scientific reports

Check for updates

OPEN

Diagnostic performance of artificial intelligence for detecting peritoneal and small bowel dissemination in epithelial ovarian cancer using preoperative contrast-enhanced CT imaging

Ria Kim[1], Toshiyuki Seki[1]✉, Katsuhiko Noda[2], Kaname Yoshida[2], Kota Yokosu[1], Rintaro Hamada[1], Erika Habuchi[1], Masahiro Takahashi[3], Hirokuni Takano[1] & Aikou Okamoto[1]

In epithelial ovarian cancer (EOC), accurate assessment of the lesions is crucial for determining the appropriate treatment strategy. We evaluated artificial intelligence (AI) for detecting peritoneal and small bowel dissemination in EOC using preoperative contrast-enhanced computed tomography (CT). We retrospectively analyzed 227 patients (254 CT scans) with histologically confirmed ovarian, fallopian tube, or peritoneal carcinoma who underwent surgery between October 2019 and December 2024 at two affiliated institutions. The mean patient age was 56.6 years. The FIGO stage was early in 118 cases and advanced in 136 cases. Histology included serous (n = 123) and non-serous (n = 131). The mean CA125 level and peritoneal cancer index (PCI) score were 2,243.9 U/mL and 5.8, respectively. Two AI models were developed: one for peritoneal dissemination (P-Model) and one for small bowel dissemination (SB-Model). Diagnostic performance was evaluated by sensitivity, specificity, and accuracy. For the P-Model, clinical factors were compared between high- and low-accuracy groups. The P-Model achieved 68.75% sensitivity, 85.84% specificity, and 77.30% accuracy; the SB-Model, 86.40%, 77.47%, and 81.93%, respectively. Low AI accuracy was linked to ascites volume, CA125 level, and PCI score. AI-based analysis of preoperative CT demonstrated favorable diagnostic performance for detecting dissemination in patients with EOC.

Keywords Artificial intelligence, Peritoneal dissemination, Small bowel dissemination, Ovarian cancer, Debulking surgery

Epithelial ovarian cancer (EOC) is the seventh most common cancer in females worldwide and the second most common gynecological malignancy[1]. Approximately 60%–70% of patients present at an advanced stage with massive peritoneal and visceral dissemination in the abdominal cavity (International Federation of Gynecology and Obstetrics [FIGO] stage III or IV)[2,3]. The optimal standard of care for patients with advanced ovarian cancer is either primary debulking surgery (PDS) or preoperative neoadjuvant platinum-based chemotherapy followed by cytoreductive surgery, which is called interval debulking surgery (IDS)[4]. Neoadjuvant chemotherapy (NAC) and IDS are considered in patients for whom complete resection at PDS is not feasible owing to tumor distribution or in those who cannot tolerate the extent of surgery required for complete cytoreduction. Strong evidence indicates that survival outcomes are significantly improved in patients who achieve no residual disease after debulking, regardless of whether PDS or IDS is performed[5]. However, advanced EOC often spreads extensively to all the peritoneal surfaces, including parietal and visceral surfaces, resulting in a high tumor burden, rendering complete resection impossible. Therefore, accurate preoperative prediction of tumor location and burden is critical for selecting patients in whom complete cytoreduction is achievable and for optimal surgical planning.

[1]Department of Obstetrics and Gynecology, The Jikei University School of Medicine, 3-25-8 Nishi-Shinbashi, Minato-Ku, Tokyo 105-8461, Japan. [2]SIOS Technology, Inc., Tokyo, Japan. [3]Department of Otorhinolaryngology, The Jikei University School of Medicine, Tokyo 105-8461, Japan. ✉email: t_seki@jikei.ac.jp

Computed tomography (CT) is widely used to assess the location and extent of peritoneal dissemination. However, its diagnostic accuracy remains suboptimal, often necessitating direct visual confirmation of tumor location via diagnostic laparotomy or laparoscopy[6].

In particular, dissemination on the surface of the small bowel and its mesentery is a common finding that often prevents complete tumor resection[7]. However, preoperative imaging for detecting such lesions remains challenging, with diagnostic accuracy rates as low as 20%–30%[5,8].

Thus, we conducted this study to optimize the diagnostic accuracy of detecting peritoneal and small bowel dissemination by integrating contrast-enhanced CT imaging studies with machine learning methods.

Results

P-Model for peritoneal dissemination diagnosis

Supplementary Table S1 summarizes the image-unit-based evaluation of the P-Model for diagnosing peritoneal dissemination, comparing single-model predictions and ensemble predictions on two types of images: entire CT images and 75% cropped images. The results showed that ensemble prediction using entire CT images achieved the highest diagnostic performance, with an accuracy of 69.49% (sensitivity = 62.26%, specificity = 76.73%).

Supplementary Fig S1 exhibits variability of accuracies of image-unit-based evaluation using P-Models. The horizontal line indicates the average accuracy of all models. The vertical line is a line plotted from the worst accuracy to the best accuracy. This chart demonstrates that the range of variability in accuracies of ensemble predictions is much smaller than that of single-model predictions.

Figure 1 and Supplementary Table S2 present the results of the case-unit-based evaluation of the P-Model for diagnosing peritoneal dissemination, comparing single-model and ensemble predictions across two types of images: the entire CT images and 75% cropped images. The rows of this table indicate the number of consecutive slice images used to calculate the positive prediction rate for each case.

The best performance was achieved with ensemble prediction on 21 consecutive slices of 75% cropped images, yielding an accuracy of 77.30% (95% confidence interval (CI) = 71.76 ~ 82.02%), sensitivity of 68.75% (95%CI = 59.86 ~ 76.43%), and specificity of 85.84% (95%CI = 79.02 ~ 90.7%).

