Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Computer-Aided Decision-Making System for Endometrial Atypical Hyperplasia based on Multimodal & Multi-Instance Deep Convolution Networks

Xin Liao (xinty927@163.com)
Sichuan University West China Second University Hospital https://orcid.org/0000-0003-1718-0389
Qingli Li
Sichuan University West China Second University Hospital
Xin Zheng
University of Electronic Science and Technology of China
Jin He
Chengdu SinoRail Electronics CO.LTD

Research Article

Keywords: computer-aided decision-making system, pattern recognition, endometrial hyperplasia, endometrioid carcinoma, cognitive intelligence, decision supporting, deep convolution network, prognostic analysi

Posted Date: September 7th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-865957/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Computer-aided Decision-making System for Endometrial Atypical Hyperplasia based on Multimodal & Multi-instance Deep Convolution Networks

Xin Liao[1, 2], Jin He[3], Qingli Li[2, 4*], Xin Zheng[5*]

(1. Department of Pathology, West China Second University Hospital of Sichuan University, Chengdu 610041, PR.China;

2. Key Laboratory of Birth Defects and Related Diseases of Women and Children (Sichuan University) Ministry of Education, Chengdu 610041, PR.China;

3.Chengdu SinoRail Electronics CO. LTD., Chengdu 610299, PR.China;

4. Department of Obstetrics and Gynecology,West China Second University Hospital of Sichuan University, Chengdu 610041, PR.China;

5. College of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, PR.China)

1st corresponding author: Qingli Li Email: liqingli73@163.com

2nd corresponding author:.Xin Zheng Email: xin_zheng0311@uestc.edu.cn

ABSTRACT

The pathological diagnosis is the gold standard for neoplasms and their precursors, which is highly relevant to the treatment planning and the prognostic analysis. Currently, deep learning networks have been used for the pathological computer-assisted diagnosis and treatment decision-makings. However, due to extremely large size of the whole slide images (WSIs) of pathological slides, the prevailing deep learning models are un-applicable directly in the WSIs analysis. Moreover, the precise exclusion of the blank regions and interfere regions, as well as the manual annotation of various lesioned and normal regions in super large WSIs are infeasible and unavailable in clinical practice. To address aforementioned problems, we develop an computer-aided decision-making system based on multimodal and multi-instance deep convolution networks (CNN) to assist in the diagnosis and treatment of endometrial atypical hyperplasia (AH)/ endometrial intraepithelial hyperplasia (EIH). Firstly, we set up the frame-work

of computer-aided decision-making system based on the WSIs image patterns of AH/EIH, and then transfer the large-scale WSI analysis to the small-scale analysis of multiple suspected lesion regions which can be accomplished the major computer vision models, and eventually the results of prognostic analysis for multiple small-scale suspected lesion regions are summarized to obtain the prognostic results of WSIs by the decision supporting algorithm based on the cognition intelligence. We validate the method via experimental analysis of 102 endometrial atypical hyperplasia patients at the West China Second University Hospital of Sichuan University. The performance achieved for endometrial AH/EIH prognostic analysis includes accuracy (85.3%), precision (84.6%), recall (86.3%). Meanwhile, the method has superior performance to prognostic judgment of a single pathologist as well as approximates to analysis results determined by three pathologists according to the majority voting method.

KEYWORDS

computer-aided decision-making system; pattern recognition; endometrial hyperplasia; endometrioid carcinoma; cognitive intelligence; decision supporting; deep convolution network; prognostic analysis.

DATA AVAILABILITY STATEMENT

All data supporting the findings of this study are obtained from West China Second University Hospital (Sichuan University) and University of Electronic Science and Technology of China. The data is available from authors upon the reasonable request to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of West China Second University Hospital (IRB No. 2020112).

CONSENT TO PARTICIPATE

The patients/participants provided their written informed consent to participate in this study. The consent is available from authors upon the reasonable request to the corresponding authors.

AUTHOR CONTRIBUTIONS

XL and QL designed the study; XL, JH and XZ analyzed and interpreted the data; XL and QL collected the clinical materials and information; XL and XZ conducted the experiments, XL,JH and XZ provide the technical and material support. All authors contributed to the writing of the manuscript and final approval.

FUNDING

This work was supported in part by Key Laboratory Open Fund of Sichuan(2017LF3008).

CONFLICTS OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CONSENT OF PUBLICATION

The copyright to the English-language article is transferred to Springer effective if and when the article is accepted for publication.

1. Introduction

Endometrioid adenocarcinoma is a gynecological malignancy, characterized by high incidence worldwide(Carlson and Mccluggage 2019). EC's precancerous lesion is the atypical hyperplasia of endometrium (AH)/endometrial intraepithelial hyperplasia (EIH) (Downing et al. 2020). The pathlogical diagnosis is the gold standard for AH/EIH, which is highly relevant to the treatment planning and prognostic analysis. However, precise differential diagnosis of AH/EIH and EC remains problematic, even for experienced pathologists, due to the common coexistence of EIH/AH and EC in the pathological slides as well as the low coincidence of AH/EIH diagnosis in biopsies.

In recent years, computer vision technology has achieved great success in the medical field of computer-assisted diagnosis and decision-making, due to the development of the deep convolution network (CNN) which can automatically learn from image data to obtain excellent feature expressions (Guo et al. 2020). Meanwhile,along with the progress in the micrography and whole-slide scanning technology, the pathological slides can be retained in the form of digital images nowadays, enabling wide applications of the computer vision technology in the field of computer-assisted pathological diagnosis. However, even if deep learning technology has shown great promise in the field image analysis, it faces a series of unique challenges when applied in the field of pathological image analysis (Pinckaers et al. 2020; Rijthoven et al. 2020; Xing et al. 2021). First of all, the size of a histopathological whole slide image (WSI) it too large to be input directly into a convolutional neural network(Ciresan et al. 2013; Shahzad et al. 2020). Secondly, the precise annotations of the lesion and normal regions are infeasible and unavailable in clinical practice, because it takes both time and effort to make manual labeling on a large number of digital image patches, especially when the lesion and normal regions are consistently mixed-up in the images(Am Mendola et al. 2020; Hashimoto et al. 2020). Furthermore, there are some unique challenges when the technology is applied in a specific field. For the analysis of WSI patterns from endometrial biopsy, there are two main problems: (1) Uterine curettage causes breakage of the endometrial tissue, resulting in a large number of blank areas without the endometrium. Therefore, it is necessary to exclude these blank areas prior to analysis; (2) Existence of blood and mucus in the endometrial biopsy tissue is inevitable, thus it is necessary to exclude blood interference areas and mucus interference areas in the image.

To solve these problems, we propose an auxilliary diagnosis and decision-making system , termed EndometrialPrognosisNet, based on multimodal and multiple instance CNN. We further validate prognostic value of the method using 102 cases of endometrial AH/EIH diagnosed in the West China Second University Hospital of Sichuan University.

The rest of this paper is organized as follows. Some related work is reviewed in Section.2. In Section.3, we present the methodology involved. We present the experiments and results in Section. 4. Section. 5 is conclusion.

