

<https://doi.org/10.1038/s41698-024-00766-9>

Deep learning for endometrial cancer subtyping and predicting tumor mutational burden from histopathological slides

Check for updates

Ching-Wei Wang¹✉, Nabila Puspita Firdi¹, Yu-Ching Lee¹, Tzu-Chiao Chu¹, Hikam Muzakky¹,
Tzu-Chien Liu¹, Po-Jen Lai¹ & Tai-Kuang Chao^{1,2,3}✉

Endometrial cancer (EC) diagnosis traditionally relies on tumor morphology and nuclear grade, but personalized therapy demands a deeper understanding of tumor mutational burden (TMB), i.e., a key biomarker for immune checkpoint inhibition and immunotherapy response. Traditional TMB prediction methods, such as sequencing exomes or whole genomes, are costly and often unavailable in clinical settings. We present the first TR-MAMIL deep learning framework to predict TMB status and classify the EC cancer subtype directly from H&E-stained WSIs, enabling effective personalized immunotherapy planning and prognostic refinement of EC patients. Our models were evaluated on a large dataset from The Cancer Genome Atlas. TR-MAMIL performed exceptionally well in classifying aggressive and non-aggressive EC, as well as predicting TMB, outperforming seven state-of-the-art approaches. It also performed well in classifying normal and abnormal p53 mutations in EC using H&E WSIs. Kaplan–Meier analysis further demonstrated TR-MAMIL’s ability to differentiate patients with longer survival in the aggressive EC.

Endometrial cancer (EC) represents a diverse landscape of histologic subtypes, making it the most common gynecologic malignancy¹. Tumor morphology and grade, determined through histopathological examination, remain crucial for EC management. Under the revised FIGO staging system, non-aggressive histological subtypes include low-grade (G1 and G2) endometrioid carcinomas, accounting for approximately 65%². They possess estrogen and progesterone receptors, show hormonal sensitivity, and are composed of low-grade cells, resulting in a favorable prognosis³. In contrast, aggressive EC mainly includes FIGO grade 3 (G3) endometrioid carcinomas, serous carcinomas (SC), and clear cell carcinomas. These malignancies are characterized by hormone independence and lack of expression of estrogen and progesterone receptors. They consist of high-grade cells, frequently presenting at advanced stages, indicative of a poorer prognosis^{4,5}. 2023 revised FIGO staging introduced significant changes and correctly distinguishing between aggressive and non-aggressive types is crucial for accurate staging.

While the traditional diagnosis of EC relies on tumor morphology and grade, which provide a foundation for treatment decisions, the rise of

personalized therapy requires a deeper understanding of tumor mutational burden (TMB), which is an established predictive biomarker for immune checkpoint inhibitor (ICI)^{6,7}. TMB provides precise and comprehensive information for determining the efficacy of immunotherapies in EC^{8,9}. TMB is usually defined as the number of somatic mutations per megabase (mut/Mb) of interrogated genomic sequence. High TMB (TMB-H) has been associated with improved patient response rates and survival benefits from ICIs, making it a promising predictive biomarker for immunotherapy^{8,10–12}. Currently, the Food and Drug Administration (FDA) in the United States is considering approving TMB-based testing as a companion diagnostic for determining the suitability of ICI therapy. As for mismatch repair deficient (dMMR) solid tumors defined as having a high TMB, recent research has shown a high objective response rate (ORR) of 53% to anti-PD-1 therapy^{13,14}. Clinical trials also indicate the suitability of using ICIs in TMB-H subtypes of EC⁸. Traditionally, TMB can be quantified through various next-generation sequencing (NGS)-based sequencing technologies. NGS provides comprehensive genomic analysis, enabling detailed TMB evaluation¹⁵. However, its sophisticated technique requires advanced

¹Graduate Institute of Biomedical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan.

²Institute of Pathology and Parasitology, National Defense Medical Center, Taipei, Taiwan. ³Department of Pathology, Tri-Service General Hospital, Taipei, Taiwan. ✉ e-mail: cweiwang@mail.ntust.edu.tw; chaotai.kuang@msa.hinet.net

equipment and infrastructure, which can increase the cost of the testing process¹⁶. Similarly, NGS requires substantial sequencing data, making it a time-consuming process¹⁷.

Recently, there has been growing interest in predicting TMB status directly from hematoxylin and eosin (H&E)-stained whole slide images (WSI) using DL. Studies have demonstrated the feasibility of DL methods in predicting TMB status from histopathological images in lung cancer^{18–20}, colorectal cancer^{21,22}, bladder cancer²³, gastric cancer²⁴, and gliomas²⁵.

Although existing DL models like CLAM²⁶, TOAD²⁷, and TransMIL²⁸ achieve promising results across various WSI tasks, their reliance on pre-trained features or unsupervised learning can limit their ability to capture the potential crucial details within patches²⁹. Xiang et al.²⁹ tackled this challenge with a multi-scale representation attention-based network (MRAN), which has demonstrated superior performance in the detection of various kinds of cancer, including lung squamous cell carcinoma (LUSC), breast invasive carcinoma (BRCA), stomach adenocarcinoma (STAD) and breast cancer lymph node metastasis.

Our literature review revealed a gap in research regarding the aggressive and non-aggressive EC classification and TMB prediction directly from H&E-stained WSIs. To address this gap, we propose a truncated ResNet-based multilayer attention multiple instance DL framework (TR-MAMIL) with four key components in the classification of aggressive and non-aggressive EC and TMB prediction for the aggressive and non-aggressive ECs, respectively (see Fig. 1(a)). Firstly, we built an effective and efficient truncated ResNet-based (TR) feature encoder for capturing the relevant morphological and molecular features. This encoder demonstrated an average of 10% improvement in metrics over the original ResNet (see Table 7), with 21% and 2% faster training and inference times, respectively (see Fig. 6(b)). Secondly, we developed a multilayered attention MIL (MAMIL) module to identify informative regions in the slide automatically and efficiently. The MAMIL module takes into account the neighboring patches or instances of each analyzed patch in a bag, leading to a more diverse feature representation of patches and a combined representation of individual patches and their neighbors³⁰, which provides comprehensive information into the decision-making process of the proposed framework, eliminating the need for extensive manual annotation. Thirdly, we integrate the gender information as a stable factor covariate with the slide-level features to refine the final slide-level probability prediction score for each WSI. This allows the model to learn more complex features and forces the model to learn more generalized relationships, thereby improving the average performances of the TR-MAMIL by 7% and 9% in the MeanSS and AUROC, respectively (see Table 9). Finally, to address the potential overfitting issue and improve the model generalizability, we devised model selection strategies with early stopping mechanisms to help produce the best models.

To assess the efficacy of our proposed framework, we compared the performance of TR-MAMIL with the seven above-mentioned state-of-the-art (SOTA) DL approaches, which have achieved notable success in the computational pathology, including ClassicMIL³¹ for the detection of prostate cancer, skin cancer, basal cell carcinoma, and breast cancer lymph node metastasis, CLAM²⁶ and TransMIL²⁸ both for subtyping of non-small cell lung cancer (NSCLC) and renal cell carcinoma (RCC) and detection of breast cancer lymph node metastasis, Wang et al.³² in patient response prediction for personalized ovarian cancer treatment, Improved_InceptionV3_MS³³ for predicting therapeutic effect and assessing MSI status in ovarian cancer, TOAD²⁷ for identifying the origin of 18 distinct metastasis tumor, and MRAN²⁹ for detection of LUSC, BRCA, STAD, and breast cancer lymph node metastasis.

Through extensive experimental evaluation, our framework demonstrated remarkable performance in the prediction of both the EC subtype and TMB status from histopathological WSIs and consistently outperformed the benchmark methods. TR-MAMIL achieved outstanding performance in the classification of the aggressive and non-aggressive EC. For TMB prediction in the aggressive EC, TR-MAMIL obtains 73% and $82 \pm 13\%$ for the MeanSS and AUROC, respectively. In predicting TMB status in the non-aggressive EC, TR-MAMIL also demonstrated the best

performance in comparison to the seven SOTA approaches. Importantly, according to the Kaplan–Meier survival analysis, the results show that TR-MAMIL successfully differentiates patients with longer disease-specific survival (DSS) and overall survival (OS) with significant difference ($p < 0.01$ for DSS, $p < 0.05$ for OS) between the TMB predicted classes in the aggressive EC. Additionally, the framework also distinguishes disease-free survival (DFS) outcomes, showing a significant difference ($p < 0.05$) when using TP53 predictions. These compelling findings highlight the potential of TR-MAMIL to guide personalized treatment decisions by accurately predicting the EC cancer subtype and the TMB status for effective immunotherapy planning for EC patients.

Results

Materials: Patient cohorts

In this study, we utilized 918 anonymized whole-section H&E-stained WSIs collected from 529 patients of the TCGA cohort. All EC images from TCGA, including 759 frozen, 144 diagnostic formalin-fixed paraffin-embedded (FFPE), and 15 error WSIs, were determined for TMB status by NGS results, providing the information on the number of somatic mutations per megabase (mut/Mb) of interrogated genomic sequence. In this study, the patients' sequencing results are classified into TMB-low or TMB-high categories, with a score of 10 mut/Mb or higher defining high-TMB status [TMB-H: TMB ≥ 10 (186 patients); TMB-low: TMB < 10 (331 patients); NA: TMB data not available (12 patients)]. The TCGA cohort was collected from 29 tissue source sites available in the public repositories at the National Institutes of Health, USA (<https://portal.gdc.cancer.gov/>), where the tissue source site is accounted for dataset sampling (see Fig. 1(a)). The dataset comprises H&E-stained pathological WSIs with varying dimensions, ranging from 7967 to 174,281 pixels in width and 11,672 to 85,452 pixels in height (see Fig. 1(b)). The resolution of the WSIs is 0.252 microns per pixel (MPP). These images were extracted from 529 patients diagnosed with EC from patients aged 31 to 91 and from over seven different races, as illustrated in Fig. 1(c, d). The dataset consists of various morphological subtypes, including endometrioid carcinoma G1 ($n = 97$), endometrioid carcinoma G2 ($n = 117$), endometrioid carcinoma G3 ($n = 185$), serous carcinoma (SC, $n = 109$), combined endometrioid carcinoma G2 and SC ($n = 1$), and combined endometrioid carcinoma G3 and SC ($n = 20$) (see Fig. 1(e, f)). As shown in Fig. 1(e), for data exclusion, we excluded the rare type with only one single sample, i.e. the hybrid type of G2+SC, from our experimental analysis.

In addition, a separate tissue microarray (TMA) cohort of patients' paraffin-embedded tissues was retrospectively retrieved from the Department of Pathology at Tri-Service General Hospital, Taipei, Taiwan. The resolution of the TMA WSIs is 0.503 microns per pixel (MPP). The TMA cohort contained 242 EC tissue cores, including 127 cores of TP53 wild type and 115 cores of TP53 mutation, as shown in Fig. 1(g, h). The TMA sections were incubated in 3% hydrogen peroxide for 10 minutes to suppress the activity of endogenous peroxidase. They were then incubated with anti-p53 primary antibody (ready-to-use; cat# D0-7, Roche) for 1 hour at room temperature, followed by incubation with horseradish peroxidase-labeled immunoglobulin (Dako, Carpinteria, CA, USA) for 1 hour at room temperature. Peroxidase activity was visualized using a solution of diaminobenzidine (DAB) at room temperature. The WSIs were then acquired with a digital slide scanner (Leica AT Turbo) with a 20x objective lens. Abnormal p53 (mutation-type) staining is defined as either strong nuclear expression in tumor cells ($>80\%$), the complete absence of expression in tumor cells, or unequivocal cytoplasmic expression. Normal p53 (wild-type) expression was defined as nuclear staining of variable intensity in the tumor cells³⁴.

The TCGA and TMA datasets were processed separately. Stratified sampling was employed to divide the individual cohorts into patient-independent training sets (2/3) and testing sets (1/3) to ensure the proportional representation of important characteristics in each group. Furthermore, for training, we have divided the whole training set into training (9/10) and validation (1/10) subsets. Ethical approvals have been obtained from the research ethics committee of the Tri-Service General Hospital (TSGHIRB No.1-107-05-171 and No.B202005070).

