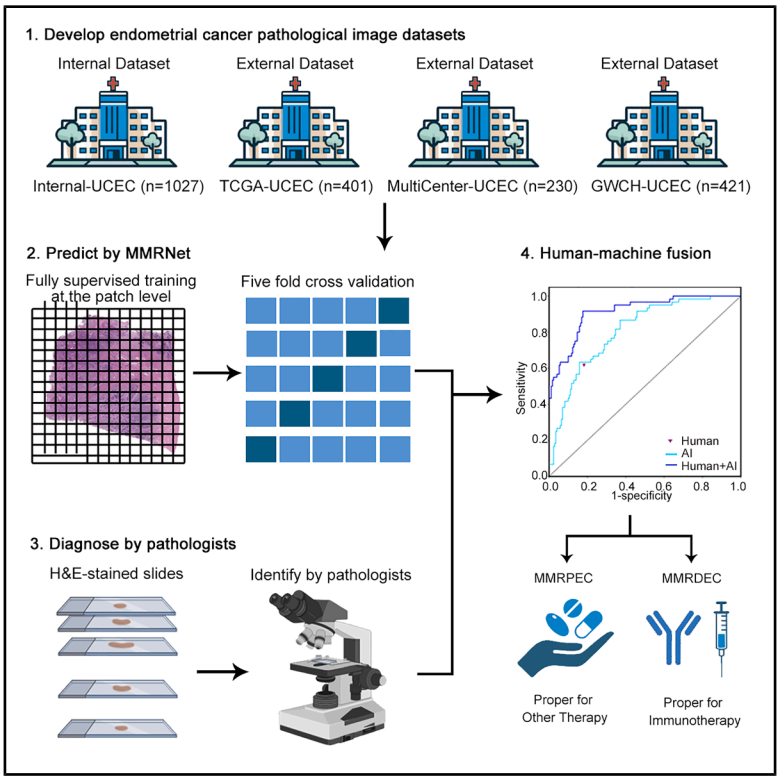
Cell Reports
Medicine

Article

MMRNet: Ensemble deep learning models for predicting mismatch repair deficiency in endometrial cancer from histopathological images

Graphical abstract



Highlights

- Mismatch repair deficiency is predictable from H&E slides with high accuracy
- MMR status is comprehensively assessed across diverse EC patient cohorts
- Simplifying the model and extracting MMR features enhance clinical applicability
- A human-machine fusion approach enhances the model's effectiveness

Authors

Li-Li Liu, Bing-Zhong Jing, Xuan Liu, ..., Hong-Yi Gao, Rong-Zhen Luo, Mu-Yan Cai

Correspondence

1522712335@qq.com (H.-Y.G.), luorzh@sysucc.org.cn (R.-Z.L.), caimy@sysucc.org.cn (M.-Y.C.)

In brief

Liu et al. leverage a diverse patient cohort across multiple centers to enhance the clinical applicability of MMRNet in EC. Simplifying the architecture and emphasizing MMR features, along with a human-machine fusion approach, improve performance and highlight its potential as a screening tool.

Liu et al., 2025, Cell Reports Medicine 6, 102099
May 20, 2025 © 2025 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.xcrm.2025.102099

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

MMRNet: Ensemble deep learning models for predicting mismatch repair deficiency in endometrial cancer from histopathological images

Li-Li Liu,[1,2,10] Bing-Zhong Jing,[1,3,10] Xuan Liu,[1,2,10] Rong-Gang Li,[5,10] Zhao Wan,[6] Jiang-Yu Zhang,[7] Xiao-Ming Ouyang,[8] Qing-Nuan Kong,[9] Xiao-Ling Kang,[4] Dong-Dong Wang,[4] Hao-Hua Chen,[1,3] Zi-Han Zhao,[1,2] Hao-Yu Liang,[1,2] Ma-Yan Huang,[1,2] Cheng-You Zheng,[1,2] Xia Yang,[1,2] Xue-Yi Zheng,[1,2] Xin-Ke Zhang,[1,2] Li-Jun Wei,[1,2] Chao Cao,[1,2] Hong-Yi Gao,[4,*] Rong-Zhen Luo,[1,2,*] and Mu-Yan Cai[1,2,11,*]

[1]State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou 510060, P. R. China
[2]Department of Pathology, Sun Yat-sen University Cancer Center, Guangzhou 510060, China
[3]Department of Information, Sun Yat-sen University Cancer Center, Guangzhou 510060, China
[4]Department of Pathology, Guangdong Women and Children Hospital, Guangzhou 511400, China
[5]Department of Pathology, Jiangmen Central Hospital, Jiangmen 529000, China
[6]Department of Pathology, Zhuhai Maternal and Child Health Care Hospital, Zhuhai 519000, China
[7]Department of Pathology, Affiliated Cancer Hospital and Institute of Guangzhou Medical University, Guangzhou 510095, China
[8]Department of Pathology, The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou 510260, China
[9]Department of Pathology, Qingdao Municipal Hospital, Qingdao 266071, China
[10]These authors contributed equally
[11]Lead contact
*Correspondence: 1522712335@qq.com (H.-Y.G.), luorzh@sysucc.org.cn (R.-Z.L.), caimy@sysucc.org.cn (M.-Y.C.)
https://doi.org/10.1016/j.xcrm.2025.102099

SUMMARY

Combining molecular classification with clinicopathologic methods improves risk assessment and chooses therapies for endometrial cancer (EC). Detecting mismatch repair (MMR) deficiencies in EC is crucial for screening Lynch syndrome and identifying immunotherapy candidates. An affordable and accessible tool is urgently needed to determine MMR status in EC patients. We introduce MMRNet, a deep convolutional neural network designed to predict MMR-deficient EC from whole-slide images stained with hematoxylin and eosin. MMRNet demonstrates strong performance, achieving an average area under the receiver operating characteristic curve (AUROC) of 0.897, with a sensitivity of 0.628 and a specificity of 0.949 in internal cross-validation. External validation using three additional datasets results in AUROCs of 0.790, 0.807, and 0.863. Employing a human-machine fusion approach notably improves diagnostic accuracy. MMRNet presents an effective method for identifying EC cases for confirmatory MMR testing and may assist in selecting candidates for immunotherapy.

INTRODUCTION

Endometrial cancer (EC), the third most prevalent gynecological cancer worldwide, continues to rise in incidence each year.[1] Recent advances in risk stratification have improved treatment algorithms through molecular classification.[2,3] The Cancer Genome Atlas (TCGA) has been instrumental in this effort,[4–6] categorizing EC into four molecular subtypes: polymerase ε ultramutated (POLEmut), p53 abnormal (p53abn), mismatch repair-deficient (MMRD), and no specific molecular profile (NSMP).[7]

Mismatch repair-deficient EC (MMRDEC), representing 20%–30% of cases, shows a favorable prognosis with immune checkpoint inhibitor treatment.[1,8] Although initially associated with Lynch syndrome,[9] the emphasis has shifted to predicting responses to these inhibitors.[10] High-grade microsatellite instability (MSI-H) is assessed using MSI-based PCR and next-generation sequencing (NGS),[11,12] while immunohistochemistry (IHC) can reveal the loss of mismatch repair protein expression, serving as a reliable MSI surrogate.[12,13] However, these tests are resource intensive and time consuming.[14,15] Despite recommendations for MMR or MSI testing in EC patients, accessibility and cost limitations have hindered widespread adoption. For example, internal data from the Ohio State University Medical Center revealed that the cost of MMR or MSI testing was $425,[16] while histomorphology classification using MMR IHC biomarkers in EC patients costs $74.32, resulting in an aggregate annual cost of $72,950 in Ontario, Canada.[17] These figures highlight the significant financial burden of traditional methods. As a result, MMR immunohistochemical testing has primarily been reserved for EC patients with a high likelihood of MMR-deficient disease, based on initial evaluations of H&E slides by pathologists. However, even with specialized gynecological

This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

expertise, pathologists face challenges in accurately identifying MMR-deficient EC from hematoxylin and eosin (H&E) slides.[18–20]

A cost-effective, efficient tool to screen MMR status in EC is urgently needed to guide patient care and reduce unnecessary costs. Deep learning models have shown potential in detecting cancer subtypes and molecular characteristics from H&E-stained whole-slide images (WSIs), potentially serving as effective cancer biomarkers.[21] These models have accurately predicted EC molecular subtypes,[22] Epstein-Barr virus status in gastric cancer,[23] and MSI status in EC, gastric, and colorectal cancers using H&E-stained WSIs.[22–25] We propose that a deep learning model could enhance MMRDEC prediction and improve confirmatory MMR test selection. In this study, we aimed to develop and validate MMRNet, an effective deep learning model for predicting MMR status in EC patients using H&E-stained slides. Our study leverages a diverse patient cohort across multiple centers, including an Asian population, to enhance the clinical applicability of MMRNet in EC. Moreover, we designed a simple yet powerful human-machine fusion approach, combining the strengths of both the deep learning model and pathologists to enhance diagnostic performance.

RESULTS

Patient cohorts

To evaluate the effectiveness of MMRNet in predicting MMR status in EC, we utilized four distinct cohorts (Figure S2). MMRNet was developed using the Internal-Uterine Corpus Endometrial Carcinoma (Internal-UCEC) dataset, which comprised 242 H&E-stained WSIs from 242 MMRDEC patients and 785 WSIs from 784 MMR-proficient endometrial cancer (MMRPEC) patients at a single medical center. Additionally, three independent external validation datasets—TCGA-UCEC, MultiCenter-UCEC, and GWCH-UCEC—were used. The TCGA-UCEC (https://portal.gdc.cancer.gov/) dataset included 401 WSIs from 369 patients, with 135 WSIs from 127 MMRDEC patients and 266 WSIs from 242 MMRPEC patients. The MultiCenter-UCEC dataset contained 230 WSIs from 230 patients, with 60 MMRDEC and 170 MMRPEC patients. The GWCH-UCEC dataset comprised 421 WSIs from 421 patients, with 88 MMRDEC and 333 MMRPEC patients. Further details of these datasets are provided in Table S1.

