Deep learning-based tumor resectability prediction model in patients with Ovarian Cancer: a preliminary evaluation

Francesca Fati[1,2,*,†], Marina Rosanu[1], Luigi De Vitis[1], Gabriella Schivardi[1],
Giovanni Damiano Aletti[1], Francesco Multinu[1], Roberto Veraldi[3], Paolo Zaffino[3],
Carlo Cosentino[3], Maria Francesca Spadea[3,4] and Elena De Momi[1,2]

[1]Department of Gynecology, European Institute of Oncology (IEO), via Giuseppe Ripamonti 435, Milan, 20142, Italy

[2]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, Milan, 20133, Italy

[3]Department of Experimental and Clinical Medicine, Università degli Studi Magna Graecia di Catanzaro, viale Europa, Catanzaro, 88100, Italy

[4]Institute of Biomedical Engineering, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany

Abstract

Ovarian cancer (OC) is the most lethal gynecologic malignancy worldwide, characterized by aggressive behavior, high relapse rate, and rapid progression. The cornerstone of OC treatment is cytoreductive surgery, targeting the removal of all detectable tumor lesions wherever feasible. In instances of widespread disease or significant perioperative morbidity risk, patients may initially receive neoadjuvant chemotherapy aimed at reducing the tumor's volume prior to surgical intervention. The pivotal decision between surgery and chemotherapy poses a significant therapeutic challenge in OC management. Our contribution is to develop an artificial intelligence-based model to support this critical decision by predicting Tumor Resectability (TR) from preoperative Computed Tomography (CT) images at the time of diagnosis.

Our study aims to develop a 3D Convolutional Neural Network capable of predicting TR in a cohort of 650 with advanced stage epithelial patients with OC who underwent surgery at the European Institute of Oncology (IEO, Milan, Italy). The model processes preoperative CT scans of the Thorax, Abdomen, and Pelvis to deliver a binary prediction: TR=0 indicates a tumor completely resected, while TR=1 indicates the presence of residual tumor after cytoreductive surgery. We design and train our model from the ground up, achieving as preliminary results an accuracy of 65%.

As far as we are aware, this is the first attempt to leverage deep learning for assessing TR in OC patients based on preoperative CT scans. Our model represents a non-invasive and preoperative tool with the potential to facilitate clinical decision making in the era of individualized and precision medicine.

The work is part of the project *Under-XAI: understanding ovarian cancer initiation and progression through explainable AI.* Project code: PNRR-MAD-2022-12376574.

Keywords

Ovarian Cancer (OC), Tumor Resecability (TR) prediction, Artificial Intelligence (AI), Precision Medicine

1. Introduction

Ovarian Cancer (OC) is the most lethal gynaecologic malignancy worldwide, ranking as the fifth deadliest cancer among women and accounting for approximately 13000 deaths in 2023 in the United States [1].

According to guidelines, suspected OC patients firstly undergo pelvic ultrasound, Computed Tomography of the Thorax, Abdomen and Pelvis (CT TAP) and CA125 measurement for staging purposes. Depending on the CT TAP results and clinical assessment, clinicians evaluate the tumor resectability. Patients likely to achieve complete tumor resection undergo primary debulking surgery followed by adjuvant chemotherapy. Otherwise, they receive neoadjuvant chemotherapy, followed by interval debulking surgery and adjuvant chemotherapy.

Although most patients initially respond positively to this standard of care, it is estimated that 70% of patients will experience a relapse. Surgical intervention aims at achieving complete tumor resection; however, it often results to be either aggressive, leading to severe postoperative complications, or ineffective, resulting in incomplete tumor removal and an associated twofold increase in the risk of death, with the latter scenario occurring in approximately 40% of cases [2].

The challenge in clinical practice is accurately predicting the success of cytoreductive surgery, critical due to the severe consequences of misjudgment, such as unnecessary invasive procedures causing significant perioperative complications and emotional distress. The complexity of predicting surgical outcomes is heightened by the varied and distinct presentations of OC - four clinical cases are shown in Figure 1 - making it difficult to assess tumor resectability from diagnostic imaging. Advancements in this area are crucial to minimize unnecessary surgeries and tailor treatments to patient-specific needs. Nowadays radiomics, a computational tool for extracting high-dimensional features from medical images, becomes
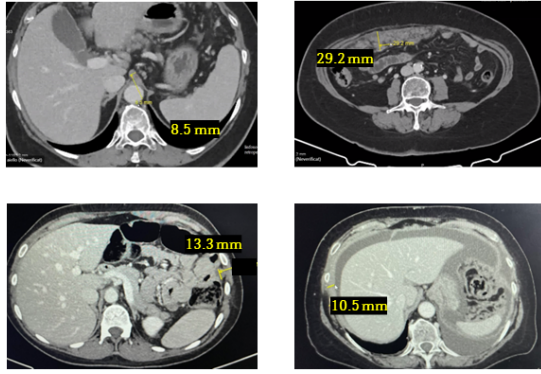
✉ francesca.fati@ieo.it (F. Fati)