Supplementary Fig S2 exhibits variability of accuracies of case-unit-based evaluation using P-Models. The horizontal line indicates the average accuracy of all models. The vertical line is a line plotted from the worst accuracy to the best accuracy. This chart demonstrates that the range of variability in accuracies of ensemble predictions is much smaller than that of single-model predictions.

Detailed sensitivities and specificities for all evaluations are provided in Supplementary Table S3.

SB-Model for small bowel dissemination diagnosis

Supplementary Table S4 summarizes the image-unit-based evaluation of the SB-Model for diagnosing small bowel dissemination, comparing single-model and ensemble predictions on entire CT images and 75% cropped images. Ensemble prediction using 75% cropped images achieved the best accuracy of 72.79% (sensitivity = 65.24%, specificity = 80.34%).

Supplementary Fig S3 exhibits variability of accuracies of image-unit-based evaluation using SB-Models. The horizontal line indicates the average accuracy of all models. The vertical line is a line plotted from the worst accuracy to the best accuracy. This chart demonstrates that the range of variability in accuracies of ensemble predictions is much smaller than that of single-model predictions.

Figure 2 and Supplementary Table S5 present the case-unit-based evaluation results. The rows of the table indicate the number of consecutive slice images used to calculate the positive prediction rate for each case. The best performance was achieved with ensemble prediction on nine consecutive slices of 75% cropped images, yielding an accuracy of 81.93% (95%CI = 76.73 ~ 86.17%), sensitivity of 86.40% (95%CI = 74.69 ~ 93.18%), and specificity of 77.47% (95%CI = 73.07 ~ 81.33%).

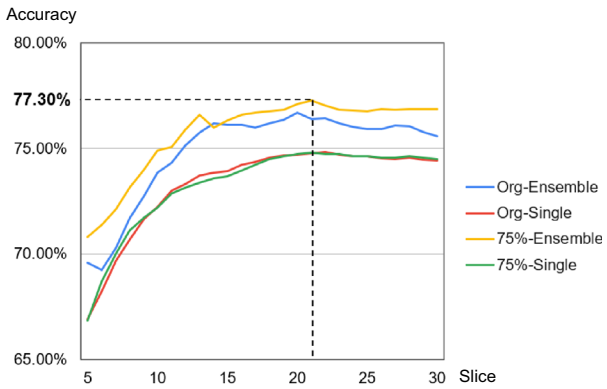

Fig. 1. Accuracy of the case-unit-based evaluation performed by the P-Model for the diagnosis of peritoneal dissemination. The y-axis represents AI accuracy, and the x-axis represents the number of consecutive CT slices analyzed. The ensemble prediction on 21 consecutive slices of 75% cropped images achieved the highest accuracy of 77.30%.

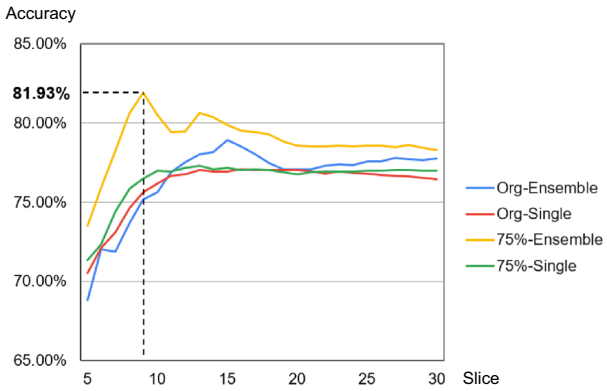

Fig. 2. Accuracy of the case-unit-based evaluation performed by the SB-Model for the diagnosis of small bowel dissemination. The y-axis represents AI accuracy, and the x-axis represents the number of consecutive CT slices analyzed. The ensemble prediction on nine consecutive slices of 75% cropped images achieved the highest accuracy of 81.93%.

Variable		Cases without peritoneal dissemination n = 22	Cases with peritoneal dissemination n = 24
Peritoneal dissemination detected on radiology report, n(%)	Yes	11 (50.0)	14 (58.3)
	No	11 (50.0)	10 (41.7)

Table 1. Radiological interpretation of cases with ≤ 25% AI accuracy in detecting the presence or absence of peritoneal dissemination.

Supplementary Fig S4 exhibits variability of accuracies of case-unit-based evaluation using SB-Models. The horizontal line indicates the average accuracy of all models. The vertical line is a line plotted from the worst accuracy to the best accuracy. This chart demonstrates that the range of variability in accuracy of ensemble predictions is much smaller than that of single-model predictions.

Detailed sensitivities and specificities of all evaluations are shown in Supplementary Table S6.

Evaluation of cases with low diagnostic accuracy by the AI model

Table 1 shows whether cases with an AI accuracy of ≤ 25% for the P-Model were correctly interpreted preoperatively by radiologists. In these cases, the sensitivity and specificity of radiologists' interpretations were 63.0% and 57.9%, respectively.

Table 2 compares the clinical characteristics of patients with AI diagnostic accuracy ≤ 25% versus > 25% for the P-Model, stratified by the presence or absence of peritoneal dissemination based on intraoperative macroscopic findings.