2. Related works

2.1 EC and AH/EIH

Endometrial adenocarcinoma is a gynecological malignancy, characterized by high incidence worldwide, of which endometrioid carcinoma (EC) represents the highest proportion (Carlson and Mccluggage 2019). EC's precancerous lesion is the atypical hyperplasia of endometrium (AH)/endometrial intraepithelial hyperplasia (EIH), the characteristic image pattern of which is defined as crowded aggregation of the tubular or branching endometrial glands with cytological atypia (Downing et al. 2020). However, variability of the image patterns for glandular and cytological atypia, coupled by complexity of the glandular structure complicated by the menstrual cycle, inevitably result in subjectivity, thus lowering inter- and intra-observer reliabilities. Meanwhile, AH/EIH and EC have been found to coexist in around 25~40% cases, while around 1/3 of all AH/EIH cases diagnosed in the biopsy are expected to eventually be diagnosed as EC in the immediate hysterectomy or a 1-year follow-up (Elke et al. 2010). Notably, precise differential diagnosis of AH/EIH and EC remains problematic, even for experienced pathologists, due to the common coexistence of EIH/AH and EC in the pathological slide as well as the low coincidence of AH/EIH diagnosis in biopsies. To date, two main options for the treatment of AH/EIH, namely hysterectomy and conservative hormone treatment, have been described (Papke et al.). Particularly, hysterectomy is recommended for patients above 40 years old, who do not require or are unresponsive to hormone treatment, as well as AH/EIH cases that clinicians find it hard to distinguish from EC in biopsy(Papke et al.). However, hysterectomy application is not applicable in patients at a fertility age, thereby necessitating conservative treatment which is accompanied by the risk of EC and fast progression. This poses a great challenge for gynecologists, necessitating development of an optimal therapy for patients.

2.2 Problem setting

An introduction is given to problem setting, while the multimodal and multiple instance learning technology is used for method development. For an arbitrary natural number N, the following definition is given: $[N] := \{1 \ ,..., \ N\}$. A set of vectors, with non-negative elements

and whose sum is one, are referred to as probability vectors. Two probability vectors p, and q are given, and their cross entropy represented by $L(p,q)$

A prognostic data set of *N* endometrial AH/EIH cases is constructed. Data corresponding to each case comprises a H&E staining pathological WSI I_n, physiological state information A_n in diagnosis and prognostic state information Y_n, with the data set represented as $\{(I_n, A_n, Y_n)\}_{n=1}^{N}$, $n \in [N]$. The size of the WSI is too large to be directly input into the CNN, thus we apply a method based on image patches. We consider image patches with 512×512 pixels . The image patch set (known as bag in multiple instance learning framework), comprises histopathological WSI corresponding to the *nth* patient and is represented by β_n, where $n \in [N]$. Each bag $b \in \beta_n$ has characteristics represented by a set of image patches τ_b, where $b \in \beta_n$. The notions of WSI, bag and image patch are shown in Fig.1.

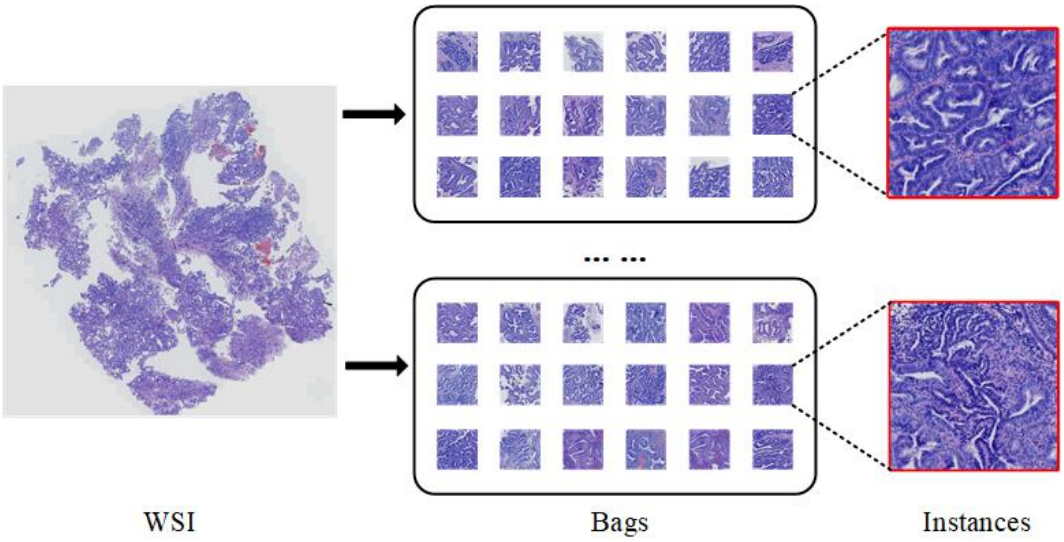


Fig. 1. Illustration of the notions of WSI, bag, and instance (image patch). Numerous image patches (512×512 pixels) are extracted from the WSI, at a magnification of ×20. Class labels of the bags (image patch sets), rather than those of the image patches, are evident. Notably, image patches in a single bag originate from the WSI in the same case so that the trained multiple instance learning model can predict prognosis of the patient according to multiple image patches in the bag.

2.3 Multiple Instance Learning (MIL)

In multiple instance learning (MIL), a multiple instance classifier is established by learning the bags (multiple instance bags) with class labels, while the classifier is applied for prediction of multiple unknown bags. The bag is a set comprising multiple samples, and only contains labels, without the samples containing the labels. Previous studies have described various models and algorithms of multiple instance learning (Campanella et al. 2018; Shamsolmoali et al. 2021), while MIL has been recently applied for pathological WSI analysis. In endometrial H&E slides, normal

and lesioned regions are mixed up, and not all image patches from the WSIs in AH cases contain the lesioned regions (Lotter et al. 2021; Vu et al. 2020; Yao et al. 2020a; Zhang et al. 2020). In a practical diagnostic scene, pathologists observe multiple areas in the pathological slides, comprehensively analyze lesioned and normal regions, then give diagnostic conclusions. In order to simulate the practical diagnosis processes performed by pathologists, thee following multiple instance deep learning algorithms are designed. Specifically, a set of image patches corresponding to the WSI in one case form a bag whose label is prognosis of the case. Its characteristics are similar to those of the image patches (those of lesioned and normal regions) contained in the bag, while prognosis of the case is determined by combining the above-mentioned two types of image patches. For example, for a binary prognostic classification (such as cancer and cancer-free) analysis of the case, image patches from both the lesioned and normal regions are included in the patient's WSI. Bags comprising image patches extracted from WSIs of patients with prognostic states of cancer/cancer-free are regarded as positive/negative bags respectively, whereas those image patches extracted from WSIs of cancer patients are regarded as positive instances and the image patches extracted from the WSIs of cancer-free patients regarded as negative instances. Thus, the above binary classification can be explained as the MIL problem, while the accurately labeled bags consisting of endometrial tissue image patches without lesioned and normal regions, are used to perform the prognosis analysis.

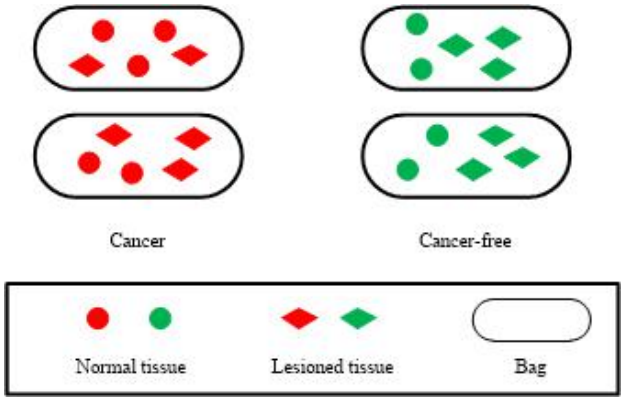


Fig. 2 Explanation of MIL. Profiles of bags of cancer and cancer-free types generated by the WSIs, corresponding to cancer and cancer-free cases, a respectively.