Development and evaluation of MMRNet

To predict whether a patient belongs to MMRD or MMRP subgroup, we utilized an ensemble binary categorizer named MMRNet through various network backbones (e.g., EfficientNet and ResNet18). The dataset was divided into five parts at the slide level, and each of the five individual classifiers underwent training using a 5-fold cross-validation strategy. Specifically, during training, each individual categorizer utilized four pieces for training purposes, while the remaining segment served as an internal validation set to determine the optimal stopping point for training the categorizer.[26] By implementing 5-fold cross-validation, the model is exposed to a diverse range of histologic patterns, which helps mitigate the impact of slide-to-slide variability and focal heterogeneity. Previous study[27] supports using cross-validation to manage datasets with high heterogeneity, as it reduces variance and prevents overfitting to specific data subsets. This approach ensures more reliable and stable performance across diverse histological presentations. To mitigate the bias due to tumor size, we designed MMRNet to aggregate patch-level predictions through simple averaging to produce slide-level results. Specifically, the large WSIs from TCGA were segmented into non-overlapping tiles at 10× magnification after applying color normalization and resizing to 224 × 224 pixels. This approach reduces computational load while minimizing bias from tissue heterogeneity, ensuring that diverse tumor regions are appropriately analyzed. Furthermore, fully supervised patch-level training allows MMRNet to focus on learning task-specific features, thereby enhancing both efficiency and specificity in identifying MMR-related patterns. Then, MMRNet underwent internal evaluation using the Internal-UCEC dataset and external assessment using three external datasets: TCGA-UCEC, MultiCenter-UCEC, and GWCH-UCEC. Further details regarding external testing and internal evaluation were described in our previous study.[28] Ultimately, the predicted MMR status of the entire set was compared against the ground truth MMR status for validation.

Diagnostic performance of MMRNet

To develop a deep learning model for determining MMR status, we initially trained ResNet18 using an internal dataset.[28] We then evaluated the model's performance across three external cohorts. Figure 1 illustrates the comprehensive workflow of MMRNet. In the internal dataset, the area under the receiver operating characteristic curve (AUROC) ranged from 0.885 to 0.916 across test folds (Table S2). Overall, MMRNet achieved an AUROC of 0.897, a sensitivity of 0.628, a specificity of 0.949, and a negative predictive value (NPV) of 0.892 (Table S2). In the TCGA-UCEC dataset, MMRNet achieved an AUROC of 0.790 (95% confidence interval [CI]: 0.784–0.795), a sensitivity of 0.808 (95% CI: 0.801–0.815), and a specificity of 0.570 (95% CI: 0.563–0.577). In the MultiCenter-UCEC dataset, the model yielded an AUROC of 0.807 (95% CI: 0.800–0.813), a sensitivity of 0.833 (95% CI: 0.824–0.843), and a specificity of 0.629 (95% CI: 0.621–0.638). In the GWCH-UCEC dataset, MMRNet achieved an AUROC of 0.863 (95% CI: 0.858–0.868), a sensitivity of 0.784 (95% CI: 0.774–0.794), and a specificity of 0.802 (95% CI: 0.797–0.806) (Table S3; Figure S3). We also conducted a health economics evaluation comparing MMRNet with conventional diagnostic methods. We assessed time, cost, manpower, and equipment requirements for each method. MMRNet (total time: 1 day; general cost: ¥60) demonstrated clear advantages in efficiency and affordability over IHC (total time: 2 days; general cost: ¥1550), PCR (total time: 3 days; general cost: ¥1610), and NGS (total time: 15 days; general cost: ¥6000) (Figure S4). To evaluate the clinical relevance of MMRNet, we analyzed overall survival (OS) and disease-free survival (DFS) in TCGA and our internal datasets. Consistent with previous studies, no statistically significant differences in prognosis were observed between MMRDEC and MMRPEC groups, whether classified by true MMR status or MMRNet predictions. We further focused on patients within our internal dataset who received immune checkpoint inhibitors. Both true MMR status or MMRNet classification showed a trend toward longer OS for MMRDEC patients compared to MMRPEC patients following

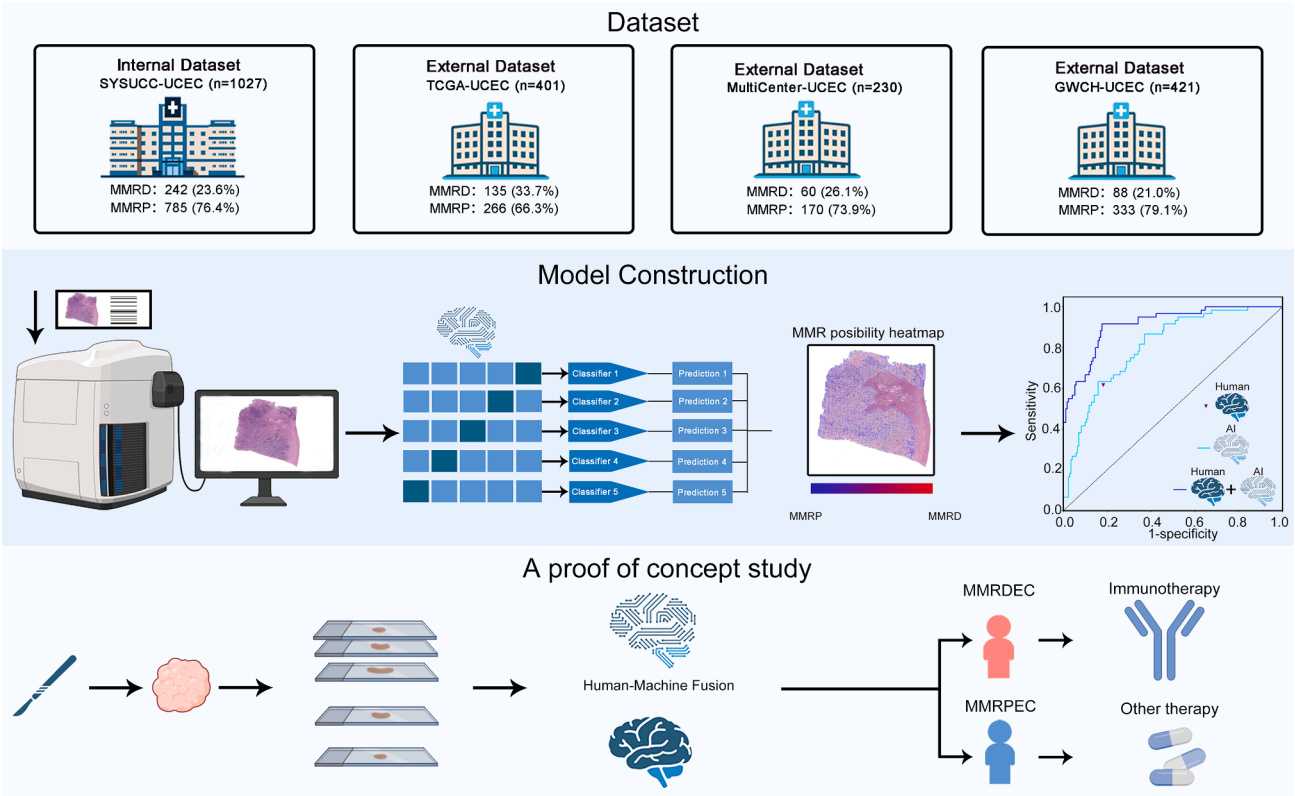


Figure 1. Workflow of MMRNet for predicting MMR status using H&E-stained WSIs

WSIs were initially segmented into non-overlapping ×10 magnification tiles. Following color normalization and resizing to 224 × 224 pixels, the tiles underwent cancer detection. Tiles identified as tumor regions were then inputted into MMRNet to derive probability of tile-level MMR status. The MMRNet ensemble model integrated outputs from five individual classifiers at the output layer. Tile-level probabilities were averaged to determine slide-level MMR status. MMR, DNA mismatch repair; WSI, whole-slide image.

immunotherapy, although these differences were not statistically significant. Figure S5 shows that MMRNet has the potential to support clinical outcome classification.

Pathologist reader's performance

To compare MMRNet's diagnostic performance in predicting MMR status in EC with that of pathologists, a reader study was conducted using external datasets: TCGA-UCEC, MultiCenter-UCEC, and GWCH-UCEC cohorts. Pathologists assessed H&E-stained slides to determine MMR status based on morphological characteristics. Notably, their diagnostic abilities improved with experience. In the TCGA-UCEC cohort, the following AUROCs were observed: Junior pathologist 1 (0.648, 95% CI: 0.597–0.701), Junior pathologist 2 (0.638, 95% CI: 0.585–0.691), Senior pathologist 1 (0.665, 95% CI: 0.607–0.723), Senior pathologist 2 (0.641, 95% CI: 0.587–0.697), Expert pathologist 1 (0.708, 95% CI: 0.659–0.759), and Expert pathologist 2 (0.702, 95% CI: 0.651–0.756). Sensitivities ranged from 0.362 to 0.614 and specificities from 0.703 to 0.814. For the MultiCenter-UCEC dataset, the AUROCs were as follows: Junior pathologist 1 (0.636, 95% CI: 0.554–0.718), Junior pathologist 2 (0.646, 95% CI: 0.565–0.727), Senior pathologist 1 (0.664, 95% CI: 0.585–0.743), Senior pathologist 2 (0.659, 95% CI: 0.579–0.739), Expert pathologist 1 (0.831, 95% CI: 0.774–0.885), and Expert pathologist 2 (0.825, 95% CI: 0.762–0.889). Sensitivities ranged from 0.450 to 0.617 and specificities from 0.782 to 0.871. In the GWCH-UCEC dataset, the AUROCs were as follows: Junior pathologist 1 (0.687, 95% CI: 0.630–0.745), Junior pathologist 2 (0.717, 95% CI: 0.666–0.767), Senior pathologist 1 (0.829, 95% CI: 0.785–0.875), Senior pathologist 2 (0.780, 95% CI: 0.726–0.834), Expert pathologist 1 (0.890, 95% CI: 0.855–0.924), and Expert pathologist 2 (0.890, 95% CI: 0.856–0.925). Sensitivities ranged from 0.432 to 0.716 and specificities from 0.757 to 0.895. MMRNet's AUROC significantly exceeded that of both junior and senior pathologists and closely matched the performance levels of expert pathologists across all three external datasets ($p < 0.001$) (Figure 2). Furthermore, MMRNet demonstrated superior sensitivity compared to all pathologists (Table 1). This consistent superiority in sensitivity, regardless of the pathologists' experience level, underscores MMRNet's utility in screening for potential MMRD, facilitating the selection of confirmatory MMR tests.

