Predicting Response to Neoadjuvant Chemotherapy in Ovarian Cancer from CT Baseline Using Multi-Loss Deep Learning

Francesco Pastori*[1], Francesca Fati*[1,2], Marina Rosanu[1], Luigi De Vitis[3], Lucia Ribero[1], Gabriella Schivardi[1], Giovanni Damiano Aletti[1,4], Nicoletta Colombo[1], Jvan Casarin[5], Francesco Multinu[1,4], and Elena De Momi[1,2]

[1] Department of Gynecologic Oncology, European Institute of Oncology, IEO, IRCCS, Milan, Italy
[2] Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy
[3] Department of Obstetrics and Gynecology, Mayo Clinic, Rochester, USA
[4] Department of Oncology and Hemato-Oncology, University of Milan, Milan, Italy
[5] Department of Medicine and Innovative Technology, Università degli Studi dell'Insubria, Varese, Italy
Corresponding author: francesco.pastori@mail.polimi.it

Abstract. Ovarian cancer is the most lethal gynecologic malignancy: around 60% of patients are diagnosed at an advanced stage, with an associated 5-year survival rate of about 30%. Early identification of non-responders to neoadjuvant chemotherapy remains a key unmet need, as it could prevent ineffective therapy and avoid delays in optimal surgical management.

This work proposes a non-invasive deep learning framework to predict neoadjuvant chemotherapy response from pre-treatment contrast-enhanced CT by leveraging automatically derived 3D lesion masks. The approach encodes axial slices with a partially fine-tuned pretrained image encoder and aggregates slice-level representations into a volumetric embedding through an attention-based module. Training combines classification loss with supervised contrastive regularization and hard-negative mining to improve separation between ambiguous responders and non-responders. The method was developed on a retrospective single-center cohort from the European Institute of Oncology (Milan, IT), including 280 eligible patients (147 responder, 133 non-responder). On the test cohort, the model achieved a ROC-AUC of 0.73 (95% CI: 0.58–0.86) and an F1-score of 0.70 (95% CI: 0.56–0.82). Overall, these results suggest that the proposed architecture learns clinically relevant predictive patterns and provides a robust foundation for an imaging-based stratification tool.

Keywords: ovarian cancer; neoadjuvant chemotherapy response prediction; contrast-enhanced CT; Vision Transformers; supervised contrastive learning

1 Introduction

Ovarian cancer is the most lethal gynecologic malignancy, 60% of patients are diagnosed when the tumor is in an advanced stage and they present a 5-year survival rate of 30% [1]. The epithelial ovarian cancer accounts for about 90% of all cases and its most prevalent and aggressive form is the high-grade serous ovarian carcinoma, which constitutes the majority of advanced-stage diagnoses and is the primary focus of this work [8]. Standard management typically involves primary debulking surgery aiming for macroscopic complete resection of the disseminated disease, followed by platinum-based adjuvant chemotherapy. When complete resection is unlikely due to extensive disease burden, or when patients cannot undergo upfront surgery due to clinical condition, an alternative pathway is neoadjuvant chemotherapy which is commonly administered as three cycles of carboplatin and paclitaxel, followed by interval debulking surgery and additional postoperative chemotherapy. [9]

A key unmet need is that, even with careful preoperative evaluation, approximately 40% of patients assigned to neoadjuvant chemotherapy fail to achieve any objective benefit from carboplatin paclitaxel therapy and would likely benefit from alternative strategies if identified earlier as non-responders [3]. Given the aggressive nature of ovarian cancer, time is a critical factor: predicting, before treatment initiation, whether a tumor will respond to neoadjuvant chemotherapy could prevent ineffective therapy and avoid delays in definitive surgical management. Addressing this challenge requires a predictive tool that is rapid, non-invasive, and widely available in routine clinical practice.

Building on this clinical need, this work proposes a deep learning based predictive framework that leverages pre-treatment contrast-enhanced CT scans to classify patients as responders or non-responders to neoadjuvant chemotherapy, with the aim of supporting treatment planning and enabling more personalized and effective management of ovarian cancer.

2 State of the Art

Recent advances in artificial intelligence for medical imaging, especially deep learning, have demonstrated the ability to extract complex imaging patterns that may be imperceptible to visual inspection, motivating radiomics and deep learning approaches for treatment response prediction from pre-treatment scans.