2.4 Multimodal Learning (MML)

Multimodal Deep Learning technology synthesizes information obtained from two or more modes during the analysis process, realizes information complement, and improves precision as well as robustness of prediction results. Previous studies have described various models, algorithms and development trends for Multimodal Learning(Bayoudh et al. 2021; Zubatiuk and Isayev 2021). In recent years, researches on curative effect and prognostic analyses have applied multimodal technology for joint feature learning and cross-modal relationship modeling(Cheerla and Gevaert 2019; Hosseini et al. 2020; Hügle et al. 2021; Yao et al. 2020b). In addition,

pathologists have also incorporated relevant clinical and physiological information in diagnostic and prognostic analyses, in addition to observing pathological slides under electronic microscopes. Overall, this affirms the relevance of different modal information during diagnostic and prognostic analyses (Hamdi et al. 2021). Our model expands coverage of information contained in the input data by integrating different modal information, such as pathological slides and physiological characteristics, with the aim of simulating the practical analytical processes performed by pathologists during prognosis. Overall, this approach improves precision and robustness of prediction results.

2.5 Deep-learning-based Cognitive Intelligence

Cognitive intelligence is defined to be the artificial intelligence of computer systems which simulate the human brain, which possesses some human cognitive capabilities to perform some specific cognitive tasks, for instance, feature learning, understanding, reasoning and decision-making and etc.(LeCun et al. 2015). Especially, the feature learning is the key point of cognitive intelligence, directly affecting the establishment of cognitive model and the effectiveness of the cognitive computing. Recently, the deep leaning technology has set-up a revolution in the field of cognitive intelligence. Compared with the traditional pattern recognition and concept learning, deep network model has more powerful feature learning ability, hence promoting the accuracy of cognitive intelligence in the fields of recognition, classification and prediction. To date, the cognitive intelligence based on the deep learning technology has been utilized in computer vision (Ji et al. 2013), speech recognition (Tompson et al. 2014.), emotion analysis (Ma et al.2015), natural language processing (He et al.2016), new drug development (Chen et al. 2016;Leung et al. 2014) and computer-aided diagnosis (Brasil et al.2021; Zeng et al.2021), and impressed by making breakthroughs in these fields.

3. Material and Methods

3.1 Data Acquisition

We experimentally validated the network using 102 cases diagnosed with AH/EIH in the biopsies between 2019 and 2020, at the pathology department of West China's Second University Hospital. Each case had 1 Formalin-fixed paraffin-embedded (FFPE) slide, with all 102 slides collected and reviewed by 2 experienced pathologists to confirm AH/EIH diagnosis. The slides

were then scanned at a magnification of ×20 ($0.5\mu m/pixel$) on a Motic®EasyScan system (Motic Electric Group Co., Ltd), to generate high-quality WSIs. Each patient's age, as well as their final diagnosis after hysterectomy applied within 1-year follow-up period, collected and well-documented . The study group comprised 51 cancer and 51 cancer-free patients.

3.2 The Proposed Method

Prognosis of endometrial AH/EIH cases is predicted by summarizing the class labels of bags extracted from WSIs, then combining the date with physiological state information. Specifically, a test WSI I_n and physiological information A_n in diagnosis are given, with class label probability predicted as $P\left(\hat{Y}_n = 1\right) = p_i / p_0 + p_1$, $i \in \{0,1\}$, then,

$$
\begin{aligned}
p_1 &= \exp\left(\frac{1}{|\beta_n|}\sum_{b\in\beta_n}\log P\left(\hat{Y}_n = 1\right)\right), \\
p_0 &= \exp\left(\frac{1}{|\beta_n|}\sum_{b\in\beta_n}\log P\left(\hat{Y}_n = 0\right)\right).
\end{aligned}
\tag{1}
$$

where, $P\left(\hat{Y}_n = 0\right)$ and $P\left(\hat{Y}_n = 1\right)$ are class label probability of the bag $b \in \beta_n$. The proposed network, EndometrialPrognosisNet, comprises the following three parts, whose outputs are class label probabilities of prognosis of cases, as shown in Fig. 3.

(1) Image Preprocessing Unit $G_p : WSI \to patch$ **.** The unit evenly slices WSIs, discards image patches at their edges, then reuses the UNET network (Ronneberger et al., 2015) to perform semantic segmentation of foreground areas (including staining tissue areas, blood interference areas, mucus interference areas) and blank areas. Moreover, it discards the image patches with the proportion in the foreground areas of less than the threshold T (set T = 0.5 by experience).

(2) Staining Tissue Region Detector $D_{tissue}\left(patch,\ \theta_D\right) : \left\{patch_i\right\}_{i\in\tau_b} \to P\left(L_n\right)$. A profile of its network structure is shown in Fig. 4. Briefly, its identification results describe blood and mucus interference areas, as well as stained tissue regions, including the lesioned and normal regions). Wherein, θ_D is the set of trainable parameters, and the feature vectors of the image patches are mapped as probabilities L_n of the image patch class labels, while $D_{tissue}\left(.;\ \theta_D\right)$ is characterized by a set of trainable parameters θ_D.

(3) Prognostic State Predictor $P_{state}\left(tissue,\ psy,\ \theta_P\right)_{i\in\tau_b} \to P\left(\hat{Y}_b\right)$. It has a multimodal convolution network with attention mechanism whose structure is shown in Fig. 5. Summarily, it combines characteristics of the staining tissue areas and physiological state information ***psy*** to

predict the corresponding prognosis $\hat{Y}_b$ which is represented by $P_{state}(tissue,\ psy;\ \theta_P)$. Where, the characteristic θ_P is a set of trainable parameters, while $(V,\ \omega) \in \theta_y$ in the equation is a parameter set of the attention network. Q-dimensional feature vectors $\{h_i'\}_{i\in\tau_b}$, generated by a full connection layer, are used to obtain the attention weighting feature vector $z \in R^{Q'}$ which is calculated by $z = \sum_{i\in\tau_b} a_i h_i'$, while the attention weight is defined as

$$a_i = \frac{\exp\left(\omega^T \tanh\left(Vh_i'\right)\right)}{\sum_{j\in\tau_b} \exp\left(\omega^T \tanh\left(Vh_j'\right)\right)}, \quad i \in \tau_b \tag{2}$$

In the EndometrialPrognosisNet network, different modules are trained separately as follows: (1) training the **Tissue Detector** (the left network in Fig. 3); (2) training the multimodal **Prognostic Analysor** (the right network in Fig. 3).