Human-machine fusion

We developed a human-machine fusion strategy to assess the clinical applicability of deep learning models by integrating

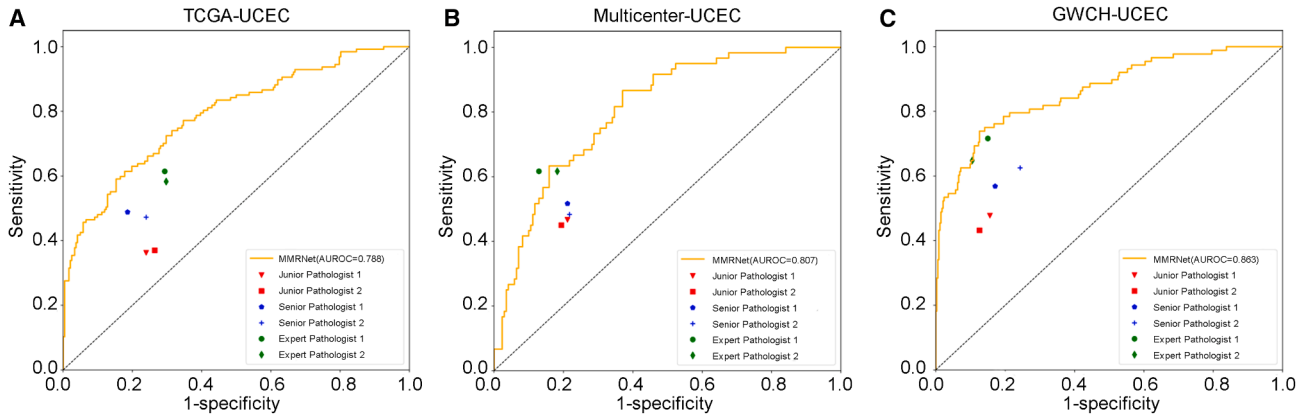


Figure 2. Diagnostic performances of MMRNet and pathologists on external datasets

AUROCs were computed for MMRNet and pathologists on the TCGA-UCEC (A), MultiCenter-UCEC (B), and GWCH-UCEC (C) datasets. MMRNet consistently exhibited significantly higher AUROCs in all three datasets ($p < 0.001$, Delong's test, two-sided). Each WSI was evaluated four times during the 5-fold validation, and the results were averaged. TCGA-UCEC, The Cancer Genome Atlas dataset; MultiCenter-UCEC, multiple medical centers dataset; GWCH-UCEC, Guangdong Women and Children Hospital dataset; AUROC, area under the receiver operating curve.

MMRNet into a universal testing framework. This approach combines MMRNet predictions with evaluations by pathologists of varying expertise levels to improve diagnostic accuracy. In the TCGA-UCEC cohort, the fusion strategy achieved AUROCs of 0.7579 (95% CI: 0.7107–0.8077), 0.7486 (95% CI: 0.7002–0.7995), 0.7853 (95% CI: 0.7373–0.8346), 0.7654 (95% CI: 0.7179–0.8157), 0.8016 (95% CI: 0.7582–0.8476), and 0.8011 (95% CI: 0.7552–0.8502) for Junior pathologist 1, Junior pathologist 2, Senior pathologist 1, Senior pathologist 2, Expert pathologist 1, and Expert pathologist 2, respectively, surpassing the performance of MMRNet alone (0.7880). The human-machine fusion approach achieved sensitivities ranging from 0.7559 to 0.9055 and specificities from 0.4421 to 0.5868 on the TCGA-UCEC dataset. Similarly, on the MultiCenter-UCEC dataset, combining MMRNet predictions with evaluations by Junior pathologist 1, Junior pathologist 2, Senior pathologist 1, Senior pathologist 2, Expert pathologist 1, and Expert pathologist 2 resulted in AUROCs of 0.7431 (95% CI: 0.6699–0.8165), 0.7501 (95% CI: 0.6770–0.8228), 0.7614 (95% CI: 0.6926–0.8322), 0.7512 (95% CI: 0.6800–0.8230), 0.9134 (95% CI: 0.8733–0.9537), and 0.8838 (95% CI: 0.8327–0.9359), respectively. The human-machine fusion approach achieved sensitivities ranging from 0.5500 to 0.7500 and specificities from 0.8000 to 0.8941 on the MultiCenter-UCEC cohort. Furthermore, on the GWCH-UCEC dataset, combining MMRNet with evaluations by Junior pathologist 1, Junior pathologist 2, Senior pathologist 1, Senior pathologist 2, Expert pathologist 1, and Expert pathologist 2 resulted in AUROCs of 0.7983 (95% CI: 0.7529–0.8443), 0.8507 (95% CI: 0.8125–0.8892), 0.8932 (95% CI: 0.8592–0.9271), 0.8751 (95% CI: 0.8337–0.9179), 0.9322 (95% CI: 0.9056–0.9586), and 0.9287 (95% CI: 0.9011–0.9561), respectively. The human-machine fusion strategy achieved sensitivities ranging from 0.5114 to 0.9205 and specificities from 0.5435 to 0.8559 on the GWCH-UCEC cohort (Table 2; Figure 3). These findings underscore the effectiveness of the human-machine fusion approach in improving the prediction of MMR status in EC.

Correlation between histopathological features and the MMRNet's prediction

To understand the intrinsic mechanisms of the deep learning model, we examined how histopathological features relate to MMRNet's predictions for MMRDEC using logistic regression models. In the TCGA-UCEC dataset, predicting MMRDEC was significantly linked to prominent tumor-infiltrating lymphocytes (TILs) (odds ratio [OR], 6.135; $p < 0.001$) and prominent peritumoral lymphocytes (OR, 3.544; $p < 0.001$). The MultiCenter-UCEC dataset revealed correlations between MMRDEC prediction and prominent peritumoral lymphocytes (OR, 2.930; $p = 0.016$), solid sheets of round-to-oval cells (OR, 9.047; $p < 0.001$), and lymphoepithelioma-like morphology (OR, 8.826; $p = 0.033$). Moreover, in the GWCH-UCEC dataset, MMRDEC prediction was strongly associated with prominent TILs (OR, 7.806; $p < 0.001$), prominent peritumoral lymphocytes (OR, 4.069; $p < 0.001$), solid sheets of round-to-oval cells (OR, 2.985; $p = 0.002$), and background endometrial hyperplasia (OR, 1.724; $p = 0.008$) (Figure 4; Table 3). Table S4 provides the frequencies of these morphological characteristics across the datasets. To directly visualize the regions MMRNet attends to, we overlaid heatmaps on WSIs. In correctly predicted cases, MMRDEC regions predominantly display red hues, while MMRPEC regions are primarily blue (Figure 4). In misclassified cases, the color patterns are reversed, suggesting discrepancies in feature attention (Figure S6). These findings offer crucial insights into how specific histopathological features play a pivotal role in MMRNet's accurate prediction of MMR status.

Misdiagnosis from MMRNet

To thoroughly assess MMRNet's performance, we conducted a comprehensive analysis of its misdiagnoses. In the TCGA-UCEC dataset, MMRNet misdiagnosed 187 out of 401 slides, comprising 10 cases of MMRDEC and 177 cases of MMRPEC. All 10 misdiagnosed MMRDEC slides were inaccurately assessed by at least one pathologist. Among the 177 misdiagnosed MMRPEC slides, 46 were incorrectly diagnosed by all

Table 1. Performance comparison between MMRNet and pathologists on TCGA-UCEC, MultiCenter-UCEC, and GWCH-UCEC

	TCGA-UCEC				MultiCenter-UCEC				GWCH-UCEC			
Methods	Sensitivity	Difference	*p*	Specificity	Difference	*p*	Sensitivity	Difference	*p*	Specificity	Difference	*p*	Sensitivity	Difference	*p*	Specificity	Difference	*p*
MMRNet	0.9843	NA	NA	0.1612	NA	NA	0.8333	NA	NA	0.6294	NA	NA	1.0000	NA	NA	0.0360	NA	NA
Junior pathologist 1	0.3622	−0.6221	<0.001	0.7603	0.5991	<0.001	0.4667	−0.3666	0.029	0.7882	0.1588	<0.001	0.4773	−0.5227	<0.001	0.8438	0.8078	<0.001
Junior pathologist 2	0.3701	−0.6142	<0.001	0.7355	0.5743	<0.001	0.4500	−0.3833	0.080	0.8059	0.1765	<0.001	0.4318	−0.5682	<0.001	0.8739	0.8379	<0.001
Senior pathologist 1	0.4882	−0.4961	<0.001	0.8140	0.6528	<0.001	0.5167	−0.3166	0.003	0.7882	0.1588	<0.001	0.5682	−0.4318	0.001	0.8288	0.7928	<0.001
Senior pathologist 2	0.4724	−0.5119	<0.001	0.7603	0.5991	<0.001	0.4833	−0.3500	0.015	0.7824	0.1530	<0.001	0.6250	−0.3750	0.017	0.7568	0.7208	0.195
Expert pathologist 1	0.5827	−0.4016	<0.001	0.7025	0.5413	<0.001	0.6167	−0.2166	<0.001	0.8176	0.1882	<0.001	0.6477	−0.3523	0.013	0.8949	0.8589	<0.001
Expert pathologist 2	0.6142	−0.3701	<0.001	0.7066	0.5454	<0.001	0.6167	−0.2166	<0.001	0.8706	0.2412	0.015	0.7159	−0.2841	0.108	0.8498	0.8138	<0.001

Note: Difference indicated the difference between each pathologist and MMRNet. TCGA-UCEC, external dataset from The Cancer Genome Atlas; MultiCenter-UCEC, external dataset from multiple medical centers; GWCH-UCEC, external dataset from Guangdong Women and Children Hospital; NA, not applicable.

pathologists, and 131 were misjudged by more than one pathologist. We analyzed the morphological features of MMRNet's misdiagnosed cases by comparing false positives with true negatives and false negatives with true positives. Interestingly, the 177 false-positive cases exhibited a significant presence of TILs (p = 0.029) (Figure S6; Table S5). Turning to the MultiCenter-UCEC dataset, MMRNet misdiagnosed 53 out of 230 slides, including 44 MMRDEC and 9 MMRPEC cases. Among these, 3 misdiagnosed MMRPEC slides were inaccurately assessed by at least one pathologist. Of the 44 misdiagnosed MMRDEC slides, 4 were misdiagnosed by all the pathologists, 31 were misdiagnosed by at least one pathologist, and 9 were inaccurately assessed by all pathologists. Compared to true negatives, the 9 false-positive cases showed papillary/polypoid ($p < 0.001$) and mucinous differentiation ($p < 0.001$). Additionally, the 44 false-negative cases lacked solid sheets of round-to-oval cells (p = 0.011) (Table S6). Finally, in the GWCH-UCEC dataset, MMRNet misdiagnosed 114 out of 421 slides, including 17 MMRDEC and 97 MMRPEC cases. Among these, 2 misdiagnosed MMRDEC slides were correctly identified by all pathologists, 2 were misdiagnosed by all pathologists, and the remaining 13 were inaccurately assessed by at least one pathologist. Of the 97 misdiagnosed MMRPEC slides, 46 were inaccurately assessed by all pathologists, and an additional 51 were misdiagnosed by at least one pathologist. Compared to true negatives, the 97 false-positive cases showed prominent peritumoral lymphocytes (p = 0.023) and solid sheets of round-to-oval cells (p = 0.001). The 17 false-negative cases lacked lower uterine segment (LUS) centered (p = 0.024) (Table S7). We also conducted a detailed analysis comparing the misdiagnoses made by human pathologists and the MMRNet model (Tables S8, S9, and S10). These findings offer critical insights into the specific histopathological characteristics contributing to MMRNet's misdiagnoses.

