Lab Invest 103 (2023) 100247

LABORATORY INVESTIGATION

journal homepage: https://laboratoryinvestigation.org/

Research Article

Deep Learning Can Predict Bevacizumab Therapeutic Effect and Microsatellite Instability Directly from Histology in Epithelial Ovarian Cancer

Ching-Wei Wang[a,b], Yu-Ching Lee[b], Yi-Jia Lin[c,d], Nabila Puspita Firdi[a], Hikam Muzakky[a], Tzu-Chien Liu[a], Po-Jen Lai[a], Chih-Hung Wang[e,f], Yu-Chi Wang[g,h], Mu-Hsien Yu[g,h], Chia-Hua Wu[d], Tai-Kuang Chao[c,d,*]

[a] *Graduate Institute of Biomedical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan;* [b] *Graduate Institute of Applied Science and Technology, National Taiwan University of Science and Technology, Taipei, Taiwan;* [c] *Department of Pathology, Tri-Service General Hospital, Taipei, Taiwan;* [d] *Institute of Pathology and Parasitology, National Defense Medical Center, Taipei, Taiwan;* [e] *Department of Otolaryngology-Head and Neck Surgery, Tri-Service General Hospital, Taipei, Taiwan;* [f] *Department of Otolaryngology-Head and Neck Surgery, National Defense Medical Center, Taipei, Taiwan;* [g] *Department of Gynecology and Obstetrics, Tri-Service General Hospital, Taipei, Taiwan;* [h] *Department of Gynecology and Obstetrics, National Defense Medical Center, Taipei, Taiwan*

ARTICLE INFO

Article history:
Received 24 May 2023
Revised 14 September 2023
Accepted 14 September 2023
Available online 22 September 2023

Keywords:
deep learning
epithelial ovarian cancer
histopathologic images
microsatellite instability
therapeutic outcome

ABSTRACT

Epithelial ovarian cancer (EOC) remains a significant cause of mortality among gynecologic cancers, with the majority of cases being diagnosed at an advanced stage. Before targeted therapies were available, EOC treatment relied largely on debulking surgery and platinum-based chemotherapy. Vascular endothelial growth factors have been identified as inducing tumor angiogenesis. According to several clinical trials, anti—vascular endothelial growth factor-targeted therapy with bevacizumab was effective in all phases of EOC treatment. However, there are currently no biomarkers accessible for regular therapeutic use despite the importance of patient selection. Microsatellite instability (MSI), caused by a deficiency of the DNA mismatch repair system, is a molecular abnormality observed in EOC associated with Lynch syndrome. Recent evidence suggests that angiogenesis and MSI are interconnected. Developing predictive biomarkers, which enable the selection of patients who might benefit from bevacizumab-targeted therapy or immunotherapy, is critical for realizing personalized precision medicine. In this study, we developed 2 improved deep learning methods that eliminate the need for laborious detailed image-wise annotations by pathologists and compared them with 3 state-of-the-art methods to not only predict the efficacy of bevacizumab in patients with EOC using mismatch repair protein immunostained tissue microarrays but also predict MSI status directly from histopathologic images. In prediction of therapeutic outcomes, the 2 proposed methods achieved excellent performance by obtaining the highest mean sensitivity and specificity score using MSH2 or MSH6 markers and outperformed 3 state-of-the-art deep learning methods. Moreover, both statistical analysis results, using Cox proportional hazards model analysis and Kaplan-Meier progression-free survival analysis, confirm that the 2 proposed methods successfully differentiate patients with positive therapeutic effects and lower cancer recurrence rates from patients experiencing disease progression after treatment ($P < .01$). In prediction of MSI status directly from histopathology images, our proposed method also achieved a decent performance in terms of mean sensitivity and specificity score even for imbalanced data sets for

* Corresponding author.
E-mail address: chaotai.kuang@msa.hinet.net (T.-K. Chao).

0023-6837/$ - see front matter
https://doi.org/10.1016/j.labinv.2023.100247

both internal validation using tissue microarrays from the local hospital and external validation using whole section slides from The Cancer Genome Atlas archive.

Introduction

Epithelial ovarian cancer (EOC) remains the deadliest gynecologic cancer each year.[1] The various histologic subgroups of EOC include high-grade serous, endometrioid, clear cell, mucinous, and low-grade serous carcinoma.[2,3] Most EOCs are type II high-grade serous carcinoma with p53 and BRCA mutations. In contrast, type I low-grade serous carcinoma is characterized by *KRAS*, *BRAF*, *PTEN*, *PIK3CA*, *CTNNB1*, *ARID1A*, and *PPP2R1A* mutations.[4] The standard treatment of EOC includes surgery and platinum-based chemotherapy. Despite advances in treatment modalities, such as extensive surgical cytoreduction and newer adjuvant treatment, the overall survival (OS) remains low, 40% for stage III and 20% for stage IV.[5] Various trials have evaluated targeted therapeutic agents and immunotherapy.[6] Angiogenesis is a mechanism of significant EOC development and progression and plays an essential role in the development of both the primary tumor and its metastases.[7,8] Vascular endothelial growth factor (VEGF), one of the major pathways involved in tumor angiogenesis, is often overexpressed in EOC. Therefore, it is also an attractive target for therapy.[9] EOC's most widely approved anti-angiogenic therapy is bevacizumab, a humanized monoclonal antibody against VEGF.[10,11] EOC exhibits several characteristic features, including elevated levels of VEGF and prompt response to the reduction of ascites through bevacizumab.[12] However, tumors in these patients display inherent resistance to VEGF or vascular endothelial growth factor receptor (VEGFR) inhibition. This inherent resistance can occur by activating alternative angiogenic pathways that promote tumor angiogenesis via mechanisms independent of VEGF.[13] Careful patient selection is required for bevacizumab therapy. However, the criteria for patient selection are still evolving.[14] There is no evidence demonstrating that serum angiogenesis biomarkers, including VEGF and VEGFR, can be used as predictors of survival.[15-18] In order to select patients who are more likely to benefit from this therapy, the identification of predictive biomarkers for bevacizumab response is crucial.

While the use of poly(adenosine diphosphate [ADP]-ribose) polymerase inhibitors in EOC cases with BRCA1/BRCA2 mutations is well-established, it is important to note that EOC can also be observed in tumors associated with the hereditary nonpolyposis colorectal cancer (Lynch syndrome). These tumors are characterized by germline mutations in the mismatch repair (MMR) genes with microsatellite instability (MSI).[12,19-21] MSI occurs in 3% to 17% of sporadic EOC.[22] Previous studies have shown that the metastases are more common in those that are MSI-high (MSI-H) tumors with a higher angiogenic capacity.[8,23] MLH1, MSH2, MSH6, and PMS2 are the MMR genes that are most frequently affected.[24] Deficient mismatch repair (dMMR) cancer is a well-defined subgroup with distinct clinicopathologic features. It exhibits a highly immune-infiltrative tumor microenvironment, and immune checkpoint programmed death receptor ligand (PDL1) expression potentially responds to immune checkpoint blockade.[25-27] A recent study showed that MSI analysis is effective as a predictive biomarker for the effect of immune checkpoint inhibitors (ICIs).[28] VEGF-mediated angiogenesis also increases PDL1 expression in T cells.[29] Tumor-infiltrating lymphocytes play a crucial role in exerting tumor-killing activity.[30] A growing body of evidence suggests that appropriate antiangiogenic administration can shift the tumor immune environment from an immunosuppressive to an immune-supportive status.[31,32] Lynch syndrome cancer is a multitumor syndrome that usually occurs in the colorectum and the ovary, which may be optimal candidates for immune checkpoint–based therapies. A rapidly growing number of clinical trials have incorporated anti-PDL1 immunotherapy in combination with other therapies, such as chemotherapy, poly(adenosine diphosphate [ADP]-ribose) polymerase inhibitor, and anti-vascular agents, into various treatment options for advanced EOC[33]; MMR status together with immune checkpoint proteins may serve as a guide for designing therapeutic strategies that effectively target EOCs.[34]

In this study, we built 2 improved deep learning (DL)–based methods without any locally annotated regions provided by the pathologist and compared them with 3 state-of-the-art DL methods not only for predicting the therapeutic efficacy of bevacizumab in patients with EOC using MMR protein immunostained tissue microarrays (TMAs) on whole slide images (WSIs) but also for predicting MSI status directly from histopathology images. The experimental results for predicting the efficacy of bevacizumab show that the 2 proposed methods using the MSH2 protein expression achieved excellent performance, obtaining a remarkably high mean of sensitivity and specificity (MSS) (100%), accuracy (100%), sensitivity (100%), and specificity (100%). On the other hand, the 2 proposed models based on MSH6 also obtained outstanding performance, achieving a notably high MSS (98%), accuracy (98%), sensitivity (100%), and specificity (96%). Moreover, both Cox proportional hazards model analysis and Kaplan-Meier (K-M) progression-free survival (PFS) analysis confirm that the 2 proposed methods using either MSH2 or MSH6 successfully differentiate patients with positive therapeutic effects and lower cancer recurrence rates from patients experiencing disease progression after treatment ($P < .01$). For predicting MSI-MLH1 status from histopathologic slides, one of the proposed methods achieved decent performance, obtaining a good MSS (75%), accuracy (77%), sensitivity (84%), and specificity (67%), even for the challenging imbalanced data set (70.98% intact vs 29.02% loss).

Material and Methods

Material

In this study, the proposed method and all benchmarked approaches were evaluated with 2 data sets, including a TMA data set from the tissue bank at the Department of Pathology, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, and an external validation set from The Cancer Genome Atlas (TCGA). For the first internal validation set, a total of 720 patient tissue core samples with 4 MSI markers—MLH1, MSH2, MSH6, and PMS2—of patients with EOC and those with peritoneal serous papillary carcinoma (PSPC) with associated clinical data were collected. The collection of digitized WSIs of TMAs representing samples of patients with PSPC or EOC, comprising 720 female

patient specimen tissue cores, will allow the development of therapeutic prediction using artificial intelligence (AI) platforms. Regarding the rules for including and excluding data, patients were qualified for the retrospective study if they had EOC or PSPC and were receiving (1) concurrent bevacizumab therapy (upfront debulking surgery, bevacizumab, and chemotherapy as first-line systemic therapy), (2) second-line bevacizumab therapy after recurrence (postupfront debulking surgery, adjuvant chemotherapy with tumor recurrence, and bevacizumab with chemotherapy), and (3) maintenance bevacizumab therapy (concurrent therapy followed by maintenance bevacizumab therapy and second-line therapy followed by maintenance bevacizumab therapy after recurrence. On the contrary, patients are ineligible if they had borderline malignant tumors. Papillary serous carcinoma (n = 444), clear cell carcinoma (n = 69), unclassified carcinoma (n = 69), endometrioid carcinoma (n = 39), mucinous carcinoma (n = 10), and PSPC (n = 89) of 12 TMAs represent the tumors that comprised the data, constructed from the tissue bank of the Department of Pathology, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan. Ethical approvals were obtained from the research ethics committee of the Tri-Service General Hospital (TSGHIRB No.1-107-05-171 and No. B202005070). The patient baseline information of the MSI and treatment status was retrospectively retrieved from March 2013 to January 2021. Regarding demographics, the patients' ages ranged from 23 to 79 years (mean, 58.75 years). Regarding recurrence time in the month after drug initiation, the time ranged from 1 to 55 months (mean, 11.92 months). Then, concerning the treatment effectiveness class distribution, a balanced set of data was guaranteed, with 412 tissue cores (57.22%) associated with effective bevacizumab treatment outcomes, whereas 308 cores (42.78%) were associated with invalid treatment outcomes. Lastly, for the MSI status class distribution, 511 tissue cores (70.98%) were associated with intact, whereas 209 cores (29.02%) were associated with loss. For quantitative evaluation, the whole data set was randomly sampled and split into 2 subsets for training and testing, specifically 472 cores (two-thirds) for training and 248 cores (one-third) for testing. Table 1 presents the baseline characteristics of the database constructed in this study.

The Gynecologic Oncology Center's data administrators recorded patients' clinicopathologic characteristics, such as pretreatment and posttreatment serum CA125 concentrations, histologic subtypes, and recurrence status. A tumor was considered resistant to bevacizumab therapy if there had been demonstrable tumor regrowth (based on computed tomography or positron emission tomography imaging) or if there had been an increase in the serum CA125 concentration during or within 6 months of bevacizumab treatment. Furthermore, they were considered sensitive to bevacizumab if there was no measurable tumor regrowth (based on computed tomography or positron emission tomography imaging) or a low serum CA-125 concentration during or within 6 months of bevacizumab treatment.

The tissues from bevacizumab-treated patients with EOC or PSPC were embedded in paraffin wax. Two pathologists (T.K.C., Y.J.L.) screened the histologic sections and selected regions of representative tumor cells. One tissue core was then taken from each representative tumor sample and placed in a new recipient paraffin block. As previously described, a mechanical device was used to construct tissue arrays consisting of 2-mm diameter tissue cores, which can generate hundreds of tissue sections for immunohistochemical staining.[35] The 720 sample tissue cores include papillary serous carcinoma (n = 444), clear cell carcinoma (n = 69), unclassified carcinoma (n = 69), endometrioid carcinoma (n = 39), mucinous carcinoma (n = 10), and PSPC (n = 89) of 12 TMAs. In short, antigen retrieval was performed with 0.01-M sodium citrate buffer (pH 6.0) and endogenous peroxidase blocked with hydrogen peroxide. Slides were incubated for 1 hour at room temperature with the antibodies anti-MSH2 (ready-to-use; Cat# 760-5092, Ventana Medical Systems), anti-MSH6 (ready-to-use; Cat# 760-5093; Ventana Medical Systems), anti-PMS2 (ready-to-use; Cat# 760-5094; Ventana Medical Systems), and anti-MLH1 (ready-to-use; Cat# 760-5091; Ventana Medical Systems) in phosphate-buffered saline. The sections were washed and incubated with horseradish peroxidase—labeled immunoglobulin (Dako). The peroxidase activity was visualized of anti-MSH2, anti-MSH6, anti-PMS2, and anti-MLH1 in phosphate-buffered saline. It was analyzed using a diaminobenzidine solution, and then the slides were further analyzed using DL-based AI. The WSIs were acquired with a digital slide scanner (Aperio) with a 20× objective lens.

