scientific reports

Check for updates

OPEN

A Radiomic-based model to predict the depth of myometrial invasion in endometrial cancer on ultrasound images

Francesca Arezzo[1,9], Annarita Fanizzi[2,9], Rosanna Mancari[3], Emiliano Cocco[4], Samantha Bove[2]✉, Maria Colomba Comes[2]✉, Mariangela Gianciotta[5], Giorgia Lanza[5], Salvatore Lopez[1], Gerardo Cazzato[6], Erica Silvestris[1], Elsa Vitale[2], Enrico Vizza[3], Gennaro Cormio[1,7], Raffaella Massafra[2,9] & Vera Loizzi[1,8,9]

In Europe, endometrial carcinoma was found to be the fourth most common tumor in the female population in 2022. The depth of myometrial invasion is a well-established and crucial prognostic risk factor in endometrial cancer patients and is important for choosing the most appropriate management for the patient. However, while the preoperative assessment of tumor invasion carried out using radiological imaging is very important, it is a subjective examination and its accuracy is based on the experience of the operator. In this scenario we proposed a radiomic-based model to predict myometrial invasion for ultrasound images. We collected clinical data and qualitative ultrasound indicators of 77 consecutive patients affected by endometrial carcinoma. After a pre-processing phase of ultrasound images, a pre-trained Inception-V3 convolutional neural network (CNN) was used as features extractor. Then, a binary classification model and a multiclass model were trained, after a double step of feature selection; the first selection stage performed feature filtering based on a nonparametric test, the second stage selected features by evaluating not only the relationship with the outcome of interest, but also the relationship with other predictive features. For the multiclass prediction model, a cascade approach has been developed. The two proposed models were validated in 100 ten-fold cross-validation rounds. In addition, to assess the effect of the potential added value of using this tool in clinical practice, we compared the operator's performance with and without the developed automated support. The binary and multiclass model reached optimal classification performances with a mean AUC value equals to 90.76 (88.63–92.89 IC95%). When the operator was assisted by the radiomic-based decision-making system increased by 10% points in terms of precision. The multiclass model showed the per-classes recall were 93.33%, 71.88% and 90.00%, for focal infiltration, infiltration less than 50%, and infiltration greater than 50% class, respectively. The performances of the operator when assisted by the radiomic-based decision-making system were statistically superior both in terms of overall accuracy and per-class recall of intermediate class (rising to 82.82% and 71.88% compared to 71.88% and 56.25%, respectively). The proposed models have the potential to standardize examinations that rely on subjective evaluations, such as ultrasound. They can represent a valid support tool for the clinicians for an accurate estimate of the deep myometrial infiltration already in the diagnosis phase with an easily accessible, low-cost examination that causes no discomfort for the patient such as ultrasound.

Keywords Endometrial Cancer, Imaging in gynecological disease, Depth of myometrial invasion, Ultrasound, Radiomic model

[1]Clinicalized Gynecological Oncology Unit , IRCCS Istituto Tumori 'Giovanni Paolo II' , Viale Orazio Flacco 65, 70124 Bari, Italy. [2]Biostatistics and Bioinformatics Laboratory, IRCCS Istituto Tumori 'Giovanni Paolo II', Viale Orazio Flacco 65, Bari 70124, Italy. [3]Gynecologic Oncology Unit, IRCCS Regina Elena National Cancer Institute, Rome 00144, Italy. [4]Sylvester Comprehensive Cancer Center, Department of Biochemistry and Molecular Biology, Miller School of Medicine, University of Miami, Miami, FL 33136, USA. [5]Obstetrics and Gynecology Unit, Department of Biomedical Sciences and Human Oncology, University of Bari "Aldo Moro", Bari, Italy. [6]Department of Emergency and Organ Transplantation, Pathology Section, University of Bari "Aldo Moro", Bari, Italy. [7]Department of Interdisciplinary

Medicine (DIM), University of Bari "Aldo Moro", Bari, Italy. [8]Department of Translational Biomedicine and Neuroscience, University of Bari, Bari, Italy. [9]Francesca Arezzo, Annarita Fanizzi, Raffaella Massafra and Vera Loizzi contributed equally. ✉email: s.bove@oncologico.bari.it; m.c.comes@oncologico.bari.it

Background

In Europe, endometrial carcinoma (EC) was found to be the fourth most incident tumor in the female population in 2022[1]. Differences in patient characteristics and histopathologic features of the disease impact both patient prognosis and the recommended treatment approach. Therefore, the diagnosis and treatment of EC patients should be evaluated in a multi-disciplinary setting[2].

