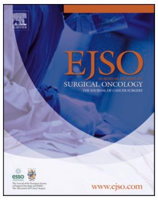
European Journal of Surgical Oncology 52 (2026) 111480

Contents lists available at ScienceDirect

European Journal of Surgical Oncology

journal homepage: www.ejso.com



Review Article

The role of artificial intelligence in pre-operative prediction of completeness of cytoreduction for peritoneal surface malignancies: a scoping review

Samuel Pau [a,*], Timothy Eglinton [b], Alan Wang [c], Jesse Fischer [a]

[a] Waikato Clinical Campus, Faculty of Medical and Health Sciences, University of Auckland, Private Bag 3200, Hamilton, 3240, New Zealand
[b] Department of Surgery and Critical Care, University of Otago, Christchurch, Christchurch Hospital Campus, 2 Riccarton Avenue, Christchurch Central, Christchurch, 8011, New Zealand
[c] Auckland Bioengineering Institute, University of Auckland, 70 Symonds Street, Auckland Central, Auckland, 1010, New Zealand

ARTICLE INFO

Keywords:
Artificial intelligence
Machine learning
Deep learning
Radiomics
Cytoreductive surgery
Peritoneal surface malignancies

ABSTRACT

Background: Complete cytoreduction is the most important prognostic factor for patients with peritoneal surface malignancies (PSM) and its prediction remains a clinical challenge. Artificial intelligence (AI) offers a novel opportunity to integrate clinical and imaging features to support surgical decision-making. This scoping review aimed to analyse applications of AI for predicting cytoreduction completeness in PSM.

Methods: A scoping review was conducted in accordance with PRISMA-ScR guidelines and registered with Open Science Framework. PubMed, Scopus, and Embase were searched from 2000 to 2025. Eligible studies applied AI models to predict cytoreduction completeness in patients with PSM using pre-operative data. Data were extracted on study design, disease type, model architecture, input predictors, performance metrics and explainability strategies.

Results: From 262 records identified from the search strategy, nine studies were included. Seven focused on ovarian cancer, one on synchronous colorectal peritoneal metastases, and one on a mixed PSM. Area under the curve (AUC) values ranged from 0.70 to 0.98. Radiomics–clinical nomograms consistently outperformed single-modality models. The DeAF deep learning framework achieved the strongest multicentre validation (AUC of 0.90), underscoring the potential of deep feature extraction. However, explainability was limited to nomograms, feature importance plots, or calibration analyses; no study adopted modern explainable AI techniques.

Conclusion: AI models demonstrate potential for pre-operative prediction of cytoreduction completeness in PSM, particularly when radiomics are combined with clinicopathological factors or when deep learning is applied. Future research should prioritise multicentre external validation, integration of multimodal data and the adoption of explainability tools to enable clinical translation.

1. Introduction

Complete cytoreduction remains the strongest determinant of survival for patients with peritoneal surface malignancies (PSM) and this appears consistent for tumour types including ovarian cancer, colorectal cancer, gastric cancer and mesothelioma [1–4]. Accurate prediction of complete cytoreduction remains a major unmet need in the management of PSM. Many pre-operative variables have been studied as potential predictors of completeness of cytoreduction (CC) but no reliable predictive models or tools have yet been adopted in clinical practice to determine that. Incomplete cytoreduction is associated with poor survival and exposes patients to substantial morbidity, mortality, and prolonged deterioration in quality of life from an extensive operation that offers little oncological benefit [1].

Conventional imaging alone provides limited accuracy, with standard prognostic tools such as the radiological peritoneal carcinomatosis index (rPCI) demonstrating notable inter-observer variability [5]. Computed tomography (CT) is the current mainstay of pre-operative staging but it demonstrates a sensitivity of only 28–51% for peritoneal disease with limited ability to detect nodules smaller than 5 mm [5].

---

[*] Corresponding author.
E-mail addresses: samuelpau_0524@hotmail.com (S. Pau), tim.eglinton@cdhb.health.nz (T. Eglinton), alan.wang@auckland.ac.nz (A. Wang), jesse.fischer@waikatodhb.health.nz (J. Fischer).

https://doi.org/10.1016/j.ejso.2026.111480
Received 29 October 2025; Received in revised form 26 January 2026; Accepted 18 February 2026
Available online 18 February 2026

Therefore, there is a pressing need for methods capable of integrating multifaceted pre-operative predictors including radiological, clinicopathological and biomarkers-based factors to predict the CC accurately, reliably, and transparently.

Artificial intelligence (AI) encompassing machine learning and deep learning, has emerged as a promising strategy to address this challenge. AI is an overarching term describing computational systems designed to perform tasks that typically require human intelligence [6]. Within this framework, machine learning provides systems with the ability to learn patterns from prior data to generate predictions without being specifically programmed [6]. Deep learning is a form of machine learning that uses multi-layered neural networks to automatically learn complex patterns from raw data [7]. In contrast to traditional machine learning, which depends on predefined features, deep learning can extract relevant features directly from imaging or clinical data, particularly when large datasets are available [7].

In surgical oncology, AI approaches have been increasingly applied to pre-operative imaging analysis, particularly through radiomics. Radiomics enables the extraction of hundreds to thousands of quantitative features from conventional cross-sectional imaging [6]. Radiomics can discriminate tumour heterogeneity beyond human perception, in contrast with the traditional practice of utilising imaging solely for visual interpretation [8]. Radiomic features include basic intensity statistics (e.g. average signal or variation), shape descriptors (e.g. volume and sphericity), texture patterns that reflect tissue organisation, and higher-order features generated through mathematical filters like wavelet transforms [8]. Machine learning algorithms can use these features to generate predictive models, while deep learning—especially convolutional neural networks (CNNs)—can autonomously learn and extract complex imaging patterns from large datasets [9]. Recent studies have demonstrated encouraging results in detection of peritoneal disease with AI-based imaging analysis [6].