For the second external validation set, hematoxylin and eosin (H&E)—stained pathologic pictures of 440 WSIs from patients

Table 1
Baseline characteristics of the first tissue microarray data set

Characteristics	Values
Tissue core	720
Age (y)	58.75 (23.00-79.00)
BMI (kg/m^2)	23.55 (16.20-38.70)
Diagnosis	
Papillary serous carcinoma	464 (64.44)
Peritoneal serous papillary carcinoma	72 (10.00)
Clear cell carcinoma	76 (10.56)
Unclassified carcinoma	68 (9.45)
Endometrioid carcinoma	32 (4.44)
Mucinous carcinoma	8 (1.11)
FIGO stage	
I	68 (9.45)
II	52 (7.22)
III	412 (57.22)
IV	188 (26.11)
TMN stage	
pT1N0M0	68 (9.44)
pT1N1M0	20 (2.78)
pT2N0M0	52 (7.22)
pT2N1M0	8 (1.11)
pT3N0M0	120 (16.67)
pT3N0M1	12 (1.67)
pT3N1M0	264 (36.67)
pT3N1M1	176 (24.44)
Surgery	
Optimal debulking	292 (40.56)
CRS + HIPEC	232 (32.22)
Suboptimal debulking	196 (27.22)
Bevacizumab treatment method	
Concurrent bevacizumab therapy	224 (31.11)
Concurrent + maintenance bevacizumab therapy	88 (12.22)
Second-line bevacizumab therapy after recurrence	344 (47.78)
Second-line + maintenance bevacizumab therapy	64 (8.89)
Recurrence time after drug (mo)	11.92 (1-55)
Treatment effectiveness	
Effective	412 (57.22)
Invalid	308 (42.78)
MSI status	
Intact	511 (70.98)
Loss	209 (29.02)

Values are reported as mean (range) or n (%).
BMI, body mass index; CRS, cytoreductive surgery; FIGO, International Federation of Gynaecology and Obstetrics; HIPEC, hyperthermic intraperitoneal chemotherapy; MSI, microsatellite instability.

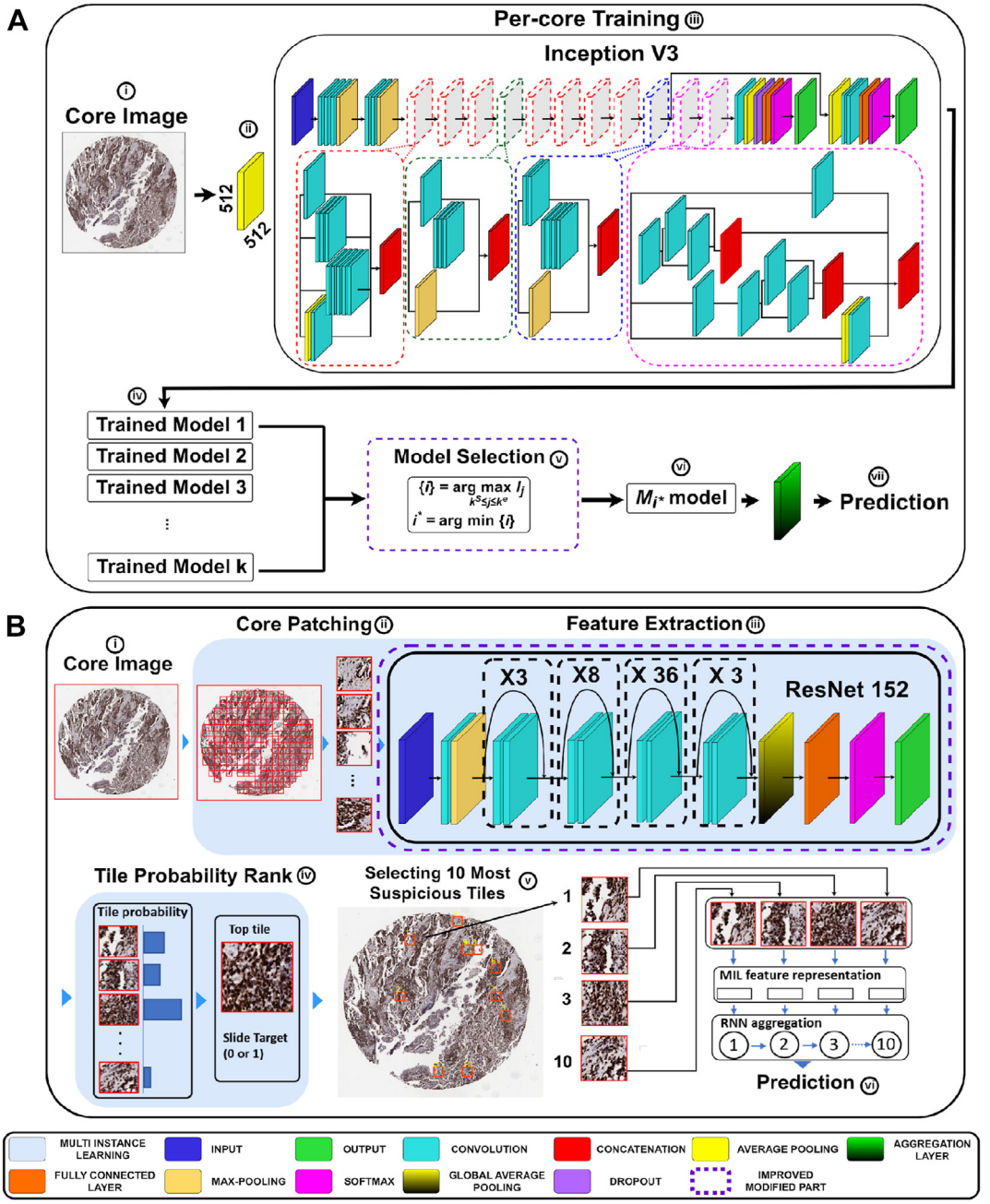

Figure 1.
Workflow of proposed deep learning networks: (A) Improved_InceptionV3_MS and (B) Improved_MIL_RNN. MIL, multiple instance learning; RNN, recurrent neural network.

with EOC were collected from TCGA data sets to develop our algorithms. All images and MSI status by next-generation sequencing (NGS) (MSI-H [MSI scores $\geq$ 10]: n = 28; MSI-low/microsatellite stable [MSI scores < 10]: n = 412) from the TCGA study can be downloaded from public repositories at the National Institutes of Health (https://portal.gdc.cancer.gov/).

Methods

In this study, we developed 2 improved models, namely, Improved InceptionV3 + Model Selection (Improved_InceptionV3_MS) and Improved Multiple Instance Learning (MIL) + Recurrent Neural Network (RNN) (Improved_MIL_RNN), for the prediction of therapeutic outcomes of bevacizumab and MSI status using MMR-immunostained TMA. For evaluation of the model performance, we adopted MSS as an evaluation criterion. By calculating the mean of both indices, the MSS reconciled the commonly observed inverse relationship between the specificity and sensitivity indices, offering a balanced assessment metric that considers both aspects of the model's performance. Sensitivity and specificity, which are essential measures in the medical field,[36-39] were utilized to evaluate the model predictions. Sensitivity refers to the ability of a test or instrument to yield a positive result for a subject with the disease, while specificity is the ability of the test or instrument to provide normal or negative results for a person

Table 2
Evaluation in therapeutic outcomes classification: mean of sensitivity and specificity

Method for all markers	Marker	Accuracy	Sensitivity	Specificity	MSS	All marker MSS ranking
InceptionV3[40]	MLH1	0.60	0.91	0.19	0.55	17
Improved_InceptionV3_MS	MLH1	0.61	0.80	0.37	0.59	15
NASNetLarge[43]	MLH1	0.55	0.20	1.00	0.6	14
MIL_RNN[41]	MLH1	0.71	0.69	0.74	0.71	12
Improved_MIL_RNN	MLH1	0.90	0.89	0.93	0.91	5
InceptionV3[40]	MSH2	0.45	0.26	0.70	0.48	19
Improved_InceptionV3_MS	MSH2	1.00	1.00	1.00	1.00	1[a]
NASNetLarge[43]	MSH2	0.87	0.89	0.85	0.87	7
MIL_RNN[41]	MSH2	0.87	0.91	0.81	0.86	8
Improved_MIL_RNN	MSH2	1.00	1.00	1.00	1.00	1[a]
InceptionV3[40]	MSH6	0.79	0.71	0.89	0.8	10
Improved_InceptionV3_MS	MSH6	0.98	1.00	0.96	0.98	2[a]
NASNetLarge[43]	MSH6	0.95	0.97	0.93	0.95	4
MIL_RNN[41]	MSH6	0.89	0.86	0.93	0.89	6
Improved_MIL_RNN	MSH6	0.97	0.94	1.00	0.97	3[a]
InceptionV3[40]	PMS2	0.42	0.00	0.96	0.48	19
Improved_InceptionV3_MS	PMS2	0.73	0.74	0.70	0.72	11
NASNetLarge[43]	PMS2	0.82	0.94	0.67	0.8	10
MIL_RNN[41]	PMS2	0.73	0.77	0.67	0.72	11
Improved_MIL_RNN	PMS2	0.81	0.07	0.93	0.82	9
Method for ensemble marker	**Ensemble marker**	**Accuracy**	**Sensitivity**	**Specificity**	**MSS**	**Ensemble marker MSS ranking**
InceptionV3[40]	MSH2 + MSH6	0.55	0.43	0.70	0.57	7
Improved_InceptionV3_MS	MSH2 + MSH6	0.98	1.00	0.96	0.98	2[a]
NASNetLarge[43]	MSH2 + MSH6	0.97	0.97	0.96	0.97	3
MIL_RNN[41]	MSH2 + MSH6	0.89	0.89	0.89	0.89	5
Improved_MIL_RNN	MSH2 + MSH6	1.00	1.00	1.00	1.00	1[a]
InceptionV3[40]	All	0.60	0.31	0.96	0.64	6
Improved_InceptionV3_MS	All	0.98	1.00	0.96	0.98	2[a]
NASNetLarge[43]	All	0.94	0.91	0.96	0.94	4
MIL_RNN[41]	All	0.89	0.89	0.89	0.89	5
Improved_MIL_RNN	All	1.00	1.00	1.00	1.00	1[a]

MSS, mean of sensitivity and specificity.
[a] Top 3 ranked models in all marker mean of sensitivity and specificity ranking, and top 2 ranked model in ensemble marker mean of sensitivity and specificity ranking.

without the disease.[36-39] High-sensitivity tests will lead to positive findings for patients with a disease, whereas high-specificity tests will show patients without a finding having no disease.[36,37] A key weakness in current AI technology to consider is limited specificity, where AI algorithms for cancer detection can sometimes be overly sensitive, leading to false positives. This can result in unnecessary biopsies and other invasive procedures for patients as well as increased health care costs. Both measures evaluate diagnostic tests and models and help guide medical decisions.[36]

Our study adopted the MSS metric as an evaluation criterion for the model, which provides a balanced evaluation by considering both the sensitivity and specificity of the model's predictions. By calculating the mean of both indices, the MSS reconciled the commonly observed inverse relationship between the specificity and sensitivity indices, offering a balanced assessment metric that considers both aspects of the model's performance.

Improved InceptionV3 + Model Selection

In 2018, the InceptionV3-based model was utilized by Coudray et al,[40] which was a fully supervised learning that requires patch-wise data annotation for classifying the images into lung adenocarcinoma, lung squamous cell carcinoma, or normal lung tissue on the WSIs from TCGA. This system attempted to solve the problem of limited data in computational pathology by dividing each instance into multiple patch units to increase the number of training samples and assigning the same label to every patch of training. The abovementioned properties are the main reason that the work by Coudray et al[40] was adopted as the baseline method to build our first improved model for this study.