Among patients without peritoneal dissemination, those with AI accuracy ≤ 25% had significantly larger volume of ascites on preoperative CT compared to those with accuracy > 25%. Ascites extending into the upper abdomen were more frequently observed in the low-accuracy group ($p < 0.001$). Additionally, preoperative CA125 levels were significantly higher in this group ($p = 0.005$). In contrast, among patients with peritoneal dissemination, those with AI accuracy ≤ 25% exhibited significantly lower ascites volumes than those with higher accuracy. Most cases in the low-accuracy group had ascites limited to the pelvic cavity on preoperative CT ($p = 0.017$), and intraoperative ascites volume was ≤ 300 mL in the majority of cases ($p = 0.002$). Furthermore, preoperative CA125 levels and PCI scores were significantly lower in the low-accuracy group compared to the high-accuracy group ($p < 0.001$ and $p = 0.005$). Regardless of the presence or absence of peritoneal dissemination, no differences in accuracy were observed by BMI or histological subtype. As shown in Supplementary Table S7, similar results were obtained when the accuracy cutoff was set at 50%. Among patients without peritoneal dissemination, the low-accuracy group had a significantly larger volume of ascites on preoperative CT than the high-accuracy group (P = 0.007). In contrast, among patients with peritoneal dissemination, the low-accuracy group had significantly smaller volumes of ascites on both preoperative CT and intraoperative assessment compared with the high-accuracy group (P = 0.002 and P < 0.001, respectively). Furthermore, preoperative CA125 levels and PCI scores were significantly lower in the low-accuracy group than in the high-accuracy group (P = 0.002 and P = 0.004, respectively).

Discussion

We found that AI-based contrast-enhanced CT imaging can predict both peritoneal and small bowel dissemination with optimal accuracy. Notably, this is the first study to apply an AI-driven machine learning approach to the preoperative diagnosis of small bowel dissemination in EOC, demonstrating an accuracy

		Cases without peritoneal dissemination n = 131			Cases with peritoneal dissemination n = 123		
		accuracy ≤ 25%	accuracy > 25%	P-value	accuracy ≤ 25%	accuracy > 25%	P-value
Variable		n = 19	n = 112		n = 27	n = 96	
Ascites volume on CT, n(%)	The pelvis	12 (63.2)	106 (94.6)	< 0.001*	19 (70.4)	42 (43.8)	0.017*
	The upper abdomen	7 (36.8)	6 (5.4)		8 (29.6)	54 (56.2)	
Intraoperative ascites(ml), n(%)	≤ 300	15 (78.9)	104 (92.9)	0.074	21 (77.8)	41 (42.7)	0.002*
	300 <	4 (21.1)	8 (7.1)		6 (22.2)	55 (57.3)	
BMI(kg/m2), mean [range]		22.7 [16.6- 30.1]	22.0 [15.4- 34.3]	0.27	22.1 [17.1- 35.2]	21.9 [14.6- 32.9]	0.77
CA125(U/ml), mean [range]		1165.2 [13- 10769]	549.7 [4- 15882]	0.005*	1437.9 [2- 23961]	4686.1 [9- 142522]	< 0.001*
PCI score, mean [range]		0 [0- 0]	0.12 [0- 8]	0.47	8.46 [1-24]	12.98 [1- 36]	0.005*
Histological subtype, n(%)	HGSC	4 (21.1)	22 (19.6)	1	18 (66.7)	79 (82.3)	0.11
	Non- HGSC	15 (78.9)	90 (80.4)		9 (33.3)	17(17.7)	

Table 2. Correlation between AI accuracy and clinical parameters in the P-Model.

exceeding 80%. This performance surpasses the diagnostic accuracy of conventional CT image interpretation by radiologists reported in previous studies.

Debulking surgery for advanced EOC usually requires resection of multiple organs involved in tumor dissemination, including extensive peritoneal surfaces, diaphragm, small and large bowel, spleen, pancreas, and other organs, depending on the tumor location. Reliable prediction would allow clinicians to make timely treatment decisions, such as choosing between primary debulking surgery and neoadjuvant chemotherapy. Moreover, identifying organs likely to require resection before surgery would facilitate multidisciplinary planning, enabling early consultation with appropriate surgical specialists (e.g., hepatobiliary, pancreatic, colorectal, thoracic, or vascular surgeons) and preparation of the necessary equipment for complex procedures.

According to the ESMO–ESGO consensus conference[9], patients should not be considered suitable candidates for primary debulking surgery if the disease demonstrates extensive dissemination patterns that preclude complete cytoreduction. Specifically, contraindications include diffuse deep infiltration of the root of the small bowel mesentery or widespread carcinomatosis of the small bowel that would necessitate extensive resection, potentially resulting in short bowel syndrome (defined as a remaining bowel length of < 1.5 m). In addition, several anatomic sites, such as the stomach, duodenum, pancreatic head, and pancreatic body, have traditionally been considered difficult areas for complete tumor resection[10]. It has been reported that, among patients with residual disease after debulking surgery for advanced ovarian cancer, the most frequent sites of residual tumor are the small bowel mesentery and serosa, with approximately 80% of patients affected[11]. However, accurate preoperative prediction of peritoneal dissemination, especially small bowel involvement, remains challenging with current imaging modalities[5,6,8,12–15].

Therefore, clinicians still rely on diagnostic laparoscopy or laparotomy to clarify tumor location and determine the feasibility of complete resection through primary debulking surgery. However, these procedures are invasive, require general anesthesia, and may delay the initiation of treatment, especially for patients who would benefit from neoadjuvant chemotherapy. Such delays can negatively affect both the physical and emotional well-being of patients and may even worsen treatment outcomes[16]. If imaging studies could accurately predict disseminated tumor burden, it would streamline the treatment flow and improve clinical decision-making.

Some argue that residual tumor measuring < 1 cm is clinically acceptable, as it is preferable to residual disease > 1 cm[17,18]. From this perspective, preoperative prediction of small bowel dissemination may be considered less meaningful, because individual nodules of small bowel involvement are typically < 1 cm in size. However, accurate prediction of small bowel dissemination remains critically important. To date, no phase III trials have established optimal treatment strategies for patients in whom residual tumor < 1 cm is anticipated. Notably, a recent phase III trial demonstrated that, in the primary treatment of advanced EOC, progression-free survival was markedly improved, with a median of 27.9 months, when complete cytoreduction was achieved by PDS[19].