Figure 1: In these CT scans, 4 different patients with OC are depicted, each presenting unique diagnostic challenges. The first patient's scan identifies a discrete retroperitoneal lymph node measuring 8.5 mm. The second patient has a conspicuous omental cake, which is considerably larger at 29.2 mm. For the third patient, a 13.3 mm nodule is present. Lastly, the fourth patient's scan shows peritoneal thickening of 10.5 mm, raising concerns about potential peritoneal carcinomatosis. Each case demonstrates the diverse presentations of OC and the inherent challenges in predicting TR preoperatively.

part of personalized oncology treatments, driven by advancements in Machine Learning (ML)[3]. However, ML requires appropriate selection among the numerous radiomic features extracted from images [4] [5].

Deep Learning (DL) has shown promising results in automatically and directly detecting valuable features from medical imaging [6], boosting the progress of computer vision in the medical field [7] [8], and demonstrating superior performance in comparison to hand-crafted image features [9].

In this paper, a 3D CNN was designed to perform binary TR classification of patients with OC. Specifically, what emerged from literature research was the absence of robust radiological indexes to select patients for total surgical resection. Hence, our primary contribution was the implementation of a non-invasive and preoperative DL model to assess whether an upfront patient could be a suitable candidate for debulking surgery, when achieving a total resection appears feasible, or the patient might be recommended to undergo neoadjuvant chemotherapy before proceeding to interval surgery, when a complete resection seems unlikely. Therefore, the proposed model might potentially assist radiologists and gynecologists to assess TR and guide therapeutic strategies for patients with OC.

2. Related Work

Artificial Intelligence (AI) has been demonstrated to enhance the effectiveness of tumor detection, classification, and treatment monitoring in cancer imaging.The integration of radiomics and DL enabled the extraction of image features and information, which might be imperceptible to the human subjective evaluations, yielding to promising medical applications [10].

In the context of OC, several noteworthy studies have been conducted.

In the domain of radiomics and ML domain, Lu et al.[11] proposed an approach to predict two-year overall survival in 364 epithelial OC patients. In this study, 657 quantitative descriptors were extracted from preoperative CT images, upon which the ML algorithm *Radiomic Prognostic Vector* was developed. The latter accurately identified the 5% of patients with a median overall survival of less than 2 years, demonstrating significant improvement over established prognostic methods. Crispin-Ortuzar et al. [12] addressed the challenge of predicting neoadjuvant chemotherapy (NACT) response of 72 high-grade serous OC (HGSOC) affected patients, presenting an ensemble ML model that, integrating baseline clinical, blood-based, and radiomic biomarkers from primary and metastatic lesions, predicted changes in total disease volume. Validation on internal and external cohorts showed that the model significantly improved prediction accuracy compared to the clinical model, highlighting the potential of radiomics in enhancing treatment response predictions.

On the front of DL, Jan et al.[10] developed an AI ensemble model combing radiomics, DL and clinical features from CT images to distinguish between benign and malignant OC. With 149 patients and 185 tumors, the model achieved 82% accuracy, 89% specificity, and 68% sensitivity. Compared to junior radiologists, the model exhibited higher accuracy and specificity while maintaining comparable sensitivity. Wang et al.[7] proposed a DL method to predict 3-year recurrence in 245 high-grade serous OC patients from preoperative CT images. The DL network, trained on 8917 CT images, extracts a 16-dimensional DL feature used to predict the outcome probability. The model achieved AUC values of 0.772 and 0.825 for high and low recurrence risk, exhibiting stronger prognostic value compared to clinical characteristics. Zheng et al.[13] proposed a Vit-based DL model for predicting overall survival in 734 high-grade serous OC patients using preoperative CT images. Analyzing 734 patients, the dataset was split into training (n = 550) and validation (n = 184) cohorts. The model demonstrated robust performance with AUC = 0.822 in the training cohort of 550 patients and AUC = 0.823 in the validation cohort of 184 women. Lei et al. [14] developed a DL model for predicting platinum sensitivity in 93 patients with

epithelial OC using contrast-enhanced magnetic resonance imaging (MRI). A pre-trained CNN were used and 1,024 features were automatically extracted from MRI sequences to predict platinum sensitivity.The model performed Area Under the Curve (AUC) of 0.97 and 0.98 in training and validation cohorts.