Here, we proposed the Improved_InceptionV3_MS framework by adding a model selection technique with an early stop mechanism into the learning framework to select the best model during training. Figure 1A shows the workflow of the proposed model. Initially, each core image is observed through the average pooling layer (Fig. 1A, i), resulting in a new 512-pixel × 512-pixel image (Fig. 1A, ii). Then, the training set is directly fed into the InceptionV3 network to fully train (Fig. 1A, iii) the model for predicting the efficacy of bevacizumab and predicting the MSI status and result in trained models (Fig. 1A, iv). Then, the model selection approach (Fig. 1A, v) computed the loss value ω_k and MSS scores l_k of the trained models on the training set S through training iteration k. In order to locate models with loss values lower than ω_s, ω_k is then continually evaluated for every δ iteration. Starting time k^S is utilized as the starting point in the model selection process and provided as an input parameter. The model then seeks the highest MSS score between k^S and k^e, where k^e is the maximum number of training iterations specified as an input parameter. MSS is a metrics evaluation, which provides a balanced evaluation by considering

Table 3
Evaluation in therapeutic outcomes classification of serous carcinoma

Method for all markers	Marker	Accuracy	Sensitivity	Specificity	MSS	All marker MSS ranking
InceptionV3[40]	MLH1	0.56	0.77	0.35	0.56	12
Improved_InceptionV3_MS	MLH1	0.53	0.91	0.17	0.54	13
NASNetLarge[43]	MLH1	0.58	0.14	1.00	0.57	11
MIL_RNN[41]	MLH1	0.69	0.59	0.78	0.69	10
Improved_MIL_RNN	MLH1	0.89	0.86	0.91	0.89	4
InceptionV3[40]	MSH2	0.47	0.23	0.70	0.46	15
Improved_InceptionV3_MS	MSH2	1.00	1.00	1.00	1.00	1[a]
NASNetLarge[43]	MSH2	0.84	0.86	0.83	0.85	6
MIL_RNN[41]	MSH2	0.87	0.86	0.87	0.87	5
Improved_MIL_RNN	MSH2	1.00	1.00	1.00	1.00	1[a]
InceptionV3[40]	MSH6	0.82	0.73	0.91	0.82	8
Improved_InceptionV3_MS	MSH6	0.98	1.00	0.96	0.98	2[a]
NASNetLarge[43]	MSH6	0.96	0.96	0.96	0.96	3
MIL_RNN[41]	MSH6	0.87	0.82	0.91	0.87	5
Improved_MIL_RNN	MSH6	0.98	0.96	1.00	0.98	2[a]
InceptionV3[40]	PMS2	0.49	0.00	0.96	0.48	14
Improved_InceptionV3_MS	PMS2	0.73	0.82	0.65	0.74	9
NASNetLarge[43]	PMS2	0.84	0.96	0.74	0.85	6
MIL_RNN[41]	PMS2	0.69	0.64	0.74	0.69	10
Improved_MIL_RNN	PMS2	0.84	0.73	0.96	0.84	7
Method for ensemble marker	**Ensemble marker**	**Accuracy**	**Sensitivity**	**Specificity**	**MSS**	**Ensemble marker MSS ranking**
InceptionV3[40]	MSH2 + MSH6	0.56	0.41	0.70	0.55	7
Improved_InceptionV3_MS	MSH2 + MSH6	0.98	1.00	0.96	0.98	2[a]
NASNetLarge[43]	MSH2 + MSH6	0.96	0.96	0.96	0.96	3
MIL_RNN[41]	MSH2 + MSH6	0.87	0.82	0.91	0.87	5
Improved_MIL_RNN	MSH2 + MSH6	1.00	1.00	1.00	1.00	1[a]
InceptionV3[40]	All	0.64	0.32	0.96	0.64	6
Improved_InceptionV3_MS	All	0.98	1.00	0.96	0.98	2[a]
NASNetLarge[43]	All	0.93	0.91	0.96	0.93	4
MIL_RNN[41]	All	0.87	0.82	0.91	0.87	5
Improved_MIL_RNN	All	1.00	1.00	1.00	1.00	1[a]

MSS, mean of sensitivity and specificity.
[a] Top 2 ranked models in all marker MSS ranking and ensemble marker MSS ranking.

both sensitivity and specificity of the model's prediction. By incorporating the MSS into the early stopping mechanism, the framework aims to optimize the model's performance on both sensitivity and specificity, leading to a more robust and effective model. Subsequently, we chose the M_i* model (Fig. 1A, vi) with the highest MSS score on the training set. We picked the model with the lowest training iterations if multiple models had the same highest score to generate the final model prediction by conducting per-core inference (Fig. 1A, vii). (k^S and δ are input parameters and are empirically defined as 0.0001 and 100, respectively [Equations 1 and 2]).

$$\{i\} = arg \max_{k^S \le j \le k^e} l_j \quad (1)$$

$$i^* = arg \min_i \{i\} \quad (2)$$

This system employs inception blocks composed of various convolutions with varying kernel sizes and a maximum pooling layer. The first 5 convolution elements are merged with 2 maximum pooling processes, and then, 11 layers of inception modules are added. The structure concludes with a fully connected layer, followed by a SoftMax layer. The InceptionV3 network was completely trained based on MMR-immunostained TMA data sets. The loss function was specified as the cross-entropy between prediction and true class labels. We utilized RMSProp refinement for training the weights using a learning rate of 0.1, weight decay of 0.9, momentum of 0.9, and ε of 1.0. Instead of only optimizing the weights of the fully connected layer, we additionally adjusted the parameters of earlier layers, including all of the convolution filters of all layers.

Improved Multiple Instance Learning + Recurrent Neural Network

In 2019, Campanella et al[41] presented MIL_RNN for prostate cancer, basal cell carcinoma, and breast cancer metastases classification. The main advantage of this weakly supervised approach is that it does not require image-wise annotation but only requires slide labels. The MIL approach trains a DL model with rich patch-wise features, aggregates the information representations across WSIs, selects the 10 most suspicious patches in each slide, and then sequentially feeds them to the RNN pooling–based mechanisms to predict a conclusive slide-level classification by the highest patch probability. The aforementioned key characteristics are the main justification for using the work by Campanella et al[41] as a base approach to build the second enhanced version for this research.

Here, we proposed the Improved_MIL_RNN model by replacing the backbone network with a deeper network, ResNet152,[42] as a feature extractor on our modified network to transform the only foreground patches from core images into sets of complex and low-dimensional feature representations. We trained the ResNet152 architecture using MIL to classify patches using our

Table 4
Multivariate Cox proportional hazards analysis on the clinical factors and top-ranked model prediction with cancer recurrence: a)Improved_InceptionV3_MS (MSH2), b) Improved_MIL_RNN (MSH2) with a and b as the first rank, c) Improved_InceptionV3_MS (MSH6) as the second rank, d) Improved_MIL_RNN (MSH6) as the third rank based on Table 2

a).	HR (95% CI)	*P* value	b).	HR (95% CI)	*P* value
Age	0.990 (0.942-1.040)	.680	Age	0.990 (0.942-1.040)	.680
BMI	0.978 (0.884-1.081)	.657	BMI	0.978 (0.884-1.081)	.657
BEV	0.978 (0.887-1.079)	.655	BEV	0.978 (0.887-1.079)	.655
FIGO	0.222 (0.014-3.566)	.288	FIGO	0.222 (0.014-3.566)	.288
TNM stage	0.859 (0.567-1.301)	.472	TNM stage	0.859 (0.567-1.301)	.472
Histology	0.710 (0.249-2.028)	.522	Histology	0.710 (0.249-2.028)	.522
Surgery			Surgery		
CRS + HIPEC	1.000 (reference)	Reference	CRS + HIPEC	1.000 (reference)	Reference
Optimal	0.979 (0.296-3.239)	.972	Optimal	0.979 (0.296-3.239)	.972
Suboptimal	0.913 (0.310-2.693)	.869	Suboptimal	0.913 (0.310-2.693)	.869
Therapy			Therapy		
Concurrent therapy	1.000 (reference)	Reference	Concurrent therapy	1.000 (reference)	Reference
Second-line therapy	0.465 (0.125-1.730)	.253	Second-line therapy	0.465 (0.125-1.730)	.253
Maintanance therapy	0.116 (0.018-0.763)	.025	Maintanance therapy	0.116 (0.018-0.763)	.025
DL model prediction (all marker MSS ranking: 1)			DL model prediction (all marker MSS ranking: 1)		
Improved_InceptionV3_MS (MSH2)	0.146 (0.042-0.509)	.003[a]	Improved_MIL_RNN (MSH2)	0.146 (0.042-0.509)	.003[a]
c).	HR (95% CI)	*P* value	d).	HR (95% CI)	*P* value
Age	0.984 (0.938-1.032)	.509	Age	0.988 (0.945-1.032)	.582
BMI	0.983 (0.891-1.083)	.725	BMI	0.971 (0.879-1.071)	.555
BEV	0.991 (0.900-1.090)	.848	BEV	0.998 (0.910-1.095)	.967
FIGO	0.314 (0.020-4.961)	.411	FIGO	0.258 (0.020-3.366)	.301
TNM stage	0.917 (0.605-1.390)	.682	TNM stage	0.943 (0.645-1.377)	.760
Histology	0.660 (0.232-1.880)	.437	Histology	0.509 (0.185-1.401)	.191
Surgery			Surgery		
CRS + HIPEC	1.000 (reference)	Reference	CRS + HIPEC	1.000 (reference)	Reference
Optimal	0.828 (0.254-2.703)	.755	Optimal	0.720 (0.229-2.257)	.573
Suboptimal	0.798 (0.271-2.344)	.681	Suboptimal	0.808 (0.274-2.380)	.699
Therapy			Therapy		
Concurrent therapy	1.000 (reference)	Reference	Concurrent therapy	1.000 (reference)	Reference
Second-line therapy	0.397 (0.109-1.444)	.161	Second-line therapy	0.401 (0.118-1.361)	.143
Maintanance therapy	0.100 (0.015-0.666)	.017	Maintanance therapy	0.063 (0.010-0.406)	.004
DL model prediction (all marker MSS ranking: 2)			DL model prediction (all marker MSS ranking: 3)		
Improved_InceptionV3_MS (MSH6)	0.209 (0.064-0.682)	.01[b]	Improved_MIL_RNN (MSH6)	0.144 (0.043-0.481)	.002[a]

BEV, bevacizumab; BMI, body mass index; CRS, cytoreductive surgery; DL, deep learning; FIGO, International Federation of Gynaecology and Obstetrics; HIPEC, hyperthermic intraperitoneal chemotherapy; HR, hazard ratio; MSS, mean of sensitivity and specificity.
[a] $P < .01$.
[b] $P < .05$.

MMR-immunostained TMA training data sets. This architecture utilizes residual functions made of shortcut connections that perform residual mapping and accumulate their results in the outputs of the stacked layers. Resnet152 consists of 1 maximum pooling layer, followed by 50 stacks of residual blocks (150 convolutional layers), and the final layers are constructed by an average pool layer, a fully connected layer, and a SoftMax output layer. A core image is predicted positive at test time if at least 1 patch inside that image is predicted positive. This image-level aggregation, known as maximum pooling, is derived from the common multiple instance assumption.

Figure 1B shows the overview process of the proposed model. To begin with, the core image (Fig. 1B, i) is patched into smaller patches (Fig. 1B, ii). After that, we utilized ResNet152 as a feature extractor to generate feature representations in a low-dimensional shape (Fig. 1B, iii). The MIL training technique involves thorough inference to rank the patches according to their probability of being positive and continuing to learn the top-ranked patches for every core image (Fig. 1B, iv). Then, the RNN technique was used for aggregating the 10 most suspicious patches in each core image and finally generating the final core-level classification (Fig. 1B, v, vi).

The model f_θ is defined as a function with parameter θ that maps input patches $b_{i,j}$ for both the positive and negative classes. Considering the bags B, we then generated a set of vectors $O = \{o_i: i = 1, 2,..., n\}$, 1 for every core image s_i, such that each hold the probabilities of the class positive for each patch bag B_{si}. The patch on every core image that has the highest probability of being positive, $k = argmax(o_i)$, is identified by its index k_i, which we then acquire. The top-ranked patch within bag B_{si} is $b_{i,k}$ for $K = 1$. The network output $\tilde{y_i} = f_\theta(b_{i,k})$ is then compared to y_i, the desired result of core image s_i, using the cross-entropy loss l, as shown in Equation 3. In a similar way, if $K > 1$, every chosen patch from a core image has the same objective y_i, and the loss may be determined for each of the K patches as described:

$$l = w_1[y_i\, log[\tilde{y_i}]] - w_0[(1 - y_i) log[1 - \tilde{y_i}]] \quad (3)$$

Assuming an imbalanced distribution of classes, the weights w_0 and w_1, denoting positive and negative classes, respectively, could be employed to highlight underrepresented cases. The overall loss is the calculated average of the losses throughout the course of a minibatch. The total loss mechanism is reduced by