In this context, treatment selection between PDS and NAC followed by IDS in clinical practice is influenced by multiple practical and patient-related factors, including performance status, histologic subtype, institutional resources, and surgical scheduling, particularly when minimal residual disease (< 1 cm) is anticipated. Accordingly, accurate preoperative prediction of the sites and extent of peritoneal and small bowel dissemination is of critical importance.

Regarding imaging modalities, several reports have demonstrated relatively favorable outcomes with MRI, ultrasound (US), and PET-CT. Michielsen et al.[12] reported that whole-abdominal MRI achieved an average accuracy of 91% for detecting peritoneal dissemination and 81.0% for detecting small bowel dissemination. Similarly, Fischerova et al.[5] reported accuracies of 88% and 70% for detecting peritoneal dissemination and small bowel dissemination, respectively. These outcomes appear superior to those of our study (see Table 3). However, whole-body MRI currently requires prolonged acquisition times of approximately 40–50 minutes[20].

	Author	n	Imaging modality	Sensitivity (%)	Specificity (%)	Accuracy (%)
Peritoneal dissemination	Fischerova, et al. 5)	67	US	86.00	88.00	87.00
			WB-MRI	93.00	79.00	88.00
			CT	95.00	58.00	82.00
	Hu, et al. 8)	1614	CT	63.00	60.00	N/A
	Michielsen, et al. 12)	32	WB-MRI	91.00	91.00	91.00
			CT	65.00	82.00	75.00
	Di Donna, et al.13)	100	CT	95.00	27.00	72.00
	Tewari, et al. 16)	280	CT	78.00	89.00	88.00
	Schmidt, et al. 22)	15	CT	90.00	91.00	90.00
		15	PET-CT	93.00	96.00	95.00
	Alcazar, et al. 23)	93	CT	60.00	94.00	86.00
		93	US	70.00	98.00	91.00
	Rizzo, et al. 24)	92	WB-MRI	76.00	87.00	N/A
	Our study	254	CT	68.75	85.84	77.30
Small bowel dissemination	Fischerova, et al. 5)	67	US	42.00	94.00	N/A
			WB-MRI	65.00	89.00	70.00
			CT	65.00	92.00	78.00
	Takeda, et al. 6)	173	CT	33.30	98.60	N/A
	Hu, et al. 8)	1614	CT	24.00	88.00	58.00
	Michielsen, et al. 12)	32	WB-MRI	83.00	80.00	79.00
			CT	49.00	91.00	81.00
	Di Donna, et al.13)	100	CT	54.00	62.00	65.90
	Bagul, et al. 14)	36	CT	35.00	93.70	76.00
	Bristow, et al.15)	41	CT	33.30	100.00	61.10
	Our study	254	CT	86.40	77.47	81.93

Table 3. Comparison between our results and previous reports of imaging diagnosis for dissemination in patients with epithelial ovarian cancer. *N/A: not available, WB-MRI: whole-body magnetic resonance imaging, CT: computed tomography, US: ultrasound.*

In the report by Michielsen et al., patients were also required to drink 1 L of pineapple juice before whole-body MRI to aid in the diagnosis of ovarian cancer dissemination. Such requirements are challenging for critically ill patients with massive carcinomatosis or bulky tumors, who may struggle to maintain the scanning position for extended periods or tolerate fluid intake. Moreover, the availability of whole-body MRI scanners remains limited, and the associated costs are high. Given these limitations, the routine clinical use of whole-body MRI in this setting remains impractical.

With regard to US, achieving high accuracy in detecting disseminated regions requires the expertise of a well-trained technician or physician. Moreover, a thorough US examination can be time-consuming, similar to whole-body MRI, which can be challenging and burdensome for patients.

By comparison, CT scans offer advantages in terms of accessibility and patient comfort over both whole-body MRI and US. However, CT also has some drawbacks. As shown in Table 3, its diagnostic ability, especially for small bowel dissemination, remains suboptimal, even when interpreted by well-trained gynecologic radiologists[6,8,12–15,21–24].

In our study, AI-based machine learning analysis of contrast-enhanced CT scans demonstrated superior diagnostic performance compared to previous reports, particularly in terms of sensitivity for detecting small bowel dissemination. The SB-Model demonstrated superior performance, with a notably high sensitivity of 86.40%. Given the critical importance of preoperative detection of small bowel dissemination—where oversight may lead to inappropriate surgical planning, this result is of substantial clinical relevance. The high sensitivity of AI in this context represents a valuable advancement in preoperative evaluation.

However, there were cases in which the AI method could not achieve high diagnostic accuracy. In cases where the P-Model exhibited ≤ 25% accuracy, preoperative CT interpretations by radiologists demonstrated a sensitivity of 63.0%, a specificity of 57.9%, and an overall accuracy of 60.5%. These findings suggest that the poor performance of AI was not solely attributable to tumor size, location, or other factors typically challenging for human readers, but may also reflect limitations inherent to the model architecture, training dataset, or data augmentation strategy.

To further explore this hypothesis, we compared AI accuracy with clinical findings. Among patients without peritoneal dissemination, those with low AI accuracy more frequently showed ascites extending into the upper abdomen on preoperative CT and had significantly higher CA125 levels. In contrast, among patients with peritoneal dissemination, low-accuracy cases more often exhibited ascites limited to the pelvic cavity, with

intraoperative ascites volumes ≤ 300 mL in most cases and significantly lower PCI scores, suggesting a smaller tumor burden.

These findings imply that the AI model may be disproportionately influenced by the features emphasized in its training data. Specifically, it may have learned to associate the presence of widespread ascites with peritoneal dissemination, potentially leading to misclassification in atypical or early-stage cases. Importantly, this limitation could likely be mitigated by expanding the training dataset to include a more diverse range of clinical presentations, thereby improving the model's robustness and generalizability. However, we believe that AI-assisted diagnosis of CT scans can support radiologic interpretation and aid physicians in clinical decision-making.