Among the 20 research papers examining OC in [15], 11 primarily aimed at classifying between benign, malignant, and/or borderline tumors. Two of these studies focused on resistance to platinum-based chemotherapy, with one extending its analysis to differentiate between high and low risks of disease survival and platinum treatment resistance [16], [17]. Additional studies targeted various classification objectives, such as differentiating HGSC from non-HGSC [18], classifying epithelial OC into type I or II [19], and identifying OC as recurrent or nonrecurrent [20]. The majority of these studies were single-center initiatives, with a sample size ranging from a minimum of 6 patients to a maximum of 758 patients. However, to the best of our knowledge, this is the first attempt to predict TR exploiting a DL-based model in OC.

3. Methodology

3.0.1. Dataset

TAP CT images with different manufacturers (GE Medical Systems, Siemens, Philips, Toshiba, Hitachi) of 650 patients with OC treated between 2016 and 2022 were retrospectively collected at the European Institute of Oncology (IEO) in Milan, Italy. In this study, the TAP CT contrast enhanced portal-venous phase acquired at the moment of diagnosis was considered. The CTs in our dataset were meticulously manually annotated by 5 expert gynecologist for the purpose of classification. This involved a thorough examination of elecronic medical records, resulting in the assignment of TR = 0 and TR = 1 labels. TR=0 means complete tumor resection with no residual tumor, and the TR=1 means no-complete tumor resection with residual tumor. In our dataset, clinicians annotated 446 cases with TR=0 and 204 cases with TR=1. During the development of the model, all data were fully anonymized to ensure the utmost privacy and data protection.

The inclusion and exclusion criteria for the study are illustrated in Table 1.

3.0.2. Image pre-processing

We performed images preprocessing techniques in Python 3.11. The following steps were performed:

1. Step #1: Segmentation and Region of Interest (ROI) selection. Identifying the ROI most affected by OC in CT images is a fundamental first step,

Table 1
Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Epithelial OC	CT slice thickness > 5 mm
Advanced stage (III-IV)	No consent to research
CT acquired before treatment	No CT or data available
Age $\geq$ 18 years	

and therefore requires segmentation in the preprocessing phase. To address the limitations associated with manual segmentation, including potential bias, time-intensive procedures, and the scarcity of annotated data, we implement automatic segmentation using *TotalSegmentator* [21]. *TotalSegmentator* is a DL segmentation model which automatically and robustly segments all major anatomical structures in body CT images. Each organ is associated with a label, which allowed to set the upper and lower bounds of the ROI as ischiopubic rami of the pelvis and left and right hemidiaphragm cupola, respectively. Afterwards, each image was cropped along the z-axis according to the interval selected.

The selection of the ROI derived from the fact that, compared to other tumors, OC metastasis occurs most frequently in the omentum or peritoneum, reporting almost 70% of patients with OC presented peritoneal cavity metastasis at the time of diagnosis.

2. Step #2: Additional standard preprocessing. Images pixels intensity was normalized between 0 and 1. Images were resized from the original dimension of 512 x 512 x n, with n varying among patients, to 128 x 128 x 128, where 128 was the average n.

The aforementioned preprocessing steps are illustrated in Figure 2.

3.0.3. Model architecture

For the classification task of predicting the binary clinical outcome TR from 3D TAP CT images, we designed a 3D CNN model. The architecture of the model comprises two fundamental components: a CNN-based Features Extractor (FE) and a feed-forward fully connected classifier. The FE is composed by a sequence of 7 convolutional blocks, each consisting of the following layers: a convolutional layer which increases the number of input channels, followed by a Rectified Linear Unit (ReLU) activation layer, a 3D batch normalization layer, another convolutional layer, which preserves the number of input feature maps, and a max-pooling layer which halves the spatial dimen-
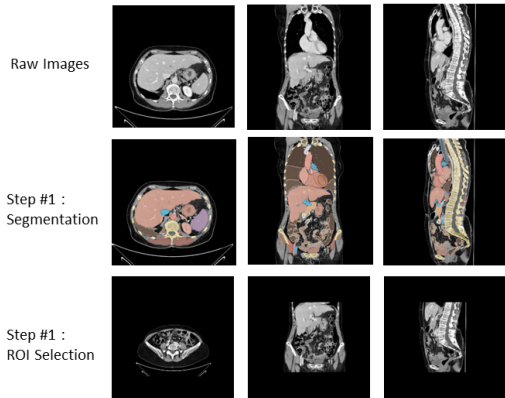
Figure 2: Preprocessing steps of TAP CT of patients with OC.