Table 5
Multivariate Cox proportional hazards analysis on the clinical factors and top-ranked model prediction in ensemble marker with cancer recurrence: a) Improved_InceptionV3_MS (Ensemble MSH2 + MSH6), b) Improved_MIL_RNN (Ensemble MSH2 + MSH6), c) Improved_InceptionV3_MS (Ensemble All Markers), d) Improved_MIL_RNN (Ensemble All Markers), with b and d as the first ranking, a and c as the second ranking in ensemble marker mean of sensitivity and specificity ranking based on Table 2

a).	HR (95% CI)	*P* value
Age	0.984 (0.938-1.032)	.509
BMI	0.983 (0.891-1.083)	.725
BEV	0.991 (0.900-1.090)	.848
FIGO	0.314 (0.020-4.961)	.411
TNM stage	0.917 (0.605-1.390)	.682
Histology	0.660 (0.232-1.880)	.437
Surgery		
CRS + HIPEC	1.000 (reference)	Reference
Optimal	0.828 (0.254-2.703)	.755
Suboptimal	0.798 (0.271-2.344)	.681
Therapy		
Concurrent therapy	1.000 (reference)	Reference
Second-line therapy	0.397 (0.109-1.444)	.161
Maintanance therapy	0.100 (0.015-0.666)	.017
DL model prediction (Ensemble marker MSS ranking: 2)		
Improved_InceptionV3_MS (ensemble MSH2 + MSH6)	0.209 (0.064-0.682)	.01[a]

b).	HR (95% CI)	*P* value
Age	0.990 (0.942-1.040)	.680
BMI	0.978 (0.884-1.081)	.657
BEV	0.978 (0.887-1.079)	.655
FIGO	0.222 (0.014-3.566)	.288
TNM stage	0.859 (0.567-1.301)	.472
Histology	0.710 (0.249-2.028)	.522
Surgery		
CRS + HIPEC	1.000 (reference)	Reference
Optimal	0.979 (0.296-3.239)	.972
Suboptimal	0.913 (0.310-2.693)	.869
Therapy		
Concurrent therapy	1.000 (reference)	Reference
Second-line therapy	0.465 (0.125-1.730)	.253
Maintanance therapy	0.116 (0.018-0.763)	.025
DL model prediction (Ensemble marker MSS ranking: 1)		
Improved_MIL_RNN (ensemble MSH2 + MSH6)	0.146 (0.042-0.509)	.003[b]

c).	HR (95% CI)	*P* value
Age	0.984 (0.938-1.032)	.509
BMI	0.983 (0.891-1.083)	.725
BEV	0.991 (0.900-1.090)	.848
FIGO	0.314 (0.020-4.961)	.411
TNM stage	0.917 (0.605-1.390)	.682
Histology	0.660 (0.232-1.880)	.437
Surgery		
CRS + HIPEC	1.000 (reference)	Reference
Optimal	0.828 (0.254-2.703)	.755
Suboptimal	0.798 (0.271-2.344)	.681
Therapy		
Concurrent therapy	1.000 (reference)	Reference
Second-line therapy	0.397 (0.109-1.444)	.161
Maintanance therapy	0.100 (0.015-0.666)	.017
DL model prediction (ensemble marker MSS ranking: 2)		
Improved_InceptionV3_MS (Ensemble All Markers)	0.209 (0.064-0.682)	.01[a]

d).	HR (95% CI)	*P* value
Age	0.990 (0.942-1.040)	.680
BMI	0.978 (0.884-1.081)	.657
BEV	0.978 (0.887-1.079)	.655
FIGO	0.222 (0.014-3.566)	.288
TNM stage	0.859 (0.567-1.301)	.472
Histology	0.710 (0.249-2.028)	.522
Surgery		
CRS + HIPEC	1.000 (reference)	Reference
Optimal	0.979 (0.296-3.239)	.972
Suboptimal	0.913 (0.310-2.693)	.869
Therapy		
Concurrent therapy	1.000 (reference)	Reference
Second-line therapy	0.465 (0.125-1.730)	.253
Maintanance therapy	0.116 (0.018-0.763)	.025
DL model prediction (ensemble marker MSS ranking: 1)		
Improved_MIL_RNN (Ensemble All Markers)	0.146 (0.042-0.509)	.003[b]

BEV, bevacizumab; BMI, body mass index; CRS, cytoreductive surgery; DL, deep learning; FIGO, International Federation of Gynaecology and Obstetrics; HIPEC, hyperthermic intraperitoneal chemotherapy; HR, hazard ratio; MSS, mean of sensitivity and specificity.
[a] $P < .05$.
[b] $P < .01$.

using stochastic gradient descent with the Adam optimization algorithm and a learning rate of 0.0001. For ResNet152, we employed minibatches of size 256 that were pretrained using ImageNet weights. To prevent overfitting, early stopping was also employed in this implementation.

The f_F output for the ResNet152 architecture is a 2048-dimensional representative vector. Assuming a slide and model f, we may generate a collection of the S most suspicious patches regarding positive class probability inside a slide. An RNN is fed with an ordered sequence of vector representations $e = e_1, e_2, \ldots, e_S$, together with a state vector h. A 0 vector is applied to initialize the state vector. Then, for step $i = 1, 2, \ldots, S$ of the recurrent forward pass, Equation 4 shows the new state vector h_i as follows:

$$h_i = ReLU(W_e e_i + W_h h_{i-1} + b) \quad (4)$$

with W_e and W_h representing the RNN model's weights. W_o maps a state vector to classify probabilities at step S, where the slide classification is denoted by $o = W_o h_s$. The model is not recurrent when $S = 1$, hence the RNN ought to learn the f_C classifier. The abovementioned technique is easily expanded to combine data of various sizes. By aggregating the predictions of all 3 models on patches collected at the exact same centered pixel location but using various magnifications, we are able to identify the S most suspicious patches from a slide provided by models $f_{20\times}, f_{10\times}$, and $f_{5\times}$, trained at 20×, 10×, and 5× magnifications, respectively. Currently, $e_{20,i}, e_{10,i}, e_{5,i}$ and the current state vector h_{i-1} represent the inputs to the RNN at step i. Equation 5 denotes the new state vector as follows:

$$h_i = ReLU(W_{20\times e_{20}\times,i} + W_{10\times e_{10}\times,i} + W_{5\times e_5\times,i} + W_h h_{i-1} + b) \quad (5)$$

We utilized 128-dimensional vectors for the recurrent unit's state representation, 10 recurrent steps ($S = 10$), and weighted the positive class to emphasize the model's sensitivity. With a batch size of 256, all RNN models were trained employing cross-entropy loss and stochastic gradient descent.

Table 6
Multivariate Cox proportional hazards analysis on the clinical factors and top-ranked model prediction with cancer recurrence on serous carcinoma: a) Improved_InceptionV3_MS (MSH2), b) Improved_MIL_RNN (MSH2) with a and b as the first rank, c) Improved_InceptionV3_MS (MSH6) as the second rank, d) Improved_MIL_RNN (MSH6) as the third rank based on Table 3

a).	HR (95% CI)	*P* value	b).	HR (95% CI)	*P* value
Age	0.995 (0.932-1.062)	.874	Age	0.995 (0.932-1.062)	.874
BMI	0.973 (0.869-1.089)	.629	BMI	0.973 (0.869-1.089)	.629
BEV	0.942 (0.842-1.054)	.298	BEV	0.942 (0.842-1.054)	.298
FIGO	1.219 (0.027-54.796)	.949	FIGO	1.219 (0.027-54.796)	.919
TNM stage	1.036 (0.627-1.712)	.891	TNM stage	1.036 (0.627-1.712)	.819
Histology	0.132 (0.024-0.722)	.019	Histology	0.132 (0.024-0.722)	.019
Surgery			Surgery		
CRS + HIPEC	1.000 (reference)	Reference	CRS + HIPEC	1.000 (reference)	Reference
Optimal	0.627 (0.171-2.304)	.482	Optimal	0.627 (0.171-2.304)	.482
Suboptimal	0.765 (0.214-2.739)	.680	Suboptimal	0.765 (0.214-2.739)	.680
Therapy			Therapy		
Concurrent therapy	1.000 (reference)	Reference	Concurrent therapy	1.000 (reference)	Reference
Second-line therapy	0.396 (0.075-2.100)	.277	Second-line therapy	0.396 (0.075-2.100)	.277
Maintanance therapy	0.421 (0.046-3.613)	.430	Maintanance therapy	0.421 (0.046-3.613)	.430
DL model prediction (all marker MSS ranking: 1)			DL model prediction (all marker MSS ranking: 1)		
Improved_InceptionV3_MS(MSH2)	0.114 (0.025-0.521)	.005[a]	Improved_MIL_RNN (MSH2)	0.114 (0.025-0.521)	.005[a]
c).	HR (95% CI)	*P* value	d).	HR (95% CI)	*P* value
Age	0.983 (0.923-1.046)	.579	Age	1.008 (0.943-1.079)	.807
BMI	0.971 (0.870-1.083)	.592	BMI	0.974(0.868-1.092)	.649
BEV	0.957 (0.859-1.066)	.422	BEV	0.919 (0.818-1.033)	.155
FIGO	1.821 (0.043-77.050)	.754	FIGO	0.973 (0.591-1.603)	.965
TNM stage	1.138 (0.692-1.872)	.611	TNM stage	1.090 (0.023-50.576)	.915
Histology	0.140 (0.026-0.757)	.022	Histology	0.110 (0.020-0.619)	.012
Surgery			Surgery		
CRS + HIPEC	1.000 (reference)	Reference	CRS + HIPEC	1.000 (reference)	Reference
Optimal	0.524 (0.146-1.875)	.321	Optimal	0.697 (0.184-2.642)	.596
Suboptimal	0.663 (0.190-2.310)	.519	Suboptimal	0.847 (0.236-3.040)	.798
Therapy			Therapy		
Concurrent therapy	1.000 (reference)	Reference	Concurrent therapy	1.000 (reference)	Reference
Second-line therapy	0.307 (0.060-1.557)	.154	Second-line therapy	0.528 (0.096-2.892)	.462
Maintanance therapy	0.353 (0.041-3.046)	.344	Maintanance therapy	0.547 (0.064-4.664)	.581
DL model prediction (all marker MSS ranking: 2)			DL model prediction (all marker MSS ranking: 2)		
Improved_InceptionV3_MS(MSH6)	0.177 (0.044-0.709)	.015[b]	Improved_MIL_RNN (MSH6)	0.063 (0.012-0.323)	.001[a]

BEV, bevacizumab; BMI, body mass index; CRS, cytoreductive surgery; DL, deep learning; FIGO, International Federation of Gynaecology and Obstetrics; HIPEC, hyperthermic intraperitoneal chemotherapy; HR, hazard ratio; MSS, mean of sensitivity and specificity.
[a] $P < .01$.
[b] $P < .05$.

Results

In this study, the experiments were conducted in 4 parts. First, we compared the 2 proposed methods with 3 state-of-the-art approaches, which had achieved notable success in the field of computational pathology, including InceptionV3 for classification of lung adenocarcinoma and lung squamous cell carcinoma[40]; NASNetLarge for prostate cancer grading[43]; MIL_RNN for classification of basal cell carcinoma, prostate cancer, and breast cancer metastases[41]; and WS_InceptionV3 for prediction of treatment outcomes to ovarian cancer.[44]

For the first experiment, by following established practices outlined in prior literature,[45-48] we employed a conventional approach of partitioning the sample into a training set with two-thirds of the data and a testing set with one-third. This widely used ratio has been consistently applied across studies, highlighting its prevalence and relevance. This division prevented training and testing models on the same data set while ensuring that the class distributions remained the same for both the training and testing data sets (see Therapeutic Outcome). Then, we used the MSS to measure the model performance, which provides a balanced evaluation by considering both the sensitivity and specificity of the model (see Methods).

In the second experiment, we performed the Cox proportional hazards regression multivariate statistical analysis on the top-ranked models to verify the model capability (Statistical Analysis: Multivariate Cox Proportional Hazards Regression Model). Third, we further conducted the K-M survival analysis to verify the performance of the effective models from the Cox hazards analysis (Statistical Analysis: Univariate Kaplan-Meier Survival Analysis). The 2 statistical analyses (Statistical Analysis: Multivariate Cox Proportional Hazards Regression Model and Statistical Analysis: Univariate Kaplan-Meier Survival Analysis) were conducted to validate the performance of the top-ranked models based on the overall MSS in the first experiment. Lastly, we assessed each model's performance in MSI prediction using the histopathology images (Microsatellite Instability Prediction).

Table 7
Multivariate Cox proportional hazards analysis on the clinical factors and top-ranked model prediction in ensemble marker with cancer recurrence on serous carcinoma: a) Improved_InceptionV3_MS (Ensemble MSH2 + MSH6), b) Improved_MIL_RNN (Ensemble MSH2 + MSH6), c) Improved_InceptionV3_MS (Ensemble All Markers), d) Improved_MIL_RNN (Ensemble All Markers), with b and d as the first ranking, a and c as the second ranking in ensemble marker MSS ranking based on Table 3

a).	HR (95% CI)	*P* value	b).	HR (95% CI)	*P* value
Age	0.983 (0.923-1.046)	.579	Age	0.995 (0.932-1.062)	.874
BMI	0.971 (0.870-1.083)	.592	BMI	0.973 (0.869-1.089)	.629
BEV	0.957(0.859-1.066)	.422	BEV	0.942 (0.842-1.054)	.298
FIGO	1.821 (0.043-77.050)	.754	FIGO	1.219 (0.027-54.796)	.919
TNM stage	1.138 (0.692-1.872)	.611	TNM stage	1.036-6.27-1.712)	.891
Histology	0.140 (0.026-0.757)	.022	Histology	0.132 (0.024-0.722)	.019
Surgery			Surgery		
CRS + HIPEC	1.000 (reference)	Reference	CRS + HIPEC	1.000 (reference)	Reference
Optimal	0.524 (0.146-1.875)	.321	Optimal	0.627 (0.171-2.304)	.482
Suboptimal	0.663 (0.190-2.310)	.519	Suboptimal	0.765 (0.214-2.739)	.680
Therapy			Therapy		
Concurrent therapy	1.000 (reference)	Reference	Concurrent therapy	1.000 (reference)	Reference
Second-line therapy	0.307 (0.060-1.557)	.154	Second-line therapy	0.369 (0.075-2.100)	.277
Maintanance therapy	0.353 (0.041-3.046)	.344	Maintanance therapy	0.421 (0.049-3.613)	.430
DL model prediction (SC ensemble marker MSS ranking: 2)			DL model prediction (SC ensemble marker MSS ranking: 1)		
Improved_InceptionV3_MS (Ensemble MSH2 + MSH6)	0.177 (0.044-0.709)	.015[a]	Improved_MIL_RNN (Ensemble MSH2 + MSH6)	0.114 (0.025-0.521)	.005[b]
c).	HR (95% CI)	*P* value	d).	HR (95% CI)	*P* value
Age	0.983 (0.923-1.046)	.579	Age	0.995 (0.932-1.062)	.874
BMI	0.971 (0.870-1.083)	.592	BMI	0.973 (0.869-1.089)	.629
BEV	0.957(0.859-1.066)	.422	BEV	0.942 (0.842-1.054)	.298
FIGO	1.821 (0.043-77.050)	.754	FIGO	1.219 (0.027-54.796)	.919
TNM stage	1.138 (0.692-1.872)	.611	TNM stage	1.036-6.27-1.712)	.891
Histology	0.140 (0.026-0.757)	.022	Histology	0.132 (0.024-0.722)	.019
Surgery			Surgery		
CRS + HIPEC	1.000 (reference)	Reference	CRS + HIPEC	1.000 (reference)	Reference
Optimal	0.524 (0.146-1.875)	.321	Optimal	0.627 (0.171-2.304)	.482
Suboptimal	0.663 (0.190-2.310)	.519	Suboptimal	0.765 (0.214-2.739)	.680
Therapy			Therapy		
Concurrent therapy	1.000 (reference)	Reference	Concurrent therapy	1.000 (reference)	Reference
Second-line therapy	0.307 (0.060-1.557)	.154	Second-line therapy	0.369 (0.075-2.100)	.277
Maintanance therapy	0.353 (0.041-3.046)	.344	Maintanance therapy	0.421 (0.049-3.613)	.430
DL model prediction (SC ensemble marker MSS ranking: 2)			DL model prediction (SC ensemble marker MSS ranking: 1)		
Improved_InceptionV3_MS (Ensemble All Markers)	0.177 (0.044-0.709)	.015[a]	Improved_MIL_RNN (Ensemble All Markers)	0.114 (0.025-0.521)	.005[b]

BEV, bevacizumab; BMI, body mass index; CRS, cytoreductive surgery; DL, deep learning; FIGO, International Federation of Gynaecology and Obstetrics; HIPEC, hyperthermic intraperitoneal chemotherapy; HR, hazard ratio; MSS, mean of sensitivity and specificity; SC, serous carcinoma.
[a] $P < .05$.
[b] $P < .01$.