The depth of myometrial invasion (DMI) is a well-established and crucial prognostic risk factor in EC[3]. It is advisable to evaluate the extent of myometrial involvement by expressing it as a percentage of the total myometrial thickness that has been infiltrated by the carcinoma[4]. The assessment should be categorized into three groups: no invasion; less than 50% invasion; or greater than or equal to 50% invasion[5].

A myometrial invasion greater than or equal to 50% of the myometrial thickness is considered deep myometrial invasion.

According to the International Federation of Gynecology and Obstetrics (FIGO)[6], the depth of myometrial invasion contributes to the staging process of the disease.

Moreover, as recommended by the International Society of Gynecological Pathology (ISGyP) and the ESGO/ESTRO/ESP guidelines, conventional pathologic features such as histotype, grade, LVSI as well as myometrial invasion are still important prognostic parameters that, together with molecular markers, contribute to the definition of prognostic risk groups, with significant implications for choosing the most suitable treatment for the patient[7,8].

Therefore, the recognized role of depth of myometrial invasion as a crucial prognostic risk and its importance in choosing the most appropriate management for the patient explains the necessity of a correct evaluation of this parameter during the patient's staging.

Currently, ultrasound (US) and magnetic resonance (MR) imaging are the most commonly used techniques for preoperative assessment of the depth of myometrial invasion in patients with EC. Several studies have shown that both modalities have similar diagnostic efficacy, however, compared with MRI, US exam has lower costs and shorter examination times to perform[9,10].

However, the preoperative assessment of tumor invasion carried out using US imaging can sometimes be extremely complex, for example when other uterine pathologies such as uterine fibroids or adenomyosis coexist[11] Moreover, compared to MR, US exam is a subjective examination method, and its accuracy relies on the experience of the operator which leads to differences in diagnostic results.

The potential of extracting a radiomics signature from radiological images for the resolution of various unmet clinical needs in different oncology settings is currently well consolidated in the literature[12].

Radiomic analysis of US images allow quantitative extraction of caraad high-dimensionality tteristics not discernible to even the most experienced operator's eye. They capture structural, intensity-based, and morphologic properties of the tumor and surrounding tissues that can help quantify heterogeneity in a way that standard image interpretation cannot. Therefore, in the context of assessing deep myometrial invasion in endometrial cancer, radiomic features offer a powerful tool that can improve diagnostic accuracy, but more importantly, standardize the assessment of the examination and ensure reproducibility.

There are several works in the literature that propose automated learning models for predicting the depth of myometrial infiltration starting from the quantitative analysis of radiological images[13], in particular Magnetic Resonance images[14–18]To the best of our knowledge, literature concerning the assessment of myometrial depth by quantitative analysis of ultrasound images is poor[19].

In this scenario, we propose models for predicting myometrial infiltration in patients suffering from EC based on the extraction of quantitative features from diagnostic ultrasound images. A well-known pre-trained convolutional neural network (CNN) architecture was used as an extractor of features subsequently used to train a machine learning model for solving the diagnostic task. Specifically, pre-trained CNNs refer to transfer learning technique. The networks have been previously trained on millions of natural non-medical images to learn how to automatically extract features of different level of abstraction. Using a pre-trained convolutional neural network as a feature extractor is a powerful approach in medical imaging analysis. Unlike traditional radiomics, which extracts predefined crafted features (e.g., intensity, texture, shape), CNNs learn hierarchical representations: the lower layers capture local features called low-level features such as edges and dots, while the deeper layers capture more complex features called high-level features such as shapes and objects. In contrast to handcrafted radiomic features that require prior knowledge for their design, features extracted from CNNs are learned automatically, minimizing subjective selection and potentially improving robustness in datasets.

In agreement with the approaches proposed in the literature for the same diagnostic problem, we first developed a binary model for predicting the depth of myometrial infiltration, i.e. infiltration less than 50% (including cases of absent or focal infiltration) vs. infiltration greater than 50%. Subsequently, with the purpose of responding to clinical and therapeutic needs, we also proposed a multiclass classification model (mod2) considering three classes, such as absent or focal infiltration vs. infiltration less than 50% vs. infiltration greater than 50%. In addition, we compared the operator's performance with and without the developed automated support to assess the effect of the potential added value of using this tool in clinical practice.

Materials and methods

Sample study

The study followed the principles of the Declaration of Helsinki and received approval from the Ethics Committee of the Azienda Ospedaliera Policlinico Consorziale - University of Bari, Italy (protocol number 6398/2020). Written informed consent for participation was required for this retrospective observational study. This retrospective observational study involved 77 patients diagnosed with endometrial carcinoma through hysteroscopic biopsy. The patients were monitored between May 2021 and May 2023 at the Azienda Ospedaliera Policlinico Consorziale - University of Bari, IRCCS Giovanni Paolo II Cancer Institute, and the IRCCS Regina Elena National Cancer Institute in Rome.

