cytoplasm and nuclei. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 2903–2906 (IEEE, 2014).
- Chen, H. et al. CytoBrain: cervical cancer screening system based on deep learning technology. *J. Comput. Sci. Technol.* **36**, 347–360 (2021).
- Liang, Y. et al. Comparison detector for cervical cell/clumps detection in the limited data scenario. *Neurocomputing* **437**, 195–205 (2021).
- Gupta, M. et al. Region of interest identification for cervical cancer images. In *Proceedings of the IEEE 17th International Symposium on Biomedical Imaging (ISBI)* 1293–1296 (IEEE, 2020).
- Nirmal-Jith, O. U. et al. DeepCerv: Deep Neural Network for Segmentation Free Robust Cervical Cell Classification. In *First International Workshop on Computational Pathology*, Lecture Notes in Computer Science 86–94 (Springer, 2018).
- Forslid, G. et al. Deep convolutional neural networks for detecting cellular changes due to malignancy. In *Proceedings of the IEEE International Conference on Computer Vision Workshops* 82–89 (IEEE, 2017).
- Lin, H., Hu, Y., Chen, S., Yao, J. & Zhang, L. Fine-grained classification of cervical cells using morphological and appearance based convolutional neural networks. *IEEE Access.* **7**, 71541–71549 (2019).
- Lin, H. et al. Dual-path network with synergistic grouping loss and evidence driven risk stratification for whole slide cervical image analysis. *Med. Image Anal.* **69**, 101955 (2021).
- Holmström, E. et al. Point-of-care digital cytology with artificial intelligence for cervical cancer screening in a resource-limited setting. *JAMA Netw. Open* **4**, e211740 (2021).
- Jantzen, J., Norup, J., Dounias, G. & Beth, B. Pap-smear benchmark data for pattern classification. In *Nat. Inspir. Smart Inf. Syst.* 1–9 (NiSIS, 2005).
- Lu, Z., Carneiro, G. & Bradley, A. P. An improved joint optimization of multiple level set functions for the segmentation of overlapping cervical cells. *IEEE Trans. Image Process.* **24**, 1261–1272 (2015).
- Lu, Z. et al. Evaluation of three algorithms for the segmentation of overlapping cervical cells. *IEEE J. Biomed. Health Inf.* **21**, 441–450 (2017).

35. Phoulady, H. A. & Mouton, P. R. A new cervical cytology dataset for nucleus detection and image classification (Cervix93) and methods for cervical nucleus detection. Preprint at <https://arxiv.org/abs/1811.09651> (2018).
36. NVIDIA TensorRT Release Notes. NVIDIA Corporation. Available at <https://docs.nvidia.com/deeplearning/tensorrt/release-notes> (2021).
37. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (IEEE, 2016).
38. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
39. Davey, D. D. et al. Bethesda 2014 implementation and human papillomavirus primary screening: practices of laboratories participating in the College of American Pathologists PAP Education Program. *Arch. Pathol. Lab. Med.* **143**, 1196–1202 (2019).
40. Ma, L. et al. Characteristics of women infected with human papillomavirus in a tertiary hospital in Beijing China, 2014–2018. *BMC Infect. Dis.* **19**, 670 (2019).
41. Van der Laak, J., Litjens, G. & Ciompi, F. Deep learning in histopathology: the path to the clinic. *Nat. Med.* **27**, 775–784 (2021).
42. Campanella, G. et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **25**, 1301–1309 (2019).
43. Srinidhi, C. L., Ciga, O. & Martel, A. L. Deep neural network models for computational histopathology: a survey. *Med. Image Anal.* **67**, 101813 (2020).
44. Diao, J. A. et al. Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. *Nat. Commun.* **12**, 1–15 (2021).
45. Bulten, W. et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol.* **21**, 233–241 (2020).
46. Zhu, X. et al. Hybrid AI-assistive diagnostic model permits rapid TBS classification of cervical liquid-based thin-layer cell smears. *Nat. Commun.* **12**, 3541 (2021).
47. Bankhead, P. et al. QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).
48. Cibas, E. S. et al. Quality assurance in gynecologic cytology. The value of cytotechnologist-cytopathologist discrepancy logs. *Am. J. Clin. Pathol.* **115**, 512–516 (2001).
49. Nakhleh, R. E. et al. *Quality Improvement Manual in Anatomic Pathology* (College of American Pathologists, 2002).
50. Deng, J. et al. ImageNet: a large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 248–255 (IEEE, 2009).
51. Kingma, D. P. & Ba, J. Adam. A method for stochastic optimization. In *the 3rd International Conference on Learning Representations* (ICLR, 2015).

## Acknowledgements

We thank the members of Britton Chance Center for Biomedical Photonics for help in experiments and comments on the paper. We thank Jiangsheng Yu of Binsheng Technology (Wuhan) Co., Ltd for suggestions on the experimental design and data processing. We thank Zhongkang Medical Examination Center and Konfoong Biotech

Information Ltd for data support. This work is supported by the NSFC projects (grant 61721092) and the director fund of the WNLO.

## Author contributions

X.L., J.H., L.C. and S.Z. conceived of the project. S.C., X.L., S.L. and J.Y. designed the aided diagnosis system. S.C., S.L., and J.Y. developed the algorithms. X.L., J.H., L.C., G.R., Y.X. and W.Z. annotated the lesion cells, evaluated the experimental results, and completed the man–machine comparison experiments. J.H., L.C., Z.W., J.C., X.F. and F.Y. provided the cervix cytopathology glass slides. X.L., G.R., Y.X., X.L. and N.L. scanned the slides into WSIs. W.H. performed the deployment and performance optimization of the system. S.C., S.L., J.Y., X.G., J.M., X.L., Z.W. and X.Z. performed the image analysis and processing. S.C., X.L., S.L. and J.Y. wrote the manuscript. T.Q. participated in the revision of the paper. All the authors revised the paper.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-25296-x>.

**Correspondence** and requests for materials should be addressed to Li Chen, Junbo Hu or Xiuli Liu.

**Peer review information** *Nature Communications* thanks Youyi Song, Masayuki Tsuneki and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021