Therapeutic Outcome

The results for prediction of therapeutic outcomes based on MMR expressions demonstrated that the 2 proposed methods, namely Improved_InceptionV3_MS and Improved_MIL_RNN, achieved excellent performance, with the highest ranking in terms of MSS scores and outperformed the benchmark approaches (Table 2). The 2 proposed methods achieved the highest MSS score using MSH2 marker, with 100% for all MSS, accuracy, sensitivity, and specificity. The 2 proposed methods utilizing MSH6 marker obtained the second-highest MSS score, with 98%, 98%, 100%, and 96% for MSS, accuracy, sensitivity, and specificity, respectively. These findings indicate that the 2 proposed methods using MSH2 or MSH6 marker can be applied to personalized treatment planning and aid in predicting the therapeutic effect of bevacizumab.

In addition, we extended our investigation through supplementary experiments to comprehensively evaluate the proposed methods on the examination of ensemble markers, specifically the combination of MSH2 with MSH6 and the combination of all available markers (Table 2). Overall, our proposed methods demonstrate comparably good performance. Notably, the proposed Improved_MIL_RNN achieved the first rank in both the kinds of ensemble models with 100% for all MSS, accuracy, sensitivity, and specificity, respectively. Moreover, the proposed Improved_InceptionV3_MS obtained the second rank in both ensemble models with 98%, 98%, 100%, and 96% for MSS, accuracy, sensitivity, and specificity, respectively. These results show that the 2 proposed ensemble models are applicable to personalized treatment planning and can help predict the therapeutic effect of bevacizumab.

Furthermore, we conducted additional sensitivity and specificity analysis focused on serous carcinoma, as shown in Table 3. The results are consistent with the overall comprehensive analysis findings as shown in Table 2, confirming that both proposed methods exhibit remarkable performances. Notably, the proposed approaches achieved the highest ranks in MSS and outperformed the benchmark approaches, specifically in the case of single

Table 8
Univariate Kaplan-Meier survival analysis on the top-ranked model prediction

Models	Marker	MSS ranking		Progression-free survival time				Overall survival time				*P* value
				Mean				Mean				
				Estimate	SE	95% CI		Estimate	SE	95% CI		
						Lower bound	Upper bound			Lower bound	Upper bound	
Improved_InceptionV3_MS	MSH2	1	0	368.444	115.126	142.797	594.092	200.000	36.062	129.318	270.682	<.001[a]
			1	581.996	69.772	445.242	718.75	462.000	61.835	340.803	583.197	
			Overall	653.523	104.569	448.567	858.478	329.000	69.373	193.030	464.970	
Improved_MIL_RNN	MSH2	1	0	368.444	115.126	142.797	594.092	200.000	36.062	129.318	270.682	<.001[a]
			1	581.996	69.772	445.242	718.75	462.000	61.835	340.803	583.197	
			Overall	653.523	104.569	448.567	858.478	329.000	69.373	193.030	464.970	
Improved_InceptionV3_MS	MSH6	2	0	381.059	121.399	143.117	619.001	225.000	34.300	157.773	292.227	.001[b]
			1	568.19	68.874	433.197	703.183	462.000	62.480	339.539	584.461	
			Overall	653.523	104.569	448.567	858.478	329.000	69.373	193.030	464.970	
Improved_MIL_RNN	MSH6	3	0	387.437	114.705	162.616	612.259	225.000	45.839	135.155	314.845	.004[b]
			1	576.084	73.856	431.326	720.842	462.000	46.447	370.963	553.037	
			Overall	653.523	104.569	448.567	858.478	329.000	69.373	193.030	464.970	
Improved_InceptionV3_MS (ensemble marker)	MSH2 + MSH6	2	0	381.059	121.399	143.117	619.001	225.000	34.300	157.773	292.227	.001[b]
			1	568.19	68.874	433.197	703.183	462.000	62.480	339.539	584.461	
			Overall	653.523	104.569	448.567	858.478	329.000	69.373	193.030	464.970	
Improved_MIL_RNN (ensemble marker)	MSH2 + MSH6	1	0	368.444	115.126	142.797	594.092	200.000	36.062	129.318	270.682	<.001[a]
			1	581.996	69.772	445.242	718.75	462.000	61.835	340.803	583.197	
			Overall	653.523	104.569	448.567	858.478	329.000	69.373	193.030	464.970	
Improved_InceptionV3_MS (ensemble marker)	All	2	0	381.059	121.399	143.117	619.001	225.000	34.300	157.773	292.227	.001[b]
			1	568.190	68.874	433.197	703.183	462.000	62.480	339.539	584.461	
			overall	653.523	104.569	448.567	858.478	329.000	69.373	193.030	464.970	
Improved_MIL_RNN (ensemble marker)	All	1	0	368.444	115.126	142.797	594.092	200.000	36.062	129.318	270.682	<.001[a]
			1	581.996	69.772	445.242	718.75	462.000	61.835	340.803	583.197	
			Overall	653.523	104.569	448.567	858.478	329.000	69.373	193.030	464.970	

MSS, mean of sensitivity and specificity.
[a] $P < .001$.
[b] $P < .01$.

marker MSH2, with 100% for MSS, accuracy, sensitivity, and specificity. Similarly, applying the 2 proposed methods alongside the MSH6 marker yielded the second-highest MSS score, with 98%, 98%, 96%, and 100% for MSS, accuracy, sensitivity, and specificity, respectively (Table 3).

For ensemble markers, the proposed methods demonstrate impressive performance in the additional serous carcinoma experiments (Table 3). The results are consistent with those of the comprehensive analysis presented in Table 2. Importantly, the Improved_MIL_RNN model achieves the first rank in terms of MSS within the ensemble model for both combinations: MSH2 with MSH6 and the combination of all markers, achieving perfect scores of 100% for MSS, accuracy, sensitivity, and specificity. Similarly, Improved_InceptionV3_MS consistently achieves the second-highest rank in both ensemble markers, with 98%, 98%, 100%, and 96% for MSS, accuracy, sensitivity, and specificity, respectively.

Statistical Analysis: Multivariate Cox Proportional Hazards Regression Model

Studies of patient treatment response over time play a critical role in understanding the impact of therapies on disease progression during survivorship. In such investigations, 2 closely associated statistical analyses, namely univariate K-M survival analysis and multivariate Cox proportional hazards model analysis, are commonly employed.[44,47-50] In this paper, we utilized both of these statistical approaches to assess the effectiveness of the proposed model prediction as an indicator of the therapeutic effect of bevacizumab, conducted with the assistance of SPSS software (SPSS Inc).[51]

The study investigated the model predictions of the 2 proposed methods by incorporating MSH2 or MSH6 markers, along with various clinical features, such as age, cancer stage (International Federation of Gynaecology and Obstetrics), body mass index, histology, number of treatments (bevacizumab), surgery type, and therapy type, in association with disease progression using multivariate Cox proportional hazards regression model in Table 4. The results verify that the 2 proposed models using either MSH2 or MSH6 markers are effective as indicators for patient selection with statistical significance ($P < .01$). Using both proposed methods with MSH2 marker, patients predicted to gain positive treatment effects by the proposed methods had a lower recurrence risk than other patients (hazard ratio [HR], 0.146; $P = .003$). Similarly, patients predicted to gain positive treatment effect by the 2 proposed models using the MSH6 marker showed a lower risk of recurrence with (HR, 0.209; $P = .01$) for the

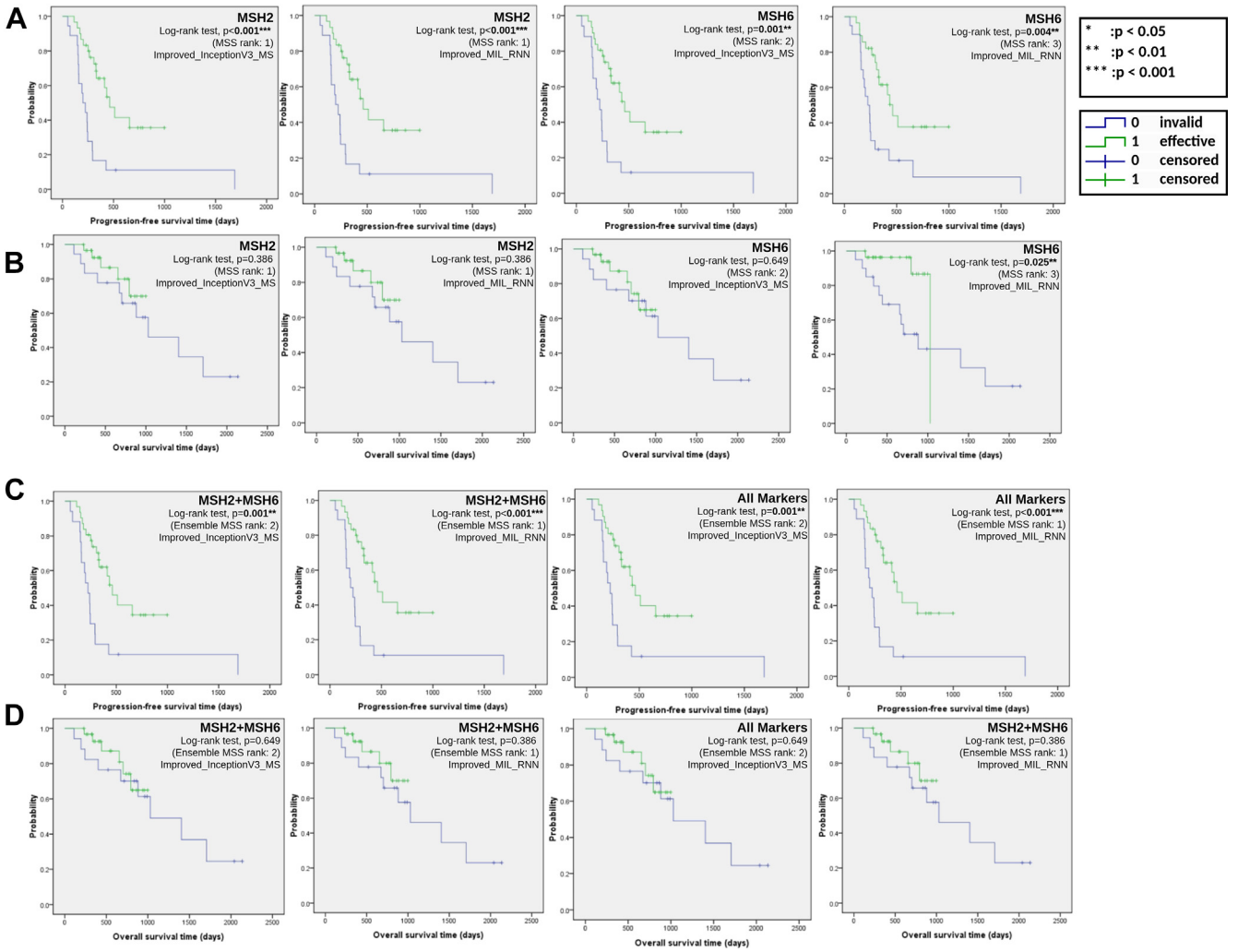

Figure 2.

Kaplan-Meier survival analysis: (A) progression-free survival (PFS) on the proposed model prediction outcomes using single marker; (B) overall survival on the proposed model prediction outcomes using single marker; (C) PFS on the proposed ensemble model prediction outcomes; (D) overall survival on the proposed ensemble model prediction outcomes. The Kaplan-Meier PFS analysis verifies the proposed methods' ability to effectively differentiate patients with favorable therapeutic outcomes, leading to lower recurrence rates, from those with disease progression after medication with statistical significance ($P < .01$) for both single marker and ensemble models. MSS, mean of sensitivity and specificity.

Improved_InceptionV3_MS models and (HR, 0.144; $P = .002$) for the Improved_MIL_RNN models if compared to other patients.

In addition to our primary investigations, we conducted a statistical analysis onto the ensemble models, including the ensemble of MSH2 and MSH6 markers and the ensemble of all 4 markers. The outcome of this analysis further validates the effectiveness of the 2 proposed models. Irrespective of utilizing the ensemble of MSH2 with MSH6 markers or the ensemble of all markers, both models emerged as indicators for patient selection with statistical significance ($P < .01$) (Table 5). Notably, when applying the proposed Improved_MIL_RNN in conjunction with either the combination of MSH2 and MSH6 markers or the combination of all markers, patients predicted to gain positive treatment effects had a lower recurrence risk than other patients (HR, 0.146; $P = .003$). This consistency was similarly observed in the proposed Improved_InceptionV3_MS models across both marker combinations and had a lower recurrence risk (HR, 0.209; $P = .01$), highlighting the predictive strength of the models.

