Development of model for identifying homologous recombination deficiency (HRD) status of ovarian cancer with deep learning on whole slide images

Ke Zhang[1†], Youhui Qiu[2†], Songwei Feng[1], Han Yin[1], Qi Liu[1], Yuxin Zhu[1], Haoyu Cui[2], Xiaoying Wei[3], Guoqing Wang[3], Xiangxue Wang[2*] and Yang Shen[1*]

Abstract

Background Homologous recombination deficiency (HRD) refers to the dysfunction of homologous recombination repair (HRR) at the cellular level. The assessment of HRD status has the important significance for the formulation of treatment plans, efficacy evaluation, and prognosis prediction of patients with ovarian cancer.

Objectives This study aimed to construct a deep learning-based classifier for identifying tumor regions from whole slide images (WSIs) and stratify the HRD status of patients with ovarian cancer (OC).

Methods The deep learning models were trained on 205 H&E-stained sections which contained 205 ovarian cancer patients, 64 were found to have HRD status while 141 had homologous recombination proficiency (HRP) status from two institutions Memorial Sloan Kettering Cancer Center (MSKCC) and Zhongda Hospital, Southeast University. The framework includes tumor regions identification by UNet++ and subtypes of ovarian cancer classifier construction. Referring to the EasyEnsemble, we classified the HRP patients into three distributed subsets. These three subsets of HRP patients were combined with the HRD patients to establish three new training groups for subsequent model construction. The three models were integrated into a single model named Ensemble Model.

Results The UNet++ algorithm segmented tumor regions with 81.8% accuracy, 85.9% recall, 83.8% dice score and 68.3% IoU. The AUC of the Ensemble Model was 0.769 (Precision=0.800, Recall=0.727, F1-score=0.762) in the study. The most discriminative features between HRD and HRP comprised S_mean_dln_obtuse_ratio, S_mean_dln_acute_ratio and mean_Graph_T-S_Betweenness_normed.

Conclusions The models we constructed enables accurate discrimination between tumor and non-tumor tissues in ovarian cancer as well as the prediction of HRD status for patients with ovarian cancer.

†Ke Zhang and Youhui Qiu contributed equally to this work.

*Correspondence:
Xiangxue Wang
xwang@nuist.edu.cn
Yang Shen
shenyang@seu.edu.cn

Full list of author information is available at the end of the article

Introduction

Ovarian cancer is one of the three major reproductive tumors that seriously threaten women's health around the world [1]. Meanwhile, although the mortality rate of ovarian cancer has decreased in recent years, its mortality rate ranks second among gynecological malignant tumors, following malignant tumor of uterine [2]. Early symptoms of ovarian cancer are often atypical and there is a lack of effective screening methods [3]. Most ovarian cancer patients are diagnosed at an advanced stage of the disease, with distant spread or metastasis of the tumor, and the prognosis is poor usually [4]. Hence, the early detection of ovarian cancer and enhancing survival rates for patients have emerged as a pivotal research focus worldwide. For patients with ovarian cancer, the predominant treatment option consists of a combination of surgery and chemotherapy primarily [5]. Regrettably, the majority of ovarian cancer patients may experience disease recurrence and progress to a platinum-resistant state, resulting in a bleak prognosis [6].

Homologous recombination deficiency (HRD) refers to the dysfunction of homologous recombination repair (HRR) at the cellular level, which can be caused by many factors such as germline or somatic mutations of HRR related genes and epigenetic inactivation [7]. Under normal circumstances, in the face of DNA damage in cells, a variety of repair methods including HRR in the human body will make up for DNA damage. However, DNA damage cannot be properly repaired by HRR if HRD is present. The accumulation of DNA double-strand breaks leads to genomic and chromosomal instability. Poly ADP-ribose polymerase (PARP) inhibitors are a class of anti-tumor drugs. PARP inhibitors can inhibit the function of PARP enzyme to block the DNA damage repair pathway, thereby promoting the apoptosis of tumor cells and playing a stronger anti-tumor effect for patients with HRD, especially for those who suffer from ovarian cancer [8]. Therefore, the assessment of HRD status has the important significance for the formulation of treatment plans, efficacy evaluation, and prognosis prediction of patients with ovarian cancer. The clinical detection methods of HRD can be classified into three categories: HRR-related gene mutation detection, genomic scar and mutation pedigree analysis, and functional detection of HRD [9]. Although these methods can determine the HRD status of patients, they are costly and time-consuming, making them less widespread, especially in underdeveloped regions. Hence, there is an imperative requirement for a more effective and expedient HRD detection system that can offer a quick and accurate determination of the HRD status for patients with ovarian cancer.

Pathology represents a crucial aspect of cancer patient information and are instrumental in providing clinical practitioners with necessary support for patient management, outcome assessment, and treatment decision-making during the clinical practice [10, 11]. In recent times, the field of digital pathology has advanced significantly due to progress in science and technology. The foundation of digital pathology involves the use of a pathology scanner to scan solid pathological section images. The digitally captured images are then stored and interpreted in a digital file format to produce digital pathological whole slide images (WSIs). This approach provides clinical researchers with the advantage of more precise and convenient diagnostic tools, which can be used for diagnosis and analysis.

The continuous advancement of deep learning technology has introduced novel prospects for pathological diagnosis and treatment and has significantly contributed to the expansion of digital pathology. Currently, deep learning technology is primarily employed in three distinct areas of digital pathology: (1) screening and diagnosis, (2) prediction of treatment response, and (3) prognosis prediction [12–15]. In previous investigations, we have effectively devised a system that utilizes deep learning algorithms to extract pathological features and predict the prognosis of patients afflicted with breast cancer [16]. Furthermore, we have also conducted investigations involving digital pathology techniques, which have enabled us to differentiate between and accurately stratify the risk associated with basal cell carcinoma [17]. These deep learning applications have demonstrated remarkable efficacy towards achieving more precise and efficient diagnosis and developing personalized treatment strategies based on a patient's specific condition.