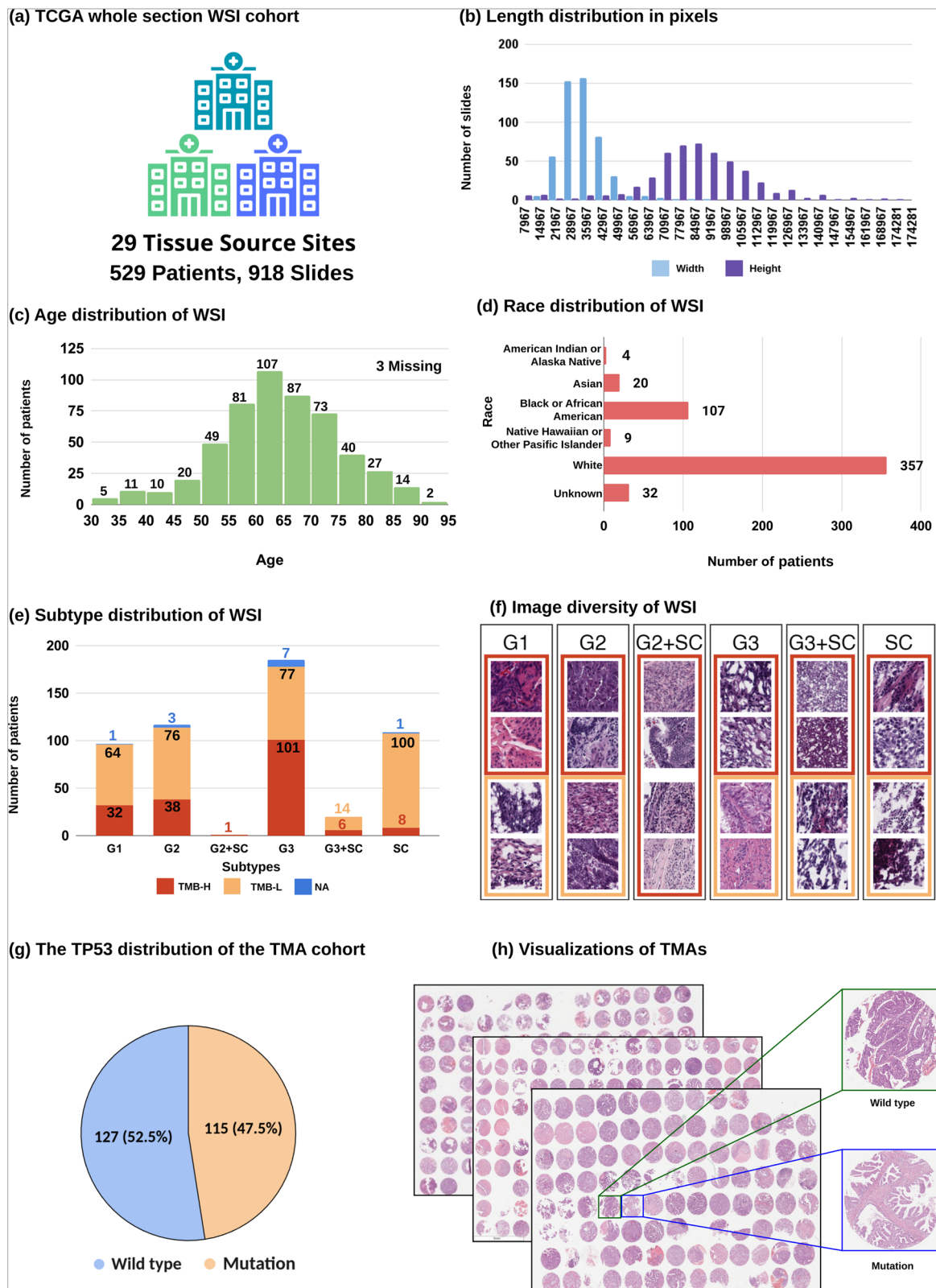


Fig. 1 | Data information of the two EC cohorts. a The TCGA whole-section WSI cohort from 29 tissue source sites, **b** Length distribution in pixels, **c** Age distribution, **d** Race distribution, **e** Subtype distribution, **f** image diversity of the data, **g** The TP53 distribution of the TMA cohort, **h** Visualizations of TMAs.

Overall results

In this study, the overall quantitative evaluation of the testing sets was performed in three parts, including by examining (1) the efficacy in EC subtyping in the first part, (2) the efficacy in TMB prediction for individual aggressive and non-aggressive ECs in the second part and (3) the efficacy as

an indicator for patient's prognosis using K-M survival analysis in the third part, with the comparison of the seven SOTA DL approaches, including ClassicMIL³¹, CLAM²⁶, Wang et al.³², Improved_InceptionV3_MS³³, TOAD²⁷, TransMIL²⁸, and MRAN²⁹, using the TCGA dataset. For the statistical analysis, we applied Fisher's exact test to examine the associations

Table 1 | Overall results with statistical analysis on the TCGA testing sets in (a) classification of aggressive and non-aggressive EC, (b) TMB prediction of the aggressive EC, and (c) TMB prediction of the non-aggressive EC

Task	Method	Acc.	Sens.	Spec.	MeanSS	AUROC ($\mu \pm \sigma$)	Fisher's test p -value
(a) Classification of aggressive and non-aggressive EC	TR-MAMIL	0.89	0.93	0.84	0.89	0.88 $\pm$ 0.05	<0.001***
	ClassicMIL ³¹	0.53	0.53	0.53	0.53	0.71 $\pm$ 0.18	0.179
	CLAM ²⁶	0.83	0.89	0.74	0.81	0.85 $\pm$ 0.04	<0.001***
	Wang et al. ³²	0.73	0.74	0.72	0.73	0.80 $\pm$ 0.05	<0.001***
	Improved_InceptionV3_MS ³³	0.69	0.90	0.40	0.65	0.80 $\pm$ 0.00	<0.001***
	TOAD ²⁷	0.80	0.76	0.86	0.81	0.81 $\pm$ 0.05	<0.001***
	TransMIL ²⁸	0.83	0.76	0.91	0.84	0.88 $\pm$ 0.05	<0.001***
	MRAN ²⁹	0.86	0.90	0.81	0.86	0.88 $\pm$ 0.05	<0.001***
(b) TMB prediction for the aggressive subtype	TR-MAMIL	0.73	0.72	0.73	0.73	0.82 $\pm$ 0.13	<0.001***
	ClassicMIL ³¹	0.61	0.53	0.65	0.59	0.62 $\pm$ 0.08	0.127
	CLAM ²⁶	0.66	0.67	0.66	0.66	0.67 $\pm$ 0.03	0.002**
	Wang et al. ³²	0.74	0.71	0.76	0.73	0.73 $\pm$ 0.02	0.436
	Improved_InceptionV3_MS ³³	0.70	0.41	0.86	0.64	0.65 $\pm$ 0.05	0.017*
	TOAD ²⁷	0.64	0.39	0.80	0.60	0.6 $\pm$ 0.05	0.038*
	TransMIL ²⁸	0.62	0.78	0.52	0.65	0.6 $\pm$ 0.03	0.189
	MRAN ²⁹	0.72	0.58	0.81	0.70	0.73 $\pm$ 0.02	<0.001***
(c) TMB prediction for the non-aggressive subtype	TR-MAMIL	0.66	0.76	0.61	0.68	0.56 $\pm$ 0.08	0.006**
	ClassicMIL ³¹	0.61	0.43	0.70	0.56	0.51 $\pm$ 0.01	0.235
	CLAM ²⁶	0.64	0.41	0.76	0.59	0.52 $\pm$ 0.07	0.339
	Wang et al. ³²	0.65	0.25	0.84	0.54	0.53 $\pm$ 0.04	0.144
	Improved_InceptionV3_MS ³³	0.62	0.55	0.65	0.60	0.52 $\pm$ 0.03	0.567
	TOAD ²⁷	0.64	0.48	0.72	0.60	0.54 $\pm$ 0.03	0.101
	TransMIL ²⁸	0.58	0.41	0.67	0.54	0.52 $\pm$ 0.01	0.297
	MRAN ²⁹	0.47	0.62	0.40	0.51	0.57 $\pm$ 0.02	0.567

MeanSS: Mean of sensitivity and specificity; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

between TR-MAMIL prediction and the actual cancer subtype or TMB status in aggressive and non-aggressive EC (see Fisher's Exact Test below). Furthermore, we investigated the potential of the models as indicators of DSS and OS using K–M survival analysis (see K–M survival analysis below).

Quantitative evaluation in the classification of aggressive and non-aggressive endometrial cancer

In the evaluation of the model performance on the testing set for the classification of aggressive (G3 endometrioid carcinoma and serous carcinoma: G3SC) and non-aggressive (G1 and G2 endometrioid carcinoma: G1G2) endometrial cancer, TR-MAMIL equipped with ResNet-50 Truncated as the backbone and F1-score metric for model selection and early stopping achieved the best performance overall and outperformed all seven SOTA benchmarked methods, obtaining the highest AUROC (88 $\pm$ 5%), sensitivity (93%), Mean of sensitivity and specificity (MeanSS) (89%) and accuracy (89%), respectively (see Table 1(a)). TR-MAMIL also demonstrated superior performances and outperformed the benchmark methods in MeanSS and AUROC. These results suggest the efficacy of TR-MAMIL for the classification of the aggressive and non-aggressive EC. Furthermore, we compared the receiver operating characteristics (ROC) curves of all top eight methods across all tasks, and Fig. 2(a) shows that TR-MAMIL achieved superior performances in the identification of aggressive and non-aggressive EC.

Quantitative evaluation in TMB and TP53 prediction

Secondly, we further evaluated the model performance on the testing sets in predicting the TMB status of the aggressive and non-aggressive EC samples.

For TMB prediction of the aggressive EC, the results demonstrated that TR-MAMIL equipped with ResNet-152 Truncated as the backbone and Cross-Entropy metric for model selection and early stopping achieved the highest AUROC (82 $\pm$ 13%) and MeanSS (73%), respectively (see Table 1(b)). In TMB prediction of the non-aggressive EC, TR-MAMIL equipped with ResNet-152 Truncated as the backbone and F1-score metric for model selection and early stopping also demonstrated the best performance in comparison to the seven SOTA approaches, achieving 76% sensitivity and 56 $\pm$ 8% of AUROC (see Table 1(c)). We also evaluated the model performance by comparing the ROC curves of all top eight methods for TMB prediction in both aggressive and non-aggressive ECs; Fig. 2(b, c) shows that TR-MAMIL achieved the highest AUROC in TMB prediction for both aggressive and non-aggressive EC, outperforming the benchmark methods. These findings demonstrate the promising ability of our methods to predict TMB status in both aggressive and non-aggressive ECs. Additionally, we also evaluated the model performance on an independent TMA testing set in predicting TP53 (mutation-type or wild type) of the EC samples (Table 2(a)). For TP53 prediction in the EC TMA testing set, the results showed that compared to seven SOTA methods, TR-MAMIL equipped with ResNet-152 Truncated as the backbone and Cross-Entropy metric for model selection and early stopping achieved the best performance with accuracy, sensitivity, specificity, MeanSS, and AUROC reaching 78%, 70%, 86%, 78%, and 78 $\pm$ 4%, respectively. For TP53 mutation prediction in the EC TCGA testing set (Table 2(b)), the TR-MAMIL framework, configured with the same ResNet-152 Truncated backbone but using the F1-score metric for model selection and early stopping, achieved an accuracy of 70%, sensitivity of 80%, specificity of 65%, MeanSS of 73%, and AUROC of 68 $\pm$ 8%.

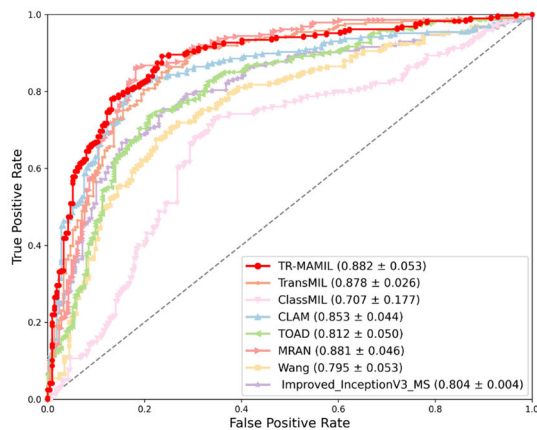
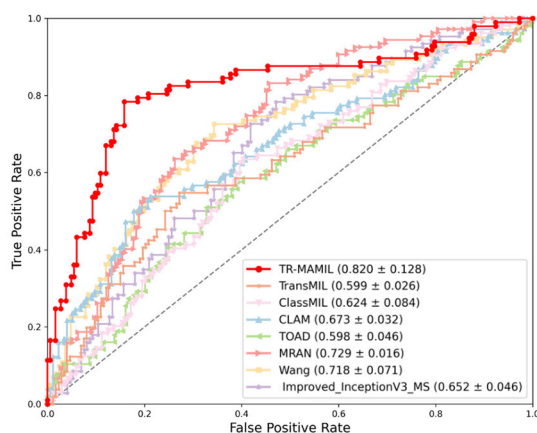
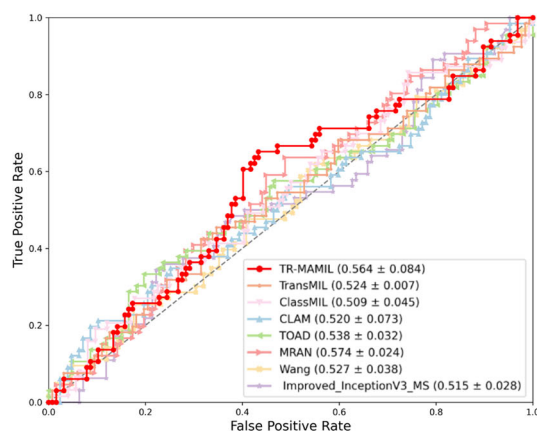
(a) ROC for EC subtyping**(b) ROC for TMB prediction in the aggressive EC****(c) ROC for TMB prediction in the non-aggressive EC**

Fig. 2 | Receiver operating characteristic curves (ROC). **a** classification of aggressive and non-aggressive EC, **b** TMB prediction in the aggressive EC, **c** TMB prediction in the non-aggressive EC.

In addition, we have further evaluated the performance of the proposed method in prediction of four MSI biomarkers using the TMA dataset. Table 3 shows that the proposed method consistently obtains excellent performance in prediction of four MSI biomarkers.

To demonstrate the model interpretability, Fig. 3(a, b) visualize the model attention maps of TR-MAMIL in the classification of aggressive and non-aggressive EC and TMB prediction on two aggressive and non-aggressive EC sample slides. Using a colormap overlaid on the images with

0.5 transparency, we highlight regions with high attention scores in red, indicating their significant contribution to the model's prediction. In contrast, blue regions correspond to low attention scores and lower predicted influence.