Furthermore, an additional statistical analysis focused on serous carcinoma was performed (Tables 6 and 7). Overall, the results of these experiments remained consistent with the broader experimental outcomes, thereby validating the efficacy of the 2 proposed methods. Whether predicated on individual markers (Table 6) or combined markers (Table 7), these approaches emerge as robust determinants for patient selection, underscored by statistical significance ($P < .01$).

Statistical Analysis: Univariate Kaplan-Meier Survival Analysis

In addition to the multivariate Cox proportional hazards regression model, we further conducted the K-M PFS analysis and OS analysis onto both the proposed methods using a single marker and the ones using ensemble markers (Table 8). The results of K-M PFS curves, as presented in Figure 2A for individual markers and Figure 2C for ensemble markers, including the combination of MSH2 with MSH6 markers and the combination of all markers, showed evidence of the efficacy and robustness of the proposed methods. These outcomes demonstrate the effectiveness of

Table 9
Univariate Kaplan-Meier survival analysis on the top-ranked model prediction for serous carcinoma

SC models	Marker	MSS ranking		Progression-free survival time				Overall survival time				*P* value
				Mean				Mean				
				Estimate	SE	95% CI		Estimate	SE	95% CI		
						Lower bound	Upper bound			Lower bound	Upper bound	
Improved_InceptionV3_MS	MSH2	1	0	368.444	115.126	142.797	594.092	200.000	36.062	129.318	270.682	<.001[a]
			1	581.996	69.772	445.242	718.75	462.000	61.835	340.803	583.197	
			Overall	653.523	104.569	448.567	858.478	329.000	69.373	193.030	464.970	
Improved_MIL_RNN	MSH2	1	0	368.444	115.126	142.797	594.092	200.000	36.062	129.318	270.682	<.001[a]
			1	581.996	69.772	445.242	718.75	462.000	61.835	340.803	583.197	
			Overall	653.523	104.569	448.567	858.478	329.000	69.373	193.030	464.970	
Improved_InceptionV3_MS	MSH6	2	0	381.059	121.399	143.117	619.001	225.000	34.300	157.773	292.227	.001[b]
			1	568.19	68.874	433.197	703.183	462.000	62.480	339.539	584.461	
			Overall	653.523	104.569	448.567	858.478	329.000	69.373	193.030	464.970	
Improved_MIL_RNN	MSH6	3	0	387.437	114.705	162.616	612.259	225.000	45.839	135.155	314.845	.004[b]
			1	576.084	73.856	431.326	720.842	462.000	46.447	370.963	553.037	
			Overall	653.523	104.569	448.567	858.478	329.000	69.373	193.030	464.970	
Improved_InceptionV3_MS (ensemble marker)	MSH2 + MSH6	2	0	321.875	94.370	136.909	506.841	225.000	41.000	144.640	305.360	.004[b]
			1	512.629	84.790	346.441	678.818	412.000	87.651	240.204	583.796	
			Overall	538.035	108.635	325.110	750.961	293.000	37.043	220.396	365.604	
Improved_MIL_RNN (ensemble marker)	MSH2 + MSH6	1	0	312.000	89.194	137.180	486.820	225.000	34.300	157.773	292.227	.011[c]
			1	529.707	87.091	359.009	700.406	462.000	93.623	278.499	645.501	
			Overall	538.035	108.635	325.110	750.961	293.000	37.043	220.396	365.604	
Improved_InceptionV3_MS (ensemble marker)	All	2	0	321.875	94.370	136.909	506.841	225.000	41.000	144.640	305.360	.004[b]
			1	512.629	84.790	346.441	678.818	412.000	87.651	240.204	583.796	
			Overall	538.035	108.635	325.110	750.961	293.000	37.043	220.396	365.604	
Improved_MIL_RNN (ensemble marker)	All	1	0	312.000	89.194	137.180	486.820	225.000	34.300	157.773	292.227	.011[c]
			1	529.707	87.091	359.009	700.406	462.000	93.623	278.499	645.501	
			Overall	538.035	108.635	325.110	750.961	293.000	37.043	220.396	365.604	

MSS, mean of sensitivity and specificity; SC, serous carcinoma.
[a] $P < .001$.
[b] $P < .01$.
[c] $P < .05$.

both proposed methods in discerning patients with favorable therapeutic outcomes, characterized by a lowered risk of cancer recurrence, from patients who experience disease progression after medication with statistical significance ($P < .01$), which are consistent with the Cox proportional hazards regression models in the previous experiment. On the other hand, Figure 2B, D comparing the K-M OS to the overall top 2 MSS ranking curves in the single markers and ensemble markers, respectively, showed no statistical significance regarding the OS time.

Furthermore, we conducted supplementary analysis focused on serous carcinoma utilizing K-M PFS and OS on our additional experiments (Table 9). The results for K-M PFS curves for serous carcinoma with a single marker (Fig. 3A) and with ensemble marker (Fig. 3C), including the ensemble of MSH2 and MSH6 and the ensemble of all markers, were consistent with the overall analysis results, which demonstrates that the 2 proposed methods possess the capability to differentiate patients attaining positive treatment outcomes in lower cancer recurrence after medication, with statistical significance ($P < .01$). These findings are consistent with the Cox proportional hazard regression model outcomes observed in the previous experiment.

Microsatellite Instability Prediction

In order to help assess the treatment, we conducted the fourth experiment to examine the capability of the DL models to predict MSI status from histopathology images. The result showed that our proposed method, Improved_MIL_RNN, achieved the best performance in terms of MSS score, followed by MIL_RNN and NASNetLarge models with MSS values >72%. Table 10 presents the evaluation results in classifying the MSI status as intact or loss and ranks the model prediction based on the MSS score instead of F-measure due to the imbalanced data set (70.98% intact vs 29.02% loss). Our proposed Improved_MIL_RNN onto MLH1 marker achieved the best results for MSS, accuracy, sensitivity, and specificity with 75%, 77%, 84%, and 67%, respectively, and outperformed all existing state-of-the-art DL models. Moreover, the proposed Improved_MIL_RNN onto PMS2 marker obtained the third highest result with 72%, 73%, 66%, and 79% for MSS, accuracy, sensitivity and specificity, respectively. The results of this experiment demonstrate that our proposed DL methods are able to predict MSI status directly from histopathologic slides, even for challenging imbalanced class data sets, that is, 511 slides of intact class (70.98%) and 209 slides of loss class (29.02%).

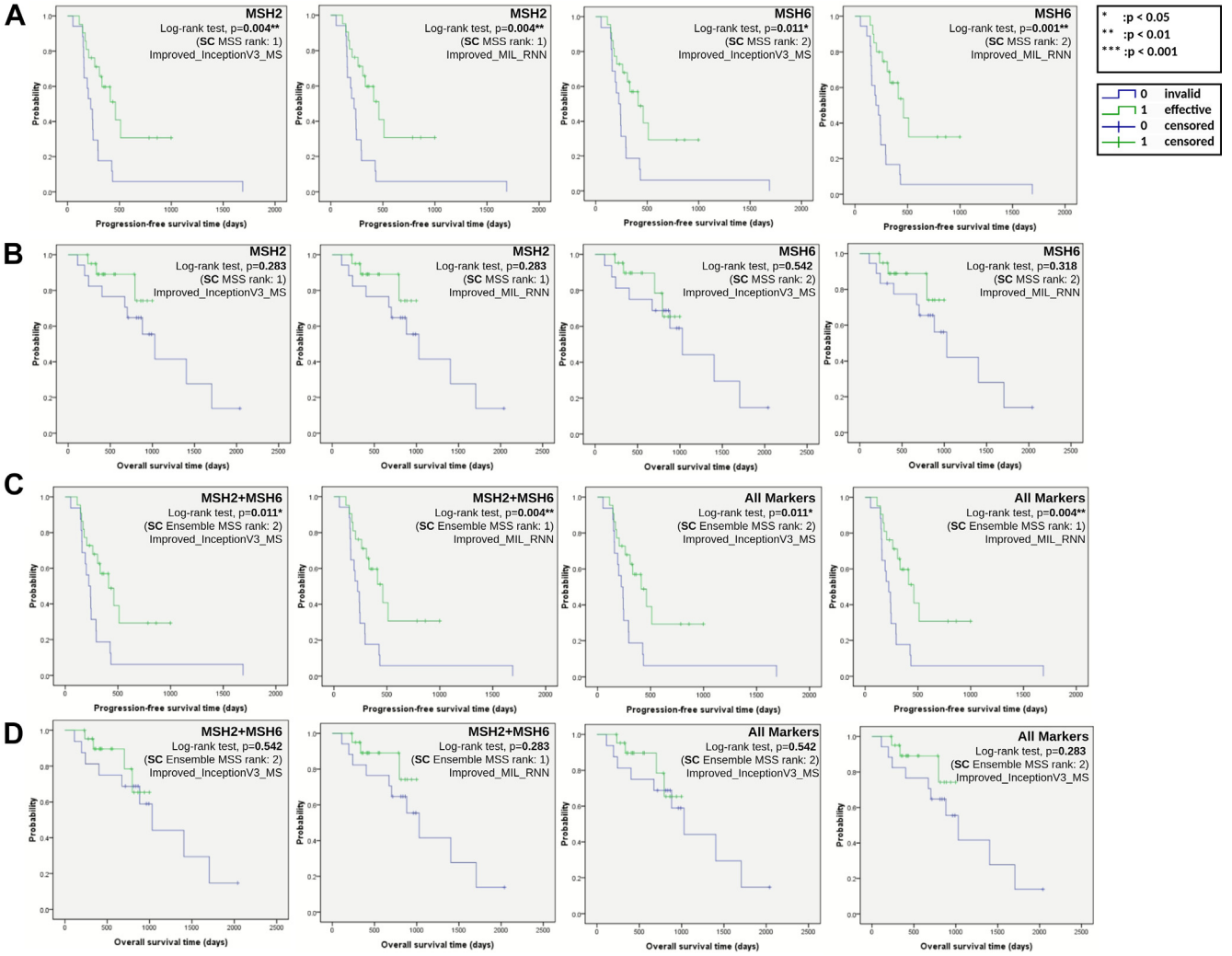

Figure 3.
Kaplan-Meier survival analysis on serous carcinoma: (A) progression-free survival (PFS) on the proposed model prediction outcomes using single marker; (B) overall survival on the proposed model prediction outcomes using single marker; (C) PFS on the proposed ensemble model prediction outcomes; (D) overall survival on the proposed ensemble model prediction outcomes. The Kaplan-Meier PFS analysis verifies the proposed methods' ability to effectively differentiate patients with favorable therapeutic outcomes, leading to lower recurrence rates, from those with disease progression after medication with statistical significance ($P < .01$) for both single marker and ensemble models.

Moreover, external validation was performed to assess the efficacy of our proposed and benchmark methods. This validation employed an independent cohort comprising 440 WSIs from TCGA ovarian cohorts. The core objective of this validation was to assess the ability of the DL models to accurately predict MSI status from WSIs (Table 11). Our proposed Improved_InceptionV3_MS model achieved superior results in accuracy (96%), sensitivity (67%), specificity (97%), and MSS (82%), outperforming all benchmark methods. The external TCGA validation set results demonstrate the robustness and generalizability of our proposed DL methods in directly predicting MSI status from histopathologic slides.

Discussion

The experimental result for predicting therapeutic outcomes showed that the 2 proposed methods obtained outstanding performance by achieving the perfect 100% MSS score using MSH2 stain and the second-best 98% MSS score using MSH6 (Table 2). Moreover, both K-M PFS analysis and Cox proportional hazards regression model analysis confirmed that the 2 proposed models using either MSH2 or MSH6 markers successfully differentiated patients with positive therapeutic effects and lower cancer recurrence rates from patients experiencing disease progression after treatment ($P < .01$). By leveraging the power of DL-based models, an exciting development has emerged in the field of MSI detection: avoiding relying solely on immunohistochemistry (IHC) analysis, which is not commonly performed for all patients except at high-volume medical centers. The proposed DL-based method, Improved_MIL_RNN, obtained the best MSS score (75%) on the MLH1 marker, and the Improved_MIL_RNN onto PMS2 achieved comparable performance with 72% MSS score. These results showed that the Improved_MIL_RNN possesses the decent capability to accurately predict MSI status from H&E-stained histopathology images, allowing cancer immunotherapy to help a wider target population even for the challenging imbalanced data set (70.98% intact and 29.02% loss). The DL-based approach remains important in the clinical management of patients with targeted therapy by assessing biomarkers that are predictive of

Table 10
Evaluation in classification of microsatellite instability status (intact vs loss)

Method	Marker	Accuracy	Sensitivity	Specificity	MSS	All marker MSS ranking
InceptionV3[40]	MLH1	0.65	0.84	0.33	0.59	9
Improved_InceptionV3_MS	MLH1	0.65	0.84	0.33	0.59	9
NASNetLarge[43]	MLH1	0.76	0.89	0.54	0.72	3[a]
MIL_RNN[41]	MLH1	0.66	0.74	0.54	0.64	4
Improved_MIL_RNN	MLH1	0.77	0.84	0.67	0.75	1[a]
InceptionV3[40]	MSH2	0.79	1	0	0.5	13
Improved_InceptionV3_MS	MSH2	0.82	0.98	0.23	0.61	7
NASNetLarge[43]	MSH2	0.79	0.98	0.08	0.53	11
MIL_RNN[41]	MSH2	0.74	0.82	0.46	0.64	4
Improved_MIL_RNN	MSH2	0.76	0.86	0.38	0.62	6
InceptionV3[40]	MSH6	0.94	1	0	0.5	13
Improved_InceptionV3_MS	MSH6	0.94	1	0	0.5	13
NASNetLarge[43]	MSH6	0.94	1	0	0.5	13
MIL_RNN[41]	MSH6	0.92	0.98	0	0.49	14
Improved_MIL_RNN	MSH6	0.9	0.95	0.25	0.6	8
InceptionV3[40]	PMS2	0.58	0.41	0.73	0.57	10
Improved_InceptionV3_MS	PMS2	0.5	0.72	0.3	0.51	12
NASNetLarge[43]	PMS2	0.65	0.38	0.88	0.63	5
MIL_RNN [41]	PMS2	0.73	0.76	0.72	0.73	2[a]
Improved_MIL_RNN	PMS2	0.73	0.66	0.79	0.72	3[a]

MSS, mean of sensitivity and specificity.
[a] Top 3 ranked models in all marker mean of sensitivity and specificity ranking.

response to targeted therapy (eg, anti-VEGF and MSI) and monitoring treatment response and resistance.