Radiomics extracts handcrafted texture and shape descriptors from segmented tumor regions of interest and applies feature selection to reduce redundancy; results can depend on segmentation quality and imaging variability. Crispin-Ortuzar et al.[3] developed an integrated radiogenomic model combining clinical and CT-derived radiomic features to predict neoadjuvant chemotherapy (NACT) response in ovarian cancer. Using an ensemble of Elastic Net, SVM, and Random Forest trained on 72 patients, the model achieved an AUC of 0.78, substantially outperforming the clinical-only baseline (AUC 0.47), highlighting the added value of radiomics.

Deep learning enables end-to-end representation learning directly from images. A key benchmark for this work is the multicenter study by Yin et al.[11], which predicted NACT response from pre-treatment CT combined with clinical features, using RECIST 1.1 as reference standard for response evaluation and including 757 patients. They developed a 2D slice-based convolutional ResNet architecture achieving an external AUC of 0.82 for response prediction.

In contrast to convolutional architectures operating on single 2D slices, emerging state-of-the-art approaches employ Vision Transformers (ViT) to model long-range dependencies, analyze complete 3D volumes, and integrate contrastive learning to strengthen representation learning. Fati et al.[5] introduced OVIT, a ViT-based model that encodes CT slices and aggregates them with attention to predict resectability. OVIT was trained on 465 patients, achieving a ROC-AUC of 0.80 on the test set and suggesting that ViT-based models can serve as clinically relevant decision support systems.

Gupta et al.[6] trained a ResNet-50 with a margin-aware contrastive loss for histopathology image classification across 5 public datasets (more than 100k images), reporting an AUC of 0.96 on binary breast histopathology classification, concluding that adding a margin constraint improves representation quality and generalization.

Building on these studies, this work jointly exploits full volumetric lesion information, Vision Transformers, and supervised contrastive learning to provide a practical foundation for imaging-based decision support, enabling early stratification of NACT responders and non-responders and potentially reducing ineffective therapy while supporting more timely, personalized treatment planning.

3 Methods

3.1 Model Architecture

An overview of the proposed pipeline is reported in Figure 1. The model predicts response to NACT from 3D pre-treatment CT scans $\mathbf{V_{CT}} \in \mathbb{R}^{H\times W\times S\times 1}$, where H, W and S denote height, width and number of axial slices, respectively. The network outputs a binary prediction $\hat{y} \in \{0, 1\}$ where 0 denotes a non-responder and 1 a responder, via a model M:

$$\hat{y} = M(\mathbf{V_{CT}}). \tag{1}$$

From the raw $\mathbf{V}_{CT}$, a lesion mask volume is obtained through an open source deep learning segmentation model M_S:

$$\mathbf{V}_{seg} = M_S(\mathbf{V}_{CT}), \qquad \mathbf{V}_{seg} \in \mathbb{R}^{H\times W\times S\times 1} \tag{2}$$

To leverage a 2D pretrained ViT encoder, the input volume is reshaped into a sequence of axial slices $\mathbf{V_{seg,i}} \in \mathbb{R}^{H\times W\times 1}$. Each slice is converted to a 3-channel representation and independently processed by a pretrained ViT encoder E_{ViT}.