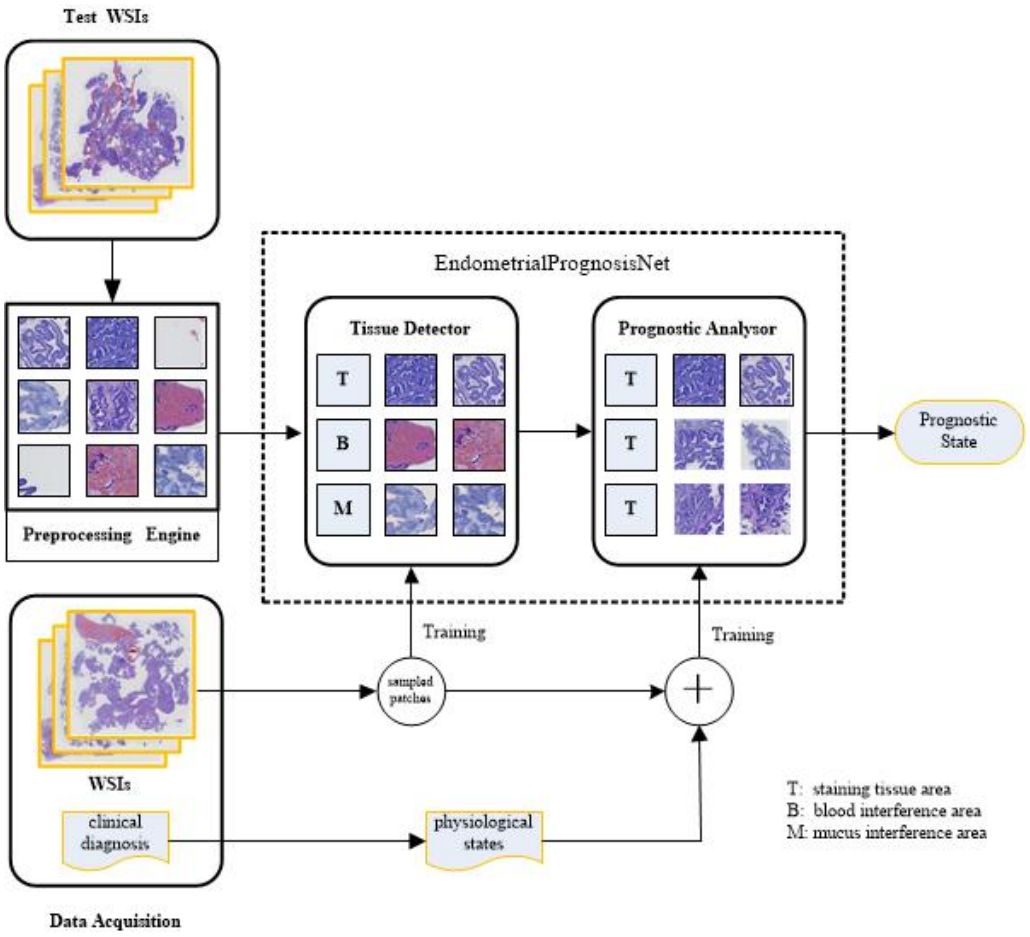


Fig. 3 Structure of the developed EndometrialPrognosisNet network. Modules for tissue region detection and multimodal prognosis are trained separately.

3.3 Staining Tissue Region Detector

Staining tissue regions are detected based on output of foreground image patches by the Image Preprocessing Unit G_p . The foreground image patches are of three types, namely the staining tissue area (which comprises lesioned and normal regions), blood interference area, and mucus interference areas. Training of the staining tissue region detector network can be formulated using the following minimization equation (Eq.(3)):

$$\left(\hat{\theta}_{Df},\hat{\theta}_{Dc}\right)\leftarrow \arg\min_{\theta_{Df},\theta_{Dc}}\sum_{n=1}^{N}L\left(L_n,\ P\left(\hat{L}_n\right)\right)+\Delta$$
$$where:\ P\left(\hat{L}_n\right)=D_{tissue}\left(\left\{D_{tissue}\left(x_i;\theta_{Df}\right)\right\}_{i\in\tau_b};\theta_{Dc}\right) \tag{3}$$

Loss function is defined as the cross entropy between a true and predicted class label probability. The first term, in Eq. (3), is used to predict loss function through the image patch class label, whereas the second is a small positive number which ensures that loss function is not zero, thus preventing gradient disappearance.

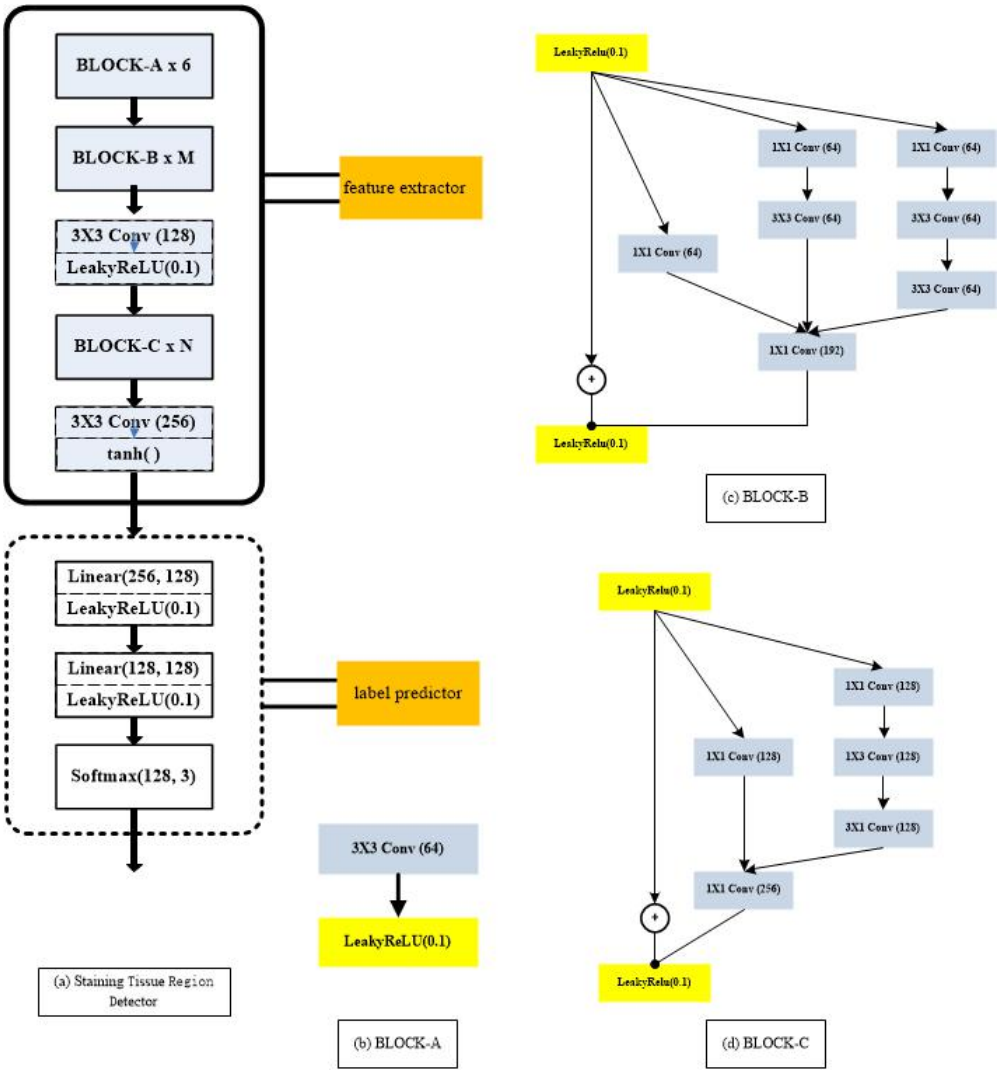


Fig. 4. Profile of network and modular structures of staining tissue region detection.