Serous, clear cell, transitional cell, endometrioid, and mucinous carcinomas are several of the EOC subtypes that represent various diseases with different pathogenic pathways.[52,53] For the purpose of treating EOC, targeted medications have been gradually incorporated into clinical trials.[12] Angiogenesis is the formation of new blood vessels from preexisting blood vessels, which led to the development of therapeutic drugs, such as bevacizumab, for antiangiogenesis that disrupt the vascular supply through the blockade of VEGF/VEGFR signaling.[54] In terms of all types of tumors, bevacizumab is the first active targeted treatment, especially for EOC.[55] The potential benefits of bevacizumab must be evaluated cautiously considering accompanying toxicities, such as gastrointestinal perforation, arterial or venous thromboembolism, proteinuria, poor wound healing, bleeding, hypertension, and posterior reversible encephalopathy syndrome.[56] Bevacizumab therapy is expensive, with the possibility of drug toxicity, and only some of these individuals will benefit from it; hence, patients should be carefully selected for bevacizumab therapy. However, the ability to identify those most likely to benefit is lacking serum biomarkers of angiogenesis.[57] As proof-of-concept, imaging biomarkers of bevacizumab responses have been investigated, with favorable results.[58] By examining histopathologic slides, we previously developed a weakly supervised DL model to assist in selecting and guiding bevacizumab-targeted therapy for patients with EOC or PSPC. Moreover, we identified the advantageous role of immune-related AIM2 biomarkers in predicting the therapeutic outcomes of bevacizumab.[44,47]

MSI-H patients with loss of the MMR gene (MSH2, MSH6, MLH1, and PMS2) are associated with overexpression of VEGF, potentially driving angiogenesis. VEGF driving abnormal tumor vasculature can reduce tissue hypoxia and enhance cytotoxic agent delivery, which polarizes inflammatory cells toward immune suppression.[59-61] Hyperactive angiogenesis influences several immunologic processes in the majority of growing solid tumors, generating a microenvironment that is immunosuppressive.[62] Antiangiogenic therapy not only prunes blood vessels that are critical for cancer growth and metastasis but also reprograms the tumor immune microenvironment.[63] Normalization of the tumor vasculature can reduce immunosuppression and promote antitumor immunity by enhancing the uptake of antigen presentation in dendritic cells and M1-associated macrophages and activation of cytotoxic $CD8^+$ T cells.[61] In addition, anti-VEGF drugs block negative immune signaling by increasing the ratio of antitumor-to-protumor immune cells, reducing the overall suppressed or revived suppressor cell pool and the expression of multiple immune checkpoints.[63,64] Human antibodies selectively targeting angiogenic vessels hold great promise for the immunotherapy of human malignancies and can help elucidate the molecular mechanisms regulating angiogenesis.[65] There is believed to be a synergistic effect between antiangiogenic agents and ICIs as abnormal tumor vasculature may prevent immune cell trafficking.[66] Antiangiogenic agents may also excessively prune

Table 11
Evaluation of microsatellite instability prediction on an external validation set from The Cancer Genome Atlas ovarian cohorts

Method	Accuracy	Sensitivity	Specificity	MSS	All marker MSS ranking
InceptionV3[40]	0.93	0.56	0.94	0.75	2
Improved_InceptionV3_MS	0.96	0.67	0.97	0.82	1
NASNetLarge[43]	0.97	0	1	0.5	4
MIL_RNN[41]	0.93	0.22	0.95	0.59	3
Improved_MIL_RNN	0.93	0.56	0.95	0.75	2

MSS, mean of sensitivity and specificity.

tumor vessels, which induces hypoxia and immunosuppression, including increased expression of PDL1.[54]

Lynch syndrome is associated with increased risks of colorectal and endometrial cancer and is also observed in EOC and a variety of other cancer types.[67,68] MMR testing is crucial for diagnosing Lynch syndrome and detecting MSI-H tumors eligible for anti-PDL1 immunotherapy.[27,69]. Via a negative feedback loop, the PD-1 pathway inhibits the immune system from dying cells. T cells can target tumor cells when the PD-1 receptor is inactivated or when the PDL1 ligand binds to it. Additionally, dMMR EOC is a distinct molecular subgroup characterized by increased tumor-infiltrating lymphocytes. These patients could be excellent anti-PDL1 therapeutic candidates.[34] Although it is unclear whether dMMR EOCs may also respond well to PDL1 blockade, as this kind of therapy has shown only modest responses in EOCs,[22,70] blocking antibodies that prevent PD-1/PDL1 interaction are under ongoing investigation in many cancer types, including EOC.[33] A previous study demonstrated a partial benefit of PD-1 blockade in some patients with EOC.[70] Patients with EOC have shown few but promising results in clinical trials with ICIs. In 2015, the first phase II clinical trial of anti−PD-1 antibody, nivolumab, in patients with EOC demonstrated encouraging clinical efficacy and tolerability.[71] In KEYNOTE-100, clinical trials revealed modest activity in recurrent EOC treated with pembrolizumab.[72] Additionally, the KEYNOTE-158 study demonstrated that the clinical benefit of anti−PD-1 therapy with pembrolizumab had an overall response rate of 33.3% in 15 patients with EOC and dMMR.[71] EOC treated with atezolizumab (anti-PDL1) also showed promising effectiveness, and MSI status might be a marker for atezolizumab effectiveness in EOC.[71] Atezolizumab combined with bevacizumab and chemotherapy is going to be investigated in 2 further phase III ovarian cancer trials; these combination studies offer the potential to determine the most effective chemotherapy regimen.[33] Considering the lack of clinical effect found in earlier EOC clinical studies, there is reason to expect that EOC will also benefit from the potential immunotherapy outcomes seen in other tumors.[73] Currently, 2 standard methods, PCR and screening for loss of MMR protein expression using IHC with antibodies directed against MLH1, MSH2, MSH6, and PMS2, are recommended for the detection of dMMR/MSI status.[21] These methods are equally valid as the initial screening test for dMMR/MSI.[21] Although IHC and PCR are both widely utilized in clinical practice for MMR/MSI identification, they were only correlated in 62.2% to 93% of cases based on previous reports.[74] NGS allows for the practical assessment of far more microsatellite loci than MSI-PCR.[69] Mutations in MMR genes leading to dMMR and MSI have been implicated in gynecologic malignancies, including EOC.[75-77] In this study, we evaluated the application of DL methods to predict MSI status in EOC based on H&E-stained WSIs from TCGA data set, with NGS MSI assessment as referenced labels for external validation, achieving good prediction outcomes. Clinical trials of anti-VEGF agents with immunotherapy should consider the appropriate dose and timing to ensure synergy.[78] Given the significance of alternative angiogenic pathways in the treatment response to bevacizumab and the potential to enhance antitumor immunity through tumor vasculature normalization, MMR proteins may be used as biomarkers to predict bevacizumab and immunotherapeutic effects in patients with EOC.

Histopathology or cytopathology is employed to verify the existence of tumor cells in an individual's specimen when diagnosing cancer. In DL, a branch of AI, massive data sets are examined for patterns and predictions using a machine learning method called artificial neural networks.[79] TCGA H&E slides were used to train and assess a convolutional neural network model that reliably identified a variety of genes and genetic mutations that could potentially be used as standard biomarkers.[80] In order to help with patient diagnosis, prognosis, and precision oncology, a number of DL techniques have the ability to infer significant genetic characteristics from readily accessible histopathology data as well as interpret the complicated heterogeneity of the tumor microenvironment.[79] Instead of focusing on the conventional tumor pathological evaluation of tumor cell identification or subclassification, we utilized the DL-based process to predict the bevacizumab treatment response.

MMR protein expression in EOC and PSPC described in this study may predict response to individualized antiangiogenic targeted therapy. In clinical practice, only some patients are tested for MSI because this requires additional genetic or immunohistochemical tests. Although assessing MSI-H/dMMR can be determined by IHC, the use of DL-based methods to predict MSI by H&E-stained WSIs of EOCs is yet to be established. We also showed that the proposed DL-based method using MLH1 or PMS2 stain could predict MSI directly from H&E histology. Our results demonstrated that the application of the DL-based method could potentially predict MSI status based only on H&E-stained WSIs in MSI-related EOCs. The proposed architecture can help the model capture more valuable morphologic features for classification and contributed to predicting a precise diagnosis of patients with EOC, which will assist in choosing appropriate treatment. However, it is still necessary to validate the MMR protein expression for bevacizumab-targeted therapy or immunotherapy prediction in patients with EOC or PSPC in more extensive population-based studies. Importantly, the 2 proposed DL models using MSH2 or MSH6 have the potential to dramatically assist treatment planning for personalized medicines and could be used in therapeutically targeted therapy drug selection. In the future, a DL-based diagnostic system will change the method of diagnosis of MSI, and the proposed framework may have the opportunity to be further extended to colorectal, endometrial, or other cancers.

Author Contributions

C.-W.W. and T.-K.C. conceived the idea of this work. C.-W.W. designed the methodology and the software of this work. Y.-C.L., N.P.F., H.M., T.-C.L., and P.-J.L. carried out the validation of the methodology and performed the formal analysis of this work. Y.-J.L., Y.-C.W., M.-H.Y., C.-H.W., and T.-K.C. participated in curation of the data set. C.-W.W., N.P.F., H.M., and T.-K.C. prepared and wrote the manuscript. C.-W.W. and T.-K.C. reviewed the manuscript. N.P.F., H.M., T.-C.L., and P.-J.L. prepared the visualization of the manuscript. C.-W.W. supervised this work. C.-W.W., C.-H.W., and T.-K.C. acquired funding for this work. All authors have read and agreed to the published version of the manuscript.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The deployment of the best proposed deep learning model, namely Improved_InceptionV3_MS, was built with the Caffe framework in Python as the primary deep learning package, and the program code has been made publicly accessible on GitHub (https://github.com/cwwang1979/Deep-Learning-Can-Predict-Bevacizumab-Therapeutic-E-ect-and-Microsatellite-Instability-Directly-from Predict-Bevacizumab-Therapeutic-Effect).

Funding

This study was supported by National Science and Technology Council, Taiwan (NSTC 112-2221-E-011-052 and NSTC 112-2321-B-016-003); Tri-Service General Hospital, Taipei, Taiwan (TSGH-D-110036, TSGH-A-111010, and TSGH-A-112008); and National Taiwan University of Science and Technology—Tri-Service General Hospital (NTUST-TSGH-112-02).

Declaration of Competing Interest

There is no competing interest that influences the work reported in this study.