Given this growing interest, it is timely to systematically review how AI has been applied to CC in PSM. This scoping review aims to (1) describe the range of model architectures, (2) describe disease types and input modalities employed, (3) summarise predictive performance of AI across cancer types and cohorts, (4) evaluate the strengths, weaknesses and transparency of the AI used and (5) identify methodological gaps and research priorities to guide future development. By synthesising the available evidence, this scoping review aims to provide a foundation for advancing AI tools into clinically robust decision-support systems that can identify patients with the highest likelihood of achieving complete cytoreduction.

2. Methods

This scoping review was performed in accordance with the preferred reporting items for systematic reviews and meta-analyses extension for scoping reviews (PRISMA-ScR) framework and checklist [10]. This methodology was elected over a systematic review due to the paucity of existing data surrounding this emerging topic, which has only relatively recently gained traction in the scientific community. The final protocol was registered retrospectively with the Open Science Framework on October 14, 2025.

2.1. *Eligibility criteria*

Peer-reviewed studies were included if they investigated patients with any type of PSM and applied artificial intelligence (AI) techniques such as machine learning or deep learning to pre-operative data for predicting cytoreduction completeness, reporting at least one model performance metric (e.g., AUC, accuracy, sensitivity, specificity). Studies relying on intra-operative or post-operative variables were excluded. Reviews, case reports, editorials, and conference abstracts lacking sufficient methodological detail were also excluded.

For this review, machine learning was defined as any algorithmic approach using automated pattern recognition or data-driven model optimization beyond traditional parameter estimation. Eligible methods included penalized regression (LASSO, Ridge, Elastic Net), ensemble algorithms (random forest, gradient boosting), support vector machines, and neural-network architectures. Studies using only standard univariate or multivariate logistic regression without regularization, automated feature selection or cross-validated model training were excluded.

2.2. *Information sources and search strategy*

A structured literature search was conducted in PubMed, Scopus and Embase using the following query:

((resectability) OR (complete cytoreduction) OR (incomplete cytoreduction))

AND ((deep learning) OR (machine learning) OR (radiomics) OR (artificial intelligence))

AND ((peritoneal) OR (peritoneal carcinomatosis) OR (peritoneal metastases)

OR (pseudomyxoma peritonei) OR (peritoneal surface malignancies))

2.3. *Selection of sources of evidence*

Literature search was conducted on July 28, 2025. All search results were exported to Rayyan AI [11]. Rayyan AI was used as a screening support tool to facilitate title and abstract review. Duplicate results were removed. Titles and abstracts were assessed to determine relevance, according to the eligibility criteria described above. All studies evaluating AI-based prediction of CC in PSM were selected for further screening to confirm relevance. Studies that did not utilise any AI model in predicting CC were subsequently excluded. Additionally, all animal studies were excluded. Two reviewers (SP and JF) performed the entire screening process on all titles and abstracts independently and sequentially to identify relevant publications, with Rayyan platform used to assist in workflow organisation and conflict identification Full-text review and final study inclusion were also performed manually by the reviewers. Disagreements were discussed and resolved without need for moderation.

2.4. *Data extraction*

A standardized data extraction form was used to collect the following information which included bibliographic details (author, year, country, and journal), population and disease type, model architecture and AI technique, input modalities, model performance metrics and explainability strategies.

2.5. *Synthesis of results*

Extracted data were synthesized narratively and grouped systematically into model architecture and radiomics integration, disease types represented, input modalities, model performance and predictive accuracy, explainability and transparency and methodological gaps and research priorities. No meta-analysis was performed due to heterogeneity in study design and outcome reporting.

3. Results

3.1. *Overview of included studies*

Nine studies, published between 2000 and 2025, met the inclusion criteria (Fig. 1). Seven solely included patients with advanced epithelial ovarian cancer (EOC), one study included patients with colorectal peritoneal metastases (CPM) and another study included patients with a range of PSM including colorectal, gynaecological and gastric cancer, pseudomyxoma peritonei (PMP), primary peritoneal malignancy and