3.4 Multimodal Prognostic State Predictor

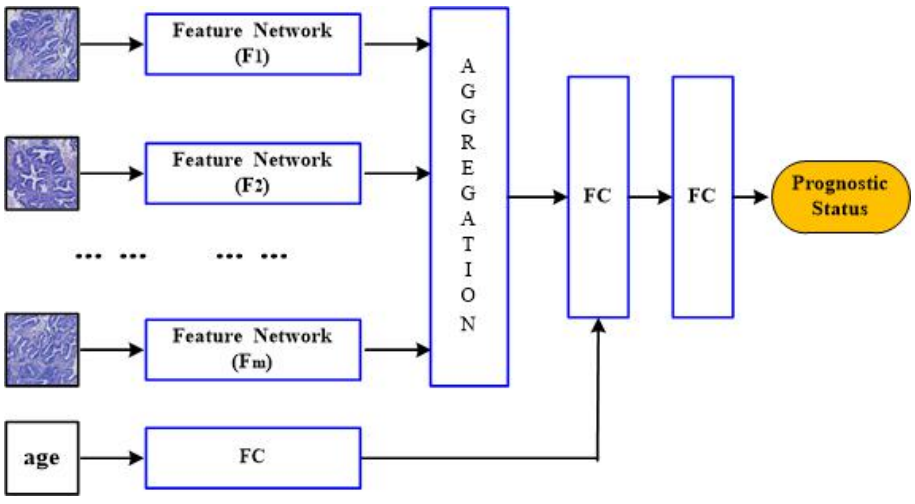
A multimodal prognosis analysis network is trained to predict class labels of the bags. Specifically, each bag contains multiple image patches and physiological state information from the same case and that corresponding to the bag, respectively. Prognosis of the case corresponding to a bag is used as a label for the bag, and comprises two types namely, cancer and cancer-free. Their prediction result is as follows:

$$P\left(\hat{Y}_n\right) = P\left(\left\{P\left(x_i;\theta_{Pf}\right)\right\}_{i\in\tau_b};\theta_{Pc}\right) \tag{4}$$

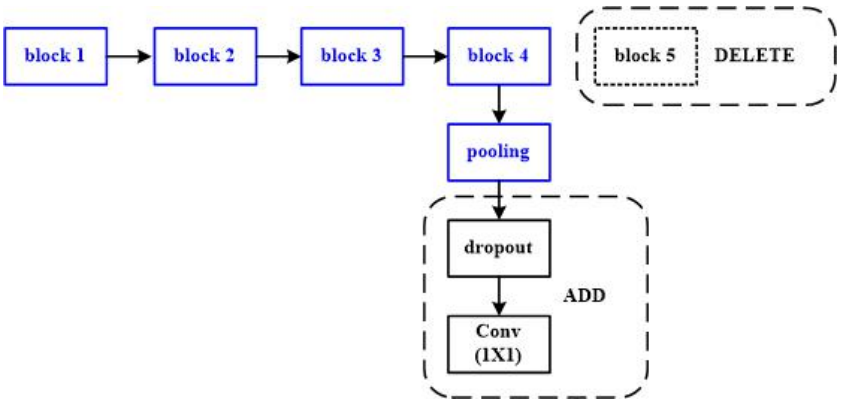
The class with the largest weighted sum, among all bag class labels in one case, is regarded as the class's prognosis type. Notably, training of the parameter set θ_P (where θ_P is the set of trainable parameters of the multimodal prognosis analysis network) can be represented by the following minimization problem:

$$\begin{gathered}\left(\hat{\theta}_{Pf},\hat{\theta}_{Pc}\right) \leftarrow \arg\min_{\theta_{Pf},\theta_{Pc}} \sum_{n=1}^{N}\sum_{b\in\beta_n} L\left(Y_n,\ P\left(\hat{Y}_n\right)\right)+\Delta_2 \\ where,\ P\left(\hat{Y}_n\right) = P\left(\left\{P\left(x_i;\theta_{Pf}\right)\right\}_{i\in\tau_b};\theta_{Pc}\right)\end{gathered} \tag{5}$$

According to Eq. (5), the first term $\sum_{n=1}^{N}\sum_{b\in\beta_n} L\left(Y_n,\ P\left(\hat{Y}_n\right)\right)$ represents loss function predicted through the bag class labels, with the attention mechanism weight of samples introduced in Eq.(2), namely with only the samples with high attention used to predict the prognosis class label corresponding to the bag. The second term Δ_2 is a small positive number, to ensure that loss function is not zero and prevent gradient disappearance. The Prognostic analysis network has a multimodal convolution network with attention mechanism whose structure is shown in Fig. 5.



(a) Profile of the prognosis analysis network, alongside its feature extraction network (Fm). M ranges from 5 ~ 20, and has an optimal value of 10.



(b) Feature Network (Fm). improved residual network based on RESNET50.

Fig. 5 The Prognostic analysis network and its backbone structure.

3.5 Algorithms

Algorithms corresponding to the related modules in the EndometrialPrognosisNet network are described in Tables 1 and 2. The algorithm for the module for detecting the tissue region is outlined in Table 1, and its parameters are updated by using a single image patch as a mini-batch training sample.

Table 1. Network parameter update in the training process of the tissue region detection network.

Inputs: the image patch training set $\{PATCH_n\}$ with region type label $\{L_n\}$, the learning rate η, the super parameter Δ, and the number of training epochs M.
Roles: training the tissue region detection network $D_{tissue}(patch,\ \theta_D)$, which includes a feature extractor $D_{tissue}(.;\theta_{Df})$ and a region class label predictor $D_{tissue}(.;\ \theta_{Dc})$.
for m=1 to M do
for n=1 to N do
$\{h_i^{'}\}_{i\in\tau_b} \leftarrow \{D_{tissue}(x_i;\theta_{Df})\}_{i\in\tau b}$
$L_{bag} \leftarrow L\left(L_n,\ D_{tissue}\left(\left(\{h_i\}_{i\in\tau_b}\right),\ \theta_{Dc}\right)\right)$
$\theta_{Df} \leftarrow \theta_{Df} - \eta\frac{\alpha L_{bag}}{\alpha\theta_{Df}}$

$\theta_{Dc} \leftarrow \theta_{Dc} - \eta * \frac{\alpha L_{bag}}{\alpha \theta_{Dc}}$
end for
end for
Outputs: the trained tissue region detection network $D_{tissue}(patch, \theta_D)$, which includes the feature extractor $D_{tissue}(.;\theta_{Df})$ and the region class label predictor $D_{tissue}(.; \theta_{Dc})$.

The algorithm for the network used in prognostic prediction is described in Table 2. Its network parameters are updated using instances (image patches) in each bag as a mini-batch.

Table 2. Network parameter update in the training process of prognosis analysis network.

Inputs: the training set $\{(I_n, A_n, Y_n)\}_{n=1}^{N}$ with the prognosis label $\{Y_n\}_{n=1}^{N}$, the learning rate η, the domain regularization parameter λ, and the number of training epochs M.
Roles: training the prognosis analysis network $P_{state}(tissue, psy; \theta_P)$, including training a feature extractor $P_{state}(.;\theta_{Pf})$ and a prognosis label predictor $P_{\text{state}}(.; \theta_{Pc})$.
for m=1 to M do
for n=1 to N do
for b=1 to $\lvert\beta_n\rvert$ do
$\{h_i'\}_{i\in\tau_b} \leftarrow \{P_{state}(x_i;\theta_{Pf})\}_{i\in\tau_b}$
$L_{bag} \leftarrow L(Y_n, P_{\text{state}}(\{h_i\}_{i\in\tau_b}); \theta_{Pc})$
$\theta_{Pf} \leftarrow \theta_{Pf} - \eta \times \frac{\alpha L_{bag}}{\alpha \theta_{Pf}}$
$\theta_{Pc} \leftarrow \theta_{Pc} - \eta \times \frac{\alpha L_{bag}}{\alpha \theta_{Pc}}$
end for
end for
end for
Outputs: the trained prognosis analysis network $P_{state}(tissue, psy; \theta_d)$ including the feature extractor $P_{state}(.;\theta_{Pf})$ and the prognosis label predictor $P_{\text{state}}(.; \theta_{Pc})$.