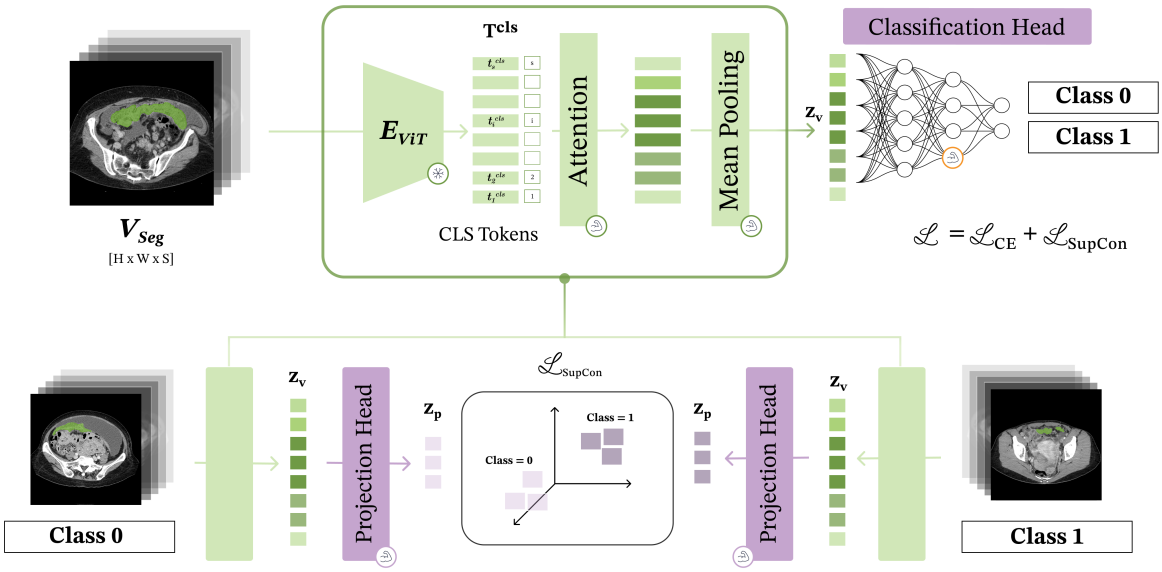

Fig. 1: Overview of the proposed model architecture for NACT response prediction. Pre-treatment 3D CT volumes are processed by a segmentation model to obtain lesion mask volumes, which are reshaped into axial slices. Each slice is encoded independently by a pretrained encoder. Slice-level representations are aggregated by an attention module and average pooling to produce a single volume embedding. This embedding is shared by two MLP heads: a classification head to predict non-responder(0) vs. responder(1), and a projection head to obtain embeddings optimized with supervised contrastive learning. Training uses a multi-loss objective combining supervised contrastive and cross-entropy loss.

For each slice i, the encoder produces token embeddings:

$$\mathbf{T}_i = E_{\text{ViT}}(\mathbf{V}_{\mathbf{seg,i}}), \qquad \mathbf{T}_i \in \mathbb{R}^{(N+1)\times d} \tag{3}$$

where N is the number of patch tokens and d is the embedding dimension.

The corresponding classification token $\mathbf{t}_i^{\text{cls}} \in \mathbb{R}^d$ is extracted as a global descriptor of the slice. The classification tokens from all S slices are concatenated to form a compact representation of the 3D input in the ViT feature space:

$$\mathbf{T}^{\text{cls}} = [\mathbf{t}_1^{\text{cls}}; \ldots; \mathbf{t}_S^{\text{cls}}], \qquad \mathbf{T}^{\text{cls}} \in \mathbb{R}^{S\times d} \tag{4}$$

A learnable positional embedding along the slice axis is added to $\mathbf{T}^{\text{cls}}$ to preserve ordering information. The resulting sequence is aggregated by an attention pooling module that models inter-slice relationships. The attention output is aggregated via mean pooling along the slice dimension, yielding a single embedding vector per input volume. The combination of slice-level ViT feature extraction,

attention pooling, and mean aggregation is denoted as the volume-encoding module M_V:

$$\mathbf{z_V} = M_V(\mathbf{V_{seg}}), \qquad \mathbf{z_V} \in \mathbb{R}^d. \tag{5}$$

From the shared volume embedding $\mathbf{z_V}$, the network branches into two heads: the classification head and the projection head. The classification head H_{cls} is implemented as an MLP composed of multiple linear layers interleaved with Layer Normalization, GELU non-linear activations, and dropout regularization. Starting from the pooled embedding dimension d, the head progressively reduces dimensionality through two hidden transformations ($d \to d/2 \to d/8$) and outputs the final 2-class logits $\hat{y} \in \mathbb{R}^2$ ($d/8 \to 2$)

$$\hat{y} = H_{\text{cls}}(\mathbf{z_V}), \qquad \hat{y} \in \mathbb{R}^2, \tag{6}$$

The logits are converted to class probabilities via softmax, the corresponding class label is then obtained by setting a decision threshold on the predicted probability.