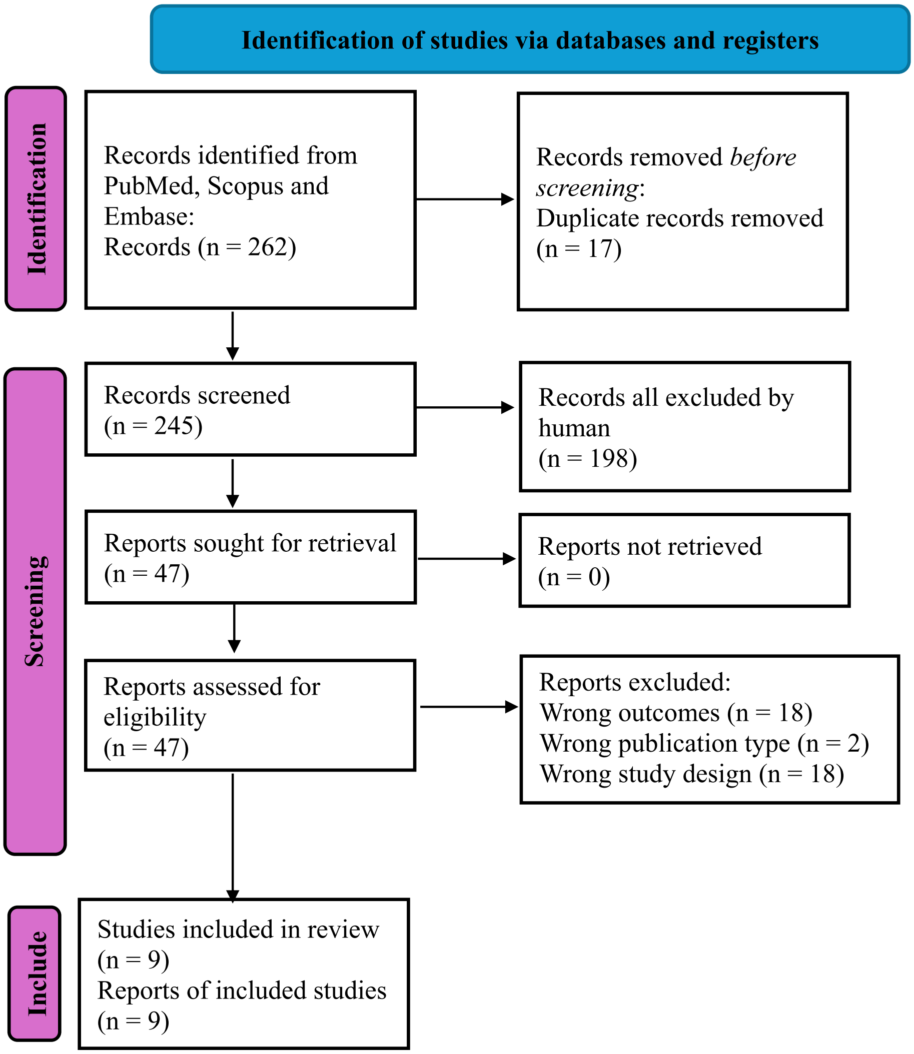


Fig. 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram outlining the identification and screening of included articles.

sarcoma. All studies employed machine learning to predict the likelihood of complete cytoreduction (CC0 or R0) and reported performance metrics such as area under the curve (AUC), accuracy, sensitivity and specificity. Although this scoping review was designed to focus on pre-operative predictors, one study included intra-operative variables within their models but was retained due to incorporation of other pre-operative variables. Another study used the Surgical Complexity Score (SCS) in the form of anticipated surgical effort rather than an intra-operative finding thus fulfilling the inclusion criteria.

3.1.1. *Model architecture: radiomics integration and machine learning strategies*

The reviewed studies exhibit a spectrum of model architectures, ranging from classical machine learning classifiers, instance-based learning, and regression-based models to deep learning frameworks. One of the key distinctions was whether radiomics were integrated into the models.

Radiomics-based models were the most common, applied in five studies as shown in Table 1. This method involves manual segmentation of regions of interest followed by a mathematical extraction of hundreds to thousands of quantitative imaging features (including intensity, shape, texture and wavelet transform) before incorporating them into a predictive nomogram [12]. Li et al. (2021), Guo et al. (2023), Lu et al. (2023) and Liu et al. (2025) utilised magnetic resonance imaging (MRI) and Li et al. (2025) utilised ultrasound scan (US) imaging [13–17]. Feature dimensionality was reduced using the Least Absolute Shrinkage and Selection Operator (LASSO), which selects the most informative variables to prevent overfitting [12]. The reduced features were then combined with clinical factors to generate the respective radiomic-clinical nomograms. In contrast, the other three studies, Maubert et al. (2019), Piedimonte et al. (2022) and Laios et al. (2020) did not utilise radiomics [18–20]. They relied on clinical and radiological scores such as PCI, surgical complexity and unresectability criteria to generate a predictive nomogram [18–20].

In terms of machine learning strategies, linear models such as regression-based LASSO logistic models were most used. This was utilised by the five radiomic-based studies [13–17]. Instance-based learning was used by Laios et al. (2020), who applied k-Nearest Neighbours (k-NN), an algorithm that classifies patients by comparing them with the most similar cases in the dataset [20]. Other classical machine learning classifiers utilised included tree-based models such decision trees (DT), random forests (RF) and gradient boosted trees (GBT) and margin-based models such as support vector machines (SVM). Maubert et al. (2019) compared various machine learning techniques; Li et al. (2025) tested 15 machine learning algorithms in their study; and Piedimonte et al. (2022) employed GBT in their study [17–19].

Deep learning specifically the Decoupling Feature Alignment and Fusion (DeAF) framework was used by Lin et al. (2025) on computed tomography (CT) scans with clinicopathological data integration [21]. DeAF differs from traditional radiomic-based learning models by

Table 1
Study model architecture including utility of radiomics, machine learning strategy and input variables.