Ethics Approval and Consent to Participate

All samples used in this study were coded anonymously in accordance with Declaration of Helsinki, and the protocol of this study was approved by the Ethics Committee of the Institutional Review Board of the Tri-Service General Hospital (TSGHIRB No.1-107-05-171 and No. B202005070). Informed consents were waived by the Committee due to the retrospective and anonymous nature.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019;69:7–34.
2. Tendulkar S, Dodamani S. Chemoresistance in ovarian cancer: prospects for new drugs. *Anticancer Agents Med Chem*. 2021;21(1):668–678.
3. Reid BM, Permuth JB, Sellers TA. Epidemiology of ovarian cancer: a review. *Cancer Biol Med*. 2017;14(6):9–32.
4. Pavlidis N, Rassy E, Vermorken JB, et al. The outcome of patients with serous papillary peritoneal cancer, fallopian tube cancer, and epithelial ovarian cancer by treatment eras: 27 years data from the SEER registry. *Cancer Epidemiol*. 2021;75, 102045.
5. Torre LA, Trabert B, DeSantis CE, et al. Ovarian cancer statistics, 2018. *CA Cancer J Clin*. 2018;68:284–296.
6. Sambasivan S. Epithelial ovarian cancer: review article. *Cancer Treat Res Commun*. 2022;33, 100629.
7. Choi HJ, Armaiz Pena GN, Pradeep S, Cho MS, Coleman RL, Sood AK. Anti-vascular therapies in ovarian cancer: moving beyond anti-VEGF approaches. *Cancer Metastasis Rev*. 2015;34:19–40.
8. De Smedt L, Lemahieu J, Palmans S, et al. Microsatellite instable vs stable colon carcinomas: analysis of tumour heterogeneity, inflammation and angiogenesis. *Br J Cancer*. 2015;113:500–509.
9. Sato S, Itamochi H. Bevacizumab and ovarian cancer. *Curr Opin Obstet Gynecol*. 2012;24:8–13.
10. Burger RA, Brady MF, Bookman MA, et al. Incorporation of bevacizumab in the primary treatment of ovarian cancer. *N Engl J Med*. 2011;365:2473–2483.
11. Perren TJ, Swart AM, Pfisterer J, et al. A phase 3 trial of bevacizumab in ovarian cancer. *N Engl J Med*. 2011;365:2484–2496.
12. Lhereux S, Braunstein M, Oza AM. Epithelial ovarian cancer: evolution of management in the era of precision medicine. *CA Cancer J Clin*. 2019;69:280–304.
13. van Beijnum JR, Nowak-Sliwinska P, Huijbers EJM, Thijssen VL, Griffioen AW. The great escape; the hallmarks of resistance to antiangiogenic therapy. *Pharmacol Rev*. 2015;67:441–461.
14. Herzog TJ, Monk BJ. Bringing new medicines to women with epithelial ovarian cancer: what is the unmet medical need? *Gynecol Oncol Res Pract*. 2017;4:13.
15. Smerdel MP, Steffensen KD, Waldstrøm M, Brandslund I, Jakobsen A. The predictive value of serum VEGF in multiresistant ovarian cancer patients treated with bevacizumab. *Gynecol Oncol*. 2010;118:167–171.
16. Raja FA, Hook JM, Ledermann JA. Biomarkers in the development of anti-angiogenic therapies for ovarian cancer. *Cancer Treat Rev*. 2012;38:662–672.
17. Wimberger P, Chebouti I, Kasimir-Bauer S, et al. Explorative investigation of vascular endothelial growth factor receptor expression in primary ovarian cancer and its clinical relevance. *Gynecol Oncol*. 2014;133:467–472.
18. Han ES, Burger RA, Darcy KM, et al. Predictive and prognostic angiogenic markers in a gynecologic oncology group phase II trial of bevacizumab in recurrent and persistent ovarian or peritoneal cancer. *Gynecol Oncol*. 2010;119:484–490.
19. Xiao X, Melton DW, Gourley C. Mismatch repair deficiency in ovarian cancer—molecular characteristics and clinical implications. *Gynecol Oncol*. 2014;132:506–512.
20. Kim YC, Zhao L, Zhang H, et al. Prevalence and spectrum of BRCA germline variants in mainland Chinese familial breast and ovarian cancer patients. *Oncotarget*. 2016;7:9600–9612.
21. Svrcek M, Lascols O, Cohen R, et al. MSI/MMR-deficient tumor diagnosis: which standard for screening and for diagnosis? Diagnostic modalities for the colon and other sites: differences between tumors. *Bull Cancer*. 2019;106:119–128.
22. Shannon C, Kirk J, Barnetson R, et al. Incidence of microsatellite instability in synchronous tumors of the ovary and endometrium. *Clin Cancer Res*. 2003;9:1387–1392.
23. Sanz-Esteban I, Calvo-Lobo C, Ríos-Lago M, Álvarez-Linera J, Muñoz-García D, Rodríguez-Sanz D. Mapping the human brain during a specific Vojta's tactile input: the ipsilateral putamen's role. *Medicine (Baltimore)*. 2018;97:e0253.
24. 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;31:625–646.
25. Llosa NJ, Cruise M, Tam A, et al. The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints. *Cancer Discov*. 2015;5:43–51.
26. Le DT, Uram JN, Wang H, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med*. 2015;372:2509–2520.
27. Therkildsen C, Jensen LH, Rasmussen M, Bernstein I. An update on immune checkpoint therapy for the treatment of lynch syndrome. *Clin Exp Gastroenterol*. 2021;14:181–197.
28. Howitt BE, Strickland KC, Sholl LM, et al. Clear cell ovarian cancers with microsatellite instability: a unique subset of ovarian cancers with increased tumor-infiltrating lymphocytes and PD-1/PD-L1 expression. *Oncoimmunology*. 2017;6, e1277308.
29. Francisco LM, Sage PT, Sharpe AH. The PD-1 pathway in tolerance and autoimmunity. *Immunol Rev*. 2010;236:219–242.
30. Hegde PS, Karanikas V, Evers S. The where, the when, and the how of immune monitoring for cancer immunotherapies in the era of checkpoint inhibition. *Clin Cancer Res*. 2016;22:1865–1874.
31. Jain RK. Antiangiogenesis strategies revisited: from starving tumors to alleviating hypoxia. *Cancer Cell*. 2014;26:605–622.
32. Fukumura D, Kloepper J, Amoozgar Z, Duda DG, Jain RK. Enhancing cancer immunotherapy using antiangiogenics: opportunities and challenges. *Nat Rev Clin Oncol*. 2018;15:325–340.
33. Hartnett EG, Knight J, Radolec M, Buckanovich RJ, Edwards RP, Vlad AM. Immunotherapy advances for epithelial ovarian cancer. *Cancers (Basel)*. 2020;12:3733.
34. Xiao X, Dong D, He W, et al. Mismatch repair deficiency is associated with MSI phenotype, increased tumor-infiltrating lymphocytes and PD-L1 expression in immune cells in ovarian cancer. *Gynecol Oncol*. 2018;149:146–154.
35. Skacel M, Skilton B, Pettay JD, Tubbs RR. Tissue microarrays: a powerful tool for high-throughput analysis of clinical specimens: a review of the method with validation data. *Appl Immunohistochem Mol Morphol*. 2002;10:1–6.
36. Shreffler J, Huecker MR. Diagnostic testing accuracy: sensitivity, specificity, predictive values *and* likelihood ratios. In: StatPearls [Internet]. StatPearls Publishing; 2023.
37. Naeger DM, Kohi MP, Webb EM, Phelps A, Ordovas KG, Newman TB. Correctly using sensitivity, specificity, and predictive values in clinical practice: how to avoid three common pitfalls. *AJR Am J Roentgenol*. 2013;200:W566–W570.
38. Glaros AG, Kline RB. Understanding the accuracy of tests with cutting scores: the sensitivity, specificity, and predictive value model. *J Clin Psychol*. 1988;44:1013–1023.
39. Bolin E, Lam W. A review of sensitivity, specificity, and likelihood ratios: evaluating the utility of the electrocardiogram as a screening tool in hypertrophic cardiomyopathy. *Congenit Heart Dis*. 2013;8:406–410.
40. Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nat Med*. 2018;24:1559–1567.
41. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med*. 2019;25:1301–1309.
42. 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*. IEEE; 2016:770–778.
43. Tolkach Y, Dohmgörgen T, Toma M, Kristiansen G. High-accuracy prostate cancer pathology using deep learning. *Nat Mach Intell*. 2020;2:411–418.
44. Wang CW, Lee YC, Chang CC, et al. A weakly supervised deep learning method for guiding ovarian cancer treatment and identifying an effective biomarker. *Cancers*. 2022;14:1651.
45. Kalderstam J, Edén P, Bendahl PO, Strand C, Fernö M, Ohlsson M. Training artificial neural networks directly on the concordance index for censored data using genetic algorithms. *Artif Intell Med*. 2013;58:125–132.

46. Foltz SM, Greene CS, Taroni JN. Cross-platform normalization enables machine learning model training on microarray and RNA-Seq data simultaneously. *Commun Biol*. 2023;6(1):222.
47. Wang CW, Chang CC, Lee YC, et al. Weakly supervised deep learning for prediction of treatment effectiveness on ovarian cancer from histopathology images. *Comput Med Imaging Graph*. 2022;99:102093.
48. Wang CW, Lee YC, Lin YJ, et al. Ensemble biomarkers for guiding anti-angiogenesis therapy for ovarian cancer using deep learning. *Clin Transl Med*. 2023;13:e1162.
49. Dudley WN, Wickham R, Coombs N. An introduction to survival statistics: Kaplan-Meier analysis. *J Adv Pract Oncol*. 2016;7:91–100.
50. Wang CW, Lee YC, Lin YJ, et al. Interpretable attention-based deep learning ensemble for personalized ovarian cancer treatment without manual annotations. *Comput Med Imaging Graph*. 2023;107, 102233.
51. SPSS Inc. *SPSS for Windows, Rel. 15.0*. SPSS; 2006.
52. Folkins AK, Jarboe EA, Saleemuddin A, et al. A candidate precursor to pelvic serous cancer (p53 signature) and its prevalence in ovaries and fallopian tubes from women with BRCA mutations. *Gynecol Oncol*. 2008;109:168–173.
53. S V, Bhagat R, PC S, PV R, Krishnamoorthy L. Microsatellite instability, promoter methylation and protein expression of the DNA mismatch repair genes in epithelial ovarian cancer. *Genomics*. 2014;104:257–263.
54. Ramjiawan RR, Griffioen AW, Duda DG. Anti-angiogenesis for cancer revisited: is there a role for combinations with immunotherapy? *Angiogenesis*. 2017;20:185–204.
55. Monk BJ, Minion LE, Coleman RL. Anti-angiogenic agents in ovarian cancer: past, present, and future. *Ann Oncol*. 2016;27(suppl 1):i33–i39.
56. Claussen C, Rody A, Hanker L. Treatment of recurrent epithelial ovarian cancer. *Geburtshilfe Frauenheilkd*. 2020;80:1195–1204.
57. Haunschild CE, Tewari KS. Bevacizumab use in the frontline, maintenance and recurrent settings for ovarian cancer. *Future Oncol*. 2020;16:225–246.
58. Ng CS, Zhang Z, Lee SI, et al. CT perfusion as an early biomarker of treatment efficacy in advanced ovarian cancer: an ACRIN and GOG Study. *Clin Cancer Res*. 2017;23:3684–3691.
59. Ferrara N, Hillan KJ, Gerber HP, Novotny W. Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nat Rev Drug Discov*. 2004;3:391–400.
60. Garcia J, Hurwitz HI, Sandler AB, et al. Bevacizumab (Avastin®) in cancer treatment: a review of 15 years of clinical experience and future outlook. *Cancer Treat Rev*. 2020;86, 102017.
61. Huang Y, Goel S, Duda DG, Fukumura D, Jain RK. Vascular normalization as an emerging strategy to enhance cancer immunotherapy. *Cancer Res*. 2013;73:2943–2948.
62. Melero I, Rouzaut A, Motz GT, Coukos G. T-cell and NK-cell infiltration into solid tumors: a key limiting factor for efficacious cancer immunotherapy. *Cancer Discov*. 2014;4:522–526.
63. Yi M, Jiao D, Qin S, Chu Q, Wu K, Li A. Synergistic effect of immune checkpoint blockade and anti-angiogenesis in cancer treatment. *Mol Cancer*. 2019;18:60.
64. Horikawa N, Abiko K, Matsumura N, et al. Expression of vascular endothelial growth factor in ovarian cancer inhibits tumor immunity through the accumulation of myeloid-derived suppressor cells. *Clin Cancer Res*. 2017;23:587–599.
65. Mutuberria R, Satijn S, Huijbers A, et al. Isolation of human antibodies to tumor-associated endothelial cell markers by in vitro human endothelial cell selection with phage display libraries. *J Immunol Methods*. 2004;287:31–47.
66. Flower H, Gallo P. Cerebellar glioblastoma in an NF1 patient. Is it surgical debulking really necessary? *Br J Neurosurg*. 2020;34:669–671.
67. Biller LH, Syngal S, Yurgelun MB. Recent advances in Lynch syndrome. *Fam Cancer*. 2019;18:211–219.
68. Cox VL, Saeed Bamashmos AA, Foo WC, et al. Lynch syndrome: genomics update and imaging review. *Radiographics*. 2018;38:483–499.
69. Dedeurwaerdere F, Claes KB, Van Dorpe J, et al. Comparison of microsatellite instability detection by immunohistochemistry and molecular techniques in colorectal and endometrial cancer. *Sci Rep*. 2021;11, 12880.
70. Hamanishi J, Mandai M, Ikeda T, et al. Safety and antitumor activity of anti-PD-1 antibody, nivolumab, in patients with platinum-resistant ovarian cancer. *J Clin Oncol*. 2015;33:4015–4022.
71. Ak N, Vatansever S. Dramatic response to single-agent atezolizumab in a patient with MSI-H serous ovarian cancer. *J Clin Pharm Ther*. 2021;46:1787–1791.
72. Matulonis UA, Shapira-Frommer R, Santin AD, et al. Antitumor activity and safety of pembrolizumab in patients with advanced recurrent ovarian cancer: results from the phase II KEYNOTE-100 study. *Ann Oncol*. 2019;30:1080–1087.
73. Wang W, Liu JR, Zou W. Immunotherapy in ovarian cancer. *Surg Oncol Clin N Am*. 2019;28:447–464.
74. Xiao J, Li W, Huang Y, et al. A next-generation sequencing-based strategy combining microsatellite instability and tumor mutation burden for comprehensive molecular diagnosis of advanced colorectal cancer. *BMC Cancer*. 2021;21:282.
75. Arora S, Velichinskii R, Lesh RW, et al. Existing and emerging biomarkers for immune checkpoint immunotherapy in solid tumors. *Adv Ther*. 2019;36:2638–2678.
76. Rasmussen M, Lim K, Rambech E, et al. Corrigendum to "Lynch syndrome-associated epithelial ovarian cancer and its immunological profile" [Gynecologic Oncology 162 (2021) 686-693]. *Gynecol Oncol*. 2022;164:242.
77. Marabelle A, Le DT, Ascierto PA, et al. Efficacy of pembrolizumab in patients with noncolorectal high microsatellite instability/mismatch repair-deficient cancer: results from the phase II KEYNOTE-158 study. *J Clin Oncol*. 2020;38:1.
78. Thompson R, Scates C. Challenges in implementing an advance care planning programme in long-term care. *Nursing Older People*. 2017;29(3):13.
79. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med*. 2021;13:152.
80. Kather JN, Heij LR, Grabsch HI, et al. Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat Cancer*. 2020;1:789–799.