Quantitative evaluation in FFPE and frozen tissue samples in the classification of aggressive and non-aggressive EC and TMB prediction

We have further evaluated the model performance on the testing sets for FFPE and frozen tissue samples, respectively. Table 4 presented the experimental results of the proposed TR-MAMIL in the three tasks: classification of the aggressive and non-aggressive EC, TMB prediction in the individual aggressive and non-aggressive EC, respectively. The TR-MAMIL method for cancer subtyping employs a truncated ResNet-50 backbone, using the F1-score metric for model selection and early stopping. For TMB prediction of aggressive subtypes, the TR-MAMIL uses a truncated ResNet-152 backbone with Cross-Entropy as the model selection metric and early stopping. In the TMB prediction of non-aggressive subtypes, a truncated ResNet-152 backbone is also employed, with Cross-Entropy as the metric for model selection and early stopping. In cancer subtyping, the results show that TR-MAMIL consistently performs well in both types of tissue slides. On the other hand, in TMB prediction, due to limited FFPE samples available in the TCGA cohort, which contains 144 (16%) FFPE WSIs and 759 (84%) frozen WSIs, the proposed TR-MAMIL obtains higher specificity on the frozen samples. The data insufficiency could have restricted the ability of the model to learn a balanced decision boundary, resulting in lower specificity for FFPE tissues, which could be resolved by adding more training data.

Statistical analysis

To assess the clinical potential of TR-MAMIL and further validate the performances of TR-MAMIL, we performed a comprehensive statistical analysis with comparison to seven SOTA DL approaches in the three tasks: classification of the aggressive and non-aggressive EC, TMB prediction in the individual aggressive and non-aggressive EC. Statistical analyses were conducted using two-tailed Fisher's exact test, K-M survival analysis of DSS and OS utilizing the SPSS software³⁵.

Fisher's exact test

According to the two-tailed Fisher's exact test, the associations between TR-MAMIL predictions and the cancer subtype or TMB status in the aggressive group) are both extremely strong ($p < 0.001$), and the association between TR-MAMIL predictions and TMB status in the non-aggressive group is strong ($p < 0.01$) (see Table 1a, b, and c). These findings convincingly validate the outstanding performance of TR-MAMIL across all tasks: classification of the aggressive and non-aggressive EC subtypes and TMB prediction in the aggressive and non-aggressive EC subtypes.

Kaplan–Meier (K–M) survival analysis

As shown in Fig. 3(c, d), the results show that the proposed TR-MAMIL framework, equipped with ResNet50 backbone and F1-score as the model selection and early stopping metric, successfully differentiates patients with longer disease-specific survival (DSS) and overall survival (OS) with significant difference ($p < 0.01$ for DSS, $p < 0.05$ for OS) between the TMB predicted classes in the aggressive EC. Furthermore, we assess the proposed model's capability in patient prognosis using K-M survival analysis of DSS, OS, and disease-free survival (DFS) for both TP53 prediction and combined TP53/TMB prediction outcomes, as presented in Tables 5 and 6, respectively. The results show that the proposed TR-MAMIL framework, equipped with ResNet152 backbone and Cross-Entropy as the model selection and early stopping metric, shows a significant difference in DFS ($p < 0.05$) using TP53 predictions in aggressive EC as shown in Fig. 4. As shown in Fig. 5, integrating TP53 mutations and TMB predictions in aggressive EC shows a marginal difference in DFS ($p = 0.087$). These compelling findings highlight the great potential of the TR-MAMIL framework in improving patient prognosis for clinical applications.

Table 2 | Overall results with statistical analysis on an independent TMA testing set and a TCGA testing set in (a) classification of abnormal p53 (mutation-type) and normal p53 (wild type), and (b) TP53 mutation prediction in EC

Task	Method	Acc.	Sens.	Spec.	MeanSS	AUROC ($\mu \pm \sigma$)	Fisher's Test p -value
(a) Classification of abnormal and normal p53 (TMA)	TR-MAMIL	0.78	0.70	0.86	0.78	0.78 $\pm$ 0.04	<0.001***
	ClassicMIL ³¹	0.71	0.73	0.69	0.71	0.73 $\pm$ 0.01	<0.001***
	CLAM ²⁶	0.75	0.68	0.81	0.74	0.77 $\pm$ 0.02	<0.001***
	Wang et al. ³²	0.66	0.97	0.38	0.68	0.75 $\pm$ 0.02	<0.001***
	Improved_InceptionV3_MS ³³	0.70	0.84	0.58	0.71	0.79 $\pm$ 0.03	0.496
	TOAD ²⁷	0.68	0.89	0.50	0.70	0.76 $\pm$ 0.04	<0.001***
	TransMIL ²⁸	0.72	0.84	0.62	0.73	0.77 $\pm$ 0.04	<0.001***
	MRAN ²⁹	0.73	0.61	0.83	0.72	0.74 $\pm$ 0.08	<0.001***
(b) TP53 mutation prediction	TR-MAMIL	0.70	0.80	0.65	0.73	0.68 $\pm$ 0.08	<0.001***

MeanSS: Mean of sensitivity and specificity; *** $p < 0.001$.**Table 3 | Evaluation in the prediction of four MSI biomarkers of EC, including (a) MLH1, (b) MSH2, (c) MSH6 and (d) PMS2**

Task	Method	Accu.	Sens.	Spec.	MeanSS	MeanSS rank	Fisher exact test
(a) MLH1 prediction	Proposed TR-MAMIL	0.92	0.91	0.93	0.92	1	<0.001***
	ClassicMIL ³¹	0.89	0.91	0.87	0.89	5	<0.001***
	CLAM ²⁶	0.91	0.97	0.87	0.92	1	<0.001***
	Wang et al. ³²	0.75	0.50	0.93	0.72	8	<0.001***
	Improved_InceptionV3_MS ³³	0.86	0.79	0.91	0.85	7	<0.001***
	TOAD ²⁷	0.89	0.97	0.82	0.90	3	<0.001***
	TransMIL ²⁸	0.90	0.88	0.91	0.90	3	<0.001***
	MRAN ²⁹	0.89	0.83	0.93	0.88	6	<0.001***
(b) MSH2 prediction	Proposed TR-MAMIL	0.83	0.81	0.85	0.83	3	<0.001***
	ClassicMIL ³¹	0.77	0.73	0.79	0.76	6	<0.001***
	CLAM ²⁶	0.74	0.78	0.72	0.75	7	<0.001***
	Wang et al. ³²	0.80	0.63	0.92	0.77	5	<0.001***
	Improved_InceptionV3_MS ³³	0.74	0.47	0.92	0.69	8	<0.001***
	TOAD ²⁷	0.85	0.79	0.89	0.84	2	<0.001***
	TransMIL ²⁸	0.80	0.69	0.87	0.78	4	<0.001***
	MRAN ²⁹	0.85	0.86	0.85	0.85	1	<0.001***
(c) MSH6 prediction	Proposed TR-MAMIL	0.85	0.80	0.86	0.83	1	<0.001***
	ClassicMIL ³¹	0.72	0.82	0.69	0.76	5	<0.001***
	CLAM ²⁶	0.76	0.71	0.77	0.74	6	<0.001***
	Wang et al. ³²	0.79	0.06	0.99	0.52	8	0.381
	Improved_InceptionV3_MS ³³	0.81	0.11	1.00	0.56	7	0.043*
	TOAD ²⁷	0.84	0.83	0.84	0.83	1	<0.001***
	TransMIL ²⁸	0.76	0.94	0.72	0.83	1	<0.001***
	MRAN ²⁹	0.88	0.72	0.92	0.82	4	<0.001***
(d) PMS2 prediction	Proposed TR-MAMIL	0.84	0.82	0.85	0.84	2	<0.001***
	ClassicMIL ³¹	0.85	0.86	0.83	0.84	2	<0.001***
	CLAM ²⁶	0.77	0.73	0.83	0.78	6	<0.001***
	Wang et al. ³²	0.79	0.86	0.70	0.78	6	<0.001***
	Improved_InceptionV3_MS ³³	0.60	0.34	0.88	0.61	8	0.023*
	TOAD ²⁷	0.80	0.89	0.70	0.79	4	<0.001***
	TransMIL ²⁸	0.79	0.77	0.80	0.79	4	<0.001***
	MRAN ²⁹	0.85	0.82	0.89	0.85	1	<0.001***

MeanSS: Mean of sensitivity and specificity; * $p < 0.05$; *** $p < 0.001$.

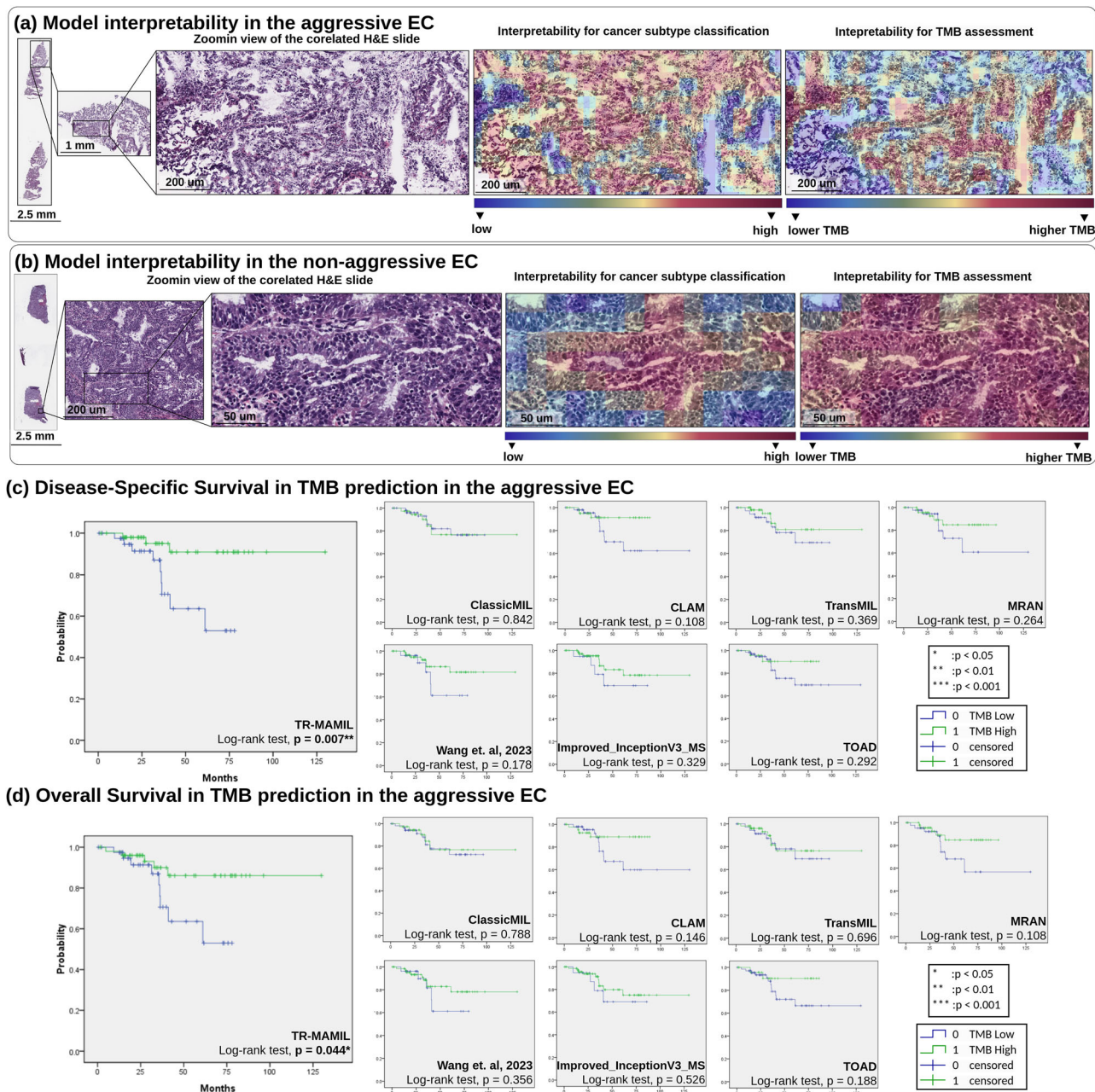


Fig. 3 | Attention heatmaps generated by the proposed model with K-M survival analysis results. a Model interpretability in an aggressive EC and **b** a non-aggressive EC. **c** Kaplan–Meier survival analysis in disease-specific survival and **d** overall Survival for TMB prediction in the aggressive EC.

Table 4 | FFPE and Frozen tissue sample results with statistical analysis on the TCGA testing sets in (a) classification of aggressive and non-aggressive EC, (b) TMB prediction of the aggressive EC, and (c) TMB prediction of the non-aggressive EC

Task	Tissue	Method	Acc.	Sens.	Spec.	MeanSS	AUROC ($\mu \pm \sigma$)	Fisher's test p -value
(a) Classification of aggressive and non-aggressive EC	FFPE	TR-MAMIL	0.88	0.96	0.71	0.84	0.86 ± 0.01	<0.001***
	Frozen		0.89	0.93	0.84	0.88	0.87 ± 0.04	<0.001***
(b) TMB prediction for the aggressive subtype	FFPE	TR-MAMIL	0.68	0.83	0.54	0.69	0.61 ± 0.04	<0.001***
	Frozen		0.72	0.71	0.73	0.72	0.69 ± 0.06	<0.001***
(c) TMB prediction for the non-aggressive subtype	FFPE	TR-MAMIL	0.75	1.00	0.57	0.79	0.65 ± 0.15	0.81
	Frozen		0.66	0.72	0.63	0.68	0.57 ± 0.08	0.022*

MeanSS: Mean of sensitivity and specificity: * $p < 0.05$; *** $p < 0.001$.