Similarly, the field of ovarian cancer research has also witnessed significant progress because of the ingenious integration of deep learning and digital pathology techniques. In the study conducted by Nero et al., a publicly available weakly supervised deep learning-based method was employed, which succeeded in accurately predicting the BRCA gene mutation status of ovarian cancer patients using digital pathology [18]. Our team is inspired by the achievements and envisages developing a novel model aimed at identifying the HRD status of ovarian cancer patients through the application of deep learning technologies that leverage informative pathology data.

Hence, the objective of our research is to develop two advanced models that employ a synergistic combination of UNet++ [19]generated tumor images and nuclear features based on the histopathologic images, which contain morphology, texture, spatial arrangement of tumor cell

Zhang *et al. Journal of Translational Medicine*   (2025) 23:267

Page 3 of 10

nucleus, interrelationships between tumor cells and other cells to differentiate ovarian cancer and stratify the HRD status of ovarian cancer.

Materials and methods

General dataset introduction
We aimed to investigate a cohort of 205 patients who had been definitively diagnosed with ovarian cancer based on postoperative pathology and HRD status, including dataset 1 (D1): 183 patients from Memorial Sloan Kettering Cancer Center (MSKCC) [20] and dataset 2 (D2): 22 patients from Zhongda Hospital, Southeast University. To effectively examine and analyze the data, all patients in the cohort had provided digital scan pathological slides alongside comprehensive clinical data. In this study, patients with no clear evidence to confirm HRD status were recognized as homologous recombination proficiency (HRP) status.

In this study, WSIs were obtained from the routine hematoxylin and eosin (H&E) diagnostic tissue slides and the quality of each slide was evaluated by two professional pathologists to ensure its reliability. Moreover, one slice of the highest quality was meticulously chosen for each patient, which was subsequently included in this study.

HRD detection
The following are the HRD detection methods used in the 2 cohorts included in this study. In D1, the research team from MSKCC employed clinical sequencing to infer HRD status, focusing specifically on variants in genes associated with the HRD DNA damage response (DDR), such as BRCA1 and BRCA2, as well as those characteristic of distinct mutational subtypes enriched for tandem duplications and foldback inversions [20, 21]. In D2, DNA extraction and HRD detection were conducted in collaboration with the laboratory personnel from Nanjing Simcere Company. The HRD assessments in this research employed next-generation sequencing (NGS) technology to identify mutations in BRCA1/2 and other homologous recombination repair (HRR)-related genes. Additionally, the HRD score was computed by evaluating loss of heterozygosity (LOH), large-scale state transitions (LST), and telomeric allelic imbalance (TAI), thereby providing a comprehensive determination of the tumor's HRD status. Tumors were classified as exhibiting HRD status when either of the following criteria was met: (1) an HRD score of ≥40, or (2) the presence of pathogenic variants in BRCA1/2 genes. Conversely, tumors not meeting these criteria were designated as HRP. The HRD Score model based on the Chinese population data can reflect the HRD status of the Chinese population compared with the

Myriad myChoice® CDx product approved by Food and Drug Administration (FDA) [22].

Tumor region identification and nuclear segmentation
The pathological sections were meticulously annotated using QuPath software by two experienced pathologists who categorized the following areas: tumor and other tissue. In cases where inconsistencies arose between the two sets of annotations, a senior pathologist reviewed the designated areas and made the final determination. The segmentation process utilized a sliding window operation of 512 pixels × 512 pixels to extract image blocks from full-field digital slices captured at 20x. These image blocks were input to the UNet++model for training, and the patches in D1 data were randomly divided into 80% for training and 20% for internal testing to evaluate the performance of the model. Ultimately, the predicted image blocks were combined to obtain the complete segmented image of the ovarian cancer tumor area.

The segmentation model developed in this research was trained on D1 and excels at distinguishing the tumor region and other tissue by making accurate pixel-level predictions in the test images. To ascertain the model's efficacy, we evaluated the performance of the model in terms of pixel precision, recall, intersection over union (IoU), and dice score. Moreover, the model we have built was applied to the pathological sections in D2 for automatic identification and extraction of images containing the tumor area.

The deep convolutional neural network method called Hover-Net is a powerful tool utilized in this study for both instance segmentation and classification of nuclei. The model was originally trained on a publicly available dataset known as CoNSeP [23]. In 2024, Pan et al. using the Hover-Net model to automatically segment different cell types in a tumor region, successfully predicted the survival of patients with lung adenocarcinoma based on pathological features [24]. In addition, in a previous study by our team, it was also by segmentation of different types of cells in the tumor region by the Hover-Net model that successfully classified the risk of basal cell carcinoma [17]. Therefore, in this research, we employed the Hover-Net model to segment cells present within the tumor area, classify each nucleus by type, and calculate pertinent morphology, texture, and spatial features.

Feature extraction
In this study, the histopathological features within the tumor region were carefully extracted and analyzed. The aim of the present study was to closely examine histopathological features in the tumor region by extracting nuclear features and tumor microenvironment (TME) profiles. A total of 240 dimensions were obtained from our feature extraction process, encompassing texture,

shape, spatial arrangement of cell nucleus characteristics and TME features.

Nuclear morphological/texture features (46 descriptors): These features focused on quantifying nuclear texture and shape using different measurements.

Nuclear spatial arrangement and TME features (194 descriptors): This set of characteristics are intended to capture the difference of nuclear topology and spatial structural relationship between tumor nuclei in corresponding conditions.

Feature selection

This study employed three different feature selection methods, namely, minimum redundancy maximum relevance (mRMR), Least Absolute Shrinkage and Selection Operator (LASSO), and Random Forest (RF), to identify the optimal features that most effectively distinguish between the two classes of ovarian cancer (HRD and HRP) in the modeling set. To avoid overfitting and the dimensionality curse, we established a maximum feature limit of 10 for the machine learning classifier. This limitation was implemented to address issues related to the generalization performance of the classifier, which can be compromised when the number of features exceeds a threshold relative to the number of training samples.

Classifier construction and evaluation

In the research, a significant disparity was observed in the number of HRD and HRP patients, indicating an imbalance in the data. In response to the observed data imbalance, referring to the EasyEnsemble [25], we classified the HRP patients into three distributed subsets. Correspondingly, these three subsets of HRP patients were combined with the HRD patients to establish three new training groups (ND1, ND2, ND3) for subsequent model construction.