sions of the input. The FE takes as input preprocessed 3D TAP CT images with spatial dimensions of 128 x 128 x 128 and number of channels of 1, and returns a final vector of 512 extracted features. The feature vector is the input of the classifier, designed as a sequential model with linear layers interleaved with ReLU activation functions and Dropout layers, each having a dropout probability of 0.3. The overall architecture is shown in Figure 3.

The model combines the 3D CNN's ability to extract useful features from input 3D TAP CT images with the classifier's power to discriminate to predict the correct class.

We designed and trained our model from scratch, for the binary target task of TR classification in patients with OC.

4. Experiment description

We split our dataset into a training and validation set, with respectively 457 and 153 patients, and we evaluate our results on an external cohort of 40 patients.

We configured a batch size $B = 8$, employing Binary Cross Entropy (BCE) as a loss function, formulated as:

$$\text{BCE}(\hat{y}, y) = -\frac{1}{B} \sum_{i=1}^{B} [y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)]$$

where y is the model output and $\hat{y}$ is the target variable. The learning rate was set to be 0.0001, with a 0.1 multiplication every 30 epochs. Optimization was performed using the Adam algorithm, and the maximum training epoch was set to 200. The entire training procedure was executed on a single NVIDIA A100 GPU with 40GB of memory.

In this study, we employed a 5-fold cross-validation method to assess the performance of different model on different dataset splits based on the accuracy on the validation sets. We then proceeded to retrain this model using the best hyperparameters as found by the cross-validation. After retraining, we evaluated its performance on a separate test set to confirm its effectiveness and generalization capabilities.

5. Results and Discussion

In figure 4, we report the confusion matrix. We obtained an accuracy value of 0.65 on a testing external cohort of 40 patients, where the class TR=0 were 20 and the class TR=1 were 20. From the results, the model correctly predicted the positive cases (class TR = 0) with an accuracy of 0.75 and the negative cases (class TR = 1) with an accuracy of 0.55. The overall accuracy of 0.65 suggests moderate general correctness, but indicate that the model had still difficulty in correctly discriminate the classes.

6. Conclusions

In this paper, we delved into the power of DL models for the classification of TR in patients with OC, utilizing 3D TAP CT scans. TR is a pivotal diagnostic factor influencing clinical treatment decisions, that would highly improve the management of OC patients if accurately predicted at diagnosis. Leveraging the capabilities of DL in the medical domain, we extend its use to address this challenge in OC. Our methodology involves employing a 3D CCN model for binary TR classification, aiming to aid clinical decision in OC care.

Previous studies have already involved and introduced DL in the context OC, but to the best of our knowledge, this is the first attempt to harness the potential of DL for specifically predicting TR. Indeed, one of the noteworthy challenges in this attempt is the absence of radiological indexes to inform total surgical tumor resection decisions. Our main contribution is to address this gap, introducing a DL model as a non-invasive preoperative tool to facilitate clinical decision making.

In conclusion, it is important to recognize the limitations of our study, notably the potential for enhanced model generalization and performance by expanding the patient cohort, and considering alternative neural network architectures, such as Vision Transformer based models. We should broaden the application of our DL approach to predict other key diagnostic factors in OC, such as platinum sensitivity, overall survival, and surgical complications. Finally, the integration of explainability techniques should be essential for interpreting the model decisions, fostering trust, and promoting wider clinical use.
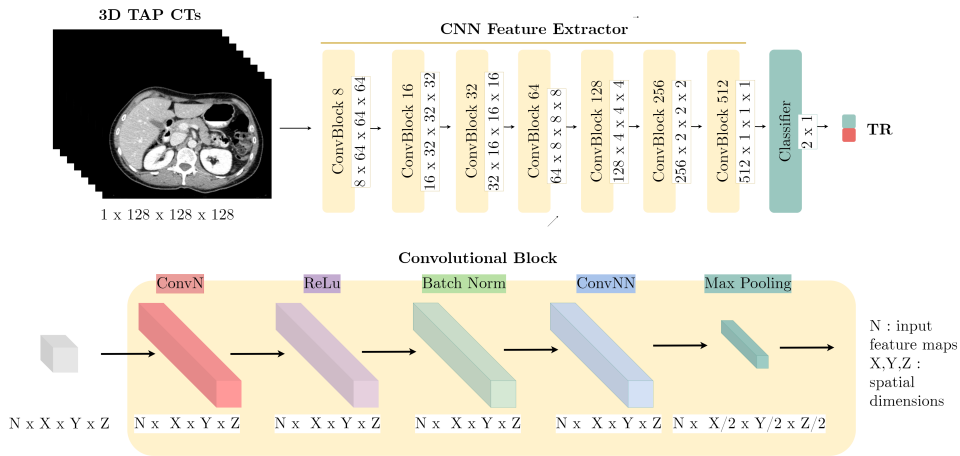
3D TAP CTs