Table 5 | Kaplan–Meier survival analysis of TP53 predictions in aggressive EC

Kaplan–Meier survival analysis (months)		Mean				p-value
		Estimate	Std. error	95% Confidence interval		
				Lower bound	Upper bound	
Disease-free survival	0	81.614	3.719	74.326	88.903	0.03*
	1	93.585	10.305	73.388	113.782	
	Overall	107.265	5.925	95.652	118.878	
Disease-specific survival	0	79.39	3.767	72.007	86.773	0.35
	1	95.713	10.512	75.11	116.315	
	Overall	107.284	5.928	95.664	118.903	
Overall survival	0	77.802	4.007	69.948	85.656	0.377
	1	93.074	10.403	72.683	113.464	
	Overall	104.608	6.047	92.756	116.46	

Cases with predicted TP53 mutations were labeled as 1, while wild-type cases were labeled as 0.; * $p < 0.05$

Table 6 | Kaplan–Meier survival analysis by integrating TMB and TP53 predictions in aggressive EC

Kaplan–Meier survival analysis (months)		Mean				p-value
		Estimate	Std. error	95% Confidence interval		
				Lower bound	Upper bound	
Disease-free survival	0	97.148	11.608	74.396	119.901	0.087
	1	84.176	4.031	76.275	92.076	
	Overall	108.418	7.484	93.748	123.088	
Disease-specific survival	0	89.436	12.969	64.017	114.856	0.188
	1	81.459	4.672	72.303	90.616	
	Overall	103.812	8.144	87.85	119.775	
Overall survival	0	86.072	12.642	61.294	110.85	0.271
	1	78.51	5.343	68.038	88.981	
	Overall	99.611	8.269	83.405	115.817	

Cases predicted to have TMB-high and wild-type TP53 were labeled as 1, while cases predicted to have TMB-low and mutated TP53 were labeled as 0.

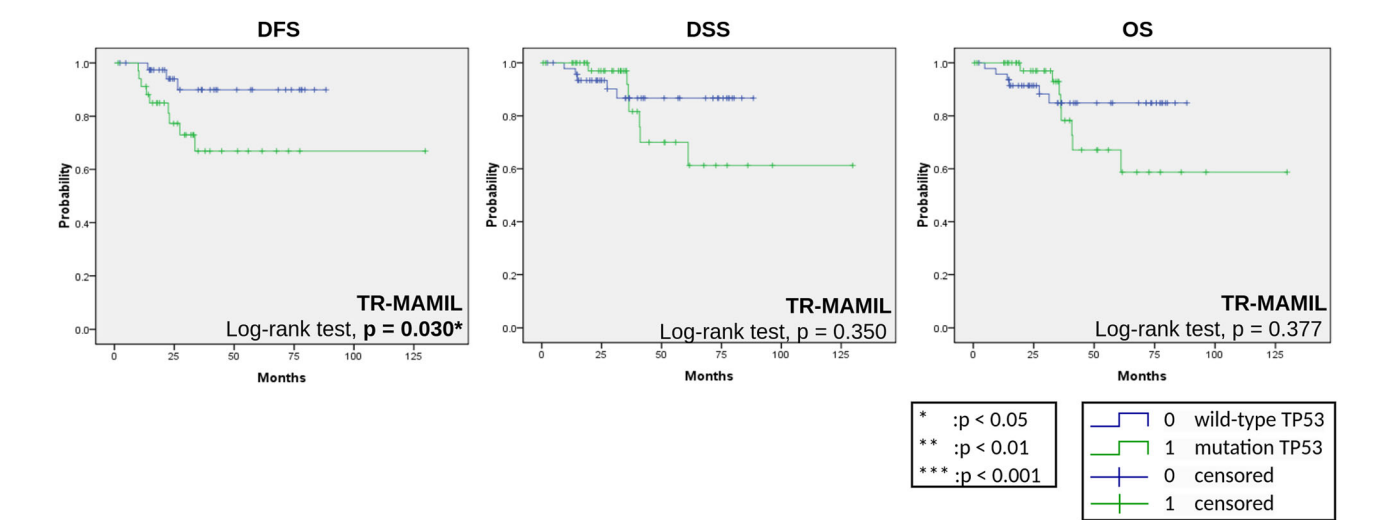


Fig. 4 | Kaplan–Meier curves for survival outcomes in aggressive EC patients stratified by the predictive TP53 mutation status (wild-type TP53 vs. mutated TP53) generated from the proposed TR-MAMIL. K–M survival analysis includes disease-free survival (DFS), disease-specific survival (DSS), and overall survival (OS).

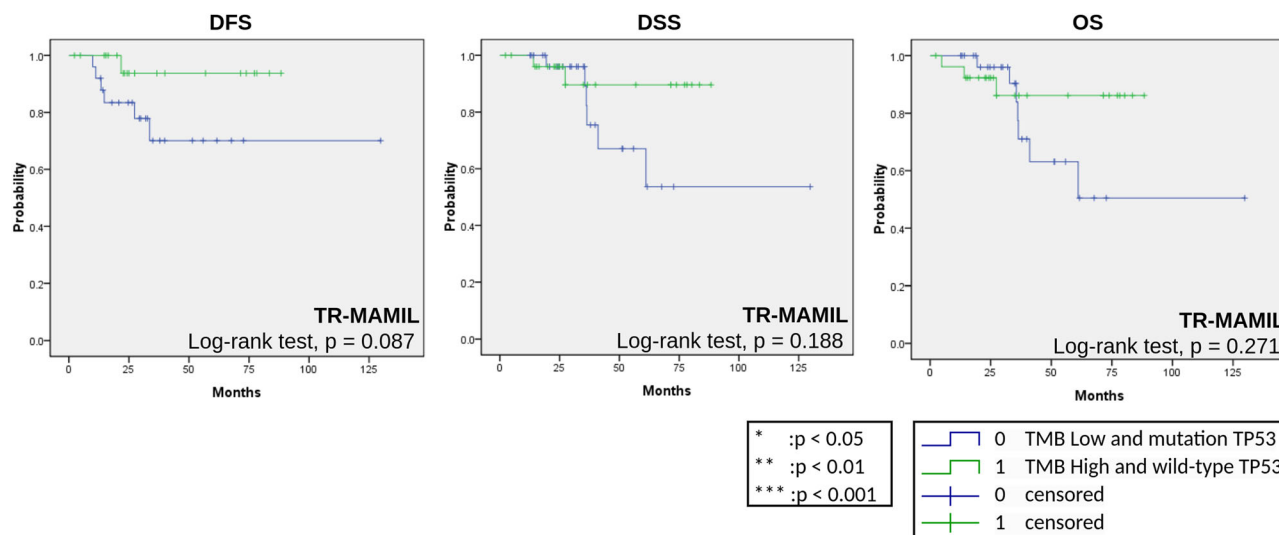


Fig. 5 | Kaplan-Meier survival analysis integrating TMB and TP53 predictions by the proposed TR-MAMIL in aggressive EC. K-M survival analysis includes disease-free survival (DFS), disease-specific survival (DSS), and overall survival (OS).

These compelling findings highlight the great potential of TR-MAMIL in TMB prediction as an indicator of a patient's DSS and OS for the aggressive EC.

Discussion

We present the first interpretable DL models to predict TMB status and classify the EC cancer subtype directly from H&E-stained WSIs, enabling effective personalized immunotherapy planning and prognostic refinement of EC patients. Among EC, although the majority are diagnosed early with a good prognosis, there are still 16% of EC patients who experience disease metastasis, with a 5-year survival rate of only 16.8%. For those patients who develop extra-uterine lesions after recurrence, systemic chemotherapy combined with radiotherapy is necessary. However, effective treatment options remain limited for patients who experience disease progression after standard therapy³⁶. The histopathological subtyping and molecular profiling analysis of EC has now become an important indicator for guiding treatment and assessing prognosis. In the 2023 revised staging system for EC, aggressive tumor types limited to the endometrium (previously classified as IA stage) are upgraded to stage IC. Alternatively, when aggressive tumor types with any myometrial invasion are clinically staged as stage IIC. (previously classified as IA or IB stage)². TR-MAMIL performs remarkably well in distinguishing between aggressive and non-aggressive EC, which can help improve the accuracy of pathological diagnosis and clinical staging. The identification of TMB, i.e. an established predictive biomarker for cancer immunotherapy, has further enhanced the precision of treatment strategies^{6,7,37} as TMB provides precise and comprehensive information for determining the efficacy of immunotherapies in EC^{8,9}. Major advances in the diagnosis and treatment of EC have been the ability to molecularly segregate and classify these carcinomas. Molecular features can be used to estimate the risk of recurrence and hence impact survival. Perhaps the most impactful molecular classification is proposed by TCGA, which classifies EC into four categories: (1) POLE/ultramutated, POLE-mutated tumors have an excellent prognosis; (2) microsatellite instability-high (MSI-H) /hypermutated, with an intermediate prognosis; (3) somatic copy-number alteration high (SCNA-high) /serous-like, nearly universal (95%) TP53 mutations, and a highly unfavorable prognosis; and (4) somatic copy-number alteration low (SCNA-low), with low copy-number alterations and low mutational burden². The pooled overall prevalence of dMMR was high for EC (26.8%)³⁸. The pooled overall prevalence of high TMB (≥ 10 mutations/Mb) was also high in EC (43.0%)³⁸. We show that TR-MAMIL can make reasonably accurate predictions, particularly in aggressive and non-aggressive EC classifications, and predict TMB status directly from H&E-stained WSIs for

EC samples. The TCGA molecular-based classification can be practically applied in clinical settings by using a simplified surrogate that includes three immunohistochemical markers (TP53, MSH6, and PMS2) and one molecular test (analysis for pathogenic POLE mutations). This surrogate approach classifies EC groups: POLE-mutant, MMR-deficient, TP53-abnormal, and NSMP2. In terms of systemic treatment for primary advanced and recurrent EC cancer, two randomized Phase III trials (ENGOT-en6/GOG-3031/RUBY and NRG-GY018/Keynote-868) have demonstrated a statistically significant and unprecedented progression-free survival (PFS) advantage by adding ICIs (dostarlimab or pembrolizumab, respectively) to standard carboplatin/paclitaxel chemotherapy, followed by ICIs maintenance therapy in dMMR patients, with hazard ratios (HR) of 0.28 (95% confidence interval [CI] 0.16–0.5) and 0.30 (95% CI 0.19–0.48), respectively. The success of ICI therapy in cancer immunotherapy has played a crucial milestone in treating advanced-stage cancers^{39–42}. Immune cells play dual roles in tumor development, promoting tumor progression while also clearing tumor cells³⁹. ICI therapy has achieved many significant positive research outcomes. It is well known that, in physiological conditions, checkpoint pathways prevent excessive T-cell activation that may result in loss of self-tolerance⁴³; for example, PD-L1 and CTLA-4 regulate the stimulation of the immune microenvironment. Malignancies are able to adapt to these responses and to exploit checkpoint pathways to promote tumor growth, and thus, ICIs can help to reactivate T-cell function, leading to the killing of the tumor cells⁴⁴. In many malignant tumors, ICI therapy can induce type I inflammation and enhance cytotoxic T cells, type 1 T helper cells, and M1 macrophages to eliminate tumor cells^{45,46}. Since CD8+ T cells are the primary direct effectors of cytotoxic responses to cancer cells, after PD-1 blockade, CD8+ T cells expand, and the change is considered the result of effective anti-tumor immunity, also correlating with positive clinical outcomes⁴⁷. TMB has emerged as a quantitative genomic biomarker capable of predicting response to ICIs beyond MSI^{18,49}.

In the study by Goodman et al.⁵⁰, the response rates to anti-PD-1/PD-L1 therapy were 58% for TMB-H patients and 20% for TMB-low patients, indicating an association between TMB and improved response to immunotherapy. Similarly, in the pivotal KEYNOTE-158 study, the ORR was 29% for TMB-H tumors compared to 6% for non-TMB-H tumors⁵¹. High TMB is prevalent across almost every type of cancer, and as a result, identifying patients with high TMB may benefit from immunotherapy in nearly all types of cancer¹⁷. Based on the findings of the KEYNOTE-158 study, Pembrolizumab has recently been approved by the U.S. FDA for use in tumors with TMB-H, regardless of histological type, also including EC¹⁰. In the molecular subtypes of EC, the TMB in the MSI-H group typically

exceeds 50 mutations/Mb, suggesting that immunotherapy may be used. In the SCNA-high group, TP53 mutations are commonly present, and the prognosis is the worst. These tumors usually have a low mutation rate and respond poorly to immunotherapy. However, a minority of specimens are microsatellite stable but exhibit exceptional TMB-H. If TMB testing is not performed, these patients may miss the opportunity to receive ICI treatment. Additionally, in POLE mutant types, TMB often exceeds 300 mutations/Mb. Due to the excellent prognosis, these patients generally do not require immunotherapy, but a small number of advanced-stage EC patients with POLE mutations and TMB-H may still have the opportunity to receive ICI treatment^{2,52,53}. Recently, pembrolizumab, a monoclonal antibody targeting PD-1, was also approved for use in any TMB-high (≥ 10) tumor, regardless of histology⁵⁴.