To accomplish the aims of this study, we utilized four distinct machine learning classifiers: Logistic Regression (LR), k-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest (RF). These classifiers were combined with three different feature selection methods, yielding a total of 12 possible combinations. For each of the ND1-3 datasets, the data were randomly partitioned into an 80% training group and a 20% internal validation group to assess model performance. The most effective feature selection method and machine learning classifier were identified for each of the ND1-3 datasets based on the area under the receiver operating characteristic curve (AUC) derived from all images in the training group, using five-fold cross-validation, and these optimal combinations were designated as Model 1, Model 2, and Model 3, respectively. Subsequently, the three models (Model 1–3) generated from the previously established ND1-3 were integrated into a single model

named Ensemble Model. This Ensemble Model was then applied to the external testing group (D2) to evaluate its performance, assessed through measures such as AUC, F1-score, recall, and specificity. The entire workflow of our research is illustrated in Fig. 1.

Results

Patients characteristics of two datasets

The study cohort consisted of 205 ovarian cancer patients, 64 were found to have HRD status while 141 had HRP status, with a mean age of 63.81 years (SD: 10.79; range 37–89 years). Meanwhile, 97.6% patients were diagnosed in the advanced stages of the disease. All patients received surgical treatment, with 142 (69.3%) patients undergoing neadjuvant chemotherapy-interal debulking surgery (NACT-IDS) and 63 (30.7%) patients received the primary debulking surgery (PDS). Only 47 patients (22.9%) were administered PARP inhibitor treatment. The clinical and pathological features of the patients are presented in Table 1.

Tumor regions identification and the performance of subtype classifier

The UNet + + algorithm segmented tumor regions with 81.8% accuracy, 85.9% recall, 83.8% dice score and 68.3% IoU (Table 2). Regarding to the differentiation classifier, the performance of the 12 combination of feature selection and classifier schemes in terms of classification performance for differentiation subtypes HRD and HRP of ovarian cancer are summarized in Table 3. The combination of RF and RF analysis achieved the best performance on WSIs in distinguishing HRD and HRP of ovarian cancer in ND1 (AUC = 0.817 ± 0.052, precision = 0.751 ± 0.117, recall = 0.736 ± 0.070, F1-score = 0.734 ± 0.050). The combination with the best performance was RF and RF in ND2 (AUC = 0.815 ± 0.043, precision = 0.814 ± 0.167, recall = 0.736 ± 0.026, F1-score = 0.763 ± 0.069). Meanwhile, the combination with the best performance was RF and RF in ND3 (AUC = 0.838 ± 0.051, precision = 0.780 ± 0.160, recall = 0.764 ± 0.065, F1-score = 0.757 ± 0.066). The results are presented in Fig. 2A-C. Moreover, our model demonstrated excellent performance in both the internal testing groups (Fig. 2D-E, Table 1S) and external testing group (Fig. 3A-C, Table 2S), indicating its reliability.

Subsequently, we integrated the three models constructed (Models 1–3) into the single model called the Ensemble Model and evaluated its effect using the external testing group. The Ensemble Model has the great performance for differentiation subtypes HRD and HRP of ovarian cancer. The AUC of the Ensemble Model was 0.769 (Precision = 0.800, Recall = 0.727, F1-score = 0.762) in the external testing group. The results are shown in Table 4; Fig. 3D.
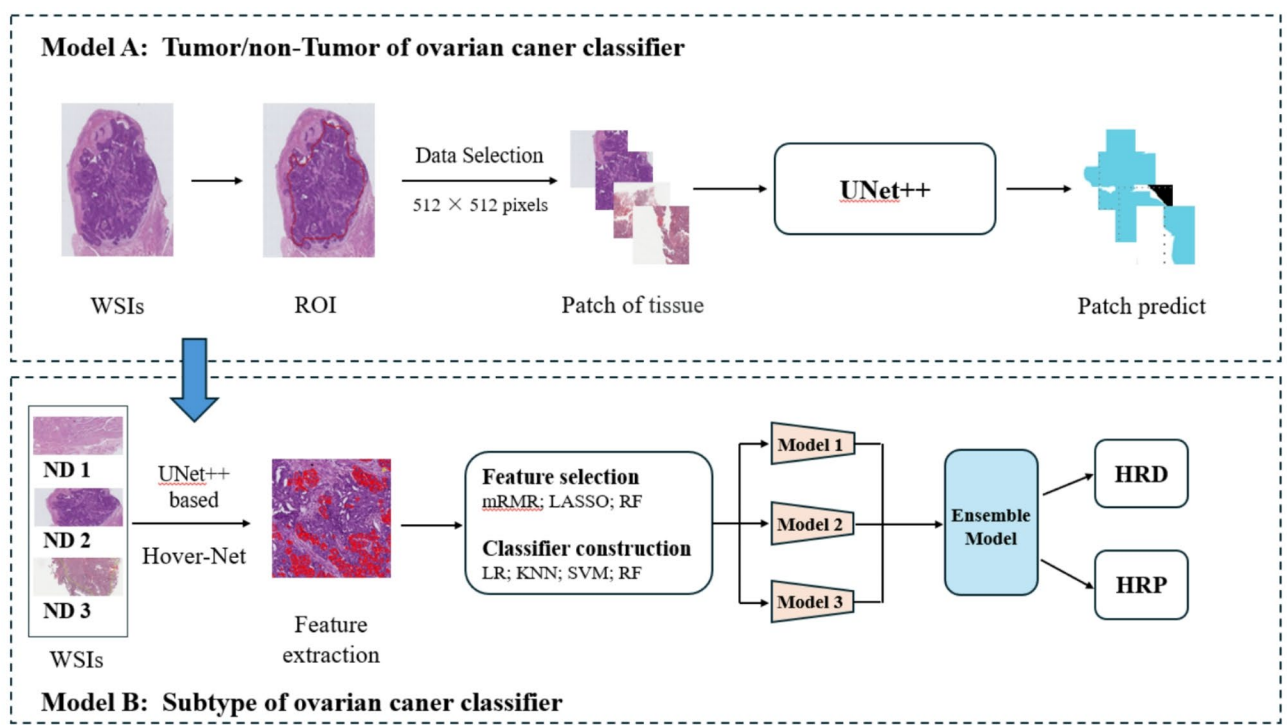
Fig. 1 Flow chart