4. Experimental validation of the network

4.1 Experimental Settings

The Tissue Region Detector and Prognosis Analysis Network in the EndometrialPrognosisNet network are first trained, 512×512-pixel image patches extracted from the WSIs, then used for both training and testing processes. In a case where certain type of image patch data is insufficient, amplification is performed by rotating the image patches by 90, 180 or 270 degrees. All the above network parameters can be optimized by the adaptive moment estimation (ADAM)algorithm(2014; Kingma and Ba, 2014), with the adaptive learning rate set at 0.0001, while super parameters β1 and β2 are set at 0.9 and 0.9999, respectively. The image patch data set is constructed after training and testing of the Tissue Region Detector. The data set has a scale of 6000 image patches, and is constructed by randomly extracting no more than 200 image patches from different type of areas in each WSI. The ratios between image patches, r from the stained tissue, mucus interference, and the blood interference areas are controlled at 2:1:1 during the extraction process. Then, the dataset is divided into training, validation and test sets, at proportions of 60, 20 and 20%, respectively, without separation of patients involved in dividing the data set. Notably, the ratio between image patches and the three types mentioned above is kept constant. During the training process, iteration parameters are updated using a single image patch, as a mini-batch training sample, with image patches shuffled every 10 training epochs. In the experiments, we use the staining tissue Region detection network $D_{tissue}\left(patch,\ \theta_D\right)$ shown in Fig. 4. This includes a feature extractor $D_{tissue}\left(.;\theta_{Df}\right)$ and an image patch class label predictor $D_{tissue}\left(patch,\ \theta_D\right)$, whose full connection layer output is a 1024 dimensional vector, that allows prediction of the image patch's class label.

To test the network's prognostic value, a dataset of bags is used for training and testing. The dataset comprises a scale of 2000 bags, with210 image patches randomly extracted from the staining tissue areas in each WSI, of which 15 are randomly set as one bag, and each WSI generates 15~25 bags. The resulting data is divided into training, validation and test sets, at proportions of 60, 20 and 20%, respectively, with separation of patients involved in dividing the data set. During the training process, iteration parameters are updated using a single bag as a mini-batch training sample, with bags in the training set shuffled every 10 training epochs. For experimental work, the prognosis label predictor is $P_{\text{state}}\left(.;\ \theta_{Pc}\right)$, as shown in Fig. 5. It also includes a feature extractor $P_{state}\left(.;\theta_{Pf}\right)$ and a prognosis label predictor $P_{\text{state}}\left(.;\ \theta_{Pc}\right)$. For label prediction, the full connection layer is used to transform output features into a 512-dimensional vector, prior to applying the attention mechanism. The numbers of input and hidden units, in the attention network, are 512 and 128, respectively, while the full connection layer of the prognosis label predictor $P_{\text{state}}\left(.;\ \theta_{Pc}\right)$ outputs a 1024-dimensional vector to predict a prognosis label.

4.2 Results and Discussions

The findings on prognosis analysis by pathologists are presented in Table 3. The first column indicates types of methods, where Human Expert 1, Human Expert 2 and Human Expert 3 represent analysis results by the three pathologists and Human Expert Majority represents analysis results produced by the three pathologists through majority voting method. In addition, MIL-MM-1 represents the method reported in the current study excluding blank areas and after removal of interference areas in the WSI, MIL-MM-2 represents the method in the current study excluding age information, and MIL-MM-3 represents the entire process of the method reported in the current study. The mean and standard deviation (***std***) $(mean,\ std)$ of various indexes determined by three fold cross validation are presented in Table 3(Efron and Bradley, 1983). Evaluation indexes including ACCURACY, PRECISION and RECALL(Townsend, 1971) were used to compare the performance of the above methods which were defined by Eq. (6):

$$\begin{aligned} ACCURACY &= \frac{TP+TN}{P+N}, \\ PRECISION &= \frac{TP}{TP+FP}, \\ RECALL &= \frac{TP}{TP+FN}. \end{aligned} \tag{6}$$

ACCURACY represents the ratio of positive samples (P) and negative samples (N) classified correctly. A high ACCURACY indicates that the classifier has a good performance. PRECISION represents the proportion of positive samples TP predicted correctly by the model in all the predicted positive samples (TP +FP), and a high PRECISION indicates that the classifier has a good performance. RECALL represents the proportion of the positive samples TP predicted correctly by the model in all the positive samples (TP +FN), and a high RECALL indicates a good performance of the classifier. Notably, TP represents the true positive sample, TN represents the true negative sample, FP represents the false positive sample, and FN represents the false negative sample.

Table 3. Comparison of prognosis analysis performed by pathologists and that performed through the proposed method. Values represent mean and standard deviation determined by the three-fold cross validation. Results produced by pathologists, results from voting of the three pathologists and results obtained through the method presented in the current study were compared. The findings showed that the method proposed in the current study exhibited high performance similar to most voting results by the three pathologists.

Method	RECALL	ACCURACY	PRECISION
Human Expert 1	(0.785, 0.034)	(0.765, 0.029)	(0.755, 0.029)
Human Expert 2	(0.785, 0.034)	(0.794, 0.029)	(0.801, 0.031)
Human Expert 3	(0.824, 0.059)	(0.794, 0.059)	(0.778, 0.056)
Human Expert Majority	(0.824, 0.059)	(0.882, 0.059)	(0.882, 0.059)
MIL-MM-1	(0.628, 0.122)	(0.618, 0.089)	(0.613, 0.078)

MIL-MM-2	(0.784, 0.089)	(0.765, 0.089)	(0.755, 0.083)
MIL-MM-3	(0.863, 0.033)	(0.853, 0.029)	(0.846, 0.031)

The findings showed that MIL-MM-3 presented similar effect to the analysis performance of Human Expert Majority (analysis indexes included mean and mean square difference of RECALL, ACCURACY, and PRECISION obtained by the three-fold cross validation experiment) (Table 3). In addition, the two methods showed higher performance compared with other methods and compared with analysis by individual pathologists (Table 3). This finding indicates that MIL-MM-3 has a higher performance compared with MIL-MM-1 and MIL-MM-2 and provides an experimental basis for removal of blank areas and interference areas, and use of multimodal inputs to perform pathological image analysis.

Intermediate steps of the proposed were analyzed and compared, which are necessary for accurately predicting the prognosis of patients with endometrial AH/EIH. Different type of image patches obtained from a WSI, including mucus interference area image patches, blood interference area image patches and staining tissue area (including the lesioned region and the normal region) image patches are presented in Fig. 6. Results for area classification performance comparison are presented in Table 4 in which the first column represents the types of models including VGG16 (Simonyan and Zisserman, 2014), RESNET50(He et al., 2016), RESNET101(He et al., 2016), INCEPTION-V4 (Szegedy et al., 2014; 2015a; b; Szegedy et al., 2016; Szegedy et al., 2016) and INCEPTION-RESNET-V2 (Szegedy et al., 2014). The testing results correspond to mucus interference area image patches, blood interference area image patches and the staining tissue area image patches and all the image patches, and performance evaluation indexes used in the analysis included RECALLs and ACCURACYs of the models. All the classification models used the pre-training weight of IIMAGENET (Jia et al., 2009; Russakovsky et al., 2014). The values in Table 4 represent the means and standard deviations of various evaluation indexes determined by three-fold cross validation. The findings indicated that the tissue region detector D_{tissue} proposed in the current study showed the best effect in the two indexes of RECALL and the ACCURACY (Table 4).