This study has several limitations. First, our AI system in this study was designed to predict only the presence or absence of dissemination and did not assess the exact tumor location, tumor burden, or the presence of residual disease of debulking surgery. An AI system capable of providing this level of detail would be more useful for surgeons in assessing the feasibility of complete resection. To further improve our system, future efforts will focus on expanding the dataset, incorporating more detailed imaging features, and developing a more generalizable model to enhance diagnostic accuracy and clinical applicability. Second, this study did not include external validation. The performance of AI models may decrease when evaluated on external datasets[25,26]. Therefore, external validation should be conducted before clinical implementation and will be addressed in future work. Finally, the results may be influenced by bias arising from individual cases, which cannot be completely eliminated. However, as indicated by the 95% confidence intervals of our results, increasing the number of cases used for training and evaluation—although a potential approach to minimizing such bias—may also lead to a substantial decline in accuracy.

In conclusion, we demonstrated that AI can detect peritoneal dissemination—particularly small bowel dissemination—with high accuracy on preoperative contrast-enhanced CT in patients with ovarian cancer. Further refinement and validation of the model are warranted to facilitate its integration into clinical practice.

Methods

Patients and clinical information

This was a retrospective, multi-institutional observational study approved by the Institutional Review Board of The Jikei University School of Medicine (approval no. 35–149[11777]). The requirement for patient consent was waived in accordance with the Ethical Guidelines for Research Involving Human Subjects in Japan[27]. We included patients diagnosed with ovarian, fallopian tube, or peritoneal cancer who underwent surgery between October 2019 and December 2024 at two affiliated hospitals: The Jikei University School of Medicine and the Jikei University Kashiwa Hospital. A total of 227 patients were included, yielding 254 contrast-enhanced CT images for analysis. Clinical data were retrieved from medical records, including age, body mass index (BMI) at surgery, date of surgery, and preoperative CA125 levels. Intraoperative findings, such as the amount of ascites and the location of peritoneal dissemination, were assessed using the Peritoneal Cancer Index (PCI) score[28,29]. Pathological data, including histological subtype and FIGO stage, were also collected.

Patient characteristics

A total of 227 patients with epithelial ovarian cancer were included in this study. For patients who underwent IDS, CT scans were obtained twice, at the time of staging surgery (laparotomy or laparoscopy) and at IDS, and both scans were used for AI training. Therefore, the number of CT cases available for AI training was 254.

Table 4 presents the characteristics of all cases. The mean age at surgery was 56.6 (range, 19.0–87.0) years, and the mean BMI was 22.0 (range, 14.6–35.2) kg/m^2. FIGO stage at diagnosis was stage I in 86 cases (33.9%), stage II in 32 cases (12.6%), stage III in 105 cases (41.7%), and stage IV in 30 cases (11.8%). Histological subtypes included high-grade serous carcinoma in 123 patients (48.4%), endometrioid carcinoma in 47 (18.5%), clear cell carcinoma in 63 (24.8%), mucinous carcinoma in 9 (3.5%), adenocarcinoma in 4 (1.6%) and other histologies in 8 (3.1%). The mean preoperative serum CA125 level was 2,243.9 (range, 2–142,522) U/mL, and the mean PCI score was 5.8 (range, 0–36). Ascites volume was ≤ 300 mL in 181 patients (71.3%) and > 300 mL in 73 patients (28.7%).

CT imaging

We collected CT imaging data acquired before debulking or diagnostic surgery. All scans were contrast-enhanced with iodine and obtained in the portal venous phase, with a slice thickness of 5 mm. Imaging data were extracted as DICOM files from the two participating institutions and subsequently converted into JPEG format. Axial plane images covering the area from the liver and right diaphragm to the pubic bone were selected for machine learning analysis.

Number of CT images

Supplementary Table S8 shows the number of cases and the total number of CT images. Cases were classified into three groups: those with both small bowel and peritoneal dissemination (sb + & p +), those with peritoneal dissemination only (sb − & p +), and those with neither small bowel nor peritoneal dissemination (sb − & p −). The presence or absence of peritoneal and small bowel dissemination was determined based on intraoperative macroscopic findings.

Types of dissemination diagnostic models

In this study, we developed two deep neural network (DNN) models: one for diagnosing peritoneal dissemination (P-Model) and one for diagnosing small bowel dissemination (SB-Model). As shown in Fig. 3, the P-Model was trained on all CT images to predict whether an image demonstrated dissemination (peritoneal and/or small

Variable		n = 254
Age(year) , mean [range]		56.6 [19.0–87.0]
BMI(kg/m^2) , mean [range]		22.0 [14.6–35.2]
CA125(U/ml) , mean [range]		2243.9 [2–142522]
FIGO stage, n(%)	I	86(33.9)
	II	32 (12.6)
	III	106 (41.7)
	IV	30 (11.8)
Residual tumor, n(%)	No gross residual disease	178 (70.1)
	< 1 cm	11 (4.3)
	1 cm ≤	65 (25.6)
Surgical procedure, n(%)	PDS*	163 (64.2)
	IDS	32 (12.6)
	diagnostic laparotomy	13 (5.1)
	diagnostic laparoscopy	46(18.1)
Histological subtype, n(%)	high grade serous carcinoma	123 (48.4)
	endometrioid carcinoma	47 (18.5)
	clear cell carcinoma	63 (24.8)
	mucinous carcinoma	9 (3.5)
	adenocarcinoma	4 (1.6)
	other	8 (3.1)
PCI score, mean[range]		5.8 [0- 36]
Intraoperative Ascites (ml), n(%)	≤ 300	181(71.3)
	300 <	73(28.7)

Table 4. Patient characteristics used for AI training (n = 254). *Primary surgery for stage I was also included as PDS. BMI: body mass index, PDS: primary debulking surgery, IDS: interval debulking surgery. PCI: peritoneal cancer index.