In parallel, the projection head H_{proj} maps the same embedding $\mathbf{z_V}$ to a low-dimensional representation $\mathbf{z_p}$ used for supervised contrastive learning:

$$\mathbf{z_p} = H_{\text{proj}}(\mathbf{z_V}), \qquad \mathbf{z_p} \in \mathbb{R}^p.$$

In the proposed model, H_{proj} is a shallow MLP with intermediate Layer Normalization and GELU activation. The projection is subsequently L2-normalized to obtain an embedding $\tilde{\mathbf{z}}_\mathbf{p}$ on the unit hypersphere:

$$\tilde{\mathbf{z}}_\mathbf{p} = \frac{\mathbf{z_p}}{\|\mathbf{z_p}\|_2}, \qquad \tilde{\mathbf{z}}_\mathbf{p} \in \mathbb{R}^p.$$

3.2 Training Strategies

A transfer learning framework was adopted, in which a pre-trained ViT encoder was fine-tuned and two task-specific heads were trained from scratch: a classification head for binary NACT response prediction and a projection head for contrastive embedding learning. Training is performed with a multi-loss function that combines a cross-entropy classification loss and a supervised contrastive loss:

$$\mathcal{L} = \mathcal{L}_{\text{CE}}(\hat{y}, y) + \alpha \ \cdot \mathcal{L}_{\text{SupCon}}(\tilde{\mathbf{z}}_\mathbf{p}, y), \tag{7}$$

where $y \in \{0, 1\}$ is the ground-truth response label and α (from 0 to 0.3) controls the relative contribution of the contrastive objective.

For the supervised contrastive loss, we consider pairs of volumes $(\mathbf{V_{seg}}_1, \mathbf{V_{seg}}_2)$ with a supervised pair label $s \in \{0, 1\}$, where $s = 1$ denotes a positive pair (same ground truth labels) and $s = 0$ a negative pair (different ground truth labels). Let $\tilde{\mathbf{z}}_{\mathbf{p_1}}, \tilde{\mathbf{z}}_{\mathbf{p_2}} \in \mathbb{R}^p$ be the embeddings produced by the projection head and normalized, and let $D = \|\tilde{\mathbf{z}}_{\mathbf{p_1}} - \tilde{\mathbf{z}}_{\mathbf{p_2}}\|_2$ be the Euclidean distance between the two embeddings. The contrastive loss is:

$$\mathcal{L}_{\text{SupCon}}(D, s) = s \cdot D^2 + (1 - s) \cdot [\max(0, m - D)]^2,$$

where $m > 0$ is the margin. Intuitively, the loss reduces the distance between positive pairs (pulling them toward $D \approx 0$) and enforces that negative pairs are separated by at least a margin m, penalizing only negatives that are "too close" ($D < m$).

The multi-loss equation (7) is a weighted combination where α controls the contribute of the contrastive component. During training, α is increased linearly from 0 up to a maximum value $\alpha_{\max}$ (0.3), so as cross-entropy provides an initial strong signal to organize the representations, while the progressive increase of α allows the contrastive term to act as a regularizer of the embedding space. At each iteration, the classification component is optimized on individual volumes using cross-entropy, while the contrastive component is optimized on volume pairs constructed from the labels. Pairs for contrastive learning are generated using two complementary strategies. In the initial epochs, both positive and negative pairs are formed by randomly sampling examples from the training set, providing a stable and low-variance contrastive signal while the representation space is still being structured. In a later stage, hard mining is enabled to build negative pairs by selecting samples with different labels that lie close to each other in the embedding space (hard negatives). This focuses optimization on the most ambiguous and informative cases, yielding a more effective contrastive gradient than "easy" negative pairs that are already well separated. In practice, the model is progressively encouraged to distinguish tumors that appear highly similar but exhibit different responses to NACT, improving its ability to capture subtle imaging cues associated with NACT response.

4 Experiments

4.1 Dataset

The study population was provided by the European Institute of Oncology (IEO) and included patients with histologically confirmed ovarian cancer, treated with NACT from 2016 to 2023. All patients had available contrast-enhanced portal venous phase CT scans at diagnosis and after three NACT cycles. RECIST 1.1 [4] criteria is used to establish target label classifying a patient as responder if the post-treatment tumor burden decreases by at least 30% compared with baseline, otherwise non-responder. After applying quality and eligibility criteria, 280 patients (147 responder, 133 non-responder) were included. Of these, 226 were used for model development, while an independent test set of 54 patients was reserved for final evaluation. For each patient all cancer lesions were segmented using an open source deep learning model OvSeg[2].