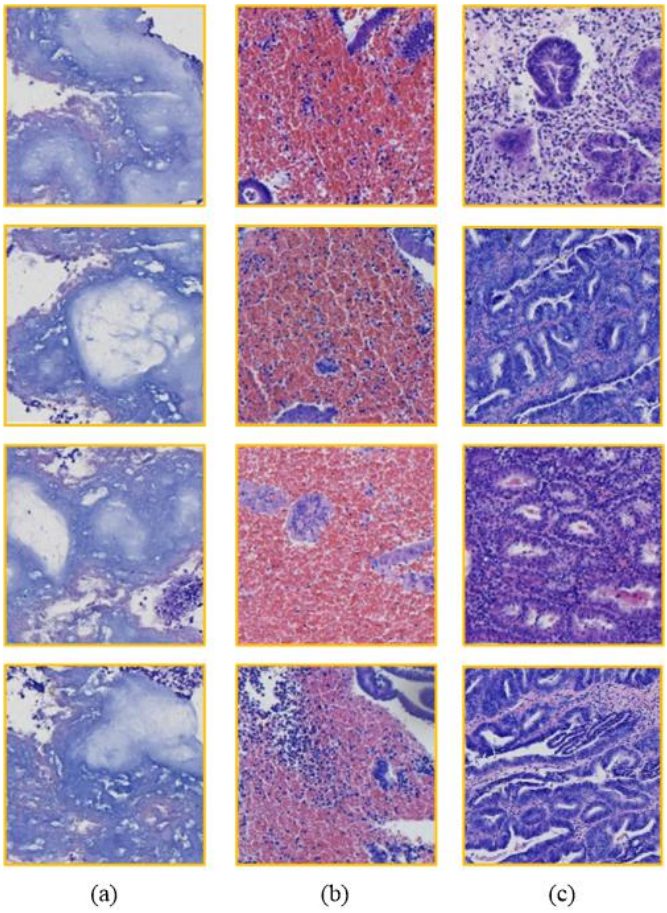


Fig. 6. Different type of image patches obtained from endometrial AH/EIH. The left column (a) shows the mucus interference area image patches, the middle column (b) shows the blood interference area image patches and the right column (c) shows the staining tissue area (including the lesioned region and the normal region) image patches. These image patches have a pixel size of 512X512 and are generated using WSI with a magnification of 20x.

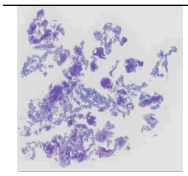
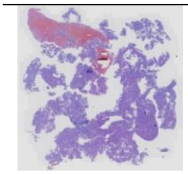
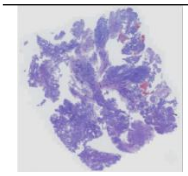
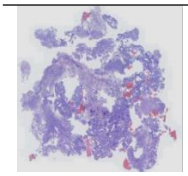
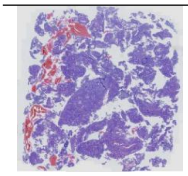
Table 4. Comparative experiments on region recognition performance

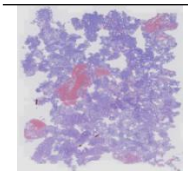
Model	**Mucus interference region**		**Blood interference region**		**Staining tissue region**		**All**	
	REC	ACC	REC	ACC	REC	ACC	REC	ACC
VGG 16	(0.93,0.029)	(0.92,0.028)	(0.92,0.025)	(0.95,0.018)	(0.95,0.017)	(0.94,0.021)	(0.94,0.023)	(0.94,0.024)
RESNET 50	(0.96,0.021)	(0.97,0.019)	(0.95,0.018)	(0.92,0.023)	(0.96,0.015)	(0.97,0.013)	(0.96,0.019)	(0.96,0.018)
RESNET101	(0.95,0.019)	(0.96,0.019)	(0.95,0.016)	(0.93,0.021)	(0.96,0.017)	(0.97,0.012)	(0.96,0.017)	(0.94,0.019)
INCEPTION V4	(0.94,0.018)	(0.94,0.021)	(0.95,0.012)	(0.94,0.019)	(0.96,0.013)	(0.96,0.015)	(0.95,0.014)	(0.95,0.018)
INCEPTION-RESNET-V2	(0.96,0.017)	(0.95,0.016)	(0.94,0.017)	(0.95,0.022)	(0.97,0.012)	(0.97,0.011)	(0.96,0.014)	(0.96,0.015)

D_{tissue}	(0.96,0.018)	(0.97,0.014)	(0.95,0.021)	(0.94,0.023)	(0.98,0.007)	(0.98,0.010)	(0.97,0.010)	(0.97,0.012)

Prognosis analysis results of six endometrial AH/EIH cases are presented in Table 5, whereby the first column shows the H&E staining tissue pathological WSIs in these cases, the second column shows the ages when the cases were diagnosed, the third column shows prognosis of the 6 cases, the fourth ~ sixth column shows the prognostic judgments of the cases by the three pathologists, the seventh column shows the voting results of the analysis conclusions by the three pathologists, and the eighth column shows the prognosis analysis conclusions using the method proposed in the current. The findings showed that under the experimental conditions described in Section 4.2, the prognosis analysis using the method proposed in the current study presented a higher ACCURACY and RECALL compared with the prognostic judgment of a single pathologist, and showed similar performance to that of most voting results of the three pathologists. This finding indicates high accuracy and effectiveness of the method proposed in the current study in prognosis analysis of endometrial AH/EIH cases.

Table 5. Examples of prognostic state analysis

WSI	AGE	PROGNSIS	Expert 1	Expert 2	Expert 3	Vote	Method
	40	cancer free	cancer free	cancer free	cancer free	cancer free	cancer free
	31	cancer	cancer	cancer	cancer	cancer	Cancer
	52	cancer	cancer free	cancer	cancer	cancer	cancer
	52	cancer	cancer	cancer free	cancer	cancer	cancer
	47	cancer	cancer	cancer free	cancer free	cancer free	cancer

	30	cancer	cancer free	cancer	cancer	cancer	cancer free

5. Conclusion

The current study presents an computer-aided system for prognostic analysis of endometrial AH/EIH based on deep convolution network developed by effectively combination of cognitive intelligence and CNN framework. The system possesses the capabilities to extract image features of endometrial AH/EIH by direct application of hemotoxylin-eosin (H&E) staining WSIs as input, and only requires simple labeling of blank, blood and mucus interference, and stained tissue areas in a small number of image patches, without the precise annotations of the normal regions and the lesioned regions in every image patch. Patient prognostic analysis is accomplished by the application of a multimodal prognosis analyzer, according to the image patterns of staining tissue areas in combination with the information on patients' physiological states. The system has the following advantages: At first, we set up the frame-work of computer-aided decision-making system based on the WSIs image patterns, and transfer the large-scale WSI analysis to the small-scale analysis of multiple suspected lesion regions, so that the analysis of WSI image patterns can be accomplished the major computer vision models. Moreover, the system, accompanied by Staining Tissue Region Detector、 Prognostic State Predictor, can accomplish effective training of prognosis analysis network without precise labeling of lesioned regions on WSIs . The system was applied for prognostic analysis of 102 endometrial AH/EIH cases, and the findings showed that the method had a higher ACCURACY, RECALL and PRECISION in prognostic analysis compared with those of prognostic judgment by individual pathologist. Notably, the performance of the methods was similar to most voting results of the three pathologists. These findings provide a basis for prognostic analysis of endometrial AH/EIH using deep learning technology. The system can be integrated into a clinical decision-making system to provide pathologists and gynecologists with an important reference for diagnosis and treatment of endometrial AH/EIH, hence the therapeutic strategy can be significantly optimized to improve prognosis of patients.

References

Am Mendola, ., Bariani E, Capitani A and Barresi V (2020) The histopathological diagnosis of atypical meningioma: glass slide versus whole slide imaging for grading assessment. Virchows Arch.: 1-10. http://doi: 10.1007/s00428-020-02988-1.