Table 1 Summary of clinical and pathological features of the patients

	All data, *N* (%)	D1, *N* (%)	D2, *N* (%)
No. of patients	205	183	22
Age	63.81 ± 10.79	65.14 ± 10.36	52.73 ± 7.72
Stage			
I–II	5(2.4)	1(0.5)	4(18.2)
III–IV	200(97.6)	182(99.5)	18(81.8)
Type of surgery			
NACT-IDS	142(69.3)	140(76.5)	2(9.1)
PDS	63(30.7)	43(23.5)	20(90.9)
Received PARPi			
Yes	47(22.9)	39(21.3)	8(36.4)
No	158(77.1)	144(78.7)	14(63.6)
Subtype			
HRD	64(31.2)	53(29)	11(50.0)
HRP	141(68.8)	130(71)	11(50.0)

Table 2 The performance of tumor region segmentation model

Class	Accuracy	Recall	IoU	Dice score
Tumor region	0.818	0.859	0.683	0.838

The discriminative features of differentiation classifier

In the training groups, there were 10-dimension discriminative features used to construct the differentiation classifier (Fig. 4). We ranked the features by the contribution of each feature to the classification target, which is obtained by the feature importances function of the RF (Table 3S).

A total of three features were present in all three models including S_mean_dln_obtuse_ratio, S_mean_dln_acute_ratio and mean_Graph_T-S_Betweenness_normed. In this study, we discovered that HRD patients exhibited a distinctive spatial arrangement of tumor-infiltrating lymphocyte nuclei, demonstrating greater dispersion in comparison to non-HRD patients. Furthermore, the distance between tumor cells and necrotic cells was found to be significantly closed in the HRD patient group. These findings suggest that the HRD status of ovarian cancer may have a significant impact on the tumor microenvironment and cell-to-cell interactions.

Discussion

For patients diagnosed with epithelial ovarian cancer, the standard approach for initial treatment involves a comprehensive evaluation of the patient's condition, followed by surgery such as NACT-IDS and PDS. Additionally, adjuvant chemotherapy with platinum-based drugs is administered after surgery [26]. Recently, PARP inhibitors have emerged as a novel class of targeted therapeutics with significant therapeutic achievements in the maintenance treatment of ovarian cancer, particularly in patients with HRD status [27]. The detection of HRD may enable the use of PARP inhibitors in approximately 50% of ovarian cancer patients who present with this molecular subtype, offering a promising avenue for targeted therapy [28].

Table 3 The performance of the 12 combination of feature selection and classifier schemes for discriminating subtypes from ovarian cancer in training groups

Database	Classifier	Feature selection	AUC	Precision	Recall	F1-score
ND1	LR	LASSO	0.654±0.141	0.575±0.082	0.584±0.160	0.559±0.037
		RF	0.745±0.053	0.704±0.078	0.739±0.049	0.718±0.050
		mRMR	0.705±0.087	0.715±0.159	0.657±0.230	0.646±0.100
	RF	LASSO	0.680±0.109	0.635±0.097	0.577±0.105	0.589±0.028
		RF	*0.817±0.052*	*0.751±0.117*	*0.736±0.070*	*0.734±0.050*
		mRMR	0.661±0.132	0.675±0.224	0.637±0.145	0.620±0.118
	SVM	LASSO	0.601±0.075	0.580±0.103	0.531±0.182	0.533±0.084
		RF	0.807±0.080	0.745±0.150	0.741±0.124	0.739±0.126
		mRMR	0.642±0.071	0.673±0.198	0.671±0.123	0.639±0.066
	KNN	LASSO	0.656±0.049	0.765±0.134	0.456±0.063	0.562±0.062
		RF	0.674±0.082	0.673±0.127	0.679±0.097	0.670±0.094
		mRMR	0.574±0.123	0.545±0.187	0.452±0.197	0.446±0.158
ND2	LR	LASSO	0.608±0.056	0.592±0.185	0.549±0.194	0.533±0.167
		RF	0.752±0.106	0.708±0.169	0.706±0.095	0.689±0.061
		mRMR	0.645±0.157	0.577±0.195	0.601±0.099	0.583±0.141
	RF	LASSO	0.672±0.124	0.602±0.221	0.552±0.111	0.558±0.166
		RF	*0.815±0.043*	*0.814±0.167*	*0.736±0.026*	*0.763±0.069*
		mRMR	0.526±0.076	0.516±0.129	0.675±0.042	0.577±0.090
	SVM	LASSO	0.658±0.113	0.685±0.242	0.644±0.127	0.649±0.177
		RF	0.802±0.037	0.693±0.102	0.683±0.126	0.675±0.082
		mRMR	0.639±0.121	0.626±0.161	0.704±0.162	0.639±0.090
	KNN	LASSO	0.563±0.109	0.573±0.191	0.512±0.119	0.521±0.147
		RF	0.745±0.087	0.736±0.155	0.628±0.184	0.656±0.124
		mRMR	0.597±0.075	0.621±0.162	0.560±0.067	0.571±0.073
ND3	LR	LASSO	0.690±0.131	0.683±0.207	0.724±0.095	0.693±0.142
		RF	0.813±0.118	0.756±0.163	0.718±0.155	0.715±0.087
		mRMR	0.715±0.056	0.683±0.149	0.675±0.101	0.662±0.057
	RF	LASSO	0.608±0.075	0.607±0.135	0.688±0.129	0.622±0.086
		RF	*0.838±0.051*	*0.780±0.160*	*0.764±0.065*	*0.757±0.066*
		mRMR	0.611±0.094	0.595±0.158	0.596±0.087	0.573±0.092
	SVM	LASSO	0.569±0.110	0.592±0.141	0.576±0.164	0.570±0.140
		RF	0.782±0.080	0.698±0.148	0.724±0.146	0.702±0.127
		mRMR	0.672±0.129	0.685±0.176	0.578±0.161	0.606±0.140
	KNN	LASSO	0.533±0.076	0.588±0.169	0.428±0.094	0.466±0.057
		RF	0.763±0.068	0.751±0.168	0.676±0.109	0.691±0.068
		mRMR	0.551±0.092	0.538±0.182	0.554±0.145	0.537±0.144

One of the current challenges faced by clinicians is the rapid identification of ovarian cancer patients with HRD status, enabling them to derive maximum benefit from therapeutic application of PARP inhibitors. It is noteworthy that there is currently no globally standardized protocol for the detection of HRD status. Presently available technologies face challenges in directly determining HRD status of ovarian cancer patients through clinical information and histopathological features. Gene testing is commonly employed to identify HRD status, with products such as Myriad myChoice® CDx and Foundation-FocusTM CDx BRCA LOH currently FDA-approved for clinical use. Nonetheless, these tests are time-consuming and expensive, presenting major drawbacks.