4.2 Implementation Details

The pretrained ViT encoder selected for the experiments is DINOv3 Base Patch16 [10] thanks to its strong semantic prior.

Fine-tuning of the pretrained encoder was performed in a parameter-efficient manner using LoRA [7] adapters, restricting adaptation to the last Transformer

block while keeping the remaining backbone frozen. This strategy updates only a small subset of additional parameters, reducing computational cost and limiting the risk of overfitting, which is particularly important in the presence of small dataset. All experiments were trained on a single NVIDIA A100-PCIE GPU with 40 GB of memory. Optimization was performed using AdamW for up to 100 epochs with early stopping based on validation performance. The learning rate was set to 1×10^{-4} and decreased following a cosine-shaped schedule, to regularize the classifier we applied a dropout rate of 0.2, the contrastive term α grows linearly up to $\alpha_{\max}$.

4.3 Performance metrics

The response prediction performance was evaluated using receiver operating characteristic (ROC) analysis, reporting the area under the curve (ROC-AUC). The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) as the threshold varies:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \qquad \mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}.$$

$$\mathrm{ROC\text{-}AUC} = \int_0^1 \mathrm{TPR}(\mathrm{FPR})\, d\mathrm{FPR}.$$

In addition, we reported accuracy, precision, recall, and F1:

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}, \qquad \mathrm{Recall} = \mathrm{TPR},$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \qquad \mathrm{F1} = \frac{2\,\mathrm{Precision}\,\mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}.$$

The classification threshold was selected on the validation set by maximizing F1, and the same threshold was then used for test evaluation. To further characterize the model, we performed an attention-weight analysis, a test-set calibration assessment, and a confidence-aware error analysis. Statistical significance versus the baseline was assessed using DeLong's test for ROC-AUC, and a paired stratified bootstrap for accuracy, precision, recall, and F1 ($p < 0.05$).

5 Results

We conducted an ablation study to quantify the impact of the main architectural and training choices on model performance. Specifically, we compared four configurations: the proposed multiloss with LoRA fine tuning (ML-FT), a cross-entropy only variant obtained by removing the contrastive loss from the previous model (CE-FT), a multiloss without fine tuning in which the backbone is fully frozen (ML-FR), and the baseline model that corresponds to our re-implementation of Yin et al.[11] architecture. All models were evaluated on the same test set and with the same bootstrapping procedure. Quantitative performance measures are summarized in Table 1.

Table 1: Quantitative performance comparison between configurations: each table entry features the median value with 95% CIs. For each metric, the best value is in bold. Asterisks (*) indicate statistically significant differences versus the baseline ($p<0.05$).

Metric	Yin et al.	MultiLoss-Frozen	CrossEntropy-FineTuned	MultiLoss-FineTuned
Confusion Matrix Elements (point estimates, 95% CI)				
True Positives [↑]	20 (13–27)	**31 (24–38)**	15 (9–22)	21 (14–28)
False Positives [↓]	12 (6–18)	15 (9–22)	**3 (0–7)**	5 (1–10)
True Negatives [↑]	8 (3–14)	5 (1–9)	**17 (10–24)**	15 (9–22)
False Negatives [↓]	14 (8–21)	**3 (0–7)**	19 (12–26)	13 (7–19)
Classification Metrics (95% CI)				
ROC-AUC [↑]	0.47 (0.31–0.62)	0.59 (0.43–0.75)*	0.66 (0.51–0.81)*	**0.73 (0.58–0.86)***
Accuracy [↑]	0.52 (0.39–0.63)	**0.67 (0.54–0.80)***	0.59 (0.46–0.72)*	**0.67 (0.54–0.80)***
Precision [↑]	0.62 (0.45–0.78)	0.67 (0.53–0.81)	**0.83 (0.65–1.00)***	0.81 (0.64–0.95)*
Recall [↑]	0.59 (0.41–0.74)	**0.91 (0.81–1.00)***	0.44 (0.28–0.61)	0.62 (0.45–0.78)*
F1 Score [↑]	0.61 (0.45–0.72)	**0.78 (0.67–0.87)***	0.58 (0.40–0.72)	0.70 (0.56–0.82)*