Bayoudh K., Knani R, Hamdaoui F and Mtibaa A (2021) A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. The Visual Computer. http://doi: 10.1007/s00371-021-02166-7.

Campanella G, Silva V, and Fuchs TJ (2018) Terabyte-scale Deep Multiple Instance Learning for Classification and Localization in Pathology. CVPR2018.arXiv:1805.06983.

Carlson J, and Mccluggage WG (2019) Reclassifying endometrial carcinomas with a combined morphological and molecular approach. Curr Opin Oncol 31: 411-419. http://doi: 10.1097/CCO.0000000000000560.

Ciresan DC, Giusti A, Gambardella LM and Schmidhuber J (2013) Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks. MICCAI 2013:411-419.

Downing MJ, Papke D, Tyekucheva S, and Mutter GL (2020) A New Classification of Benign, Premalignant, and Malignant Endometrial Tissues Using Machine Learning Applied to 1413 Candidate Variables. Int J Gynecol Pathol 39: 333-343. http://doi: 10.1097/PGP.0000000000000615.

Efron and Bradley (1983) Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation. JASS 78: 316-331. http://doi: 10.1080/01621459.1983.10477973.

Elke A, Jarboe and George L, and Mutter (2010) Endometrial intraepithelial neoplasia. Semin Diagn Pathol. doi: 10.1053/j.semdp.2010.09.007.

Hamdi A, Aboeleneen A, and Shaban K (2021) MARL: Multimodal Attentional Representation Learning for Disease Prediction. CVPR 2021:143-154.

Hashimoto N, Fukushima D, Koga R, Takagi Y, Ko K, Kohno K, Nakaguro M, Nakamura S, Hontani H, and Takeuchi I (2020) Multi-scale Domain-adversarial Multiple-instance CNN for Cancer Subtype Classification with Unannotated Histopathological Images. CVPR 2020. arXiv:2001.01599.

He K, Zhang X, Ren S, and Sun J (2016) Deep Residual Learning for Image Recognition. IEEE. CVPR 2016:770-778. http://doi: 10.1109/CVPR.2016.90.

Hosseini MP, Tran TX, Pompili D, Elisevich K and Soltanian-Zadeh H (2020) Multimodal Data Analysis of Epileptic EEG and rs-fMRI via Deep Learning and Edge Computing. Artif Intell Med 104: 108-113. http://doi: 10.1016/j.artmed.2020.101813.

Hügle M, Kalweit G, Hügle T, and Boedecker J (2021) A Dynamic Deep Neural Network for Multimodal Clinical Data Analysis. Studies in Computational intelligence11(3):79-92. http://doi: 10.1007/978-3-030-53352-6_8.

Jia D, Wei D, Socher R, Li LJ, Kai L, and Li FF (2009) ImageNet: A large-scale hierarchical image database. CVPR 2009:248-255.

Kingma D, and Ba J (2014). Adam: A Method for Stochastic Optimization. Computer Science. doi.

Lotter W, Diab AR, Haslam B, Kim JG and Sorensen AG (2021) Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. Nat Med. http:// doi: 10.1038/s41591-020-01174-9.

Papke DJ, Lohmann S, Downing M, Hufnagl P and Mutter GL Computational augmentation of neoplastic endometrial glands in digital pathology displays. J of Pathol. http://doi: 10.1002/path.5586.

Pinckaers H, Bulten W, Jeroen V and Litjens G (2020) Detection of prostate cancer in whole-slide images through end-to-end training with image-level labels. IEEE T Med Imagine 40:1817-1826.

Rijthoven MV, Balkenhol M, Silia K, Laak J and Ciompi F (2020) HookNet: multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images. Med Image Anal 68(8):1080-1089.

Ronneberger O, Fischer P and Brox T (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. Springer International Publishing. http://doi:

10.1007/978-3-319-24574-4_28.
Russakovsky O, Deng J, Su H, Krause J, Satheesh S and et al (2014) ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vision 24: 1-42. http://doi: 10.1007/s11263-015-0816-y.
Shahzad M, Umar AI, Khan MA and et al (2020) Robust Method for Semantic Segmentation of Whole-Slide Blood Cell Microscopic Image. Comput. Math Method M 2020: 1-13. http://doi: 10.1155/2020/4015323.
Shamsolmoali P, Wang R and Sadka AH (2021) Introduction to the Special Issue on Advanced Approaches for Multiple Instance Learning on Multimedia Applications. ACM T Multim Comput. 17: 1-2. http://doi: 10.1145/3459603.
Simonyan K and Zisserman A (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. Computer Science. CVPR 2014: 1409-1556.
Szegedy C, Liu W, Jia Y and et al (2014) Going Deeper with Convolutions. IEEE Computer Society. http://doi: 10.1109/CVPR.2015.7298594.
Szegedy C, Vanhoucke V, Ioffe S and et al (2016) Rethinking the Inception Architecture for Computer Vision. CVPR 2016: 2818-2826.
Townsend JT (1971) Theoretical analysis of an alphabetic confusion matrix. Atten Percept Psycho 9: 40-50. http://doi: 10.3758/BF03212817.
Vu T, Lai P, Raich R and et al (2020) A Novel Attribute-Based Symmetric Multiple Instance Learning for Histopathological Image Analysis. IEEE T Med Imaging. http://doi: 10.1109/TMI.2020.2987796.
Xing F, Zhang X, and Cornish TC(2021) Artificial intelligence for pathology. Artif Intell Med 218: 183-221.
Yao J, Zhu X, Jonnagaddala J and et al (2020) Whole Slide Images based Cancer Survival Prediction using Attention Guided Deep Multiple Instance Learning Networks. Med. Image Anal. http://doi: 10.1016/j.media.2020.101789.
Zhang X, Lu D, Gao P and et al (2020) Survival-relevant high-risk subregion identification for glioblastoma patients: the MRI-based multiple instance learning approach. Eur. Radiol.. http://doi: 10.1007/s00330-020-06912-8.
Zubatiuk T and Isayev O (2021) Development of Multimodal Machine Learning Potentials: Toward a Physics-Aware Artificial Intelligence. Accounts Chem. Res. http:// doi: 10.1021/acs.accounts.0c00868.
LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature521(7553): 436−444.
Ji S W, Xu W, Yang M, Yu K (2013). 3D convolutional neural networks for human action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 35(1): 221−231.
Tompson J, Jain A, LeCun Y, Bregler C (2014) Joint training of a convolutional network and a graphical model for human pose estimation.Eprint Arxiv 2014:1799-1807.
Ma JS, Sheridan RP, Liaw A, Dahl G E, Svetnik V (2015) Deep neural nets as a method for quantitative structure-activity relationships. Journal of Chemical Information and Modeling 55(2): 263−274.
Zeng N, Li H, Peng Y (2021) A new deep belief network-based multi-task learning for diagnosis of Alzheimer's disease. NEURAL COMPUT APPL 12:210-224.
He Y H, Xiang S M, Kang C C, Wang J, Pan C H (2016). Crossmodal retrieval via deep and bidirectional representation learning. IEEE T Multimedia 18(7): 1363−1377.
Chen YF, Li Y, Narayan R, Subramanian A, Xie XH (2016) Gene expression inference with deep

learning. Bioinformatics 32(12): 1832−1839.
Leung MK, Xiong HY, Lee LJ, Frey BJ (2014) Deep learning of the tissue regulated splicing code. Bioinformatics 30(12): 121−129.
Guo K, Ren S, Bhuiyan M, et al (2020) MDMaaS: Medical-Assisted Diagnosis Model as a Service With Artificial Intelligence and Trust. IEEE T Ind Inform 16(3): 2102-2114.