Evaluated in TCGA EC data, TR-MAMIL demonstrated superior performances compared to seven state-of-the-art methods^{26–29,31–33}. TR-MAMIL achieved outstanding performance in the classification of the aggressive and non-aggressive EC, with 97%, 93%, 89%, and 89% for the area under the receiver operating characteristic curve (AUROC), sensitivity, mean of sensitivity and specificity (MeanSS) and accuracy, respectively. For TMB prediction in the aggressive EC, TR-MAMIL achieves 73% and 78% for the MeanSS and AUROC, respectively. In predicting TMB status in the non-aggressive EC, TR-MAMIL also demonstrated the best performance compared to the seven SOTA approaches, achieving 76% sensitivity and 70% AUROC. Furthermore, according to Fisher's exact test, the associations between TR-MAMIL prediction and the actual cancer subtype or TMB status in the aggressive group are both extremely strong ($p < 0.001$), and the association between the prediction of TR-MAMIL and the actual TMB status in the non-aggressive group is strong ($p < 0.01$). According to the 2020 edition of the World Health Organization Classification, in EC, TP53 mutation indicates a poor prognosis. High-grade endometrioid carcinoma (G3) exhibits diversity in prognosis, clinical presentation, and molecular characteristics, and it is also the tumor type most likely to benefit from molecular classification. In our study, analysis of the high-grade endometrioid carcinoma (G3) group can predict whether it belongs to the group with poor prognosis (TP53 mutation)^{2,55}. Notably, our methods consistently outperformed seven state-of-the-art benchmarked approaches in computational pathology. Importantly, according to the Kaplan–Meier survival analysis, the results show that TR-MAMIL successfully differentiates patients with longer DSS and OS with significant differences ($p < 0.01$ for DSS, $p < 0.05$ for OS) between the TMB predicted classes in the aggressive EC. These compelling findings highlight the potential of TR-MAMIL to guide personalized treatment decisions by accurately predicting the EC cancer subtype and the TMB status for effective immunotherapy planning for EC patients. Moreover, a run-time analysis demonstrates that TR-MAMIL achieves high efficiency in inference time, taking only 26.21 seconds per slide on average, which makes TR-MAMIL feasible for practical clinical usage.

TR-MAMIL was evaluated using a variety of backbones for the feature encoder and a variety of metrics for the model selection with early stopping mechanisms. We found that employing a 1024-dimensional truncated ResNet outperformed the original 2048-dimensional ResNet with an average of 10% improvement in metrics over the original ResNet, with 21% and 2% faster training and inference times, respectively. This observation aligns with the tendency of deeper neural network layers to specialize in filters particular to features relevant to the source task and data. Interestingly, our study revealed a task-dependent optimal network depth for classifying endometrial cancer. For morphological classification (i.e., classification of aggressive and non-aggressive EC), a shallower network architecture achieved superior performance. This likely arises from the inherent simplicity of the features used in this task, such as cell morphology, texture, and spatial patterns. Shallower networks effectively capture these patterns without requiring complex non-linear transformations⁵⁶. Conversely, molecular classification, predicting TMB within both aggressive and non-aggressive ECs, benefited from the increased representational power of deeper networks. These architectures excel at extracting subtle relationships

between diverse genomic features and phenotypic expression, uncovering complex non-linear patterns and hidden relationships within the data. Additionally, integrating a stable covariate with the deep features extracted from the histological slide proved to be beneficial for enhancing the model performance, resulting in an average performance improvement of 7%, 10%, and 9% in terms of the mean of sensitivity and specificity, AUPRC and AUROC, respectively. While this study provides compelling evidence for the benefits of incorporating a stable covariate, additional research is encouraged to conduct larger-scale clinical validation to fully understand the specific mechanisms and to optimize integration for future research.

Given the relatively high cost of pembrolizumab and other potential targeted treatments, a key question to inform health system planning and budget impact evaluations is how many patients might be eligible for these treatments based on the presence of the biomarkers³⁸. Overall, our framework demonstrates the potential to accurately identify aggressive and non-aggressive ECs and predict TMB status in the aggressive and non-aggressive EC subtypes directly from H&E slides. This capability holds promise for improving patient care and treatment outcomes. Here, we utilized DL to analyze WSIs and found the ability to distinguish between non-aggressive and aggressive ECs. Simultaneously, we leveraged TMB data provided by TCGA to train the model for predicting TMB status. Within the molecular classification of EC, the dMMR group shows a response to ICI. In clinical practice, dMMR can be assessed through relatively simple immunohistochemical staining (IHC) and polymerase chain reaction (PCR), replacing the need for complex and expensive NGS. MSI-H/dMMR can be easily assessed through IHC for the loss of expression of one of the four MMR proteins (MLH1, PMS2, MSH2, and MSH6). For MSI testing through PCR, MSI at ≥ 2 loci is defined as MSI-high, instability at a single locus is defined as MSI-low, and no instability at any of the tested loci is defined as MeanSS. However, TMB analysis requires the use of NGS, making the use of DL to analyze pathology WSIs a promising approach⁵⁷. While TR-MAMIL offers promising results for direct TMB prediction from H&E-stained WSIs of EC, the inherent complexity of this task necessitates further refinement. Future studies exploring novel approaches and data integration strategies hold the key to unlocking improved performance. Beyond the successful application in this study, TR-MAMIL holds promise for broader utilization. We aim to explore their potential in various histopathological image analysis tasks. Further validation on a larger and more diverse database is crucial to demonstrate the robustness and generalizability of TR-MAMIL across clinical settings. Overcoming these challenges could lead to more effective treatment decisions and improved patient outcomes. Successful development of these models could revolutionize cancer care, enabling personalized and precise treatment strategies.

Methods

This study introduced three highly effective and efficient multilayer attention multiple instance DL approaches in the classification of aggressive and non-aggressive EC and TMB prediction for the aggressive and non-aggressive EC directly from H&E-stained WSIs. Ethical approvals have been obtained from the research ethics committee of the Tri-Service General Hospital (TSGHIRB No.1-107-05-171 and No.B202005070). All the slides were directly downloaded from the TCGA platform, and we did not apply any pre-processing techniques like stain normalization or data augmentation. Firstly, we established a uniform data representation by consistently segmenting and extracting tissue patch coordinates at $20 \times$ magnification and employ a simple thresholding method for foreground segmentation and extract the patches (see Fig. 7(a)). Secondly, two effective and efficient truncated ResNet-based feature encoders, including truncated ResNet50 and truncated ResNet152, were devised for capturing the relevant morphological and molecular features (see Fig. 7(b)). These encoders demonstrated an average of 10% improvement in all metrics over the original ResNet (see Table 7), with 21% and 2% faster training and inference times, respectively (see Fig. 6(b)). Thirdly, a multilayered attention MIL module was proposed to identify informative regions in the slide automatically and efficiently, and fourthly, a stable factor was integrated with the slide-level

Table 7 | Performance comparison among the original and truncated ResNets on the testing set in the classification of aggressive and non-aggressive EC

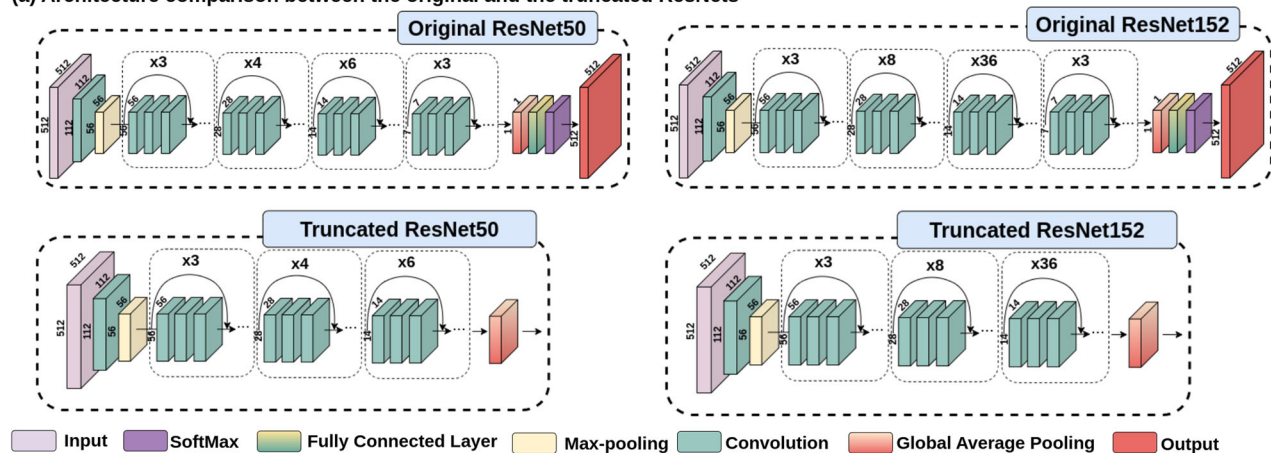
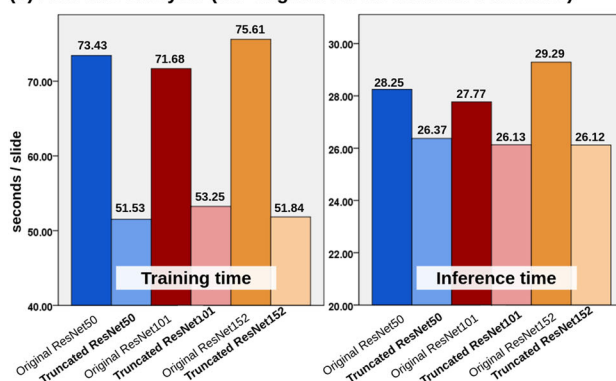
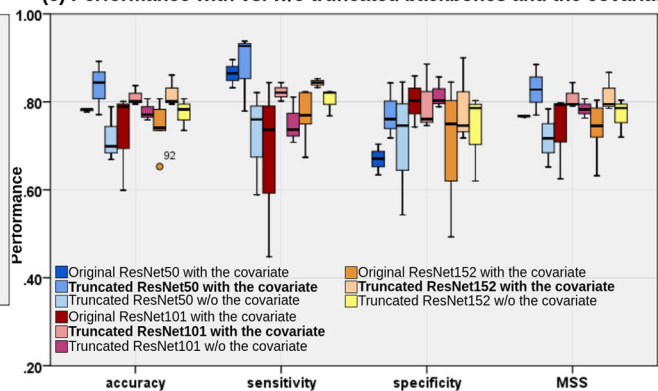
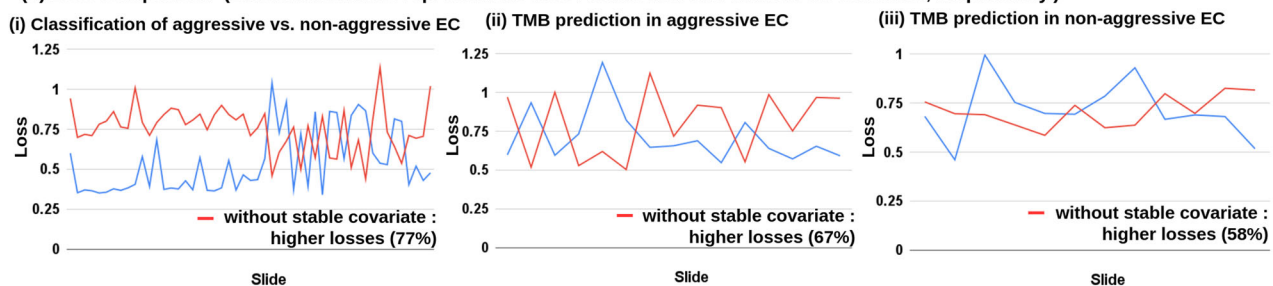
Backbone		Acc.	Sens.	Spec.	MeanSS	AUROC
ResNet50		0.89	0.93	0.84	0.89	0.97
ResNet101	Truncated	0.84	0.80	0.89	0.84	0.86
ResNet152		0.86	0.83	0.90	0.87	0.93
ResNet50		0.78	0.87	0.67	0.77	0.87
ResNet101	Original	0.80	0.84	0.74	0.79	0.85
ResNet152		0.74	0.75	0.71	0.73	0.80

MeanSS: Mean of sensitivity and specificity

features as a covariate not only to refine the final slide-level probability prediction score for each WSI and but also to allow the model to learn more complex features by forcing the model to learn more generalized relationships, thereby improving the average performances of TR-MAMIL by 7% and 9% in the MeanSS and AUROC, respectively (see Table 9 and Fig. 7(c)). Finally, to address the potential overfitting issue, which can affect the model generalizability, we devised model selection strategies with early stopping mechanisms to help produce the best models (see Fig. 7(d)). The workflow diagram in this study is provided in Fig. 7.

Tissue segmentation and patching

In this study, our pipeline begins by automatically segmenting tissue regions in each digitized slide. We established a uniform data representation by consistently segmenting and extracting tissue patches at 20× magnification.