From an interpretative perspective, introducing the contrastive loss improves the discriminative ability compared to CE alone (ML-FT: AUC 0.73 vs CE-FT: AUC 0.65), consistent with the hypothesis that explicitly supervising the geometry of the embedding space promotes a more robust separation between responders and non-responders. The fine tuning ablation shows a marked impact on AUC (ML-FT: AUC 0.73 vs ML-FR: AUC 0.58), suggesting that parameter-efficient fine tuning, even when limited to the last backbone block, contributes substantially to domain transfer toward mask lesion inputs. At the same time, the configuration without fine tuning ML-FR exhibits very high recall (0.91) and thus a higher F1 (0.77), but with lower precision (0.67) and, most importantly, a lower AUC: this profile is compatible with a decision behavior more oriented toward predicting the positive class, which can increase true positives but reduce the overall separation between the scores of the two classes. Finally, ML-FT achieved a higher ROC-AUC than the reproduced Yin et al. baseline when both were evaluated within the same pipeline and on the same test set (ML-FT:0.73 vs CE-FT: 0.47), indicating stronger discriminative performance of our approach under these experimental conditions. All ROC-AUC differences between the ML-FT model and the other models were statistically significant according to DeLong's test.

5.1 Confidence-aware error analysis

Confidence-aware error analysis consisted of manually inspecting uncertain correct cases (correct but low-confidence predictions) and confident wrong cases (incorrect but high-confidence predictions), revealing that many uncertain correct predictions lie close to the intrinsic RECIST 1.1 cutoff used to define a patient as responder or non-responder, indicating that some borderline decisions reflect ambiguity in the reference standard rather than purely model uncertainty.

Meanwhile, confident wrong cases cannot be explained solely by proximity to the RECIST 1.1 cutoff and were treated as potential failure modes: cases in which the model produces a clear but incorrect decision, warranting targeted review to identify possible systematic factors (e.g., acquisition noise).

5.2 Calibration Analysis

Calibration was assessed via a reliability diagram, which shows the calibration curve predominantly above the perfect calibration line: for mean predicted probabilities of 0.13 (n=9), 0.31 (n=18), and 0.49 (n=12), the empirical fraction of positives is 0.33, 0.50, and 0.75, respectively; similarly, at higher-confidence intervals the observed values are 0.74 vs 0.75 (n=8) and 0.87 vs 1.00 (n=7). Here, n denotes the number of test samples falling within each predicted probability interval. Overall, the systematic tendency of the observed positive fractions to exceed the corresponding mean predicted probabilities indicates that, on the test set, the model produces under-confident estimates for the positive class.

5.3 Analysis of attention

Attention weights quantify the relative contribution of each axial slice to the aggregated volumetric representation and, consequently, to the final prediction. In Figure 2, weights are visualized with a color map, where cooler tones denote lower contribution and warmer tones denote higher contribution. Figure 2a shows a correctly classified example, where higher weights are predominantly assigned to slices intersecting the lesion core, whereas slices near the superior and inferior boundaries receive lower weights. This pattern aligns with the intuition that central slices, typically containing a larger tumor cross-section, convey richer and more discriminative morphological information than peripheral slices. Overall, this qualitative observation suggests that the attention module learns a meaningful spatial prior, prioritizing the most informative regions along the axial extent. In the misclassified example (Figure 2b), attention appears more diffuse and less selectively concentrated on the lesion core, indicating that the central-slice emphasis observed in correct predictions may be reduced in harder cases.

6 Discussion and Conclusion

In this work, we developed a non-invasive deep learning framework to predict response to NACT from pre-treatment 3D lesion masks, with the goal of anticipating non-responder patients before therapy initiation. The proposed multi-loss model with LoRA adaptation achieved the best overall performance on the test set, reaching a ROC-AUC of 0.73 (95% CI [0.58, 0.85]) with an accuracy of 0.66 and an F1-score of 0.70. Ablation results support the central design choices of the approach: the supervised contrastive margin term improves global separability over a cross-entropy-only variant and the evaluated state-of-the-art baseline.