Study	Radiomics	Segmentation	Feature extraction	Machine Learning Strategies	Deep Learning	Disease Type	Input Modalities
Maubert et al., 2019	No radiomics extraction and selection carried out	None	None	Tested multiple classical machine learning (CT, RF, DT, SVM)	None	Colorectal cancer Gynaecological Gastric Pseudomyxoma Peritonei Peritoneal primitive origin Sarcoma	**Clinical factors**: Age, ECOG, Disease type, BMI, ASA, Sex **Radiological factors:** Nine criteria of non-resectability reflecting organ involvement based on CT assessment of disease extent
Laios et al., 2020	No radiomics extraction and selection carried out	None	None	Instance-based (k-nearest neighbours) logistic regression	None	Advanced epithelial ovarian cancer (AOC)	**Clinical factors:** Age, BMI, pre-treatment CA 125, resection margins (R0), Charlson Comorbidity Index, type of surgery **Radiological factors:** Surgical complexity score (anticipated extent of surgery) Disease score (imaging and CA 125)
Li et al., 2021	Yes (MRI)	Manual (radiologists; software not specified)	Handcrafted radiomics (MRI derived features)	Regression-based (LASSO logistic regression)	None	High grade Serous Ovarian Carcinoma (HGSOC)	**Clinical factors:** Resection margins, CA 125, LDH, NLR and age **MRI Radiomic features:** 396 features – first order, shape, texture and wavelet features
Piedimonte et al., 2022	No radiomics extraction and selection carried out	None	None	Gradient-Boosted Trees (GBT) model	None	AOC	**Clinical factors**: Age, ECOG, BMI, BRCA status, Albumin, Ascites, CA-125 **Radiological factors:** Elements of a modified surgical resectability score derived from radiological assessment of disease burden and elements of the pre-operative surgical complexity score derived from anticipated procedures.
Guo et al., 2023	Yes (MRI)	Manual (radiologists; ITK SNAP)	Handcrafted radiomics (CT derived features)	Regression-based (LASSO logistic regression)	None	HGSOC	**Clinical factors:** Age, CA125, HE-4, LDH, NLR, BMI, ascites, height, weight, body surface area, tumour size, history and laterality **MRI radiomics features:** 1305 features – first order, shape, texture and wavelet features
Lu et al., 2023	Yes (MRI)	Manual (radiologists; ITK SNAP)	Handcrafted radiomics (CT derived features)	Regression-based (LASSO logistic regression)	None	HGSOC	**Clinical factors:** Age, MAP score – metastases in abdomen and pelvis, CA 125, Serum HE-4, NLR, FIGO stage, ascites **MRI radiomic features:** 1316 features – first order, shape, texture and wavelet features
Lin et al., 2025	Deep learning using CNN extraction (CT)	Manual (radiologists; ITK SNAP)	Automatic CNN-based deep feature extraction (ResNet3D with self-attention)	DeAF framework with self- attention	Yes	Synchronous colorectal peritoneal metastases	**Clinical factors:** Sex, age, CEA, CA 19-9, CA125, tumour location, liver metastasis, T and N stage, KRAS mutation, BRAF mutation, histological subtype, tumour differentiation and radiological PCI scores **CNN automatically extracted deep features**
Li et al., 2025	Yes (US)	Manual (radiologists; software not specified)	Handcrafted radiomics (US derived features)	Tested multiple classical machine learning (logistic regression, random forest, SVM, Light GBM)	No	AOC	Clinical factors: Clinical variables, demographics, symptoms, Laboratory tests, ultrasound derived doppler metrics, MAP scores **US radiomic features:** 1561 features - first order, shape, texture and wavelet features
Liu et al., 2025	Yes (MRI)	Manual segmentation by radiologists and handcrafted radiomics extracted	Handcrafted radiomics (MRI-derived features)	Regression-based (LASSO logistic regression)	None	Serous Ovarian Carcinoma (SOC)	**Clinical factors:** Age, CA-125, HE4, LDH, NLR, ASA **MRI radiomics features:** 1218 features – first order, shape, texture and wavelet features

Abbreviations: ASA- American Society of Anaesthesiologists physical status; BMI - body mass index; BRCA - breast cancer gene; CA 125– cancer antigen 125, CA 19-9 – carbohydrate antigen 19-9; CEA – carcinoembryonic antigen; CNN - convolutional neural network; ECOG - Eastern Cooperative Oncology Group performance status; FIGO - International Federation of Gynaecology and Obstetric; HE4 – human epididymis protein 4; LDH - lactate dehydrogenase; MAP – metastases in abdomen and pelvis; NLR - neutrophil-to-lymphocyte ratio.

extracting radiomics features automatically with deep-learned representations using CNNs [20]. This was the only study that has utilised deep learning in the prediction of CC.

Table 2
Comparison of the model predictive performance between the studies.

Study	Outcomes	Comparison groups	Training	Internal validation	External validation	AUC (validation group)	Accuracy	Sens	Spec	Best Performing Group
Maubert et al., 2019	Prediction of "open and close" procedures for peritoneal carcinomatosis	CT; DT; RF; SVM	218 patients 70% training	92 patients 30% Testing	None	N/R	CT 90.22% DT 85.86% SVM 97.1% RF 97.82%	N/R	N/R	RF
Laios et al., 2020	Predict complete cytoreduction in AOC	k-NN (k = 15–20); Logistic regression	96 patients	51 patients	None	N/R	k-NN 65.8% LR 63.4%	N/R	N/R	k-NN
Li et al., 2021	Prediction of residual disease in advanced HGSOC	Clinical; Radiomics; Combined nomogram	160 patients	57 patients	None	0.623 (clinical) 0.744 (radiomics) 0.803 (combined)	73.6%	90.3%	53.8%	Combined nomogram
Piedimonte et al., 2022	Prediction of optimal cytoreduction (<1 cm) in advanced ovarian cancer	Prediction of optimal cytoreduction (<1 cm) in advanced ovarian cancer	92 patients	30 patients validation 29 patients test	None	0.896 (validation) 0.89 (test)	N/R	N/R	N/R	Good performance in optimal cytoreduction (<1 cm)
	Prediction of no gross residual (RD = 0)	Prediction of no gross residual (RD = 0)			None	0.52 (validation) 0.84 (test)	N/R	N/R	N/R	Poor performance in predicting RD = 0
Guo et al., 2023	Prediction of miliary small bowel disease in HGSOC	Clinical; Radiomics; Combined nomogram	91 patients	37 patients	None	0.705 (clinical) 0.842 (radiomics) 0.858 (combined)	N/R	80.0%	77.3%	Combined nomogram
Lu et al., 2023	Prediction of gross residual disease in HGSOC (R0 vs non-R0)	Clinical; Radiomics; Combined nomogram	106 patients	5-fold cross validation 22 patient internal validation cohort	None	0.708 (clinical) 0.850 (radiomics) 0.900 (combined)	81.8%	80.0%	83.3%	Combined nomogram
Lin et al., 2025 (DeAF)	Prediction of CC0-1 and CC2-3 for synchronous colorectal peritoneal metastases	Clinical; DeAF (deep learning fusion)	SYSU 84 patients	SYSU 30 patients	FPTH 33 patients FUSCC 13 patients SMU 26 patients	0.305 – 0.805 (clinical) 0.906 – 0.960 (DeAF)	86.7%	80.0%	90.0%	DeAF
Li et al., 2025	AOC, residual tumor ≥1 cm	Clinical; Ultrasound radiomics; Combined	78 patients included in the development model using 10-fold cross validation		34 patients	0.723 (clinical) 0.704 (radiomics) 0.817 (combined)	81.7%	81.7%	81.7%	Combined nomogram
Liu et al., 2025	Prediction of suboptimal debulking surgery for SOC	Clinical; Radiomics; Combined	158 patients	70 patients	45 patients	0.826 (clinical) 0.747 (radiomics) 0.854 (combined)	75.6%	76.9%	73.7%	Combined nomogram