The ultrasound examinations were performed transvaginally (or transrectally, in case of the patient's inability to perform transvaginal ultrasound) and where necessary, the examination was completed through transabdominal evaluation. However, the images analyzed for feature extraction were always longitudinal scans of the uterus evaluated transvaginally or transrectally in order to use the most accurate approach and standardize the method of feature extraction[20]. The ultrasound scans were conducted using either a 5.0–9.0 MHz vaginal probe or a 3.5–5.0 MHz abdominal probe. All ultrasound reports and images were made available for review.

The ultrasound examinations were performed by two operators for each center with 5–15 years of experience in gynecological ultrasound. All ultrasound examiners were experienced level II or III according to the European Federation of Societies for Ultrasound in Medicine and Biology[21]. The images were then centralized, where an operator with more than ten years of experience confirmed the labeling performed by the previous operators. The radiomic features were extracted and subsequently analyzed by physicists. This was done in order to allow a uniform and accurate evaluation since a non-standardized definition of features would make the results derived from the application of AI algorithms misleading.

Clinical data and qualitative ultrasound findings were collected from 77 consecutive patients diagnosed with endometrial carcinoma (EC). Table 1 provides a summary of the sample characteristics. Fifteen patients

Characteristics	Distribution
Age at diagnosis	
Median (1 th quartile; 3 th quartile)	60.00 (55.00; 69.00)
BMI	
Median (1 th quartile; 3 th quartile)	30.00 (28.00; 69.00)
Menopausal status	
Yes (Abs; %)	63; 81.82%
No (Abs; %)	14; 18.18%
Endometrial echostructure	
Uniform (Abs; %)	4; 62.73%
Not uniform (Abs; %)	73; 11.82%
Endometrial echogenicity	
Hyperechoic (Abs; %)	58; 75.32%
Hypoechoic (Abs; %)	3; 3.89%
Isoechogenic	15; 9.48%
Inhomogeneous with irregular cystic areas	1; 1.30%
Endometrial midline	
Linear (Abs; %)	18; 23.38%
Not linear (Abs; %)	8; 10.39%
Irregular (Abs; %)	17; 22.08%
Non definable	34; 44.15%
Endometrial–myometrial junction	
Regular (Abs; %)	21; 27.27%
Irregular (Abs; %)	28; 36.36%
interrupted (Abs; %)	24; 31.17%
Non definable (Abs; %)	2; 2.60%
Cervical involvement	
Yes (Abs; %)	65; 84.42%
No (Abs; %) Uncertain (Abs; %)	9; 11.69% 3; 3.89%
Myometrium infiltration	
Not infiltrated/Focal infiltration (Abs; %)	15; 19.48%
Infiltration less than 50% (Abs; %)	32; 41.56%
Infiltration greater than 50% (Abs; %)	30; 38.96%

Table 1. Characteristics of the 77 patients analyzed in the study.

(19.48%) had either focal or no myometrial invasion, 32 patients (41.56%) exhibited myometrial invasion of less than 50%, and 30 patients (38.96%) had myometrial invasion of 50% or more. The histological analysis conducted on the uterus after total hysterectomy, based on definitive paraffin sections, served as the reference standard for diagnosis.

Subjective assessment

A comprehensive ultrasound assessment of the endometrium is crucial for both diagnosing and staging patients with endometrial carcinoma (EC). This evaluation should follow the guidelines established by the IETA consensus and include both quantitative and qualitative assessments. A key aspect of staging EC is determining the extent of myometrial invasion. When conducted by an experienced sonographer, this evaluation aligns more closely with histopathological findings. The process involves a dynamic, subjective examination, moving the ultrasound probe across the uterus to assess the endo-myometrial junction, detect any signs of infiltration, and measure the depth of invasion into the uterine wall[22]. According to IETA consensus, the depth of myometrial invasion was assessed by evaluating the tumor/uterine anteroposterior (AP) diameter ratio.

The assessment of the depth of myometrial invasion, both binary and multiclass, was then re-evaluated by the same operator who was asked to confirm or re-evaluate his decision in light of the suggestions provided by the prediction model. It should be noted that the reevaluation was performed blind, meaning that the operator knew neither the patient's name nor the ground truth. Specifically, during re-evaluation, for each patient, the operator was provided with the system-generated infiltration class identified by majority voting in the 100 validation rounds.