In recent years, the increasing development of deep learning technology has highlighted its potential in the field of medical image processing [29]. There is growing enthusiasm for the use of deep learning to address clinical challenges and improve patient outcomes. Histopathologic scanning of cancer tissue typically involves capturing images at extremely high magnifications, resulting in file sizes that can reach into the billions of pixels. As a result, these pathological images contain a vast amount of tumor microenvironment data, which is crucial in studying the progression of cancer. Nero et al. utilized an openly available deep learning-based weakly supervised method called clustering-constrained-attention multiple-instance learning (CLAM) to construct a diagnostic model capable of identifying the BRCA1/2
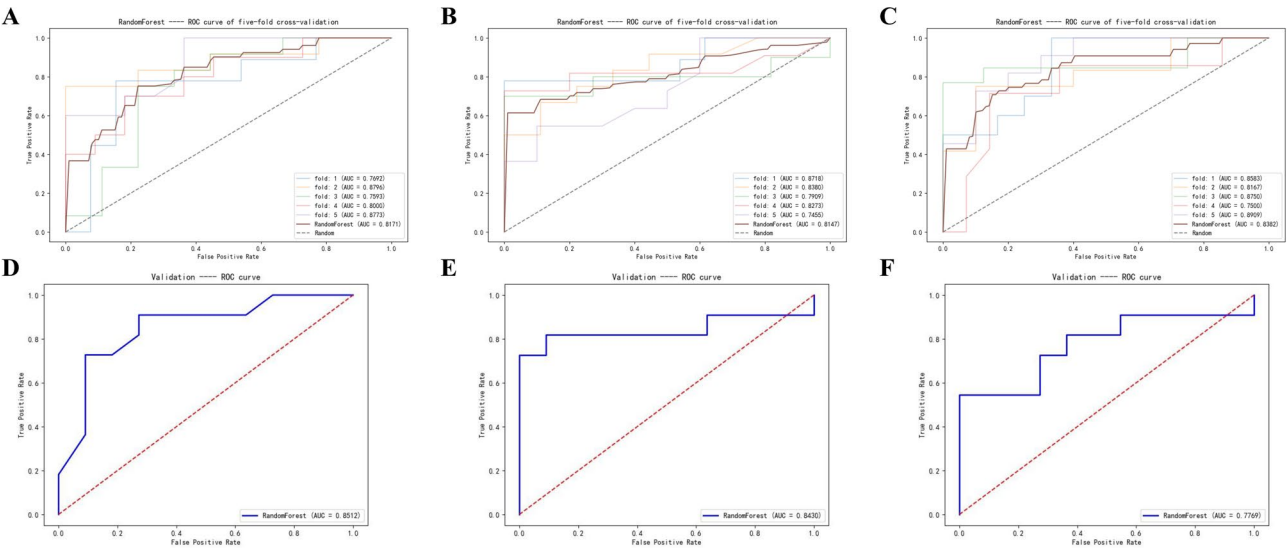
Zhang *et al. Journal of Translational Medicine*   (2025) 23:267

Page 7 of 10

Fig. 2 Performance of the subtype classifier on training groups. Training group in ND1-3 (A-C). Internal testing group in ND1-3 (D-F)
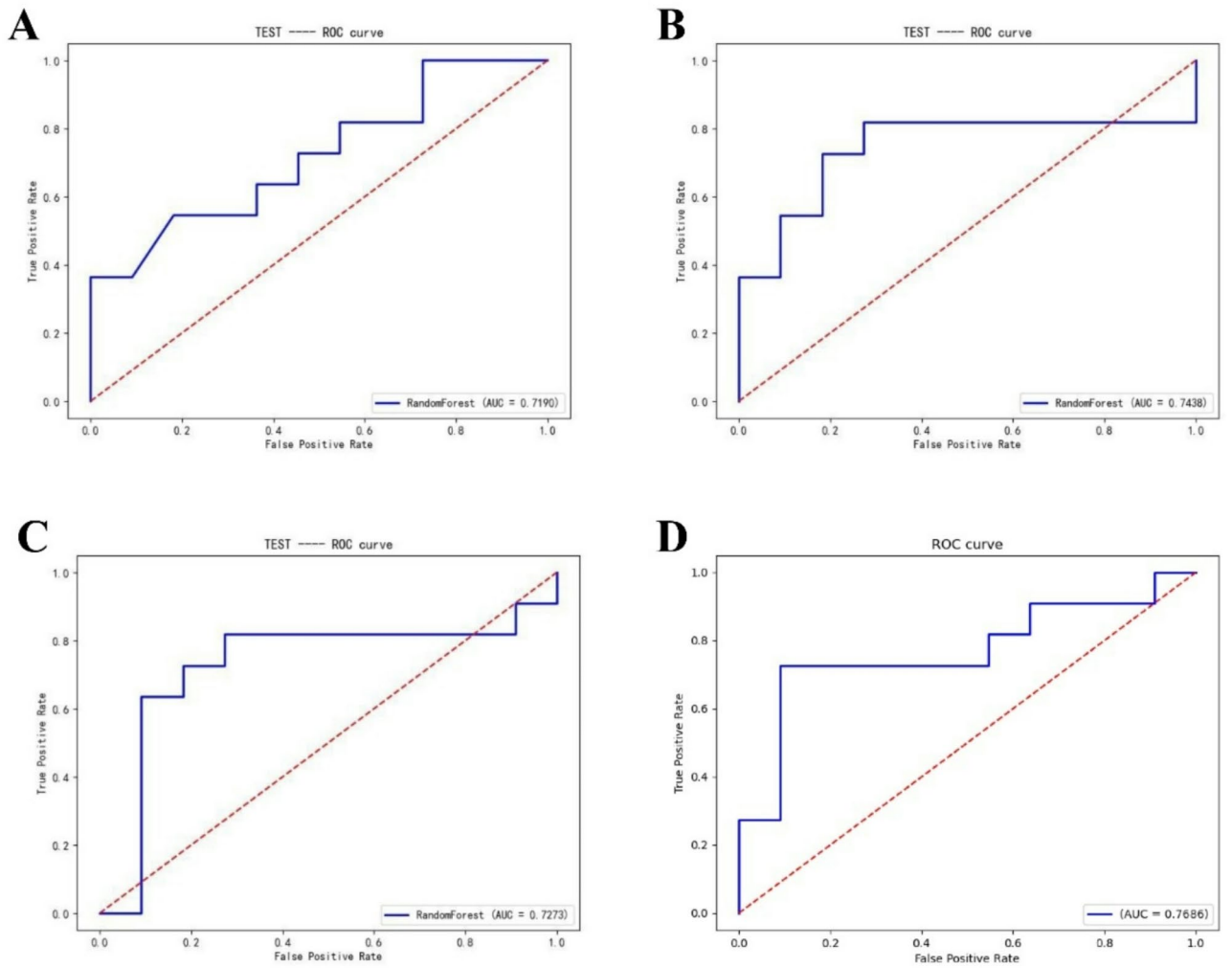