Abbreviation: HGSOC – high grade serous ovarian cancer; AOC- advanced epithelial ovarian cancer; SOC – serous ovarian cancer; RD = 0 – no residual disease; CC0-1 – complete cytoreduction; CC2-3 – incomplete cytoreduction; LR – logistic regression; CT - conventional trees; DT - decision trees; RF - random forests; SVM - support vector machine; k-NN - k-nearest neighbours; N/R - not reported; SYSU - Sun Yat-sen University; FPTH - Fujian Province tumour hospital; FUSCC - Fudan University Shanghai Cancer Center; SMU - Southern Medical University.

3.1.2. Disease type and input predictors

Across the reviewed literature, the predominant disease type assessed was advanced EOC, particularly the high-grade serous subtype. Laios et al. (2020), Li et al. (2021), Guo et al. (2023), Piedimonte et al. (2022), Li et al. (2025), and Liu et al. (2025) all focused on ovarian cancer where multiple clinical variables, radiological assessments and radiomics features were utilised to predict the likelihood of complete cytoreduction [13,14,16,17,19]. A smaller subset of studies included other types of peritoneal metastases. Maubert et al. (2019) included colorectal, gastric, and appendiceal primaries, though notably without radiomics integration, relying instead on operative and imaging scores such as the rPCI [18]. Lin et al. (2025) applied deep learning specifically to the prediction of cytoreduction for synchronous CPM [21].

Input variables spanned a broad spectrum and varied by disease type but are broadly classified into radiological, radiomics and clinical factors**.** Clinical features included age, body mass index (BMI), comorbidity indices, ASA and ECOG performance status, and tumour markers such as CEA, CA-125, CA 19-9, and HE-4, alongside other biomarkers including NLR, LDH, and albumin. Radiological variables included ascites, rPCI, MRI sequence data, surgical complexity scores, and mapping of disease sites. Radiomics features input were central to Li (2021), Lu (2023), Guo (2023), Liu (2025), and Li (2025) in their prediction model, like the deep learning approach of Lin et al. (2025), they applied CNN to CT images to automatically extract high-dimensional imaging features without handcrafted radiomics.

3.1.3. Model performance and predictive accuracy

Across the reviewed studies, seven reported AUCs from their validation groups as shown in Table 2). Internal validation was carried out in all studies. External validation was uncommon, with only three studies testing it and the true multicentre validation carried out by Lin et al. (2025) [16,17,21].

In the validation groups, radiomics–clinical nomograms consistently outperformed single-modality models. Li et al. (2021), Lu et al. (2023), and Guo et al. (2023) each demonstrated that integrating radiomics with clinical variables improved discrimination, typically achieving AUCs between 0.80 and 0.90 [13–15]. Radiomics-only models outperformed clinical-only models in Li et al. (2021), Guo et al. (2023), and Lu et al. (2023) [13–15], although the reverse was observed in Li et al. (2025) and Liu et al. (2025) [15,16], where clinical predictors retained stronger discriminatory power. Among radiomics-based studies, Liu et al. (2025) reported the strongest regression-based results, with a combined MRI radiomics–clinical nomogram achieving an AUC of 0.90 [16]. Li et al. (2025) found that LightGBM, applied to ultrasound radiomics and clinical features, outperformed regression-based approaches, with an external validation AUC of 0.82 [17].

From a non-radiomics perspective, classical machine learning models showed mixed performance. Maubert et al. (2019) reported very high apparent accuracy (97.8%) using random forests compared with decision trees and support vector machines, although the absence of AUC reporting, reliance on intraoperative inputs, and small dataset raise concerns of overfitting [18]. Laios et al. (2020) achieved only modest accuracy (65.8%) with k-Nearest Neighbours in predicting complete cytoreduction in advanced ovarian cancer [20]. Piedimonte et al. (2022) reported strong results when using residual disease <1 cm as the endpoint (AUC 0.89), but performance fell dramatically when no residual disease (RD $= 0$) was applied as the outcome, dropping to an AUC of 0.52 [19].

Deep learning delivered robust and generalisable performance. Lin et al. (2025) developed the DeAF framework using a ResNet3D CNN with self-attention, integrated with clinical features and achieved a consistent multicentre validation with AUCs of 0.90 across three independent cohorts, the strongest externally validated result in this review [21]. The DeAF framework also significantly outperformed the standalone clinical model, mirroring trends seen in radiomics-based studies. In addition, it achieved high overall accuracy (86.7%), sensitivity (80.0%), and specificity (90.0%) [21], underscoring its potential scalability.

3.1.4. Machine learning strengths, weaknesses and transparency assessment

The LASSO logistic regression machine learning offers explainability through feature coefficients, calibration curves or nomograms, which clinicians could interpret as relative contributions of imaging and clinical factors [22,24]. Although linear models offer explainability and reduce overfitting through regularization, its reliance on linear relationships and instability with correlated features limit its scalability for complex, multimodal prediction tasks [22–24]. Other simpler machine learning models such as single decision tree and instance-based models such as k-NN were inherently interpretable as they resemble clinical decision rules, but they are prone to instability and overfitting, especially in small datasets [23]. A key limitation of k-NN is the need to determine the optimal number of neighbours (k), which critically influences model bias–variance balance and overall predictive performance [23].