CNN Feature Extractor

ConvBlock 8
8 x 64 x 64 x 64

ConvBlock 16
16 x 32 x 32 x 32

ConvBlock 32
32 x 16 x 16 x 16

ConvBlock 64
64 x 8 x 8 x 8

ConvBlock 128
128 x 4 x 4 x 4

ConvBlock 256
256 x 2 x 2 x 2

ConvBlock 512
512 x 1 x 1 x 1

Classifier
2 x 1

TR

1 x 128 x 128 x 128

Convolutional Block

ConvN   ReLu   Batch Norm   ConvNN   Max Pooling

N x X x Y x Z   N x X x Y x Z   N x X x Y x Z   N x X x Y x Z   N x X x Y x Z   N x X/2 x Y/2 x Z/2

N : input feature maps
X,Y,Z : spatial dimensions

Figure 3: The model architecture consists of two main components: a CNN-based Features Extractor (FE) and a feed-forward fully connected classifier. The Features Extractor includes 7 convolutional blocks. Each convolutional block is composed by a convolutional layer to increase the number of input channels, followed by ReLU activation, 3D batch normalization, another convolutional layer which does not change the number of input feature maps, and max-pooling which halves the spatial dimensions. The FE outputs a final 512-features vector. The classifier is a sequential model of linear layers with ReLU activations and Dropout layers (dropout probability: 0.3) which process the feature vector and returns the final probability to belong either to class TR=0 or to class TR=1.

Figure 4: Confusion matrix evaluating our model's performances on a testing external cohort of 40 patients. The model correctly predicted the positive cases (class TR = 0) with an accuracy of 0.75 and the negative cases (class TR = 1) with an accuracy of 0.55. The model shows an overall accuracy of 0.65.

7. Declarations

7.0.1. Conflict of interest

The authors declare that they have no conflict of interest.

7.0.2. Ethical approval

This work is part of the PNRR-MAD-2022-12376574 project *Under-XAI: understanding ovarian cancer initiation and progression through explainable AI*, being exempted from the ethical committee approval by the National Ministry of Health. Furthermore, the European Institute of Oncology has implemented a broad consent which allows to include in the study all the institute's patients, except those that refused explicitly to sign the informed consent.

7.0.3. Informed consent

Informed consent was obtained from all individual participants included in the study.

References

[1] R. Siegel, D. Naishadham, A. Jemal, Cancer statistics, 2013., CA: a cancer journal for clinicians 63 (2013) 11−30.

[2] F. Heitz, P. Harter, P. F. Alesina, M. K. Walz, D. Lorenz, H. Groeben, S. Heikaus, A. Fisseler-Eckhoff, S. Schneider, B. Ataseven, et al., Pattern of and reason for postoperative residual disease in

patients with advanced ovarian cancer following upfront radical debulking surgery, Gynecologic oncology 141 (2016) 264–270.

[3] B. J. Erickson, P. Korfiatis, Z. Akkus, T. L. Kline, Machine learning for medical imaging, radiographics 37 (2017) 505–515.

[4] R. J. Gillies, P. E. Kinahan, H. Hricak, Radiomics: images are more than pictures, they are data, Radiology 278 (2016) 563–577.

[5] M. R. Tomaszewski, R. J. Gillies, The biological meaning of radiomic features, Radiology 298 (2021) 505–516.

[6] B. J. Erickson, P. Korfiatis, T. L. Kline, Z. Akkus, K. Philbrick, A. D. Weston, Deep learning in radiology: does one size fit all?, Journal of the American College of Radiology 15 (2018) 521–526.

[7] S. Wang, Z. Liu, Y. Rong, B. Zhou, Y. Bai, W. Wei, M. Wang, Y. Guo, J. Tian, Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer, Radiotherapy and Oncology 132 (2019) 171–177.