Fig. 3 The Receiver Operating Characteristic Curve (ROC) of the subtype classifiers on external testing groups. Model 1 (**A**), Model 2 (**B**), Model 3 (**C**), Ensemble Model (**D**)

Zhang *et al. Journal of Translational Medicine*   (2025) 23:267

Page 8 of 10

Table 4 The performance of Ensemble Model

Model	AUC	Precision	Recall	F1-score
Ensemble Model	0.769	0.800	0.727	0.762

gene status in ovarian cancer patients from WSIs that

had undergone H&E staining in 2022 [18]. However, the model developed by the aforementioned team demonstrated suboptimal performance, with a validation ROC AUC of 0.59, indicating the necessity for further optimization. In contrast, our team has developed a more
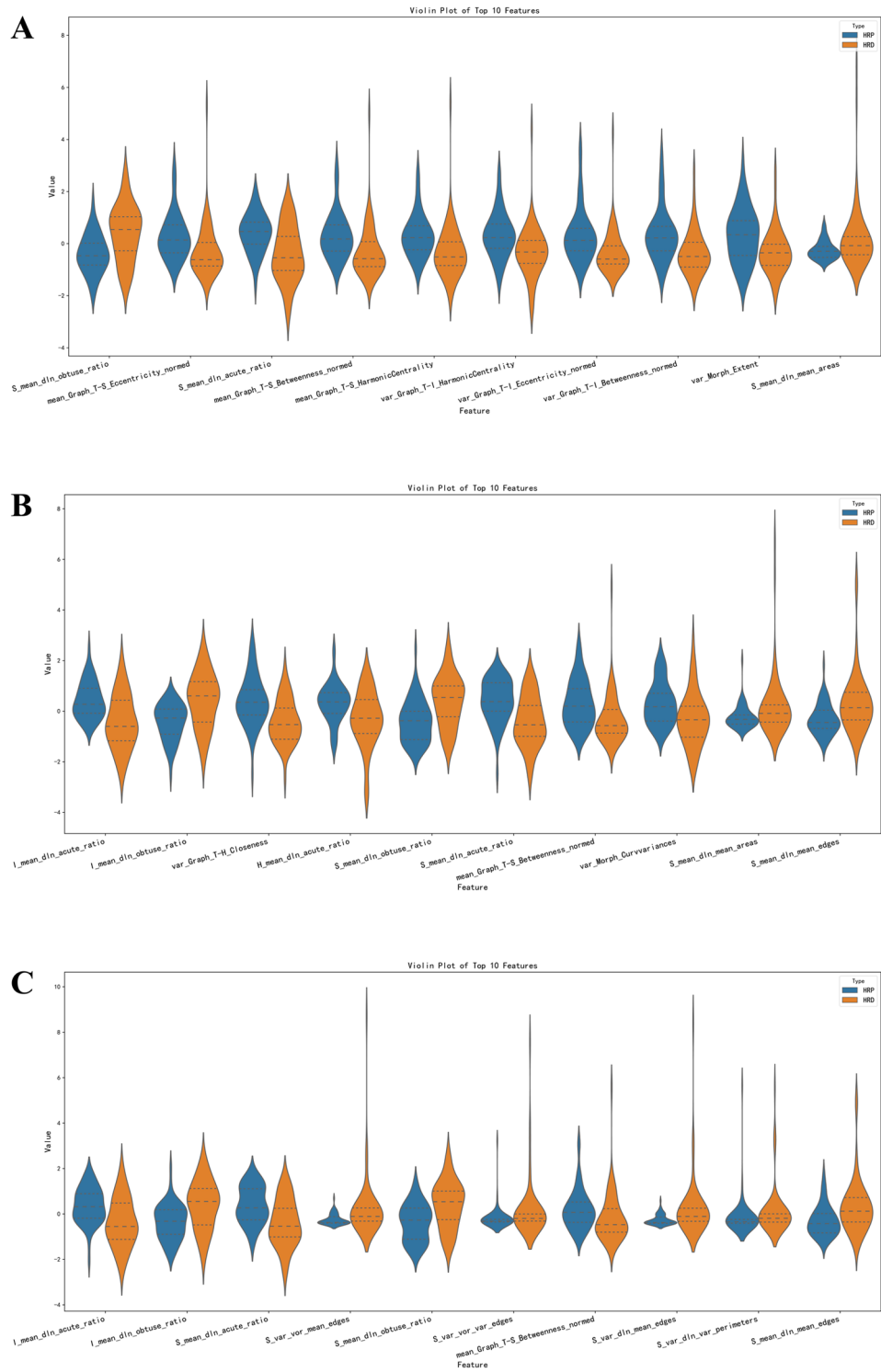

Fig. 4 Discriminative top 10 features of the subtype classifier of ovarian cancer. Model 1 (**A**), Model 2 (**B**), Model 3 (**C**)

excellent model based on WSIs through deep learning approaches. By incorporating nuclear morphological/texture features along with nuclear spatial arrangement and TME features, our model exhibits superior performance in discriminating HRD status in ovarian cancer patients. The determination of HRD status holds significant clinical implications for therapeutic outcome prediction in ovarian cancer and compared to single-gene mutation status assessment, HRD status evaluation offers enhanced potential for patient stratification and optimization of therapeutic decision-making [27, 30, 31].