Tree-based ensemble algorithms such as RF and GBT demonstrate a major advantage over linear models in their ability to capture complex non-linear relationships and high-order interactions between radiological, clinicopathological and biomarker variables. These models can learn subtle patterns in disease distribution, tumour biology and host factors that influence the likelihood of achieving complete cytoreduction in PSM. However, tree-based models are inherently difficult to explain because their predictions are derived from the combined output of numerous individual decision trees, each built on different subsets of data [23]. RF and GBT both improve model stability and predictive accuracy—the former by aggregating multiple trees and the latter by sequentially refining residual errors—but at the expense of model transparency and interpretability [23,24]. Margin-based models such as SVM can handle high-dimensional data effectively and perform well in various settings except when the data set contains more noise [23]. SVM decision boundaries are less intuitive as they depend on complex kernel transformations and weighted combinations of support vectors, offering little transparency about how individual features drive the prediction [24]. The deep learning CNN model used by Lin et al. (2025) incorporated a self-attention mechanism within the DeAF framework, which allowed the model to weight features adaptively and improved interpretability at a structural level [21]. Deep learning utilising CNN enables accurate predictive power but, similarly, understanding its reason for specific predictions can be complex or even impossible [25]. Consequently, RF, GBT, SVM, and deep learning architectures (e.g., CNNs) operate as "black-box" classifiers, offering limited interpretability and minimal transparency regarding how input features contribute to model predictions [24,25].

This barrier highlights the need to utilise accurate explainability tools that enable clinicians to understand, for instance, why a patient was predicted by AI to have incomplete cytoreduction and how individual input variables (clinical, radiological, biomarker-based or radiomic) contributed to that prediction. Explainable artificial intelligence (XAI) was developed to address the challenges of model transparency and interpretability, and may be achieved either through post-hoc explanation techniques applied to complex "black-box" models or through intrinsically self-explainable model architectures such as logistic regression or small decision trees in which predictions are transparent by design [26]. In this review, none of the included studies implemented modern XAI techniques such as Shapley additive explanations (SHAP), local interpretable model-agnostic explanations (LIME), or gradient-weighted class activation mapping (Grad-CAM) to provide patient-level insights [24]. Although several models demonstrated partial interpretability, true explainability was absent across all studies, representing a key barrier to clinical translation and adoption.

3.1.5. Methodological gaps and research priorities to guide future development

This scoping review identifies several methodological limitations which constrain the clinical translation of AI models in predicting complete cytoreduction for peritoneal malignancy. While there are more studies examining CC [27–29], they exclusively analysed intra-operative and post-operative factors and thus were deliberately excluded as they do not address the central clinical challenge: the need for accurate pre-operative decision-making. This scoping review was also limited by the small number of eligible studies and potential exclusion of non-English or unpublished data; however, it provides a comprehensive synthesis of the current pre-operative AI literature.

Radiomics formed the foundation of many AI-based predictive models. These approaches depend critically on accurate manual tumour segmentation and robust feature extraction, as the quality of these steps directly determines model performance [8]. PSM can be inherently heterogeneous even within the same histological subtype such as ovarian cancer [30]. Radiomics can capture this complexity by quantifying tumour shape, texture, first-order intensity, and higher-order features such as wavelet transforms—patterns invisible to the human eye [8]. This review has identified several radiomic-based models and a deep learning model with promising predictive potential. This underlines the influence of radiomic features in predicting CC in PSM; however, as most available studies focus predominantly on ovarian cancer, it remains elusive whether these findings can be generalised to other malignancies, underscoring the need for further research in colorectal, gastric, appendiceal and other peritoneal primaries.

Radiomics feature extraction is performed either in a handcrafted fashion—most commonly using PyRadiomics—or in an automated fashion through deep learning [21,29]. Handcrafted radiomics rely on predefined mathematical formulas, making features more standardised and interpretable [31] but potentially missing subtle or non-linear patterns. By contrast, deep CNN-based automated extraction can capture more complex representations but at the cost of reduced interpretability and explainability [32]. Prior to feature extraction, all studies relied on manual segmentation, which remains the current gold standard for accuracy but is time-consuming, prone to inter-observer variability, and difficult to scale in routine practice [33]. These challenges are amplified in PSM, where ascites, irregular deposits, and diffuse disease can make segmentation particularly difficult and inaccurate. According to Ma et al., future work is likely to involve more sophisticated prompt-based auto-segmentation tools (rather than fully automated) [33] to help delineate tumour boundaries efficiently while reducing variability and expediting workflow.

Another consistent theme is that models combining radiomics with clinicopathological variables consistently outperform single-modality approaches. This indicates that future research should focus on multimodal integration, building models that combine radiomics and established clinicopathological factors. Beyond this, the next step will be to extend such integration to other variables such as genomic and molecular data and potentially other sources of data such as intra-operative photos from diagnostic laparoscopy or digitised pathology slides. These multimodal fusion models have the potential to capture even more morphological and biological determinants of CC.

The variation in machine learning algorithms across studies represents a key source of heterogeneity, with no consensus on the optimal strategy. Model choice is often dictated by sample size: most current studies were single-centre with relatively small cohorts, which favoured the use of simpler approaches such as linear regression-based radiomics nomograms [23]. These remain the most accessible and interpretable strategies in limited datasets. By contrast, more advanced algorithms such as RF, SVM, GBT and deep learning frameworks generally require larger, more diverse datasets to avoid overfitting and to fully utilise their ability to capture non-linear relationships and complex feature interactions [23,32]. This underscores the importance of multicentre collaboration in the context of rare diseases, which will be essential to increase sample size, improve data heterogeneity and provide the statistical power needed to validate more sophisticated models. Future research should therefore aim to systematically benchmark different classifiers within large datasets to identify the most interpretable and clinically feasible strategies. Another gap lies in the limited use of external validation. Most studies relied on internal validation cohorts drawn from the same centre, which risks model overfitting and undermines generalisability. Only a small number incorporated external validation, with true multicentre testing achieved only by Lin et al. (2025). The lack of external validation is a major barrier to clinical implementation. To establish confidence in AI tools, future work should prioritise external validation datasets to ensure generalisability and reproducibility across institutions [34].