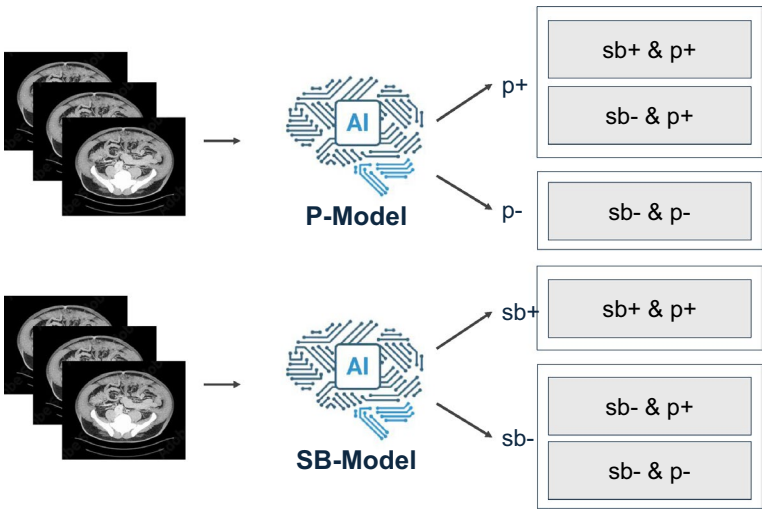

Fig. 3. Overall study workflow. The P-Model was trained on all CT images to predict whether an image demonstrated dissemination (peritoneal and/or small bowel; p +) or no dissemination (p –). The SB-Model was trained on all CT images to predict the presence (sb +) or absence (sb –) of small bowel dissemination. Images containing peritoneal dissemination without small bowel involvement were classified as sb – .

bowel; p +) or no dissemination (p –). The SB-Model was trained on all CT images to predict the presence (sb +) or absence (sb –) of small bowel dissemination. Images containing peritoneal dissemination without small bowel involvement were classified as sb – .

Shape of DNN model

In this study, we adopted the MobileNetV2 architecture[30], a relatively compact DNN consisting of 88 layers, with a fixed input image size of 224 × 244 pixels and 3,538,984 learning parameters. We selected MobileNetV2 because its compact architecture helps reduce overfitting under limited data conditions. In addition, this architecture has already been widely applied in medical image analysis, demonstrating versatility and robustness across clinical imaging tasks[31].

Training and cross-validation of the DNN model

We adopted an eight-fold cross-validation approach by dividing the cases into eight groups, using seven groups for training (training dataset) and the remaining group for evaluation (evaluation dataset), maintaining a 7:1 ratio. As a result, eight DNN models were generated, each corresponding to one cross-validation set. To increase dataset diversity, images in each training dataset were augmented 960-fold. During training, One out of 480 augmented images were randomly selected for each epoch. A total of 20 epochs were performed for each cross-validation set. The augmentation was performed using general augmentation techniques including rotation, flipping, scaling, etc. and it was applied only to the training data but not for the evaluation.

Because DNN models exhibit variability in performance when trained on large volumes of augmented data derived from a small number of patients, 24 sets of cross-validation sets were created to verify fluctuations in model accuracy. Each set consisted of 8 models, resulting in a total of 192 models (8 datasets × 24) for both the P-Model and SB-Model.

Supplementary Table S9 shows the number of cases and images of each group used for cross-validation during P-Model training.

Supplementary Table S10 shows the number of cases and images in the training and evaluation sets for each cross-validation of the P-Model.

Supplementary Table S11 shows the number of cases and images in each group used for cross-validation during SB-Model training.

Supplementary Table S12 shows the number of cases and images in the training and evaluation sets for each cross-validation of the SB-Model.

Evaluation of DNN models

The performance of the DNN models was evaluated using two methods: image-unit-based evaluation and case-unit-based evaluation, applied to two types of images: the entire CT image and images cropped to the central 75% of the original field (equivalent to 133% zoom). Both the P-Model and SB-Model were assessed using these methods.

For image-unit-based evaluation, predictions were performed in two ways: single-model predictions and ensemble predictions using 23 models, all derived from 24 cross-validation sets.

For case-unit-based evaluation, the highest positive prediction rate was calculated across consecutive slice images for each case. The number of consecutive slices examined ranged from 5 to 30. Predictions were again performed using both single and ensemble approaches on the two types of images (entire and 75% cropped).

We adopted the Wilson method to calculate the 95% confidence interval of accuracy, sensitivity and specificity[32].

Evaluation of CT images by radiologists

Furthermore, we evaluated whether radiologists could make a diagnosis in cases where the P-model showed less than 25% accuracy. The diagnostic performance of radiologists was assessed by retrospectively reviewing preoperative CT interpretation reports that had been generated by board-certified radiologists at our institutions. These reports were independently reviewed by gynecologic oncologists. Radiologists were consulted regarding the feasibility of diagnosing small bowel dissemination. They indicated that such a diagnosis is nearly impossible because individual nodules on the surface of the small bowel are very small. Therefore, radiologist-based evaluation of small bowel dissemination was not conducted. In addition, we examined the features of the images associated with low AI performance in the P-Model, defining low AI performance as below 25% and 50%. We analyzed the associations between diagnostic accuracy and preoperative CA125 levels, BMI, intraoperative PCI score, and intraoperative ascites volume.

Statistical analysis

Continuous variables were expressed as means with ranges and analyzed using either the Mann–Whitney *U* test or the Student's *t*-test, as appropriate. Categorical variables were reported as absolute numbers and analyzed using Fisher's exact test. A p-value of less than 0.05 was considered statistically significant.