DISCUSSION

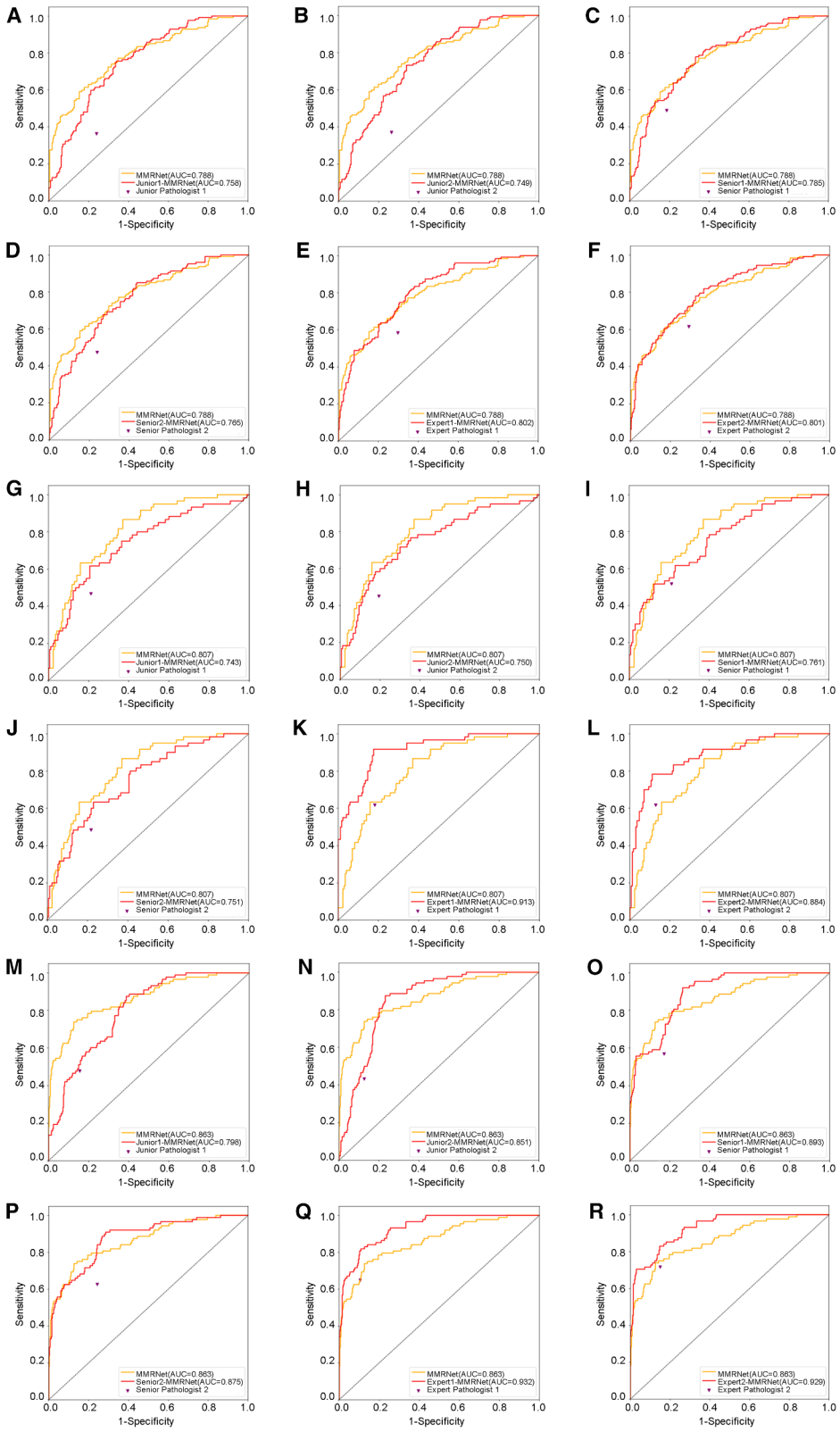
In this study, we utilized an internal dataset and three external datasets to evaluate MMRNet, an effective deep learning model, for predicting MMR status in H&E-stained WSIs of EC. Our results demonstrate that MMRNet surpasses pathologists in diverse clinical settings and institutions. Furthermore, integrating human-machine interaction enhances MMRNet's capability to identify the MMRDEC subtype, potentially serving as a biomarker for selecting EC patients suitable for immunotherapy.

While AI-based systems have shown promise in MSI testing across various tumor types,[22–25] their effectiveness in EC has been limited. For instance, Fremond et al. identified four molecular subtypes of EC using the im4MEC model, validated exclusively on high-risk EC cases from the PORTEC-3 trial.[22] In contrast, our study introduces several remarkable contributions to address existing gaps and extend the clinical applicability of MMR status prediction. Unlike the im4MEC model,[22] MMRNet was validated across an internal dataset and three external datasets, encompassing a broader spectrum of EC stages from multiple centers. This diverse cohort enhances the model's clinical relevance by providing a robust and comprehensive assessment of MMR status in EC. Most previous studies, including that of

Table 2. Prediction fusion from MMRNet and pathologists with varying levels of expertise

	TCGA-UCEC			MultiCenter-UCEC			GWCH-UCEC		
Methods	Sensitivity	Specificity	AUROC	Sensitivity	Specificity	AUROC	Sensitivity	Specificity	AUROC
MMRNet	0.9843 (0.94, 1.00)	0.1612 (0.12, 0.21)	0.7880 (0.7400, 0.8386)	0.8333 (0.71, 0.92)	0.6294 (0.55, 0.70)	0.8066 (0.7471, 0.8677)	1.0000 (0.96, 1.00)	0.0360 (0.02, 0.06)	0.8631 (0.8186, 0.9079)
Junior 1-MMRNet fusion	0.7795 (0.70, 0.85)	0.5868 (0.52, 0.65)	0.7579 (0.7107, 0.8077)	0.5500 (0.42, 0.68)	0.8118 (0.74, 0.87)	0.7431 (0.6699, 0.8165)	0.5909 (0.48, 0.69)	0.7898 (0.74, 0.83)	0.7983 (0.7529, 0.8443)
Junior 2-MMRNet fusion	0.7559 (0.67, 0.83)	0.5826 (0.52, 0.65)	0.7486 (0.7002, 0.7995)	0.5667 (0.43, 0.69)	0.8235 (0.76, 0.88)	0.7501 (0.6770, 0.8228)	0.5114 (0.40, 0.62)	0.8559 (0.81, 0.89)	0.8507 (0.8125, 0.8892)
Senior 1-MMRNet fusion	0.8425 (0.77, 0.90)	0.5372 (0.47, 0.60)	0.7853 (0.7373, 0.8346)	0.5500 (0.42, 0.68)	0.8000 (0.73, 0.86)	0.7614 (0.6926, 0.8322)	0.7614 (0.66, 0.85)	0.7988 (0.75, 0.84)	0.8932 (0.8592, 0.9271)
Senior 2-MMRNet fusion	0.8583 (0.79, 0.91)	0.5000 (0.44, 0.56)	0.7654 (0.7179, 0.8157)	0.5500 (0.42, 0.68)	0.8000 (0.73, 0.86)	0.7512 (0.6800, 0.8230)	0.9205 (0.84, 0.97)	0.5435 (0.49, 0.60)	0.8751 (0.8337, 0.9179)
Expert 1-MMRNet fusion	0.8976 (0.83, 0.94)	0.4793 (0.41, 0.54)	0.8016 (0.7582, 0.8476)	0.7500 (0.62, 0.85)	0.8647 (0.80, 0.91)	0.9134 (0.8733, 0.9537)	0.8636 (0.77, 0.93)	0.7808 (0.73, 0.82)	0.9322 (0.9056, 0.9586)
Expert 2-MMRNet fusion	0.9055 (0.84, 0.95)	0.4421 (0.38, 0.51)	0.8011 (0.7552, 0.8502)	0.7333 (0.60, 0.84)	0.8941 (0.84, 0.94)	0.8838 (0.8327, 0.9359)	0.8750 (0.79, 0.94)	0.7538 (0.70, 0.80)	0.9287 (0.9011, 0.9561)

Note: 95% confidence intervals are included in brackets. Junior 1-MMRNet fusion, the prediction fusion from the MMRNet and Junior pathologist 1; Junior 2-MMRNet fusion, the prediction fusion from the MMRNet and Junior pathologist 2; Senior 1-MMRNet fusion, the prediction fusion from the MMRNet and Senior pathologist 1; Senior 2-MMRNet fusion, the prediction fusion from the MMRNet and Senior pathologist 2; Expert 1-MMRNet fusion, the prediction fusion from the MMRNet and Expert pathologist 1; Expert 2-MMRNet fusion, the prediction fusion from the MMRNet and Expert pathologist 2; TCGA-UCEC, external dataset from The Cancer Genome Atlas; MultiCenter-UCEC, external dataset from multiple medical centers; GWCH-UCEC, external dataset from Guangdong Women and Children Hospital; AUROC, area under the receiver operating curve.



(legend on next page)