Explainability of AI represents a critical gap and was often limited to nomograms, feature importance rankings or calibration analyses. While these provide some transparency, they fall short of what is needed for clinical confidence, especially in deep learning models where decision-making processes are otherwise opaque. Without transparent explanation frameworks, clinicians may remain reluctant to integrate AI predictions into pre-operative planning. Overcoming this barrier will require deliberate integration of XAI approaches based on the type of machine learning utilised to clearly demonstrate why a model predicts incomplete or complete cytoreduction for each patient.

Finally, there remains a skew in disease types investigated. Most studies have focused on ovarian cancer, with comparatively little attention to colorectal, gastric, appendiceal, and primary peritoneal malignancies, despite their substantial clinical burden. To ensure broader clinical applicability, future research should expand into these underrepresented disease groups. Predictive models based on AI hold substantial promise and are likely to become integrated into routine clinical workflows soon. Incorporating these predictive tools into multidisciplinary decision-making could reduce non-therapeutic laparotomies and improve patient selection for CRS and HIPEC.

4. Conclusion

AI shows strong promise for predicting completeness of cytoreduction in peritoneal surface malignancies using pre-operative clinical and imaging features. Radiomics–clinical nomograms and deep learning frameworks both achieved encouraging accuracy, with multimodal integration consistently outperforming single-modality approaches. To enable clinical translation, future work must prioritise multicentre external validation, adoption of explainable AI methods, and expansion beyond ovarian cancer to other peritoneal malignancies.

CRediT authorship contribution statement

Samuel Pau: conceptualization, data curation, formal analysis, methodology, writing of the original draft. Tim Eglinton: supervision, writing – review and editing. Alan Wang: supervision, writing – review and editing. Jesse Fischer: conceptualization, data curation, methodology, supervision, writing – review and editing.

Use and declaration of AI and AI-assisted technologies

The authors acknowledge the use of Rayyan, an AI-assisted tool, during the study selection and screening process for the systematic review. No generative AI technologies were used in the writing or editing of the manuscript.

Funding

None.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

None.

References

[1] Qin RX, Lim JH, Ly J, Fischer J, Smith N, Karalus M, et al. Long-term survival following cytoreductive surgery and hyperthermic intraperitoneal chemotherapy in Waikato, aotearoa New Zealand: a 12-year experience. ANZ J Surg 2024;94(4): 621–7. https://doi.org/10.1111/ans.18777. PMID: 37994292.

[2] Yan TD, Deraco M, Baratti D, Kusamura S, Elias D, Glehen O, et al. Cytoreductive surgery and hyperthermic intraperitoneal chemotherapy for malignant peritoneal mesothelioma: multi-institutional experience. J Clin Oncol 2009;27(36):6237–42. https://doi.org/10.1200/JCO.2009.23.9640. PMID: 19917862.

[3] du Bois A, Reuss A, Pujade-Lauraine E, Harter P, Ray-Coquard I, Pfisterer J. Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: a combined exploratory analysis of 3 prospectively randomized phase 3 multicenter trials: by the Arbeitsgemeinschaft Gynaekologische Onkologie Studiengruppe Ovarialkarzinom (AGO-OVAR) and the Groupe d'Investigateurs Nationaux Pour Les Etudes des Cancers de l'Ovaire (GINECO). Cancer 2009;115(6):1234–44. https://doi.org/10.1002/cncr.24149.

[4] Coccolini F, Catena F, Glehen O, Yonemura Y, Sugarbaker PH, Piso P, et al. Complete versus incomplete cytoreduction in peritoneal carcinosis from gastric cancer, with consideration to PCI cut-off: systematic review and meta-analysis. Eur J Surg Oncol 2015;41(7):911–9. https://doi.org/10.1016/j.ejso.2015.03.231. PMID: 25936764.

[5] Flicek K, Ashfaq A, Johnson CD, Menias C, Bagaria S, Wasif N. Correlation of radiologic with surgical peritoneal cancer index scores in patients with pseudomyxoma peritonei and peritoneal carcinomatosis: how well can we predict resectability? J Gastrointest Surg 2016;20(2):307–12. https://doi.org/10.1007/s11605-015-2880-6. PMID: 26162922.

[6] Wei GX, Zhou YW, Li ZP, Qiu M. Application of artificial intelligence in the diagnosis, treatment, and recurrence prediction of peritoneal carcinomatosis. Heliyon 2024;10(7):e29249. https://doi.org/10.1016/j.heliyon.2024.e29249. PMID: 38601686; PMCID: PMC11004411.

[7] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44. https://doi.org/10.1038/nature14539. PMID:26017442.

[8] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology 2016;278(2):563–77. https://doi.org/10.1148/radiol.2015151169.

[9] Li M, Jiang Y, Zhang Y, Zhu H. Medical image analysis using deep learning algorithms. Front Public Health 2023;11:1273253. https://doi.org/10.3389/fpubh.2023.1273253.

[10] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467–73. https://doi.org/10.7326/M18-0850. PMID: 30178033.

[11] Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan — a web and Mobile app for systematic reviews. Syst Rev 2016;5:210. https://doi.org/10.1186/s13643-016-0384-4.

[12] Attenberger UI, Langs G. How does radiomics actually work?—Review. Röfo 2021; 193(6):652–7. https://doi.org/10.1055/a-1293-8953. PMID: 33264805.

[13] Li H, Zhang R, Li R, Xia W, Chen X, Zhang J, et al. Noninvasive prediction of residual disease for advanced high-grade serous ovarian carcinoma by MRI-based radiomic-clinical nomogram. Eur Radiol 2021;31(12):7855–64. https://doi.org/10.1007/s00330-021-07902-0.