Ethics

This study was approved by the Institutional Review Board of The Jikei University School of Medicine (approval no. 35–149[11777]). In patient application forms, the need for informed consent was waived by these above institutional review boards. It was clearly stated that patients were allowed to opt out of the study at any time. Information on how they could opt out was provided on our website, and arrangements were made for patients to opt out. All methods were performed in accordance with the relevant guidelines and regulations.

Data availability

All data generated and analyzed during this study are included in this published article and its supplementary files. Additional datasets are available from the corresponding author upon reasonable request. Requests for data should be addressed to Toshiyuki Seki, M.D., Ph.D. (e-mail: t_seki@jikei.ac.jp).

Received: 24 October 2025; Accepted: 23 February 2026
Published online: 23 March 2026

References

1. Ferlay, J. et al. (ed IARC) (Lyon, France, 2013).
2. Heintz, A. P. et al. Carcinoma of the ovary. *Int. J. Gynaecol. Obstet.* **83**(Suppl 1), 135–166. https://doi.org/10.1016/s0020-7292(03)90118-4 (2003).
3. Pecorelli, S., Favalli, G., Zigliani, L. & Odicino, F. Cancer in women. *Int. J. Gynaecol. Obstet.* **82**, 369–379. https://doi.org/10.1016/s0020-7292(03)00225-x (2003).
4. Chang, S. J., Bristow, R. E., Chi, D. S. & Cliby, W. A. Role of aggressive surgical cytoreduction in advanced ovarian cancer. *J. Gynecol. Oncol.* **26**, 336–342. https://doi.org/10.3802/jgo.2015.26.4.336 (2015).
5. Fischerova, D. et al. Preoperative staging of ovarian cancer: Comparison between ultrasound, CT and whole-body diffusion-weighted MRI (ISAAC study). *Ultrasound Obstet. Gynecol.* **59**, 248–262. https://doi.org/10.1002/uog.23654 (2022).
6. Takeda, T. et al. Evaluation of preoperative prediction of intestinal invasion in patients with ovarian cancer. *Int. J. Gynaecol. Obstet.* **153**, 398–404. https://doi.org/10.1002/ijgo.13492 (2021).
7. González-Martín, A. et al. Newly diagnosed and relapsed epithelial ovarian cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann. Oncol.* **34**, 833–848. https://doi.org/10.1016/j.annonc.2023.07.011 (2023).
8. Hu, T. W. Y., Nie, D., Gou, J. H. & Li, Z. Y. Predictive significance of preoperative CT findings for suboptimal cytoreduction in advanced ovarian cancer: A meta-analysis. *Cancer Manag. Res.* **10**, 2019–2030. https://doi.org/10.2147/cmar.s166658 (2018).
9. Colombo, N. et al. ESMO-ESGO consensus conference recommendations on ovarian cancer: Pathology and molecular biology, early and advanced stages, borderline tumours and recurrent disease. *Int. J. Gynecol. Cancer* **29**, 728–760. https://doi.org/10.1136/ijgc-2019-000308 (2019).
10. Querleu, D. et al. European Society of Gynaecological Oncology (ESGO) Guidelines for Ovarian Cancer Surgery. *Int. J. Gynecol. Cancer* **27**, 1534–1542. https://doi.org/10.1097/igc.0000000000001041 (2017).
11. Heitz, F. et al. Pattern of and reason for postoperative residual disease in patients with advanced ovarian cancer following upfront radical debulking surgery. *Gynecol. Oncol.* **141**, 264–270. https://doi.org/10.1016/j.ygyno.2016.03.015 (2016).
12. Michielsen, K. et al. Whole-body MRI with diffusion-weighted sequence for staging of patients with suspected ovarian cancer: A clinical feasibility study in comparison to CT and FDG-PET/CT. *Eur. Radiol.* **24**, 889–901. https://doi.org/10.1007/s00330-013-3083-8 (2014).
13. Di Donna, M. C. et al. Concordance of radiological, laparoscopic and laparotomic scoring to predict complete cytoreduction in women with advanced ovarian cancer. *Cancers* **15**, 500. https://doi.org/10.3390/cancers15020500 (2023).
14. Bagul, K., Vijaykumar, D. K., Rajanbabu, A., Antony, M. A. & Ranganathan, V. Advanced primary epithelial ovarian and peritoneal carcinoma—Does diagnostic accuracy of preoperative CT scan for detection of peritoneal metastatic sites reflect into prediction of suboptimal debulking? A prospective study. *Indian J. Surg. Oncol.* **8**, 98–104. https://doi.org/10.1007/s13193-016-0601-6 (2017).
15. Bristow, R. E. et al. A model for predicting surgical outcome in patients with advanced ovarian carcinoma using computed tomography. *Cancer* **89**, 1532–1540. https://doi.org/10.1002/1097-0142(20001001)89:7%3c1532::aid-cncr17%3e3.0.co;2-a (2000).
16. Tewari, K. S., Java, J. J., Eskander, R. N., Monk, B. J. & Burger, R. A. Early initiation of chemotherapy following complete resection of advanced ovarian cancer associated with improved survival: NRG Oncology/Gynecologic Oncology Group study. *Ann. Oncol.* **27**, 114–121. https://doi.org/10.1093/annonc/mdv500 (2016).
17. Fagotti, A. et al. Randomized trial of primary debulking surgery versus neoadjuvant chemotherapy for advanced epithelial ovarian cancer (SCORPION-NCT01461850). *Int. J. Gynecol. Cancer* **30**, 1657–1664. https://doi.org/10.1136/ijgc-2020-001640 (2020).
18. Onda, T. et al. Comparison of survival between primary debulking surgery and neoadjuvant chemotherapy for stage III/IV ovarian, tubal and peritoneal cancers in phase III randomised trial. *Eur. J. Cancer* **130**, 114–125. https://doi.org/10.1016/j.ejca.2020.02.020 (2020).
19. Reuss, A. et al. TRUST: Trial of Radical Upfront Surgical Therapy in advanced ovarian cancer (ENGOT ov33/AGO-OVAR OP7). *Int. J. Gynecol. Cancer* **29**, 1327–1331. https://doi.org/10.1136/ijgc-2019-000682 (2019).
20. Tunariu, N. et al. What's new for clinical whole-body MRI (WB-MRI) in the 21st century. *Br. J. Radiol.* **93**, 20200562. https://doi.org/10.1259/bjr.20200562 (2020).
21. Tempany, C. M. et al. Staging of advanced ovarian cancer: Comparison of imaging modalities--report from the Radiological Diagnostic Oncology Group. *Radiology* **215**, 761–767. https://doi.org/10.1148/radiology.215.3.r00jn25761 (2000).
22. Schmidt, S., Meuli, R. A., Achtari, C. & Prior, J. O. Peritoneal carcinomatosis in primary ovarian cancer staging. *Clin. Nucl. Med.* **40**, 371–377. https://doi.org/10.1097/rlu.0000000000000768 (2015).
23. Alcázar, J. L. et al. Pre-operative assessment of intra-abdominal disease spread in epithelial ovarian cancer: A comparative study between ultrasound and computed tomography. *Int. J. Gynecol. Cancer* **29**, 227–233. https://doi.org/10.1136/ijgc-2018-000066 (2019).
24. Rizzo, S. et al. Pre-operative evaluation of epithelial ovarian cancer patients: Role of whole body diffusion weighted imaging MR and CT scans in the selection of patients suitable for primary debulking surgery. A single-centre study. *Eur. J. Radiol.* **123**, 108786. https://doi.org/10.1016/j.ejrad.2019.108786 (2020).
25. Yu, A. C., Mohajer, B. & Eng, J. External Validation of Deep Learning Algorithms for Radiologic Diagnosis: A Systematic Review. *Radiol.: Artif. Intell.* https://doi.org/10.1148/ryai.210064 (2022).
26. Kim, D. W., Jang, H. Y., Kim, K. W., Shin, Y. & Park, S. H. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: Results from recently published papers. *Korean J. Radiol.* **20**, 405. https://doi.org/10.3348/kjr.2019.0025 (2019).
27. Eba, J. & Nakamura, K. Overview of the ethical guidelines for medical and biological research involving human subjects in Japan. *Jpn. J. Clin. Oncol.* **52**, 539–544. https://doi.org/10.1093/jjco/hyac034 (2022).
28. Lampe, B., Kroll, N., Piso, P., Forner, D. M. & Mallmann, P. Prognostic significance of Sugarbaker's peritoneal cancer index for the operability of ovarian carcinoma. *Int. J. Gynecol. Cancer* **25**, 135–144. https://doi.org/10.1097/igc.0000000000000327 (2015).
29. Climent, M. T., Serra, A., Gilabert-Estellés, J., Gilabert-Aguilar, J. & Llueca, A. Comparison of peritoneal carcinomatosis scoring methods in predicting resectability and prognosis in gynecologic malignancies. *J. Clin. Med.* **10**, 2553. https://doi.org/10.3390/jcm10122553 (2021).
30. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L.-C. in *2018 IEEE/CVF conference on computer vision and pattern recognition*. 4510–4520 (IEEE).
31. Takahashi, Y. et al. Automated system for diagnosing endometrial cancer by adopting deep-learning technology in hysteroscopy. *PLoS ONE* **16**, e0248526. https://doi.org/10.1371/journal.pone.0248526 (2021).
32. Wilson., E. B. Probable Inference, the Law of Succession, and Statistical Inference. *J. Am. Stat. Assoc.* **22**, 209 (1927).