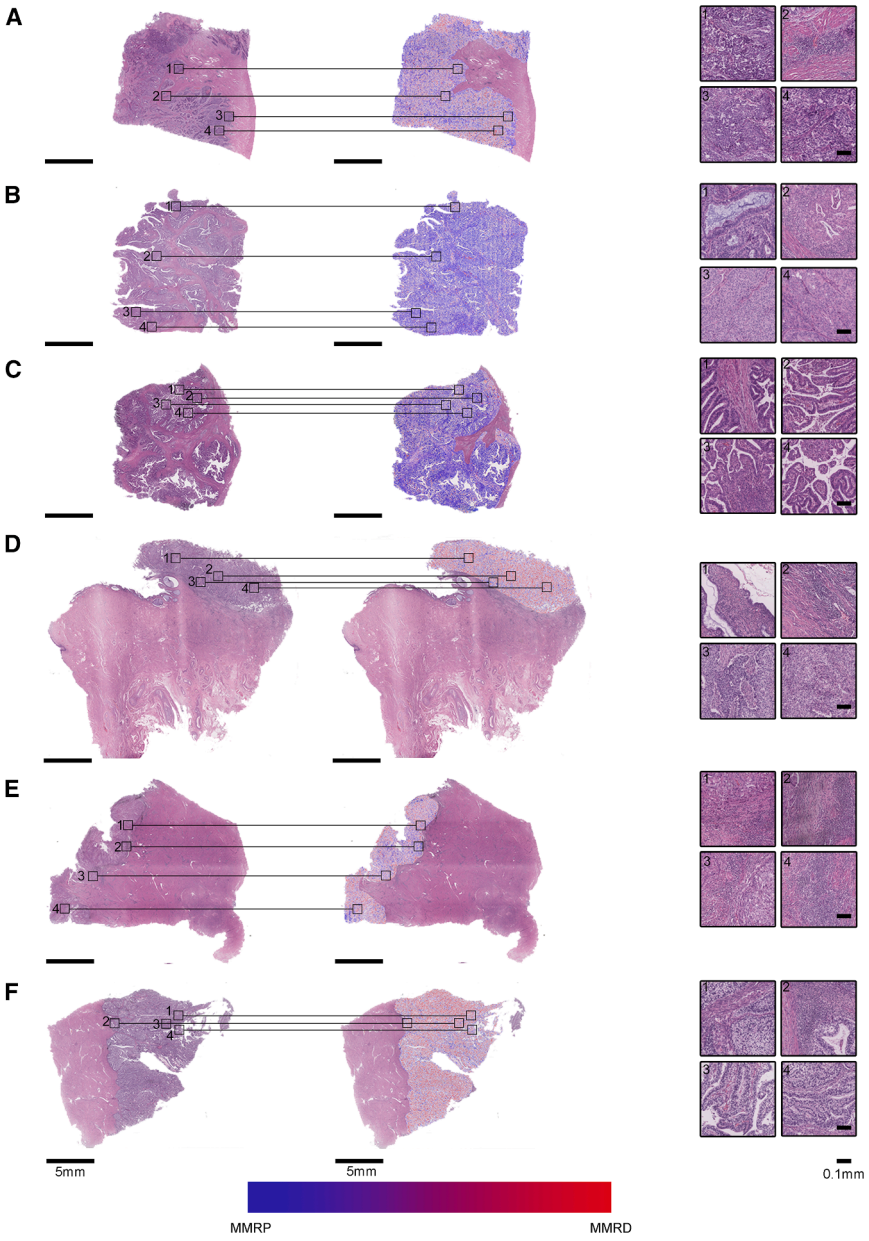


Figure 4. Successful cases predicted by MMRNet

(A–C) Histological images (left column, scale bar, 5 mm) from patients with MMRPEC in the Internal-UCEC dataset. Overlaid heatmaps on WSIs (middle column, scale bar, 5 mm) indicate MMRNet's prediction of MMRPEC with low MMR scores (bluish color) (right column, tiles at 10× magnification, scale bar, 0.1 mm).

(D–F) Histological images (left column, scale bar, 5 mm) from patients with MMRDEC in the Internal-UCEC dataset. Overlaid heatmaps on WSIs (middle column, scale bar, 5 mm) indicate MMRNet's prediction of MMRDEC with high scores (reddish color) (right column, tiles at 10× magnification, scale bar, 0.1 mm).

Fremond et al.,[22] have primarily focused on European cohorts. By incorporating a predominantly Chinese patient cohort, including two internal datasets from China and an external validation set from TCGA, our study broadens the generalizability of MMRNet, especially for underrepresented populations. This focus on an Asian demographic addresses a significant gap in existing research and offers outstanding insights into MMR status prediction across diverse populations. Additionally, other deep learning models, such as those trained on combined TCGA and Clinical Proteomic Tumor Analysis Consortium datasets, achieved an AUROC of 0.827, with significant performance decline on external cohorts (AUROC of 0.667).[4] By comparison, MMRNet maintained superior performance with an AUROC of 0.897, surpassing previous models (AUROC of 0.799)[6] and demonstrating greater consistency across external datasets. Furthermore, MMRNet achieved higher sensitivity than pathologists across three external datasets, highlighting its potential as a reliable screening tool for identifying EC patients appropriate for confirmatory MMR IHC testing and reducing the risk of misdiagnosis in MMR-deficient EC cases. Our study introduces a binary classification model based on ResNet18, distinct from previous work in terms of task setting, model architecture, and computational efficiency. Unlike the self-supervised approach employed by Fremond et al.,[22] which may overlook specific MMR-related features, our fully supervised training at the patch level enables MMRNet to prioritize task-specific characteristics from the outset. This method enhances both efficiency and specificity in MMR feature learning. Slide-level predictions are achieved through a straightforward aggregation of patch-level results via averaging, a streamlined approach that demonstrates high

Figure 3. Integration of predictions from MMRNet and individual pathologists on external datasets

(A–F) depict AUROCs achieved by the fusion of MMRNet and pathologists with varying expertise on the TCGA-UCEC dataset. (G–L) illustrate AUROCs attained by the fusion of MMRNet and pathologists with varying expertise on the MultiCenter-UCEC dataset. (M–R) demonstrate AUROCs generated by the fusion of MMRNet and pathologists with varying expertise on the GWCH-UCEC dataset. TCGA-UCEC, The Cancer Genome Atlas dataset; MultiCenter-UCEC, multiple medical centers dataset; GWCH-UCEC, Guangdong Women and Children Hospital dataset; AUROC, area under the receiver operating curve; 95% CI, 95% confidence interval.

Table 3. The logistic regression models for the association between the clinicopathological features and MMRNet's prediction on the external datasets

	TCGA-UCEC				MultiCenter-UCEC				GWCH-UCEC			
	Univariate		Multivariate		Univariate		Multivariate		Univariate		Multivariate	
Features	OR	*p*	OR	*p*	OR	*p*	OR	*p*	OR	*p*	OR	*p*
Lower uterine segment (LUS) centered	NA	0.999	NA	0.999	NA	1.000	NA	1.000	3.061 (0.554, 16.903)	0.199	1.434 (0.211, 9.743)	0.713
Papillary/polypoid	0.955 (0.506, 1.803)	0.887	1.050 (0.542, 2.033)	0.885	1.311 (0.417, 4.124)	0.643	0.814 (0.201, 3.301)	0.773	0.847 (0.520, 1.379)	0.504	0.851 (0.496, 1.458)	0.556
Heterogeneity/ dedifferentiated	NA	0.999	NA	0.999	2.806 (0.281, 28.050)	0.380	0.734 (0.046, 11.735)	0.827	NA	0.999	NA	0.999
TILs	6.135 (2.172, 17.335)	<0.001	3.895 (1.209, 12.541)	0.023	1.996 (0.734, 5.430)	0.176	0.676 (0.167, 2.741)	0.584	7.806 (3.895, 15.646)	<0.001	4.236 (1.853, 9.686)	<0.001
Prominent peritumoral lymphocytes	3.544 (1.760, 7.135)	<0.001	2.128 (0.980, 4.620)	0.056	2.930 (1.222, 7.023)	0.016	2.444 (0.798, 7.480)	0.117	4.069 (2.556, 6.477)	<0.001	2.149 (1.202, 3.841)	0.010
Solid sheets of round-to-oval cells	1.472 (0.847, 2.558)	0.170	1.141 (0.632, 2.060)	0.661	9.047 (3.317, 24.677)	<0.001	8.517 (2.705, 26.820)	<0.001	2.985 (1.503, 5.928)	0.002	1.821 (0.805, 4.116)	0.150
Mucinous differentiation	2.638 (0.775, 8.983)	0.121	2.050 (0.577, 7.286)	0.267	0.762 (0.168, 3.462)	0.725	0.924 (0.188, 4.533)	0.923	1.678 (0.768, 3.667)	0.194	1.759 (0.739, 4.186)	0.202
Lymphoepithelioma-like	2.667 (0.329, 21.590)	0.358	0.446 (0.042, 4.783)	0.505	8.826 (1.186, 65.664)	0.033	3.112 (0.283, 34.172)	0.353	NA	0.999	NA	0.999
Background with endometrium hyperplasia	1.343 (0.568, 3.178)	0.502	1.069 (0.433, 2.639)	0.886	1.383 (0.544, 3.517)	0.496	1.799 (0.617, 5.246)	0.282	1.724 (1.156, 2.570)	0.008	2.355 (1.510, 3.671)	<0.001

Note: 95% confidence intervals are included in brackets. TCGA-UCEC, external dataset from The Cancer Genome Atlas; MultiCenter-UCEC, external dataset from multiple medical centers; GWCH-UCEC, external dataset from Guangdong Women and Children Hospital; OR, odds ratio; NA, not applicable; TILs, tumor-infiltrating lymphocytes.

predictive accuracy and suggests that fully supervised training effectively captures essential MMR features with robust generalizability. The choice of ResNet18, a shallower architecture requiring fewer parameters than Fremond et al.'s ResNet50,[22] also reduces computational demands, enhancing MMRNet's suitability for large-scale pathology slide processing and making it practical for deployment in resource-constrained settings. To further validate MMRNet, we constructed a comparative analysis with the im4MEC model by Fremond.[22] We trained the im4MEC model using tile data from our internal datasets with default parameters from its public repository and tested it on three additional datasets (Table S11). MMRNet consistently outperformed im4MEC, achieving higher AUROC across all datasets (Figure S7). Additionally, recognizing the uncertainties inherent in both MMRNet predictions and pathologists' assessments, we implemented a human-machine fusion approach, as detailed in our previous investigation.[28] Our results illustrated that this fusion approach outperformed both MMRNet and pathologists across varying expertise levels, suggesting that less experienced pathologists could achieve expert-level diagnostic accuracy with support from MMRNet predictions. Clinically, this combined approach could serve as a decision-support tool for pathologists seeking additional confidence in their diagnoses. Further research should explore the potential application of human-machine fusion in diagnostic settings.

MMRNet demonstrates high sensitivity (90%), but its specificity (59%) is relatively lower. However, this trade-off results in an AUROC that meets the acceptance standard for predictive models. In developing MMRNet, we maintained the cutoff value derived from the internal training dataset, which optimized the balance between sensitivity and specificity to achieve the highest AUROC. Given the clinical context, we intentionally prioritized high sensitivity to minimize false negatives. Sensitivity reflects the model's ability to correctly identify patients with MMR deficiency, which is crucial for conditions where missing a diagnosis could delay effective treatment, such as immunotherapy for MMR-deficient EC. While this approach may increase the false-positive rate, it ensures that patients who could benefit from further diagnostic evaluation or treatment are not overlooked.[29] In clinical practice, this strategy is particularly relevant for diseases with poor prognoses, where early detection can significantly improve outcomes. From a clinical practice perspective, one potential use of MMRNet is for preliminary screening, where patients predicted to be MMRPEC could benefit from cost savings by avoiding further MMR tests. In this context, high sensitivity, which reduces the risk of false negatives, is particularly valuable. Patients predicted to be MMRDEC should also be validated through MMR IHC, PCR, or NGS before proceeding with treatment for the patient based on their personal preferences and individual circumstances. MMR IHC testing for all patients in our study is performed in well-established hospitals, but, in some economically disadvantaged areas, it is not feasible to conduct MMR IHC screening for everyone. In these cases, MMRNet could be used for initial screening. We acknowledge that specificity was notably lower in the TCGA (57%) and multicenter datasets (62.9%). In the TCGA dataset, poor staining quality and low-resolution WSIs likely contributed to reduced specificity. Similarly, variability in image quality across centers in the multicenter dataset may have impacted performance. This challenge is not unique to our model. For example, the im4MEC model by Fremond et al.[22] exhibited low sensitivity (23%–51%) but high specificity (82%–93%) (Table S11). Similarly, other deep learning models in related studies have prioritized high sensitivity with lower specificity. Wang et al.[30] demonstrated a sensitivity of 100.00% for the endometrioid carcinoma G1/G2 and 94.00% for G3 using TCGA dataset. Zhang et al.[6] developed a model achieving an AUROC of 0.799 and sensitivity of 0.857 for microsatellite instability prediction in EC, emphasizing high sensitivity. Goyal et al.[31] also developed a gynecologic pathology model with high sensitivity (0.90) and lower specificity (0.67). In this context, MMRNet's high sensitivity, despite lower specificity, aligns with clinical priorities by effectively identifying patients at risk for MMR deficiency. To address the issue of accuracy, it is necessary to optimize the model and expand the sample size for improvement in the future. This approach could also be applied in future clinical scenarios, where both MMRNet and MMR IHC are used together to continuously optimize the model's accuracy. On one hand, this would improve the accuracy of MMRNet, and on the other hand, it could help minimize missed or misdiagnosed cases in MMR IHC, complementing each other.