[14] Guo Q, Lin Z, Lu J, Li R, Wu L, Deng L, et al. Preoperative prediction of miliary changes in the small bowel mesentery in advanced high-grade serous ovarian cancer using MRI radiomics nomogram. Abdom Radiol 2023;48(4):1119–30. https://doi.org/10.1007/s00261-023-03802-7.

[15] Lu J, Cai S, Wang F, Wu PY, Pan X, Qiang J, et al. Development of a prediction model for gross residual in high-grade serous ovarian cancer by combining preoperative assessments of abdominal and pelvic metastases and multiparametric MRI. Acad Radiol 2023;30(9):1823–31. https://doi.org/10.1016/j.acra.2022.12.009.

[16] Liu L, Zhang W, Wang Y, Wu J, Fan Q, Chen W, et al. Radiomics combined with clinical and MRI features may provide preoperative evaluation of suboptimal debulking surgery for serous ovarian carcinoma. Abdom Radiol 2025;50:496–512. https://doi.org/10.1007/s00261-024-04343-3.

[17] Li S, Ding Q, Li L, Liu Y, Zou H, Wang Y, et al. Ultrasonic radiomics-based nomogram for preoperative prediction of residual tumor in advanced epithelial ovarian cancer: a multicenter retrospective study. Front Oncol 2025;15:1540734. https://doi.org/10.3389/fonc.2025.1540734.

[18] Maubert A, Birtwisle L, Bernard JL, Benizri E, Bereder JM. Can machine learning predict resectability of a peritoneal carcinomatosis? Surg Oncol 2019;29:120–5. https://doi.org/10.1016/j.suronc.2019.04.008. PMID: 31196475.

[19] Piedimonte S, Erdman L, So D, Bernardini MQ, Ferguson SE, Laframboise S, et al. Using a machine learning algorithm to predict outcome of primary cytoreductive surgery in advanced ovarian cancer. J Surg Oncol 2023;127(3):465–72. https://doi.org/10.1002/jso.27137.

[20] Laios A, Gryparis A, DeJong D, Hutson R, Theophilou G, Leach C. Predicting complete cytoreduction for advanced ovarian cancer patients using nearest-neighbor models. J Ovarian Res 2020;13(1):117. https://doi.org/10.1186/s13048-020-00700-0. PMID: 32993745; PMCID: PMC7526140.

[21] Lin Q, Chen C, Li K, Cao W, Wang R, Fichera A, et al. A deep-learning model to predict the completeness of cytoreductive surgery in colorectal cancer with peritoneal metastasis. Eur J Surg Oncol 2025;51(7):109760. https://doi.org/10.1016/j.ejso.2025.109760.

[22] Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. Nat Methods 2018;15(4):233–4. https://doi.org/10.1038/nmeth.4642. PMID: 30100822; PMCID: PMC6082636.

[23] Sarker IH. Machine learning: algorithms, real-world applications and research directions. SN Comput Sci 2021;2(3):160. https://doi.org/10.1007/s42979-021-00592-x. PMID: 33778771; PMCID: PMC7983091.

[24] Saranya A, Subhashini R. A systematic review of explainable artificial intelligence models and applications: recent developments and future trends. Decis Anal J 2023;7:100230. https://doi.org/10.1016/j.dajour.2023.100230.

[25] Barbierato E, Gatti A. The challenges of machine learning: a critical review. Electronics (Basel) 2024;13(2):416. https://doi.org/10.3390/electronics13020416.

[26] Gunning D, Vorm E, Wang JY, Turek M. DARPA's explainable AI (XAI) program: a retrospective. Appl AI Lett 2021;2(4):e61. https://doi.org/10.1002/ail2.61.

[27] Laios A, Kalampokis E, Johnson R, Munot S, Thangavelu A, Hutson R, et al. Development of a novel intra-operative score to record diseases' anatomic fingerprints (ANAFI score) for the prediction of complete cytoreduction in advanced-stage ovarian cancer by using machine learning and explainable artificial intelligence. Cancers (Basel) 2023 Jul 25;15(15):3886. https://doi.org/10.3390/cancers15153886.

[28] Laios A, Kalampokis E, Mamalis ME, Tarabanis C, Nugent D, Thangavelu A, et al. RoBERTa-assisted outcome prediction in ovarian cancer cytoreductive surgery using operative notes. Cancer Control 2023;30:10732748231209892. https://doi.org/10.1177/10732748231209892. PMID: 37915208; PMCID: PMC10624075.

[29] Laios A, Kalampokis E, Johnson R, Munot S, Thangavelu A, Hutson R, et al. Predicting surgical effort using explainable artificial intelligence in advanced-stage epithelial ovarian cancer. Cancers (Basel) 2022 Jul 1;14(14):3447. https://doi.org/10.3390/cancers14143447.

[30] Wang Y, Zhu N, Liu J, Chen F, Song Y, Ma Y, et al. Role of tumor microenvironment in ovarian cancer metastasis and clinical advancements. J Transl Med 2025;23(1): 539. https://doi.org/10.1186/s12967-025-06508-0. PMID: 40369674; PMCID: PMC12079989.

[31] van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. Cancer Res 2017;77(21):e104–7. https://doi.org/10.1158/0008-5472.CAN-17-0339. PMID: 29092951; PMCID: PMC5672828.

[32] Mienye ID, Swart TG, Obaido G, Jordan M, Ilono P. Deep convolutional neural networks in medical image analysis: a review. Information 2025;16(3):195. https://doi.org/10.3390/info16030195.

[33] Ma J, He Y, Li F, Han L, You C, Wang B. Segment anything in medical images. Nat Commun 2024;15(1):654. https://doi.org/10.1038/s41467-024-44824-z. PMID: 38253604; PMCID: PMC10803759.

[34] Steyerberg EW, Harrell Jr FE. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 2016;69:245–7. https://doi.org/10.1016/j.jclinepi.2015.04.005. PMID: 25981519; PMCID: PMC5578404.