(a) Correctly classified (b) Misclassified

Fig. 2: Attention-weight visualization across axial slices for two cases. Colors encode slice contribution (cooler: lower, warmer: higher). (a) Correct classification with weights concentrated on the lesion core. (b) Misclassification with a more diffuse attention distribution.

Qualitative analysis of the slice-attention suggests that the model aggregates volumetric information coherently, prioritizing the central lesion region. Beyond discrimination, calibration analysis highlighted a systematic under-confidence for the positive class, this suggests that the model's probabilities are conservative rather than over-optimistic, which is desirable in a decision-support setting and can be further improved through post-hoc calibration. Finally, a confidence-aware error analysis showed that several uncertain-but-correct predictions concentrate near the intrinsic RECIST cutoff used to define responders versus non-responders, indicating that a meaningful fraction of ambiguity is driven by borderline ground truth rather than unstable modeling; in contrast, confident wrong cases emerge as actionable failure modes that can guide targeted review and future refinements. Overall, these findings strongly support the main objective of the work: using only pre-treatment lesion morphology, the proposed architecture learns clinically relevant predictive patterns and provides a robust foundation for an imaging-based stratification tool capable of anticipating NACT response, with clear potential to reduce ineffective therapy and avoid delays in optimal surgical management.

Conflict of Interest All authors have no conflicts of interest.

Acknowledgements The work is part of the project *Under-XAI: understanding ovarian cancer initiation and progression through explainable AI*, which received funding from the National Recovery and Resilience Plan (PNRR), Mission 6–Health, as part of the initiative *Strengthening and Enhancement of Biomedi-*

cal Research of the National Health Service, under the EU's NextGenerationEU program. Project code: PNRR-MAD-2022-12376574. CUP: J47G22000530001.

References

1. Cancer stat facts: ovarian cancer. available at: seer.cancer.gov/statfacts/html/ovary.html. [accessed october 2025]. Surveillance, Epidemiology, and End Results Program. (2025)
2. Buddenkotte, T., Rundo, L., Woitek, R., Escudero Sanchez, L., Beer, L., Crispin-Ortuzar, M., Etmann, C., Mukherjee, S., Bura, V., McCague, C., et al.: Deep learning-based segmentation of multisite disease in ovarian cancer. European radiology experimental **7**(1), 77 (2023)
3. Crispin-Ortuzar, M., Woitek, R., Reinius, M.A., Moore, E., Beer, L., Bura, V., Rundo, L., McCague, C., Ursprung, S., Escudero Sanchez, L., et al.: Integrated radiogenomics models predict response to neoadjuvant chemotherapy in high grade serous ovarian cancer. Nature communications **14**(1), 6756 (2023)
4. Eisenhauer, E.A., Therasse, P., Bogaerts, J., Schwartz, L.H., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., Mooney, M., et al.: New response evaluation criteria in solid tumours: revised recist guideline (version 1.1). European journal of cancer **45**(2), 228–247 (2009)
5. Fati, F., Rosanu, M., De Vitis, L., Rota, A., Traversa, A., Ribero, L., Schivardi, G., Petralia, G., Aletti, G.D., Colombo, N., et al.: Deep learning for decision support in ovarian cancer treatment planning (2025)
6. Gupta, E., Gupta, V.: Margin-aware optimized contrastive learning for enhanced self-supervised histopathological image classification. Health Information Science and Systems **13**(1), 2 (2024)
7. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. Iclr **1**(2), 3 (2022)
8. Lisio, M.A., Fu, L., Goyeneche, A., Gao, Z.h., Telleria, C.: High-grade serous ovarian cancer: basic sciences, clinical and therapeutic standpoints. International journal of molecular sciences **20**(4), 952 (2019)
9. Matulonis, U.A.: Management of newly diagnosed or recurrent ovarian cancer. Clinical Advances in Hematology and Oncology (2018)
10. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al.: Dinov3. arXiv preprint arXiv:2508.10104 (2025)
11. Yin, R., Guo, Y., Wang, Y., Zhang, Q., Dou, Z., Wang, Y., Qi, L., Chen, Y., Zhang, C., Li, H., et al.: Predicting neoadjuvant chemotherapy response and high-grade serous ovarian cancer from ct images in ovarian cancer with multitask deep learning: a multicenter study. Academic Radiology **30**, S192–S201 (2023)