Acknowledging the perceived complexity of deep learning models,[21,32,33] we employed logistic regression to uncover MMRNet's predictive features. MMRNet showed promise in identifying previously undefined characteristics of MMRDEC,[34] as demonstrated by our analysis of misdiagnosed cases. To directly visualize the regions MMRNet attends to, we overlaid heatmaps on WSIs. In these heatmaps, MMR-proficient (MMRPEC) regions with low MMR scores are highlighted in blue, while MMR-deficient (MMRDEC) regions with high MMR scores are shown in red. Rectangular markers (0.5 $\times$ 0.5 mm) were placed on the WSIs to highlight morphological regions influencing MMR status prediction. Our approach is consistent with findings in similar studies. For example, Zhang et al.[6] used heatmaps to highlight MSI-H features with red-marked regions. Moreover, lymphocytes infiltrating signature score (LISS) regression and classification models have demonstrated that a high homologous recombination deficiency (HRD) prediction score correlated with lymphocyte-rich regions in rectangular markers.[35] Similarly, attention maps in EC grading models have emphasized histological features with high-attention areas marked in red.[31] By focusing on the spatial distribution of MMRNet's predictions and their correlation with known histopathological features, we provide stronger evidence that MMRNet attends to relevant tissue characteristics. Zooming into specific patches may further enhance the identification of subtle but meaningful histomorphological patterns. However, these cases presented challenging features that posed difficulties for both MMRNet and pathologists, highlighting the need for efficient approaches to address these challenges and improve diagnostic accuracy in the future.

We conducted a detailed analysis comparing the misdiagnoses made by human pathologists and the MMRNet model: (1) Analysis of misdiagnosed cases: we examined morphological features associated with misdiagnoses by comparing false positives with true negatives and false negatives with true positives.

In the TCGA dataset, human viewers showed a higher tendency for misclassification, particularly in cases with a significant presence of TILs ($p = 0.029$) (Table S5). (2) Comparison across pathologist's experience levels: when examining errors made by human pathologists, the misdiagnoses of expert pathologists are related to papillary/polypoid, heterogeneity/dedifferentiated, prominent TILs, prominent peritumoral lymphocytes, solid sheets of round-to-oval cells, and background endometrium with hyperplasia (Table S8). Senior pathologists' misdiagnoses are related to heterogeneity/dedifferentiated, prominent TILs, prominent peritumoral lymphocytes, solid sheets of round-to-oval cells, and lymphoepithelioma-like morphology (Table S8). And the misdiagnoses of junior pathologists are correlated with papillary/polypoid, prominent TILs, prominent peritumoral lymphocytes, solid sheets of round-to-oval cells, and lymphoepithelioma-like morphology (Table S8). The comparison of expert, senior, and junior pathologists' misdiagnosed and correctly diagnosed cases on MultiCenter-UCEC dataset is shown in Table S9. And the comparison of expert, senior, and junior pathologists' misdiagnosed and correctly diagnosed cases on GWCH-UCEC dataset is shown in Table S10. All the tables are added to the supplemental information, indicating errors are especially common among junior pathologists. (3) MMRNet's performance advantage: compared to human pathologists, MMRNet demonstrated reduced reliance on these challenging histopathological features, leading to more consistent and accurate classifications across the TCGA, MultiCenter-UCEC, and GWCH-UCEC datasets. This suggests that MMRNet can effectively support human pathologists by reducing diagnostic errors caused by subtle or ambiguous features, ultimately improving diagnostic accuracy.

In summary, our study introduced MMRNet, a deep learning model tailored for predicting MMR status in EC using H&E-stained WSIs. Significantly, our human-machine fusion approach enhanced MMRNet's performance, suggesting its potential as an effective preliminary screening tool. This improvement highlights its promising role in identifying cases suitable for confirmatory MMR/MSI testing, particularly for MMR-deficient EC subtypes.

Limitations of the study

Despite promising results, our study has limitations. First, the retrospective nature of MMRNet's training underscores the need for prospective clinical validation. While logistic regression supports its biological plausibility, improved visualization techniques are required for better interpretability. Second, the human-AI fusion approach, though modest in overall impact, significantly benefits junior pathologists (e.g., AUROC increase from 0.6480 to 0.7579 in TCGA). Further refinement could enhance its clinical utility. Third, while MMRNet accurately identifies MSI status, it does not replace molecular classification. The retrospective dataset limits comprehensive molecular characterization, necessitating future integration with molecular profiling. Fourth, for Lynch syndrome-associated cases, our small dataset (14 cases) limits conclusions. Larger studies, particularly comparing MMRNet to NGS, are needed to assess its accuracy. Finally, variability in sensitivity and specificity across datasets, influenced by sample heterogeneity and data quality, challenges real-world adoption. Future advancements should focus on integrating clinical and histopathological data, multimodal learning, and advanced deep learning architectures. While not yet ready for clinical deployment, MMRNet provides a foundation for future AI-driven diagnostics.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources and reagents should be directed to and will be fulfilled by the lead contact, Mu-Yan Cai (caimy@sysucc.org.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All the diagnostic whole slides and associated labels from TCGA are accessible via the NIH Genomic Data Commons (https://portal.gdc.cancer.gov/). However, limitations exist concerning the availability of the Internal-UCEC, MultiCenter-UCEC, and GWCH-UCEC WSIs due to institutional permissions obtained via IRB approval for this specific study. These slides are not publicly accessible due to obligations related to patient privacy. The source code for MMRNet is available at GitHub (https://github.com/luorzh/MMRNet-CRM; https://doi.org/10.5281/zenodo.15049484) and will be made publicly available as of the date of publication. Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (No. 82072853) and 308-Program for Clinical Research of Sun Yat-sen University Cancer Center (PCR308-SYSUCC-2016001).

AUTHOR CONTRIBUTIONS

M.-Y.C., R.-Z.L., and L.-L.L. designed the study. L.-L.L., H.-Y.L., X.L., X.Y., M.-Y.H., and C.-Y.Z. collected the samples and acquired the image data. H.-Y.G., R.-G.L., Z.W., J.-Y.Z., X.-M.Y., Q.-N.K., X.-L.K., and D.-D.W. provided the clinical and pathological data of multiple medical centers. B.-Z.J. and C.-Y.Z. performed the machine learning. M.-Y.C., R.-Z.L., X.-K.Z., L.-L. L., L.-J.W., and C.C. conducted the reader study. L.-L.L., Z.-H.Z., X.L., and H.-Y.L. did the statistical analyses. All authors vouch for the data, analyses, and interpretations. L.-L.L., X.L., and M.-Y.C. wrote the first draft of the manuscript, and all authors reviewed, contributed to, and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Patients and samples
- METHOD DETAILS
  - Scanning and annotation of H&E-stained slides
  - Assessment of MMR status
  - Pathology reader study
  - Human-machine fusion
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2025.102099.

Received: July 13, 2024
Revised: November 5, 2024
Accepted: April 8, 2025
Published: April 29, 2025