Afterwards, in 2023, Bourgade's research team proposed a novel deep learning classifier with the ability to predict the BRCA mutation status from Whole Slide Images (WSIs) of ovarian cancer patients, following H&E staining [32]. This approach exhibited significantly improved performance relative to previous methods, highlighting the potential of deep learning in digital pathology to achieve high levels of precision and efficiency in cancer diagnosis. In addition, Wang's team developed two deep learning methods capable of predicting the therapeutic response to Bevacizumab in ovarian cancer and the microsatellite instability status of ovarian cancer patients on WSIs, without the need for pathologists to conduct detailed image annotation [33]. This innovative approach offers significant benefits, including the potential to enhance the efficiency of cancer diagnosis and treatment, while also reducing the workload of pathologists in analyzing WSIs.

In this investigation, we created two models based on digitized H&E-stained histopathological images of ovarian cancer patients. The first model was developed to identify the tumor region in ovarian cancer patients, while the second model was designed to discriminate whether the patient of ovarian cancer with HRD status or not. The features obtained from the WSIs are closely related to various aspects of the image, such as nuclear orientation, shape, and texture. Our study reveals for the first time the most representative pathological features in HRD and HRP ovarian cancer, which may help distinguish the clinicopathological diagnosis in the future.

Notwithstanding the significant findings of this study, some limitations are worth noting. Firstly, the utilization of patient image data obtained from various datasets without controlling for typical confounding variables such as pathological section quality, diverse patient demographics, and distinct treatment measures. These variables can significantly impact both the quality of images and the subsequent feature analysis procedures, thus affecting the overall accuracy and reliability of the results. Consequently, future studies should aim to address these confounding variables to improve the quality of image data and accurately reflect the patient population under investigation. Furthermore, it is imperative

to recognize that the current study is constrained by a relatively limited sample size. To augment the robustness and generalizability of the developed model, it is crucial to expand the patient cohort in subsequent investigations. As our research advances, we are committed to the ongoing refinement and optimization of the constructed models. We have commenced the integration of additional external validation cohorts from diverse geographical regions and healthcare institutions. This expansion will encompass a more comprehensive array of patients and clinical data, including overall survival, status of recurrence, tumor pathological types and so on. Such an approach will facilitate a more thorough evaluation of the model's performance across a wider range of patient demographics and clinical scenarios, thereby enhancing its applicability and reliability in real-world settings.

Conclusion
In summary, our study demonstrates that employing deep learning-based pathological image analysis can effectively perform routine pathological tasks. Our models we constructed enables accurate discrimination between tumor and non-tumor tissues in ovarian cancer as well as the prediction of HRD status for patients with ovarian cancer. In future research, the adoption of deep learning can potentially identify further pathological relevant features and facilitate the construction of more precise models, thus contributing to advancing healthcare and promoting medical knowledge more rapidly.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12967-025-06234-7 .

Supplementary Material 1

Data availability
The following information was supplied regarding data availability: Data in the Memorial Sloan Kettering Cancer Center (MSKCC) is available from the Synapse (Sage Bionetworks) (Accession Code: syn25946117). The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
The studies involving human participants were reviewed and approved by the Ethics Committees and Institutional Review Boards of Zhongda Hospital Southeast University (2024ZDSYLL268-P01). The patients provided their written informed consent to participate in this study.

Conflict of interest
There are no conflicts of interest regarding the publication of this article.

Author details
[1]Department of Obstetrics and Gynaecology, School of Medicine, Zhongda Hospital, Southeast University, Nanjing 210009, China
[2]Institute for AI in Medicine, School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing, China
[3]Department of Pathology, School of Medicine, Zhongda Hospital, Southeast University, Nanjing, China