(a) Architecture comparison between the original and the truncated ResNets**(b) Run time analysis (the original vs. the truncated ResNets)****(c) Performance with vs. w/o truncated backbones and the covariate****(d) Loss comparison (Blue and red lines represent the loss results with and without the covariate, respectively.)****Fig. 6 | Ablation studies on truncated backbones and the stable covariate.**

a Architecture comparison between the original and truncated ResNets, **b** Run-time analysis between the original and truncated ResNets in classification of aggressive and non-aggressive EC, **c** Performance comparison with or without truncated

backbones and the covariate in classification of aggressive and non-aggressive EC, **d** Loss comparison on the validation sets for: **d-i** classification of aggressive and non-aggressive EC, **d-ii** TMB prediction in the aggressive EC, and **d-iii** TMB prediction in the non-aggressive EC.

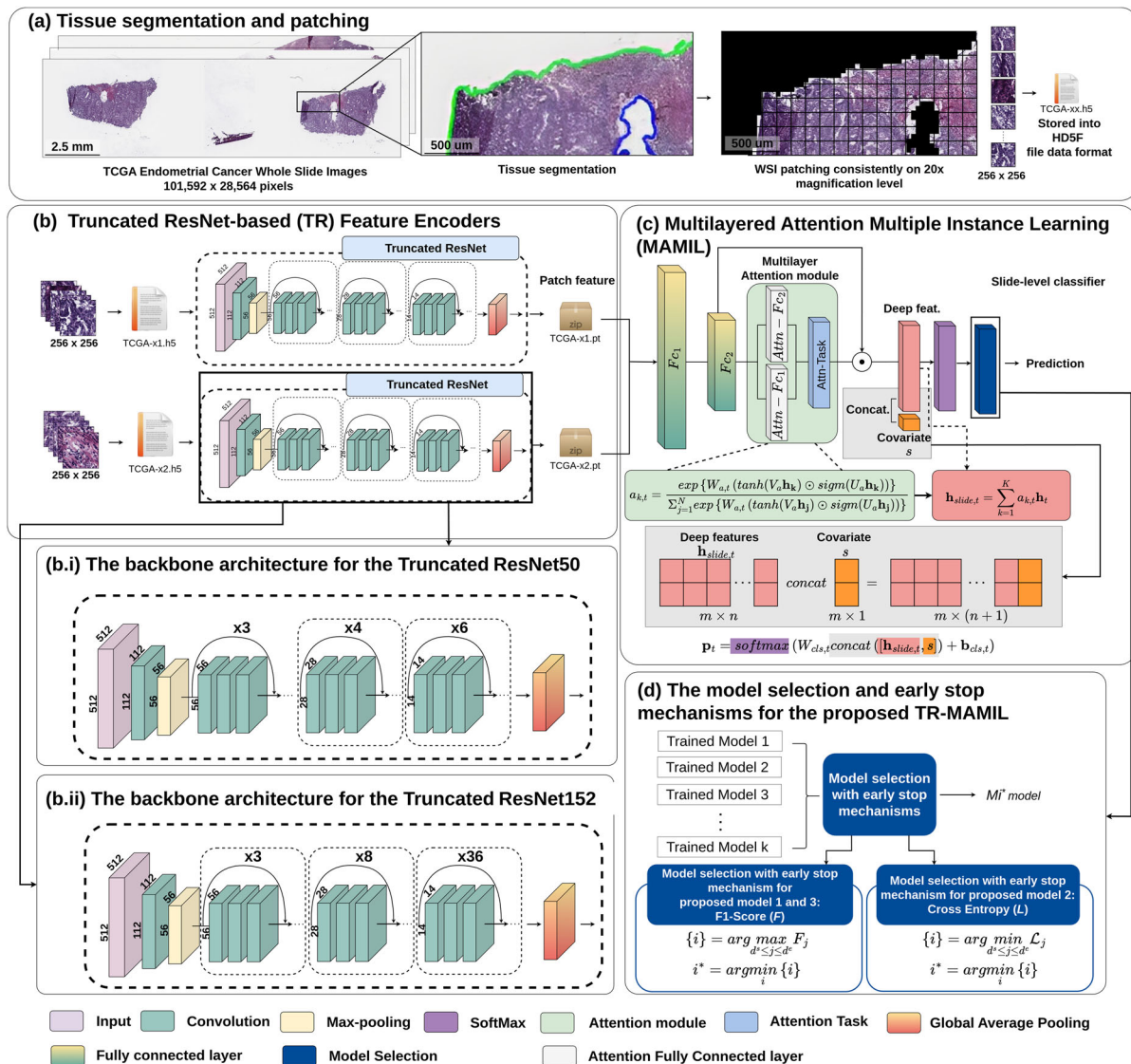


Fig. 7 | Overview of the whole process for TR-MAMIL. a Foreground segmentation and patching process, **b** feature extraction with truncated ResNet backbone: **b-i** the backbone architecture of the truncated ResNet50, **b-ii** the backbone architecture of

the truncated ResNet152, **c** the multilayer attention-based MIL and the final classification process, **d** the model selection and early stop mechanism for the proposed methods.

This greatly helps TR-MAMIL achieve optimal model performances. The WSI is initially loaded into memory at a downsampled resolution (e.g., 32× downscale). We then convert the image from RGB to the HSV color space and generate a binary mask outlining tissue regions (foreground) by thresholding the saturation channel. To refine the mask, we apply median blurring and morphological closing operations to fill gaps and holes. Detected foreground objects with contours above a specified area threshold are stored for downstream processing. The segmentation mask is also available for visual inspection. After segmentation, our algorithm extracts 256 × 256-pixel patches from the segmented foreground contours at a user-defined magnification level. The coordinates of these patches and the slide metadata are stored in the HDF5 hierarchical data format (see Fig. 7(a)).

Feature extraction

After patching, a deep CNN is utilized to compute a concise feature representation for each image patch within every slide (see Fig. 7(b)). We build effective and efficient truncated ResNet-based feature encoders pre-trained on ImageNet for capturing the relevant morphological and molecular features. Specifically, a truncated ResNet50 serves as the backbone in the classification of aggressive and non-aggressive EC (see Fig. 7(b-i)), while

truncated ResNet152 is employed for TMB prediction for both aggressive and non-aggressive EC, respectively (see Fig. 7(b-ii)). These models are truncated after the third residual block, followed by an adaptive average-spatial pooling layer, transforming each 256 × 256-pixel patch into a 1024-dimensional feature vector (see Fig. 6(a) for the detailed comparison between the proposed 1024-dimensional truncated ResNet and the 2048-dimensional original ResNet for ResNet50 and ResNet152, respectively). Utilizing these features in supervised learning provides advantages such as accelerated training times, reduced computational costs, and the ability to train on thousands of WSIs efficiently within a few hours. Moreover, using low-dimensional features allows simultaneous processing of all patches within a slide (up to 150,000) on a single consumer-grade GPU, eliminating the need for patch sampling and addressing issues related to noisy labels.

Ablation studies

We conducted ablation experiments to validate the efficacy and the efficiency of the core components in TR-MAMIL, including (1) the comparison of employing the original and the proposed truncated ResNet as the backbone in the feature extraction process for the EC subtyping, (2) the runtime analysis with the comparison of the original and the proposed

truncated ResNet, and (3) the comparison of different feature extraction methods

Comparison of truncated and original ResNets

Firstly, we compared the performances of TR-MAMIL using the original or truncated ResNet with three different depths, including ResNet50, ResNet101, and ResNet152, in the classification of aggressive and non-aggressive EC to investigate the most suitable backbone architecture. The original ResNet architecture with four residual blocks generates 2048-dimensional feature vectors, leading to high computational costs in feature extraction for both model training and inference. To address this, we truncated the ResNet by removing the fourth residual block, resulting in 1024-dimensional feature vectors (see Fig. 6(a) for a comparison of the truncated and the original ResNets). The results evaluated on the testing set demonstrated that the proposed 1024-dimensional truncated ResNet models consistently outperformed the original full 2048-dimensional ResNets with 9%, 3%, 18.67%, 13.67%, and 8% improvement in accuracy, sensitivity, specificity, MeanSS, and AUROC, respectively, and an average of 10% overall improvement in all metrics (see Table 7). The outcome could be attributed to the observation that later layers of a deep neural network tend to learn patterns, which are increasingly specific to features relevant to the source task and data of the pre-trained model that is onto the natural image classification with ImageNet datasets in this study. In contrast, features from earlier layers are characterized by their generality and applicability to diverse datasets. Since histopathology images differ significantly from natural images, using features from later ResNet layers did not necessarily improve performance compared to earlier layers. Hence, the worse performance of full ResNets may be caused by features of the later layers being too specialized for natural images and not suitable for the different textures and patterns in histopathology. The refinement strategy of truncated ResNets not only enhances informative representations for downstream tasks but also reduces the feature dimension significantly and lowers computational costs (run-time analysis is provided in the next section).

Run-time analysis with comparison of truncated and original ResNets

Secondly, we performed a run-time analysis with a comparison of the original 2048-dimensional ResNet and the 1024 truncated ResNet on a single local workstation using an Intel(R) Xeon(R) CPU E5-2650 v2 operating at a clock speed of 2.60GHz and NVIDIA GeForce GTX 1080 Ti GPU. In evaluation, the training and inference time includes the time to perform tissue segmentation, patch extraction, feature extraction, and model training or inference. The run-time analysis presented in Fig. 6(b) reveals that all truncated ResNet models consistently take up to 1.46 times faster than the original ResNets, with an overall average of 52.21 seconds/slide for training and 26.21 seconds/slide for inference. In contrast, the 2048-dimensional original ResNet takes 73.57 seconds/slide for training and 28.44 seconds/slide for inference, which is 21% and 2% longer than the training and inference time of the 1024-dimensional truncated ResNet, respectively. These findings indicate that using 1024-dimensional truncated ResNet features from an earlier convolution layer of the pre-trained ResNet offers more efficient data processing and training and reduces disk storage requirements. Therefore, truncated ResNets are used as the backbone for TR-MAMIL in feature extraction.

Compare with SSL-based backbones (ResNets and ViTs)

In the third evaluation. Self-supervised learning (SSL) has demonstrated its effectiveness in utilizing unlabeled data, and its application to pathology could greatly benefit downstream tasks⁵⁸. Recent studies also highlight the growing popularity of Transformer-based models in medical applications^{28,59–62}. Therefore, we compare our proposed truncated ResNet feature encoders with five SSL-based feature encoders, including two SSL-based ResNets and three SSL-based Vision Transformer (ViT) feature encoders. For the two SSL-based Resnets, we compared two clustering-guided contrastive learning (CCL)-based feature encoders, i.e., CCL-

ResNet50⁶³ and CCL-ResNet152. For SSL-based ViTs, we evaluated three models of simple self-supervised methods, namely self-distillation with no labels (DINO), including (1) a DINO-ViT-S/16 pre-trained on ImageNet⁶⁴, (2) a Lunit's DINO-ViT-S/16 pre-trained on pathology images⁵⁸ and (3) a Lunit's DINO-ViT-S/16 with additional data normalization. All models were evaluated on the testing sets with identical model selection and early stopping setups on each task, including F1-score for both EC subtyping and TMB prediction in the non-aggressive EC, and cross-entropy for TMB prediction in the aggressive EC.

As demonstrated in Table 8, the feature encoder with truncated-ResNet as the backbone outperforms the SSL-based feature encoders in the classification of aggressive and non-aggressive ECs (see Table 8(a)), TMB prediction in the aggressive EC (see Table 8(b)) and TMB prediction in the non-aggressive EC (see Table 8(c)). The performance of the proposed framework with a CCL-based feature encoder, i.e., CCL-ResNet50⁶³ and CCL-ResNet152 tends to have high sensitivity with very low specificity. Lunit's DINO-ViT-S/16 with and without data normalization⁵⁸ tend to have good performances in the classification of aggressive and non-aggressive EC and TMB prediction for the aggressive EC but have a worse performance in TMB prediction for the non-aggressive EC. Meanwhile, DINO-ViT-S/16⁶⁴ tends to have a stable performance in all of the tasks but still cannot surpass the proposed truncated ResNet feature encoders. These results further underscore the efficacy of our proposed truncated ResNet feature encoders as part of our proposed framework.

Multilayered attention module

Each slide is processed as a bag of feature embeddings from its constituent patches during training. The entire bag, containing k patches, is fed as a single input of dimension $k \times 1, 024$ to the MIL network, where 1024 is a fixed vector representation, produced previously in the feature extraction step. The network employs two stacked fully connected layers, F_{C1} and F_{C2} , to transform these patch embeddings into histology-specific feature representations $\{\mathbf{h}_k\}$. These layers, characterized by weight matrices and bias vectors of $W_1 \in \mathbb{R}^{512 \times 1,024}$, $\mathbf{b}_1 \in \mathbb{R}^{512}$ and $W_2 \in \mathbb{R}^{512 \times 512}$, $\mathbf{b}_2 \in \mathbb{R}^{512}$, respectively, followed by rectifier linear unit (ReLU) activation functions. This multilayer architecture allows the model to learn deep features suitable for WSI analysis by tuning the representations extracted through transfer learning, mapping the set of patch feature embeddings $\{\mathbf{z}_k\}$, $\mathbf{z}_k \in \mathbb{R}^{1,024}$ in a given WSI to 512-dimensional vectors:

$$\mathbf{h}_k = \text{ReLU}(W_2(\text{ReLU}(W_1\mathbf{z}_k + \mathbf{b}_1)) + \mathbf{b}_2) \quad (1)$$

Our approach utilized a multilayered attention module consisting of Attention Fully Connected 1 ($\text{Attn} - F_{C1}$) layer and Attention Fully Connected 2 ($\text{Attn} - F_{C2}$) layer with weight parameters $V_a \in \mathbb{R}^{384 \times 512}$ and $U_a \in \mathbb{R}^{384 \times 512}$ and task-specific weights, denoted as $W_{a,t} \in \mathbb{R}^{1 \times 384}$ for each task t (see Fig. 7(c)). The module is trained to assign attention scores $a_{k,t}$ to each patch k . After softmax activation, a high score (near 1) indicates the importance of the region for the slide-level classification task, while a low score (near 0) suggests the region lacks of the diagnostic or prognostic value.

$$a_{k,t} = \frac{\exp\{W_{a,t}(\tanh(V_a\mathbf{h}_k) \odot \text{sigm}(U_a\mathbf{h}_k))\}}{\sum_{j=1}^N \exp\{W_{a,t}(\tanh(V_a\mathbf{h}_j) \odot \text{sigm}(U_a\mathbf{h}_j))\}} \quad (2)$$

where the bias parameters are omitted for simplicity, and $\odot$ denotes the element-wise product. The sigmoid activation function is represented as 'sigm,' and N represents the total number of patch embeddings in each slide.