REFERENCES

1. Crosbie, E.J., Kitson, S.J., McAlpine, J.N., Mukhopadhyay, A., Powell, M. E., and Singh, N. (2022). Endometrial cancer. Lancet *399*, 1412–1428. https://doi.org/10.1016/S0140-6736(22)00323-3.
2. Lefebvre, T.L., Ueno, Y., Dohan, A., Chatterjee, A., Vallières, M., Winter-Reinhold, E., Saif, S., Levesque, I.R., Zeng, X.Z., Forghani, R., et al. (2022). Development and Validation of Multiparametric MRI-based Radiomics Models for Preoperative Risk Stratification of Endometrial Cancer. Radiology *305*, 375–386. https://doi.org/10.1148/radiol.212873.
3. León-Castillo, A., de Boer, S.M., Powell, M.E., Mileshkin, L.R., Mackay, H. J., Leary, A., Nijman, H.W., Singh, N., Pollock, P.M., Bessette, P., et al. (2020). Molecular Classification of the PORTEC-3 Trial for High-Risk Endometrial Cancer: Impact on Prognosis and Benefit From Adjuvant Therapy. J. Clin. Oncol. *38*, 3388–3397. https://doi.org/10.1200/JCO.20.00549.
4. Hong, R., Liu, W., DeLair, D., Razavian, N., and Fenyö, D. (2021). Predicting endometrial cancer subtypes and molecular features from histopathology images using multi-resolution deep learning models. Cell Rep. Med. *2*, 100400. https://doi.org/10.1016/j.xcrm.2021.100400.
5. Riasatian, A., Babaie, M., Maleki, D., Kalra, S., Valipour, M., Hemati, S., Zaveri, M., Safarpoor, A., Shafiei, S., Afshari, M., et al. (2021). Fine-Tuning and training of densenet for histopathology image representation using TCGA diagnostic slides. Med. Image Anal. *70*, 102032. https://doi.org/10.1016/j.media.2021.102032.
6. Zhang, Y., Chen, S., Wang, Y., Li, J., Xu, K., Chen, J., and Zhao, J. (2023). Deep learning-based methods for classification of microsatellite instability in endometrial cancer from HE-stained pathological images. J. Cancer Res. Clin. Oncol. *149*, 8877–8888. https://doi.org/10.1007/s00432-023-04838-4.
7. Kandoth, C., Schultz, N., Cherniack, A.D., Akbani, R., Liu, Y., Shen, H., Robertson, A.G., Pashtan, I., Shen, R., et al.; Cancer Genome Atlas Research Network (2013). Integrated genomic characterization of endometrial carcinoma. Nature *497*, 67–73. https://doi.org/10.1038/nature12113.
8. Le, D.T., Uram, J.N., Wang, H., Bartlett, B.R., Kemberling, H., Eyring, A.D., Skora, A.D., Luber, B.S., Azad, N.S., Laheru, D., et al. (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N. Engl. J. Med. *372*, 2509–2520. https://doi.org/10.1056/NEJMoa1500596.
9. Harrold, E.C., Foote, M.B., Rousseau, B., Walch, H., Kemel, Y., Richards, A.L., Keane, F., Cercek, A., Yaeger, R., Rathkopf, D., et al. (2023). Neoplasia risk in patients with Lynch syndrome treated with immune checkpoint blockade. Nat. Med. *29*, 2458–2463. https://doi.org/10.1038/s41591-023-02544-9.
10. Vikas, P., Messersmith, H., Compton, C., Sholl, L., Broaddus, R.R., Davis, A., Estevez-Diz, M., Garje, R., Konstantinopoulos, P.A., Leiser, A., et al. (2023). Mismatch Repair and Microsatellite Instability Testing for Immune Checkpoint Inhibitor Therapy: ASCO Endorsement of College of American Pathologists Guideline. J. Clin. Oncol. *41*, 1943–1948. https://doi.org/10.1200/JCO.22.02462.
11. Stelloo, E., Jansen, A.M.L., Osse, E.M., Nout, R.A., Creutzberg, C.L., Ruano, D., Church, D.N., Morreau, H., Smit, V.T.H.B.M., van Wezel, T., and Bosse, T. (2017). Practical guidance for mismatch repair-deficiency testing in endometrial cancer. Ann. Oncol. *28*, 96–102. https://doi.org/10.1093/annonc/mdw542.
12. Kato, A., Sato, N., Sugawara, T., Takahashi, K., Kito, M., Makino, K., Sato, T., Shimizu, D., Shirasawa, H., Miura, H., et al. (2016). Isolated Loss of PMS2 Immunohistochemical Expression is Frequently Caused by Heterogenous MLH1 Promoter Hypermethylation in Lynch Syndrome Screening for Endometrial Cancer Patients. Am. J. Surg. Pathol. *40*, 770–776. https://doi.org/10.1097/PAS.0000000000000606.
13. Lu, K.H., and Broaddus, R.R. (2020). Endometrial Cancer. N. Engl. J. Med. *383*, 2053–2064. https://doi.org/10.1056/NEJMra1514010.
14. Shia, J. (2021). The diversity of tumours with microsatellite instability: molecular mechanisms and impact upon microsatellite instability testing and mismatch repair protein immunohistochemistry. Histopathology *78*, 485–497. https://doi.org/10.1111/his.14271.
15. Amato, M., Franco, R., Facchini, G., Addeo, R., Ciardiello, F., Berretta, M., Vita, G., Sgambato, A., Pignata, S., Caraglia, M., et al. (2022). Microsatellite Instability: From the Implementation of the Detection to a Prognostic and Predictive Role in Cancers. Int. J. Mol. Sci. *23*, 8726. https://doi.org/10.3390/ijms23158726.
16. Levine, M.D., Barrington, D.A., Hampel, H., Goodfellow, P.J., and Cohn, D. E. (2022). Implementing universal upfront multi-gene panel testing in endometrial cancer: From cost to practical considerations. Gynecol. Oncol. *166*, 538–542. https://doi.org/10.1016/j.ygyno.2022.07.016.
17. Plotkin, A., Olkhov-Mitsel, E., Nofech-Mozes, S., Djordjevic, B., Mirkovic, J., Fitzpatrick, M., Krizova, A., and Look Hong, N.J. (2023). Budget impact analysis of molecular subtype profiling in endometrial cancer. Gynecol. Oncol. *178*, 54–59. https://doi.org/10.1016/j.ygyno.2023.09.006.
18. Garg, K., Leitao, M.M., Jr., Kauff, N.D., Hansen, J., Kosarin, K., Shia, J., and Soslow, R.A. (2009). Selection of endometrial carcinomas for DNA mismatch repair protein immunohistochemistry using patient age and tumor morphology enhances detection of mismatch repair abnormalities. Am. J. Surg. Pathol. *33*, 925–933. https://doi.org/10.1097/PAS.0b013e318197a046.
19. Shia, J., Black, D., Hummer, A.J., Boyd, J., and Soslow, R.A. (2008). Routinely assessed morphological features correlate with microsatellite instability status in endometrial cancer. Hum. Pathol. *39*, 116–125. https://doi.org/10.1016/j.humpath.2007.05.022.
20. Fremond, S., Koelzer, V.H., Horeweg, N., and Bosse, T. (2022). The evolving role of morphology in endometrial cancer diagnostics: From histopathology and molecular testing towards integrative data analysis by deep learning. Front. Oncol. *12*, 928977. https://doi.org/10.3389/fonc.2022.928977.
21. van der Laak, J., Litjens, G., and Ciompi, F. (2021). Deep learning in histopathology: the path to the clinic. Nat. Med. *27*, 775–784. https://doi.org/10.1038/s41591-021-01343-4.
22. Fremond, S., Andani, S., Barkey Wolf, J., Dijkstra, J., Melsbach, S., Jobsen, J.J., Brinkhuis, M., Roothaan, S., Jurgenliemk-Schulz, I., Lutgens, L.C.H.W., et al. (2023). Interpretable deep learning model to predict the molecular classification of endometrial cancer from haematoxylin and eosin-stained whole-slide images: a combined analysis of the PORTEC randomised trials and clinical cohorts. Lancet. Digit. Health *5*, e71–e82. https://doi.org/10.1016/S2589-7500(22)00210-2.
23. Muti, H.S., Heij, L.R., Keller, G., Kohlruss, M., Langer, R., Dislich, B., Cheong, J.H., Kim, Y.W., Kim, H., Kook, M.C., et al. (2021). Development and validation of deep learning classifiers to detect Epstein-Barr virus and microsatellite instability status in gastric cancer: a retrospective multicentre cohort study. Lancet. Digit. Health *3*, e654–e664. https://doi.org/10.1016/S2589-7500(21)00133-3.
24. Yamashita, R., Long, J., Longacre, T., Peng, L., Berry, G., Martin, B., Higgins, J., Rubin, D.L., and Shen, J. (2021). Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. Lancet Oncol. *22*, 132–141. https://doi.org/10.1016/S1470-2045(20)30535-0.
25. Li, D., Udaltsova, N., Layefsky, E., Doan, C., and Corley, D.A. (2021). Natural Language Processing for the Accurate Identification of Colorectal Cancer Mismatch Repair Status in Lynch Syndrome Screening. Clin.

Gastroenterol. Hepatol. *19*, 610–612.e1. https://doi.org/10.1016/j.cgh.2020.01.040.

26. Saeb, S., Lonini, L., Jayaraman, A., Mohr, D.C., and Kording, K.P. (2017). The need to approximate the use-case in clinical machine learning. GigaScience *6*, 1–9. https://doi.org/10.1093/gigascience/gix019.
27. Vu, H.L., Ng, K.T.W., Richter, A., and An, C. (2022). Analysis of input set characteristics and variances on k-fold cross validation for a Recurrent Neural Network model on waste disposal rate estimation. J. Environ. Manage. *311*, 114869. https://doi.org/10.1016/j.jenvman.2022.114869.
28. Zheng, X., Wang, R., Zhang, X., Sun, Y., Zhang, H., Zhao, Z., Zheng, Y., Luo, J., Zhang, J., Wu, H., et al. (2022). A deep learning model and human-machine fusion for prediction of EBV-associated gastric cancer from histopathology. Nat. Commun. *13*, 2790. https://doi.org/10.1038/s41467-022-30459-5.
29. Altman, D.G., and Bland, J.M. (1994). Diagnostic tests. 1: Sensitivity and specificity. BMJ *308*, 1552. https://doi.org/10.1136/bmj.308.6943.1552.
30. Wang, C.W., Muzakky, H., Firdi, N.P., Liu, T.C., Lai, P.J., Wang, Y.C., Yu, M.H., and Chao, T.K. (2024). Deep learning to assess microsatellite instability directly from histopathological whole slide images in endometrial cancer. npj Digit. Med. *7*, 143. https://doi.org/10.1038/s41746-024-01131-7.
31. Goyal, M., Tafe, L.J., Feng, J.X., Muller, K.E., Hondelink, L., Bentz, J.L., and Hassanpour, S. (2024). Deep Learning for Grading Endometrial Cancer. Am. J. Pathol. *194*, 1701–1711. https://doi.org/10.1016/j.ajpath.2024.05.003.
32. Kleppe, A., Skrede, O.J., De Raedt, S., Liestøl, K., Kerr, D.J., and Danielsen, H.E. (2021). Designing deep learning studies in cancer diagnostics. Nat. Rev. Cancer *21*, 199–211. https://doi.org/10.1038/s41568-020-00327-9.
33. Sun, R., Limkin, E.J., Vakalopoulou, M., Dercle, L., Champiat, S., Han, S. R., Verlingue, L., Brandao, D., Lancia, A., Ammari, S., et al. (2018). A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. Lancet Oncol. *19*, 1180–1191. https://doi.org/10.1016/S1470-2045(18)30413-3.
34. Bera, K., Schalper, K.A., Rimm, D.L., Velcheti, V., and Madabhushi, A. (2019). Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. Nat. Rev. Clin. Oncol. *16*, 703–715. https://doi.org/10.1038/s41571-019-0252-y.
35. El Nahhas, O.S.M., Loeffler, C.M.L., Carrero, Z.I., van Treeck, M., Kolbinger, F.R., Hewitt, K.J., Muti, H.S., Graziani, M., Zeng, Q., Calderaro, J., et al. (2024). Regression-based Deep-Learning predicts molecular biomarkers from pathology slides. Nat. Commun. *15*, 1253. https://doi.org/10.1038/s41467-024-45589-1.
36. Humphries, M.P., Maxwell, P., and Salto-Tellez, M. (2021). QuPath: The global impact of an open source digital pathology system. Comput. Struct. Biotechnol. J. *19*, 852–859. https://doi.org/10.1016/j.csbj.2021.01.022.
37. Zhou, W., Yang, Y., Yu, C., Liu, J., Duan, X., Weng, Z., Chen, D., Liang, Q., Fang, Q., Zhou, J., et al. (2021). Ensembled deep learning model outperforms human experts in diagnosing biliary atresia from sonographic gallbladder images. Nat. Commun. *12*, 1259. https://doi.org/10.1038/s41467-021-21466-z.
38. Youden, W.J. (1950). Index for rating diagnostic tests. Cancer *3*, 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
TCGA	https://portal.gdc.cancer.gov/	RRID: SCR_003193
Internal Dataset Internal-UCEC	Sun Yat-sen University Cancer Center	N/A
External Dataset MultiCenter-UCEC	Jiangmen Central Hospital, Zhuhai Maternal and Child Health Care Hospital, Affiliated Cancer Hospital and Institute of Guangzhou Medical University, The Second Affiliated Hospital of Guangzhou Medical University, Qingdao Municipal Hospital	N/A
External Dataset GWCH-UCEC	Guangdong Women and Children Hospital	N/A
Software and algorithms		
QuPath version 0.2.3	https://qupath.github.io/	RRID:SCR_018257
IBM SPSS Statistics 27.0	https://www.ibm.com/spss	RRID:SCR_002865
Medcalc 15.2.2	https://www.medcalc.org/	RRID:SCR_015044
Python version 3.9.6	https://www.python.org/	RRID:SCR_008394
PyTorch version 1.9	https://pytorch.org/	RRID:SCR_018536
BioRender	https://www.biorender.com/	RRID:SCR_018361
Our source code	https://github.com/luorzh/MMRNet-CRM	https://doi.org/10.5281/zenodo.15049484
The source code of im4MEC	https://github.com/AIRMEC/im4MEC	N/A