[8] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, et al., Identifying medical diagnoses and treatable diseases by image-based deep learning, cell 172 (2018) 1122–1131.

[9] M. S. Choi, B. S. Choi, S. Y. Chung, N. Kim, J. Chun, Y. B. Kim, J. S. Chang, J. S. Kim, Clinical evaluation of atlas-and deep learning-based automatic segmentation of multiple organs and clinical target volumes for breast cancer, Radiotherapy and Oncology 153 (2020) 139–145.

[10] Y.-T. Jan, P.-S. Tsai, W.-H. Huang, L.-Y. Chou, S.-C. Huang, J.-Z. Wang, P.-H. Lu, D.-C. Lin, C.-S. Yen, J.-P. Teng, et al., Machine learning combined with radiomics and deep learning features extracted from ct images: a novel ai model to distinguish benign from malignant ovarian tumors, Insights into Imaging 14 (2023) 68.

[11] H. Lu, M. Arshad, A. Thornton, G. Avesani, P. Cunnea, E. Curry, F. Kanavati, J. Liang, K. Nixon, S. T. Williams, et al., A mathematical-descriptor of tumor-mesoscopic-structure from computed-tomography images annotates prognostic-and molecular-phenotypes of epithelial ovarian cancer, Nature communications 10 (2019) 764.

[12] M. Crispin-Ortuzar, R. Woitek, M. A. Reinius, E. Moore, L. Beer, V. Bura, L. Rundo, C. McCague, S. Ursprung, L. Escudero Sanchez, et al., Integrated radiogenomics models predict response to neoadjuvant chemotherapy in high grade serous ovarian cancer, Nature communications 14 (2023) 6756.

[13] Y. Zheng, F. Wang, W. Zhang, Y. Li, B. Yang, X. Yang, T. Dong, Preoperative ct-based deep learning model for predicting overall survival in patients with high-grade serous ovarian cancer, Frontiers in Oncology 12 (2022) 986089.

[14] R. Lei, Y. Yu, Q. Li, Q. Yao, J. Wang, M. Gao, Z. Wu, W. Ren, Y. Tan, B. Zhang, et al., Deep learning magnetic resonance imaging predicts platinum sensitivity in patients with epithelial ovarian cancer, Frontiers in Oncology 12 (2022) 895177.

[15] P. Shrestha, B. Poudyal, S. Yadollahi, D. E. Wright, A. V. Gregory, J. D. Warner, P. Korfiatis, I. C. Green, S. L. Rassier, A. Mariani, et al., A systematic review on the use of artificial intelligence in gynecologic imaging–background, state of the art, and future directions, Gynecologic Oncology 166 (2022) 596–605.

[16] H. Veeraraghavan, H. A. Vargas, A. Jimenez-Sanchez, M. Micco, E. Mema, Y. Lakhman, M. Crispin-Ortuzar, E. P. Huang, D. A. Levine, R. N. Grisham, et al., Integrated multi-tumor radio-genomic marker of outcomes in patients with high serous ovarian carcinoma, Cancers 12 (2020) 3403.

[17] X.-p. Yu, L. Wang, H.-y. Yu, Y.-w. Zou, C. Wang, J.-w. Jiao, H. Hong, S. Zhang, Mdct-based radiomics features for the differentiation of serous borderline ovarian tumors and serous malignant ovarian tumors, Cancer Management and Research (2021) 329–336.

[18] H. An, Y. Wang, E. M. Wong, S. Lyu, L. Han, J. A. Perucho, P. Cao, E. Y. Lee, Ct texture analysis in histological classification of epithelial ovarian carcinoma, European Radiology 31 (2021) 5050–5058.

[19] L. Qian, J. Ren, A. Liu, Y. Gao, F. Hao, L. Zhao, H. Wu, G. Niu, Mr imaging of epithelial ovarian cancer: a combined model to predict histologic subtypes, European Radiology 30 (2020) 5815–5825.

[20] Y. Liu, Y. Zhang, R. Cheng, S. Liu, F. Qu, X. Yin, Q. Wang, B. Xiao, Z. Ye, Radiomics analysis of apparent diffusion coefficient in cervical cancer: a preliminary study on histological grade evaluation, Journal of Magnetic Resonance Imaging 49 (2019) 280–290.

[21] J. Wasserthal, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. T. Boll, J. Cyriac, S. Yang, et al., Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images, Radiology: Artificial Intelligence 5 (2023).