Attention pooling aggregates the feature representations $\{\mathbf{h}_k\}$ of all patches in the slide through averaging, with weights determined by their respective predicted attention scores $\{a_{k,t}\}$. The resulting feature vector, $\mathbf{h}_{\text{slide},t} \in \mathbb{R}^{512}$, serves as the histology deep features, representing the entire

Table 8 | Performance comparison of TR-MAMIL using different feature extractor methods on the testing sets in (a) classification of aggressive and non-aggressive EC, (b) TMB prediction for the aggressive EC, and (c) TMB prediction for the non-aggressive EC

Task	Backbone	Model selection	Acc.	Sens.	Spec.	MeanSS	AUROC
(a) Classification of aggressive and non-aggressive EC	Truncated ResNet50	F1-score	0.89	0.93	0.84	0.89	0.97
	Truncated ResNet101		0.84	0.80	0.89	0.84	0.86
	Truncated ResNet152		0.86	0.83	0.90	0.87	0.93
	CCL-ResNet50 ⁶³		0.58	1.00	0.00	0.50	0.53
	CCL-ResNet152		0.58	1.00	0.00	0.50	0.52
	DINO-ViT-S/16 ⁶⁴		0.84	0.90	0.77	0.83	0.90
	Lunit's DINO-ViT-S/16 ⁵⁸		0.86	0.82	0.90	0.86	0.91
	Lunit's DINO-ViT-S/16 with data normalization ⁵⁸		0.85	0.83	0.87	0.85	0.91
(b) TMB prediction for aggressive ECs	Truncated ResNet50	Cross-Entropy	0.67	0.78	0.61	0.69	0.69
	Truncated ResNet101		0.71	0.75	0.68	0.71	0.78
	Truncated ResNet152		0.73	0.63	0.73	0.73	0.78
	CCL-ResNet50 ⁶³		0.58	1.00	0.00	0.50	0.50
	CCL-ResNet152		0.58	1.00	0.00	0.50	0.48
	DINO-ViT-S/16 ⁶⁴		0.65	0.83	0.54	0.69	0.70
	Lunit's DINO-ViT-S/16 ⁵⁸		0.70	0.75	0.66	0.71	0.76
	Lunit's DINO-ViT-S/16 with data normalization ⁵⁸		0.72	0.69	0.73	0.71	0.79
(c) TMB prediction for non-aggressive ECs	Truncated ResNet50	F1-score	0.64	0.57	0.67	0.62	0.69
	Truncated ResNet101		0.69	0.53	0.84	0.61	0.65
	Truncated ResNet152		0.66	0.76	0.61	0.68	0.70
	CCL-ResNet50 ⁶³		0.58	1.00	0.00	0.50	0.67
	CCL-ResNet152		0.58	1.00	0.00	0.50	0.43
	DINO-ViT-S/16 ⁶⁴		0.59	0.43	0.67	0.55	0.51
	Lunit's DINO-ViT-S/16 ⁵⁸		0.48	0.57	0.44	0.51	0.51
	Lunit's DINO-ViT-S/16 with data normalization ⁵⁸		0.55	0.38	0.63	0.50	0.58

The bold case highlights the recommended setup of the proposed framework.

slide for task t .

$$\mathbf{h}_{slide,t} = \sum_{k=1}^K a_{k,t} \mathbf{h}_k \quad (3)$$

This trainable aggregation function intuitively enables the network to automatically identify informative regions in the slide, facilitating the classification of aggressive and non-aggressive EC and TMB prediction without the need for detailed annotations outlining precise tumor regions.

Stable covariate integration and classification

A stable factor s is encoded as an additional covariate with a binary value and concatenated to the slide-level deep features extracted from the histology slide (see Fig. 7(c)), generating a feature vector $\text{concat}([\mathbf{h}_{slide,t}, s])$. For the final classification cls layer, the slide-level probability prediction score p_t for task t is computed with a softmax function as follows.

$$p_t = \text{softmax}(W_{cls,t}(\text{concat}([\mathbf{h}_{slide,t}, s])) + \mathbf{b}_{cls,t}) \quad (4)$$

where softmax denotes the softmax activation and concat denotes concatenation. This study treats all tasks, including classification of the aggressive and non-aggressive EC and TMB prediction for the aggressive

and non-aggressive ECs, as binary classification problems. Therefore, the task-specific classification layers are defined by $W_{cls,t} \in \mathbb{R}^{2 \times 513}$.

Impact of incorporating a stable factor as a covariate

We further investigated the impact of incorporating the gender information as a stable factor covariate with the slide-level features by comparing model performance with or without the covariate in EC subtyping and TMB prediction on the testing sets. As shown in Table 9, adding the covariate improved the model performance by an average of 7% and 9% in terms of MeanSS and AUROC, respectively. The improvement in the aggressive and non-aggressive EC classification was even more significant, with increases of 24% in MeanSS and AUROC, respectively. Furthermore, as demonstrated in Fig. 6(c), the performance of TR-MAMIL with the covariate consistently performed better in all metrics when compared with the one without using the covariate on the same backbone.

Integrating a stable factor adds a concatenation layer, potentially allowing the model to learn more complex features by combining information from different network layers. This additional concatenation layer could enhance the expressive power of the model and improve performance in various tasks⁶⁵. Further investigation focused on cases where both models (with and without the covariate) failed in prediction. Figure 6(d) demonstrates that models with the stable covariate consistently gain lower losses than those without the covariate on the validation sets for all three applications. This lower loss indicates that the additional covariate acts as a

Table 9 | Performance comparison of models with or without the covariate on the testing sets in (a) classification of aggressive and non-aggressive EC, (b) TMB prediction for aggressive ECs, and (c) TMB prediction for non-aggressive ECs

Task	Model selection	Backbone	Acc.	Sens.	Spec.	MeanSS	AUROC
(a) Classification of aggressive and non-aggressive EC	F1-score	Truncated ResNet50 with the covariate	0.89	0.93	0.84	0.89	0.97
		Truncated ResNet101	0.84	0.80	0.89	0.84	0.86
		Truncated ResNet152	0.86	0.83	0.90	0.87	0.93
		Truncated ResNet50 without the covariate	0.67	0.76	0.54	0.65	0.73
		Truncated ResNet101	0.77	0.71	0.86	0.78	0.87
		Truncated ResNet152	0.81	0.82	0.79	0.80	0.85
(b) TMB prediction for aggressive ECs	Cross-Entropy	Truncated ResNet50 with the covariate	0.67	0.78	0.61	0.69	0.69
		Truncated ResNet101	0.71	0.75	0.68	0.71	0.78
		Truncated ResNet152	0.73	0.72	0.73	0.73	0.78
		Truncated ResNet50 without the covariate	0.66	0.56	0.73	0.64	0.62
		Truncated ResNet101	0.72	0.61	0.79	0.70	0.67
		Truncated ResNet152	0.67	0.75	0.63	0.69	0.71
(c) TMB prediction for non-aggressive ECs	F1-score	Truncated ResNet50 with the covariate	0.64	0.57	0.67	0.62	0.69
		Truncated ResNet101	0.69	0.38	0.84	0.61	0.65
		Truncated ResNet152	0.66	0.76	0.61	0.68	0.70
		Truncated ResNet50 without the covariate	0.53	0.52	0.54	0.53	0.58
		Truncated ResNet101	0.69	0.48	0.79	0.63	0.60
		Truncated ResNet152	0.63	0.48	0.70	0.59	0.60

The bold case highlights the recommended setup of the proposed framework.

regularizer, preventing overfitting to the training data. Overfitting occurs when the model memorizes training data too well and fails to generalize to unseen examples. By incorporating the stable covariate, the model is forced to learn more generalizable relationships, resulting in lower training and testing losses.

Model selection with early stop mechanism

We devised two model selection strategies with early stop mechanisms for different tasks (see Fig. 7(d)). The performance of the model was evaluated on the validation set in each epoch d , generating a performance measurement indicator Q_d . If Q_d on the task t had not improved beyond δ_1 epochs, the early stop mechanism was triggered at the epoch d^* , and then the training continues for another δ_2 epochs and stops at the epoch d^e where $\delta_1 = 50$, $\delta_2 = 20$ in this study. The proposed model selection selected the model M_i^* with the optimal Q_d between d^* and d^e . If multiple models had the same best score, the model with the earliest/lowest training epochs was chosen for the final model prediction through per-slide inference.

For the tasks in prediction of both cancer subtyping and TMB prediction in the non-aggressive type, the F1-score (F_d) is used as the model performance measurement indicator Q_d for model selection and early stop.

$$\{i\} = \arg \max_{d^* \leq j \leq d^e} F_j \quad (5)$$

$$i^* = \arg \min_i \{i\} \quad (6)$$

Comparison of backbones and strategies for model selection and early stop mechanisms

we examined various combinations of three truncated backbones with different depths and three different metrics for both model selection and

early stop of TR-MAMIL for EC subtyping and TMB prediction. In the classification of aggressive and non-aggressive EC, TR-MAMIL obtained excellent results in all setups with AUROC greater than 90% mostly, and the optimal setup with truncated ResNet50 as the backbone and F1-score as the metric for model selection and early stopping (see Table 10(a)).

In TMB prediction of the aggressive EC, the best performance was observed when using truncated ResNet152 as the backbone and cross-entropy value as the metric (see Table 10(b)). In TMB prediction of the non-aggressive EC, we determined that the optimal strategy by employing the truncated ResNet152 backbone with F1-score as the evaluation metric (see Table 10(c)).

For the task in prediction of TMB in the aggressive type, cross-entropy ($\mathcal{L}_d$) is used as the model performance measurement indicator Q_d for model selection and early stop.

$$\{i\} = \arg \min_{d^* \leq j \leq d^e} \mathcal{L}_j \quad (7)$$

$$i^* = \arg \min_i \{i\} \quad (8)$$

Implementation details

For all experiments, we performed a random sampling of slides with a mini-batch size of one. The sampling frequency for each slide was determined based on the inverse relative proportion of one class to another in the training set. The model parameters were updated using the Adam optimizer with weight decay of 1×10^{-5} , a learning rate of 2×10^{-4} , the first and second moments of the gradient were set to 0.9 and 0.999, and the epsilon was set to 1×10^{-8} . We applied dropout layers with a dropout rate of 0.25 after every hidden layer to prevent the potential overfitting issue.

Table 10 | Comparison of backbones and strategies for model selection with early stop mechanisms on the testing sets in (a) the classification of aggressive and non-aggressive EC, (b) TMB prediction for the aggressive EC, and (c) TMB prediction for the non-aggressive EC

Task	Metric	Backbone	Acc.	Sens.	Spec.	MeanSS	AUROC
(a) Classification of aggressive and non-aggressive EC	Cross-Entropy	ResNet50 Truncated	0.85	0.81	0.90	0.86	0.89
		ResNet101 Truncated	0.87	0.92	0.80	0.86	0.92
		ResNet152 Truncated	0.86	0.81	0.91	0.86	0.91
	MeanSS	ResNet50 Truncated	0.87	0.90	0.83	0.86	0.91
		ResNet101 Truncated	0.89	0.91	0.90	0.88	0.91
		ResNet152 Truncated	0.87	0.87	0.87	0.87	0.91
	F1-score	ResNet50 Truncated	0.89	0.93	0.84	0.89	0.97
		ResNet101 Truncated	0.84	0.80	0.89	0.84	0.86
		ResNet152 Truncated	0.86	0.83	0.90	0.87	0.93
(b) TMB prediction for aggressive ECs	Cross-Entropy	ResNet50 Truncated	0.67	0.78	0.61	0.69	0.69
		ResNet101 Truncated	0.71	0.75	0.68	0.71	0.78
		ResNet152 Truncated	0.73	0.72	0.73	0.73	0.78
	MeanSS	ResNet50 Truncated	0.66	0.67	0.66	0.66	0.67
		ResNet101 Truncated	0.71	0.75	0.68	0.71	0.78
		ResNet152 Truncated	0.66	0.86	0.54	0.70	0.73
	F1-score	ResNet50 Truncated	0.64	0.75	0.57	0.66	0.74
		ResNet101 Truncated	0.70	0.71	0.70	0.70	0.70
		ResNet152 Truncated	0.66	0.85	0.55	0.70	0.80
(c) TMB prediction for non-aggressive ECs	Cross-Entropy	ResNet50 Truncated	0.61	0.32	0.76	0.54	0.49
		ResNet101 Truncated	0.67	0.09	0.98	0.53	0.58
		ResNet152 Truncated	0.67	0.10	0.95	0.52	0.57
	MeanSS	ResNet50 Truncated	0.64	0.57	0.67	0.62	0.66
		ResNet101 Truncated	0.72	0.33	0.91	0.62	0.66
		ResNet152 Truncated	0.66	0.46	0.76	0.61	0.62
	F1-score	ResNet50 Truncated	0.64	0.57	0.67	0.62	0.69
		ResNet101 Truncated	0.69	0.38	0.84	0.61	0.65
		ResNet152 Truncated	0.66	0.76	0.61	0.68	0.70

The bold case highlights the recommended setup of the proposed framework.