References

1. Peres LC, Sinha S, Townsend MK, Fridley BL, Karlan BY, Lutgendorf SK, Shinn E, Sood AK, Tworoger SS. Predictors of survival trajectories among women with epithelial ovarian cancer. Gynecol Oncol. 2020;156:459–66.
2. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024;74:12–49.
3. Lalwani N, Prasad SR, Vikram R, Shanbhogue AK, Huettner PC, Fasih N. Histologic, molecular, and cytogenetic features of ovarian cancers: implications for diagnosis and treatment. Radiographics. 2011;31:625–46.
4. Reid BM, Permuth JB, Sellers TA. Epidemiology of ovarian cancer: a review. Cancer Biol Med. 2017;14:9–32.
5. Stewart C, Ralyea C, Lockwood S. Ovarian Cancer: an Integrated Review. Semin Oncol Nurs. 2019;35:151–6.
6. Giornelli GH. Management of relapsed ovarian cancer: a review. Springerplus. 2016;5:1197.
7. Lau CH, Seow KM, Chen KH. The Molecular mechanisms of actions, effects, and clinical implications of PARP inhibitors in epithelial ovarian cancers: a systematic review. Int J Mol Sci 2022;23.
8. Li J, Li Q, Zhang L, Zhang S, Dai Y. Poly-ADP-ribose polymerase (PARP) inhibitors and ovarian function. Biomed Pharmacother. 2023;157:114028.
9. Ledermann JA, Drew Y, Kristeleit RS. Homologous recombination deficiency and ovarian cancer. Eur J Cancer. 2016;60:49–58.
10. Flanigan JS, Silkensen SL, Wolf NG. Global Health Pathology Research: purpose and funding. Clin Lab Med. 2018;38:21–35.
11. Moch H, Blank PR, Dietel M, Elmberger G, Kerr KM, Palacios J, Penault-Llorca F, Rossi G, Szucs TD. Personalized cancer medicine and the future of pathology. Virchows Arch. 2012;460:3–8.
12. Wei JW, Suriawinata AA, Vaickus LJ, Ren B, Liu X, Lisovsky M, Tomita N, Abdollahi B, Kim AS, Snover DC, et al. Evaluation of a Deep Neural Network for Automated Classification of Colorectal Polyps on histopathologic slides. JAMA Netw Open. 2020;3:e203398.
13. Zhang X, Gleber-Netto FO, Wang S, Martins-Chaves RR, Gomez RS, Vigneswaran N, Sarkar A, William WN Jr., Papadimitrakopoulou V, Williams M, et al. Deep learning-based pathology image analysis predicts cancer progression risk in patients with oral leukoplakia. Cancer Med. 2023;12:7508–18.
14. Wang K, Ren Y, Ma L, Fan Y, Yang Z, Yang Q, Shi J, Sun Y. Deep learning-based prediction of treatment prognosis from nasal polyp histology slides. Int Forum Allergy Rhinol. 2023;13:886–98.
15. Zhou G, Zheng J, Chen Z, Hu D, Li S, Zhuang W, He Z, Lin G, Wu B, Zhang W, et al. Clinical significance of tumor-infiltrating lymphocytes investigated using routine H&E slides in small cell lung cancer. Radiat Oncol. 2022;17:127.
16. Lu C, Romo-Bucheli D, Wang X, Janowczyk A, Ganesan S, Gilmore H, Rimm D, Madabhushi A. Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers. Lab Invest. 2018;98:1438–48.
17. Lan X, Guo G, Wang X, Yan Q, Xue R, Li Y, Zhu J, Dong Z, Wang F, Li G, et al. Differentiation and risk stratification of basal cell carcinoma with deep learning on histopathologic images and measuring nuclei and tumor microenvironment features. Skin Res Technol. 2024;30:e13571.
18. Nero C, Boldrini L, Lenkowicz J, Giudice MT, Piermattei A, Inzani F, Pasciuto T, Minucci A, Fagotti A, Zannoni G et al. Deep-learning to predict BRCA mutation and survival from Digital H&E slides of epithelial ovarian Cancer. Int J Mol Sci 2022, 23.
19. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: redesigning skip connections to exploit Multiscale features in image segmentation. IEEE Trans Med Imaging. 2020;39:1856–67.
20. Boehm KM, Aherne EA, Ellenson L, Nikolovski I, Alghamdi M, Vázquez-García I, Zamarin D, Long Roche K, Liu Y, Patel D, et al. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. Nat Cancer. 2022;3:723–33.
21. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, Srinivasan P, Gao J, Chakravarty D, Devlin SM, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med. 2017;23:703–13.
22. Scaglione GL, Pignata S, Pettinato A, Paolillo C, Califano D, Scandurra G, Lombardo V, Di Gaudio F, Pecorino B, Mereu L et al. Homologous recombination Deficiency (HRD) Scoring, by means of two different shallow whole-genome sequencing pipelines (sWGS), in Ovarian Cancer patients: a comparison with myriad MyChoice assay. Int J Mol Sci 2023, 24.
23. Graham S, Vu QD, Raza SEA, Azam A, Tsang YW, Kwak JT, Rajpoot N. Hover-Net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med Image Anal. 2019;58:101563.
24. Pan X, Feng S, Wang Y, Chen J, Lin H, Wang Z, Hou F, Lu C, Chen X, Liu Z, et al. Spatial distance between tumor and lymphocyte can predict the survival of patients with resectable lung adenocarcinoma. Heliyon. 2024;10:e30779.
25. Liu XY, Wu J, Zhou ZH. Exploratory undersampling for class-imbalance learning. IEEE Trans Syst Man Cybern B Cybern. 2009;39:539–50.
26. Kuroki L, Guntupalli SR. Treatment of epithelial ovarian cancer. BMJ. 2020;371:m3773.
27. Ray-Coquard I, Pautier P, Pignata S, Pérol D, González-Martín A, Berger R, Fujiwara K, Vergote I, Colombo N, Mäenpää J, et al. Olaparib plus Bevacizumab as First-Line maintenance in Ovarian Cancer. N Engl J Med. 2019;381:2416–28.
28. Konstantinopoulos PA, Ceccaldi R, Shapiro GI, D'Andrea AD. Homologous recombination Deficiency: exploiting the fundamental vulnerability of Ovarian Cancer. Cancer Discov. 2015;5:1137–54.
29. Barragán-Montero A, Javaid U, Valdés G, Nguyen D, Desbordes P, Macq B, Willems S, Vandewinckele L, Holmström M, Löfman F, et al. Artificial intelligence and machine learning for medical imaging: a technology review. Phys Med. 2021;83:242–56.
30. Ray-Coquard I, Leary A, Pignata S, Cropet C, González-Martín A, Marth C, Nagao S, Vergote I, Colombo N, Mäenpää J, et al. Olaparib plus Bevacizumab first-line maintenance in ovarian cancer: final overall survival results from the PAOLA-1/ENGOT-ov25 trial. Ann Oncol. 2023;34:681–92.
31. DiSilvestro P, Banerjee S, Colombo N, Scambia G, Kim BG, Oaknin A, Friedlander M, Lisyanskaya A, Floquet A, Leary A, et al. Overall survival with maintenance olaparib at a 7-Year Follow-Up in patients with newly diagnosed Advanced Ovarian Cancer and a BRCA mutation: the SOLO1/GOG 3004 Trial. J Clin Oncol. 2023;41:609–17.
32. Bourgade R, Rabilloud N, Perennec T, Pécot T, Garrec C, Guédon AF, Delnatte C, Bézieau S, Lespagnol A, de Tayrac M, et al. Deep learning for detecting BRCA mutations in high-Grade Ovarian Cancer based on an innovative tumor segmentation method from whole slide images. Mod Pathol. 2023;36:100304.
33. Wang CW, Lee YC, Lin YJ, Firdi NP, Muzakky H, Liu TC, Lai PJ, Wang CH, Wang YC, Yu MH, et al. Deep learning can predict Bevacizumab therapeutic effect and microsatellite instability directly from Histology in Epithelial Ovarian Cancer. Lab Invest. 2023;103:100247.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.