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Patients and samples

The present study received approval from the Ethics Committee of Sun Yat-sen University Cancer Center (SYSUCC, B2023-019-01). As patients were not directly recruited for our research, informed consent was waived. Four pathological image datasets were procured to develop MMRNet, including an internal dataset from SYSUCC (Internal-UCEC), as well as the external datasets from The Cancer Genome Altas (TCGA-UCEC), multiple medical centers including Jiangmen Central Hospital, Zhuhai Maternal and Child Health Care Hospital, Affiliated Cancer Hospital and Institute of Guangzhou Medical University, The Second Affiliated Hospital of Guangzhou Medical University and Qingdao Municipal Hospital (MultiCenter-UCEC), and Guangdong Women and Children Hospital (GWCH-UCEC). The intention was to ensured broad patient representation and generalizability. The Internal-UCEC dataset consisted of 242 H&E-stained WSIs from 242 MMRDEC patients and 785 WSIs from 784 MMRPEC patients. The TCGA-UCEC dataset included 401 WSIs from 369 patients, comprising 135 WSIs from 127 MMRDEC patients and 266 WSIs from 242 MMRPEC patients. The MultiCenter-UCEC dataset contained 230 WSIs from 230 patients, including 60 MMRDEC and 170 MMRPEC patients. The GWCH-UCEC dataset comprised 421 WSIs from 421 patients, with 88 MMRDEC and 333 MMRPEC patients. All EC patients are female, and patients from the Internal-UCEC, MultiCenter-UCEC, and GWCH-UCEC datasets are all of Yellow Asian descent. The TCGA-UCEC dataset includes 287 White patients, 46 Black or African American patients, 16 Asian patients, 7 Native Hawaiian or Other Pacific Islander patients, 3 American Indian or Alaska Native patients, and 10 patients with unknown ethnicity. The influence of gender and ethnicity on the study results was found to be insignificant ($p > 0.05$). Further details of these datasets are provided in Table S1. The internal-UCEC served as the primary training dataset, comprising assessable slides from MMRDEC patients and randomly selected slides from the pool of MMRPEC patients at SYSUCC. TCGA-UCEC consisted of patients from the TCGA database with known MSI status, while MultiCenter-UCEC and GWCH-UCEC included EC patients with available MMR status randomly enrolled from distinct centers. Inclusion criteria involved EC patients who underwent primary hysterectomy between January 1, 2005 and March 1, 2021, at the respective center, had known MMR status, and had accessible clinical data and H&E-stained tumor slides. Exclusion criteria comprised patients with incomplete clinical data, those who underwent preoperative therapy (such as neoadjuvant chemotherapy or radiotherapy), and slides with unsatisfactory scanning quality (e.g., visible tissue folds or out of focus).

CellPress
OPEN ACCESS

METHOD DETAILS

Scanning and annotation of H&E-stained slides

In Internal-UCEC, MultiCenter-UCEC and GWCH-UCEC cohort, one or two typical H&E-stained slides from each patient's resection were scanned at ×40 magnification (0.25 μm/pixel) with an Aperio AT2 scanner (Leica Biosystems; Wetzlar, Germany), generating one or two WSIs in SVS format. Diagnostic slides from TCGA-UCEC were obtained via the Genomic Data Commons portal (http://portal.gdc.cancer.gov/). Junior pathologists, blind to patient data and MMR status, utilized the open-source software QuPath[36] (version 0.2.3) to annotate the slides by outlining the regions of interest around the cancerous areas. Subsequently, these annotations were reviewed and refined, serving as reference criterion for tumor detection.

Assessment of MMR status

The ground truth MMR status within Internal-UCEC, MultiCenter-UCEC, and GWCH-UCEC datasets was measured by IHC, targeting MMR proteins in histopathological specimens obtained from the respective institutions (Figure S1). In the case of TCGA-UCEC, MSI status was determined through formerly conducted research.[7] Prior investigations have revealed comparable sensitivity and reliability between MSI testing and MMR-IHC, indicating the interchangeability of MMR-IHC with MSI testing for identifying MMRD/MSI-H EC.[12]

Pathology reader study

To assess the pathologists' ability to identify MMR status from H&E-stained slides, six pathologists with varying years of experience (two junior pathologists with less 5 years' experience, two senior pathologists with approximately 10 years' experience, alongside two experts specializing in EC with at least than 15 years' experience) evaluated the slides from external datasets: MultiCenter-UCEC, GWCH-UCEC and TCGA-UCEC. Blind to the clinical data and MMRNet performance, the pathologists reviewed the slides and categorized each case as MMRDEC or MMRPEC based on their expertise.

To gain a deeper insight into the relationship between particular histopathological characteristics and MMRNet predictions (i.e., MMRDEC or MMRPEC), we constructed a logistic regression model to evaluate this relationship. Previous studies have linked MMRDEC with certain morphological characteristics such as prominent peritumoral lymphocytes, prominent tumor-infiltrating lymphocytes and mucinous differentiation.[18,19] Additionally, based on pathologists' experience with EC, MMRDEC is associated with other positive morphologic traits including papillary/polypoid structures, heterogeneous/differentiated patterns, lower uterine segment (LUS) centrality, dense sheets of round to oval cells, lymphoepithelioma-like features, and endometrial background with hyperplasia. Therefore, we focused on examining these nine histopathological characteristics. To better understand MMRNet, we further assessed the relationship between morphological characteristics and MMRNet diagnosis errors in this study. Specifically, we compared the morphological characteristics of false-positive cases with true-negative cases, and the features of false-negative cases with true-positive cases.

Human-machine fusion

MMRNet not only serves as an independent predictor of a patient's MMR status but can also be integrated with pathologists' predictions through a human-machine fusion approach. This fusion method, described in our previous study,[28] involves combining the predictions generated by MMRNet deep learning model with those made by pathologists of varying experience levels, primarily minimizing the uncertainty associated with the predictions.

The details of the human-machine fusion are described below, with the overall fusion strategy introduced first, followed by the design of prediction uncertainty for both MMRNet and pathologists. It is worth noting that the human-machine fusion is predefined and not involved in the training of MMRNet. Suppose an MMRNet has been well trained and prepared to collaborate with a pathologist to predict the MMR status of a patient. Based on the slide data of the patient, denote by P_m and P_h the two-dimensional output probability vectors from the MMRNet and the pathologist, respectively, and um and uh the prediction uncertainties from the MMRNet and the pathologist, respectively. Then, $1/U_m$ and $1/U_h$ can represent prediction certainties (or confidence) from the MMRNet and the pathologist, respectively. Based on the MMR predictions and prediction certainties from both the MMRNet and the pathologist, the output of the human-machine fusion, i.e., the fused prediction Pf from the MMRNet and the pathologist, can be defined as,

$$P_f = a \cdot P_m + (1 - a) \cdot P_h \quad \text{(Equation 1)}$$

where $a = \frac{1/U_m}{1/U_m + 1/U_h}$ here represents the relative importance of the prediction from the MMRNet for the final fusion prediction P_f, and similarly (1-a) represents the relative importance of the prediction from the pathologist. Intuitively, when the MMRNet is more certain than the pathologist for the MMR prediction, the final prediction P_f will be more dependent on the model prediction P_m, and vice versa (From Equation 1). The main challenge is to obtain the pathologist's prediction probability P_h and the two prediction uncertainties U_m and U_h. For the uncertainty um of the MMRNet prediction, a deep learning model may be uncertain under two conditions, either when the new slide data are very different from all those used for model training (called knowledge uncertainty), or when the new slide data is similar to one or more slides from both the MMR and non-MMR classes used for model training (called data uncertainty). Fortunately, when the deep learning model is an ensemble of multiple individual models (as MMRNet), both types of prediction

uncertainties can be captured by the entropy of the ensemble model's probability output P_m. Therefore, the uncertainty um for MMRNet prediction can be estimated by,

$$Um = -\left(P_{m,1} \cdot \ln P_{m,1} + P_{m,2} \cdot \ln P_{m,2}\right) \quad \text{(Equation 2)}$$

where $P_{m,1}$ and $P_{m,2}$ are respectively the first and second component of the probability output P_m of the MMRNet. In order to obtain the pathologist's prediction probability P_h and the associated uncertainty U_h, we collected not only his or her diagnosis result ("MMRD" or "MMRP"), but also his or her 5-scale self-confidence on the diagnosis, with "1" to "5",respectively, representing "surely MMRP", "likely MMRP", "unsure", "likely MMRD",and "surely MMRD". The five self-confidence scales [1, 2, 3, 4, 5] were simply linearly transformed to the corresponding probabilities [0.2, 0.35, 0.5, 0.65, 0.8], where the smallest and largest probability are respectively set to 0.2 and 0.8 (rather than 0 and 1) by considering the potential over-confidence or noise in reported selfconfidence. With this transformed probability, a two-dimensional probability output P_h can easily be obtained based on the pathologist's diagnostic result and his or her self-confidence report. For example, for the diagnosis result "MMRD" and selfconfidence "5", P_h would be [0.8, 0.2] where the first component represents the probability of being "MMR"; and for the diagnosis result "MMRP" and self-confidence "2", P_h would be [0.35, 0.65]. Once P_h is obtained, the associated prediction uncertainty uh can be easily estimated by the entropy of P_h, similarly based on Equation 2. Finally, after obtaining the pathologist's prediction P_h and prediction uncertainty U_h, together with the MMRNet prediction P_m and prediction uncertainty U_m, the human-machine fusion output can be obtained based on the designed fusion strategy (From Equation 1).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

Considering the ground truth MMR status as the reference criteria, MMRNet's AUROC was calculated based on its predictive scores, while pathologists' AUROCs were determined from the dichotomous categorization into MMRDEC or MMRPEC.[37] Consequently, within the ROC space, each pathologist corresponds to a point, whereas MMRNet yields a consecutive curve.[24] Youden's J statistic was applied to establish the thresholds for MMRNet ROC curves,[38] enabling the dichotomization of MMRNet probabilities into binary predictions to calculate the sensitivity, specificity, and NPV. McNemar's test was employed to compare specificity and sensitivity. Comparative analysis of baseline data among participant groups from different datasets was conducted using Chi-square tests or analysis on variance. Chi-square and t-tests were used to compare morphological characteristics of misdiagnosed and correctly diagnosed cases. The relationship between MMRNet predictions and morphological characteristics was analyzed with a logistic regression model. Statistical significance was defined as a p value less than 0.05 for two-sided tests. IBM SPSS Statistics (27.0) and Medcalc (15.2.2) were utilized for statistical analysis. In addition, Python (version 3.9.6) and the deep learning platform PyTorch (version 1.9) were employed for model development and data interpretation.