Data availability

The whole-section WSIs of EC patients from the TCGA platform in application to the cancer subtype and TMB prediction that support the findings of this study are publicly available online through the TCGA's Genomic Data Commons (<https://portal.gdc.cancer.gov/>) with the project ID as TCGA-UCEC. The exact case IDs and associated labels can be accessed on (<https://docs.google.com/spreadsheets/d/1FUBoG5nSZLBd3Xgz4r8DJhutv3cIHFmb/edit?gid=1311982976&gid=1311982976>). Full clinical information could be downloaded from (https://www.cbioportal.org/study/summary?id=ucec_tcga_pan_can_atlas_2018) (Uterine Corpus Endometrial Carcinoma (TCGA, PanCancer Atlas)). All reasonable requests for academic use of the independent EC TMAs dataset can be addressed to the corresponding authors.

Code availability

The proposed TR-MAMIL was deployed using the pytorch framework in Python, and the program code has been made publicly accessible on GitHub (https://github.com/cwwang1979/TR-MAMIL_EndometrialCancer_Subtype_TMB). For the reviewers, please use the password 'WangLabCSTMBENDO' to decompress the program file.

Received: 1 March 2024; Accepted: 19 November 2024;
Published online: 21 December 2024

References

- Raglan, O. et al. Risk factors for endometrial cancer: an umbrella review of the literature. *Int. J. Cancer* **145**, 1719–1730 (2019).
- Berek, J. S. et al. FIGO staging of endometrial cancer: 2023. *Int. J. Gynecol. Obstet.* **162**, 383–394 (2023).
- Lax, S. F., Pizer, E. S., Ronnett, B. M. & Kurman, R. J. Comparison of estrogen and progesterone receptor, Ki-67, and p53 immunoreactivity in uterine endometrioid carcinoma and endometrioid carcinoma with squamous, mucinous, secretory, and ciliated cell differentiation. *Hum. Pathol.* **29**, 924–931 (1998).
- Bokhman, J. V. Two pathogenetic types of endometrial carcinoma. *Gynecol. Oncol.* **15**, 10–17 (1983).
- Voss, M. A. et al. Should grade 3 endometrioid endometrial carcinoma be considered a type 2 cancer—a clinical and pathological evaluation. *Gynecol. Oncol.* **124**, 15–20 (2012).
- de Bortoli, T. et al. Tumour mutational burden and survival with molecularly matched therapy. *Eur. J. Cancer* **190**, 112925 (2023).
- Rieke, D. T. et al. Tumor mutational burden as a predictive biomarker for molecularly matched therapy in two independent pan-cancer cohorts. *J. Clin. Oncol.* **41**, 3066–3066 (2023).
- Cao, W. et al. Immunotherapy in endometrial cancer: rationale, practice and perspectives. *Biomark. Res.* **9**, 1–30 (2021).
- Lee, S., Lara, O., Karpel, H. & Pothuri, B. The association of tumor mutational burden, microsatellite stability, and mismatch repair

- deficiency in an endometrial cancer patient cohort (194). *Gynecol. Oncol.* **166**, S111 (2022).
10. Palmeri, M. et al. Real-world application of tumor mutational burden-high (TMB-high) and microsatellite instability (MSI) confirms their utility as immunotherapy biomarkers. *ESMO Open* **7**, 100336 (2022).
11. Hill, B. L. et al. Mismatch repair deficiency, next-generation sequencing-based microsatellite instability, and tumor mutational burden as predictive biomarkers for immune checkpoint inhibitor effectiveness in frontline treatment of advanced stage endometrial cancer. *Int. J. Gynecol. Cancer* **33**, 504–513 (2023).
12. Zhang, J., An, L., Zhou, X., Shi, R. & Wang, H. Analysis of tumor mutation burden combined with immune infiltrates in endometrial cancer. *Ann. Transl. Med.* **9**, 1–13 (2021).
13. Le, D. T. et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357**, 409–413 (2017).
14. Le, D. T. et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).
15. Büttner, R. et al. Implementing TMB measurement in clinical practice: considerations on assay requirements. *ESMO Open* **4**, e000442 (2019).
16. Meléndez, B. et al. Methods of measurement for tumor mutational burden in tumor tissue. *Transl. Lung Cancer Res.* **7**, 661 (2018).
17. Chalmers, Z. R. et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* **9**, 1–14 (2017).
18. Sadhwani, A. et al. Comparative analysis of machine learning approaches to classify tumor mutation burden in lung adenocarcinoma using histopathology images. *Sci. Rep.* **11**, 16605 (2021).
19. Niu, Y. et al. Predicting tumor mutational burden from lung adenocarcinoma histopathological images using deep learning. *Front. Oncol.* **12**, 927426 (2022).
20. Dammak, S., Cecchini, M. J., Breadner, D. & Ward, A. D. Using deep learning to predict tumor mutational burden from scans of H&E-stained multicenter slides of lung squamous cell carcinoma. *J. Med. Imaging* **10**, 017502–017502 (2023).
21. Huang, K. et al. Predicting colorectal cancer tumor mutational burden from histopathological images and clinical information using multi-modal deep learning. *Bioinformatics* **38**, 5108–5115 (2022).
22. Liu, Y., Huang, K., Yang, Y., Wu, Y. & Gao, W. Prediction of tumor mutation load in colorectal cancer histopathological images based on deep learning. *Front. Oncol.* **12**, 906888 (2022).
23. Xu, H. et al. Spatial heterogeneity and organization of tumor mutation burden with immune infiltrates within tumors based on whole slide images correlated with patient survival in bladder cancer. *J. Pathol. Inform.* **13**, 100105 (2022).
24. Li, J. et al. Predicting gastric cancer tumor mutational burden from histopathological images using multimodal deep learning. *Brief. Funct. Genom.* **23**, 228–238 (2024).
25. Sun, C. et al. Tumor mutation burden-related histopathologic features for predicting overall survival in gliomas using graph deep learning. *Am. J. Pathol.* **193**, 2111–2121 (2023).
26. Lu, M. Y. et al. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng.* **5**, 555–570 (2021).
27. Lu, M. Y. et al. Ai-based pathology predicts origins for cancers of unknown primary. *Nature* **594**, 106–110 (2021).
28. Shao, Z. et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Adv. Neural Inf. Process. Syst.* **34**, 2136–2147 (2021).
29. Xiang, H. et al. Multi-scale representation attention based deep multiple instance learning for gigapixel whole slide image analysis. *Med. Image Anal.* **89**, 102890 (2023).
30. Konstantinov, A. V. & Utkin, L. V. Multi-attention multiple instance learning. *Neural Comput. Appl.* **34**, 14029–14051 (2022).
31. Campanella, G. et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **25**, 1301–1309 (2019).
32. Wang, C.-W. et al. Interpretable attention-based deep learning ensemble for personalized ovarian cancer treatment without manual annotations. *Comput. Med. Imaging Graph.* **107**, 102233 (2023).
33. Wang, C.-W. et al. Deep learning can predict bevacizumab therapeutic effect and microsatellite instability directly from histology in epithelial ovarian cancer. *Lab. Investig.* **103**, 100247 (2023).
34. Vermij, L. et al. p53 immunohistochemistry in endometrial cancer: clinical and molecular correlates in the PORTEC-3 trial. *Mod. Pathol.* **35**, 1475–1483 (2022).
35. Inc, S. *SPSS for Windows, rel. 15.0. 1* (Inc, S, 2006).
36. Connor, E. V. & Rose, P. G. Management strategies for recurrent endometrial cancer. *Expert Rev. Anticancer Ther.* **18**, 873–885 (2018).
37. Lawlor, R. T. et al. Tumor mutational burden as a potential biomarker for immunotherapy in pancreatic cancer: systematic review and still-open questions. *Cancers* **13**, 3119 (2021).
38. Kang, Y.-J. et al. A scoping review and meta-analysis on the prevalence of pan-tumour biomarkers (DMMR, MSI, high TMB) in different solid tumours. *Sci. Rep.* **12**, 20495 (2022).
39. Yang, Y. et al. Cancer immunotherapy: harnessing the immune system to battle cancer. *J. Clin. Investig.* **125**, 3335–3337 (2015).
40. Bray, F. et al. Global Cancer Statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer J. Clin.* **68**, 394–424 (2018).
41. Brooks, R. A. et al. Current recommendations and recent progress in endometrial cancer. *Cancer J. Clin.* **69**, 258–279 (2019).
42. Odunsi, K. Immunotherapy in ovarian cancer. *Ann. Oncol.* **28**, viii1–viii7 (2017).
43. Boussiotis, V. A., Chatterjee, P. & Li, L. Biochemical signaling of PD-1 on T cells and its functional implications. *Cancer J.* **20**, 265–271 (2014).
44. Taube, J. M. et al. Association of PD-1, PD-1 ligands, and other features of the tumor immune microenvironment with response to anti-PD-1 therapy. *Clin. Cancer Res.* **20**, 5064–5074 (2014).
45. Mellman, I., Coukos, G. & Dranoff, G. Cancer immunotherapy comes of age. *Nature* **480**, 480–489 (2011).
46. Wolf, M. T. et al. A biologic scaffold-associated type 2 immune microenvironment inhibits tumor formation and synergizes with checkpoint immunotherapy. *Sci. Transl. Med.* **11**, eaat7973 (2019).
47. Zuazo, M. et al. Functional systemic CD4 immunity is required for clinical responses to PD-1/PD-1 blockade therapy. *EMBO Mol. Med.* **11**, e10293 (2019).
48. Panda, A. et al. Identifying a clinically applicable mutational burden threshold as a potential biomarker of response to immune checkpoint therapy in solid tumors. *JCO Precis. Oncol.* **1**, 1–13 (2017).
49. Litchfield, K. et al. Meta-analysis of tumor-and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* **184**, 596–614 (2021).
50. Goodman, A. M. et al. Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol. Cancer Therapeut.* **16**, 2598–2608 (2017).
51. Marabelle, A. et al. Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 keynote-158 study. *Lancet Oncol.* **21**, 1353–1365 (2020).
52. Mahdi, H., Chelariu-Raicu, A. & Slomovitz, B. M. Immunotherapy in endometrial cancer. *Int. J. Gynecol. Cancer* **33**, 351–357 (2023).
53. Bogani, G. et al. Uterine serous carcinoma. *Gynecol. Oncol.* **162**, 226–234 (2021).
54. Choucair, K. et al. Tmb: a promising immune-response biomarker, and potential spearhead in advancing targeted therapy trials. *Cancer Gene Ther.* **27**, 841–853 (2020).

55. Bosse, T. et al. Molecular classification of grade 3 endometrioid endometrial cancers identifies distinct prognostic subgroups. *Am. J. Surg. Pathol.* **42**, 561–568 (2018).
56. Perone, C. S. & Cohen-Adad, J. Promises and limitations of deep learning for medical image segmentation. *J. Med. Artif. Intell.* **2**, 1–2 (2019).
57. Boland, C. R. et al. A national cancer institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* **58**, 5248–5257 (1998).
58. Kang, M., Song, H., Park, S., Yoo, D. & Pereira, S. Benchmarking self-supervised learning on diverse pathology datasets, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 3344–3354 (Vancouver, BC, Canada, 2023).
59. Zheng, Y. et al. Kernel attention transformer for histopathology whole slide image analysis and assistant cancer diagnosis. *IEEE Trans. Med. Imaging* **42**, 2726–2739 (2023).
60. Zhou, H.-Y. et al. A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics. *Nat. Biomed. Eng.* **7**, 743–755 (2023).
61. Shamshad, F. et al. Transformers in medical imaging: a survey. *Med. Image Anal.* **88**, 1–41 (2023).
62. Zhang, Y., Wang, J., Gorriz, J. M. & Wang, S. Deep learning and vision transformer for medical image analysis. *J. Imaging* **9**, 147 (2023).
63. Wang, X. et al. RetCCL: clustering-guided contrastive learning for whole-slide image retrieval. *Med. Image Anal.* **83**, 102645 (2023).
64. Caron, M. et al. Emerging properties in self-supervised vision transformers. IEEE/CVF International Conference on Computer Vision (ICCV). 9630–9640 (Montreal, Canada, 2021).
65. Cengil, E. & Çınar, A. The effect of deep feature concatenation in the classification problem: an approach on covid-19 disease detection. *Int. J. Imaging Syst. Technol.* **32**, 26–40 (2022).

Acknowledgements

This study is supported by National Science and Technology Council, Taiwan (NSTC 113-2222-E-011-003-MY3, NSTC 113-2321-B-016-003), Tri-Service General Hospital, Taipei, Taiwan (TSGH-A-112018, TSGH-A-113012 and TSGH-A-114007) and National Taiwan University of Science and Technology - Tri-Service General Hospital (NTUST-TSGH-114-03).

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

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Ching-Wei Wang or Tai-Kuang Chao.

Reprints and permissions information is available at

<http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024