Acknowledgements
We thank all the team members at The Jikei University School of Medicine for their contributions to patient care and data acquisition, and SIOS Technology, Inc., for their cooperation. We are also grateful to the Otolaryngology-led AI Research Team at Jikei University, led by Dr. Masahiro Takahashi and consisting primarily of otolaryngologists, for their expertise in AI-based medical research and for facilitating collaboration with the Department of Obstetrics and Gynecology, which enabled this interdisciplinary study. We further thank Dr. Kazuhiko Morikawa, Dr. Youhei Ohki, and Dr. Akira Baba of the Department of Radiology, Jikei University, for their valuable assistance with the study design. Finally, we thank Oxford Science Editing (https://www.oxfordscience.org/) for English language editing. The graphical abstract was created using BioRender (Grant No. WZ28OBEZPR, YU28S6DBAR).

Author contributions
Conception and design: Toshiyuki Seki, Ria Kim. Acquisition of clinicopathological data: Toshiyuki Seki, Ria Kim, Kota Yokosu, Rintaro Hamada, Erika Habuchi. Analysis and interpretation of data: Katsuhiko Noda, Kaname Yoshida, Ria Kim, Toshiyuki Seki. Writing, review: Ria Kim, Toshiyuki Seki, Katsuhiko Noda, Kaname Yoshida Study supervision: Toshiyuki Seki, Katsuhiko Noda, Kaname Yoshida, Masahiro Takahashi, Hirokuni Takano, Aikou Okamoto.

Funding
This work was supported by the Department of Obstetrics and Gynecology at the Jikei University School of Medicine, and by JSPS KAKENHI (Grant Number JP80569559). JSPS KAKENHI had no role in the study, the Department of Obstetrics and Gynecology at the Jikei University School of Medicine contributed by providing clinical data.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-41728-4.

Correspondence and requests for materials should be addressed to T.S.