Image processing and feature extraction

The images used were in PNG format. Firstly, we normalized the images before the feature extraction stage to unify the acquired images and improve their rendering and contrast. To remove the markers that the radiologist uses during the diagnostic process to measure the lesion, a preprocessing step was necessary for the ultrasound images. an inpainting algorithm was used. (Fig. 1). This method involves replacing pixels in the targeted area through exemplar-based texture synthesis, which creates a new texture image that visually resembles the original sample[23]. In this study, we utilized an inpainting technique based on coherence transport. Once the inpainting area is defined, coherence transport propagates image values according to a transport equation in a coherent direction, maintaining a specified degree of consistency among gray-level values[24].

From the pre-processed images, radiomic features were extracted using a transfer learning approach with a convolutional neural network (CNN). This method is commonly used in literature for small sample sizes to bypass the extensive data requirements for training deep neural networks[25–27]. It involves using a pre-trained CNN, which has been trained on millions of images of various types, to extract radiomic features from our images. For this purpose, we utilized the highly effective Inception-V3 architecture[28].

Inception-V3 is a deep CNN with 48 layers. Features were extracted from the max_pooling2 d layer, which is the second layer after the second convolutional layer in this architecture. In a network architecture, features extracted from the first layers, called low-level features, representations of local details of an image, such as edges, dots, and curves. In contrast, high-level features are extracted from the last classification layers, which allow the extraction of global features, e.g., shapes and objects. We have considered low-level features here to dissect information derived from all local image structures that could be obscured instead by considering only global information.

The pre-trained CNN architecture used required an image of size. Therefore, we resized the original image to 299×299 pixels by using the imresize function that maintains the original aspect ratio of the image; we we have set a bilinear interpolation kernel and left all other parameters as default. This max_pooling2 d layer outputs a tensor with dimensions of $35 \times 35 \times 192$, which is then flattened into a vector of 235,200 elements. Consequently, each ultrasound image for every patient yielded a total of 235,200 extracted features.

The second pooling layer, being one of the earlier layers in the network, captures low-level features such as edges, dots, and curves. These details might be lost if only global information from later layers were considered.

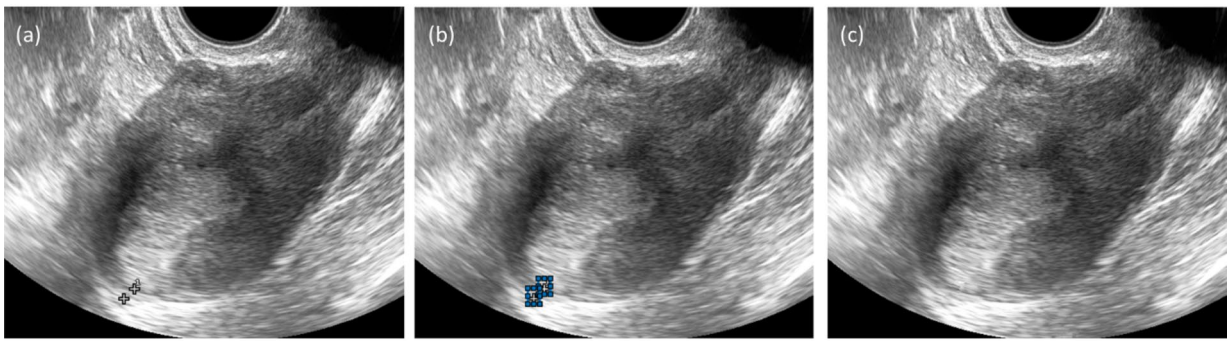

Fig. 1. Ultrasound image inpainting process. Given the input image (**a**) and selected one or more target regions (**b**), these are removed and replaced through a standard inpainting technique (**c**).

Moreover, extracting features from a pooling layer rather than a convolutional layer helps maintain invariance to truncation, occlusion, and translation.

Learning model

An overview of the study design is presented in Fig. 2. All steps were performed using MATLAB R2023a software (MathWorks, Inc., Natick, MA, USA).

A first model (*mod1*) for predicting the level of myometrial infiltration was developed to distinguish tumors with infiltration less than 50% (including cases of absent or focal infiltration) and infiltration greater than 50%. Subsequently, to respond to clinical and therapeutic needs, we proposed a second multiclass classification model (*mod2*) considering three classes of infiltration, such as absent or focal infiltration vs. infiltration less than 50% vs. infiltration greater than 50%.

The clinical necessity for developing a binary model system from the possibility to better personalize the patient's treatment. The preoperative evaluation of myometrial invasion through imaging techniques currently available to us allows modulation of the extent of surgery. The depth of myometrial invasion, together with other histological elements such as histopathological type, grade, and lymphovascular space invasion allows definition of the prognostic risk group for that patient. In case of low risk, the patient undergoes sentinel lymph node technique. In case of high risk, there is instead an indication to perform systematic lymphadenectomy. The development of a multiclass model could also be helpful in achieving further personalization of treatment for

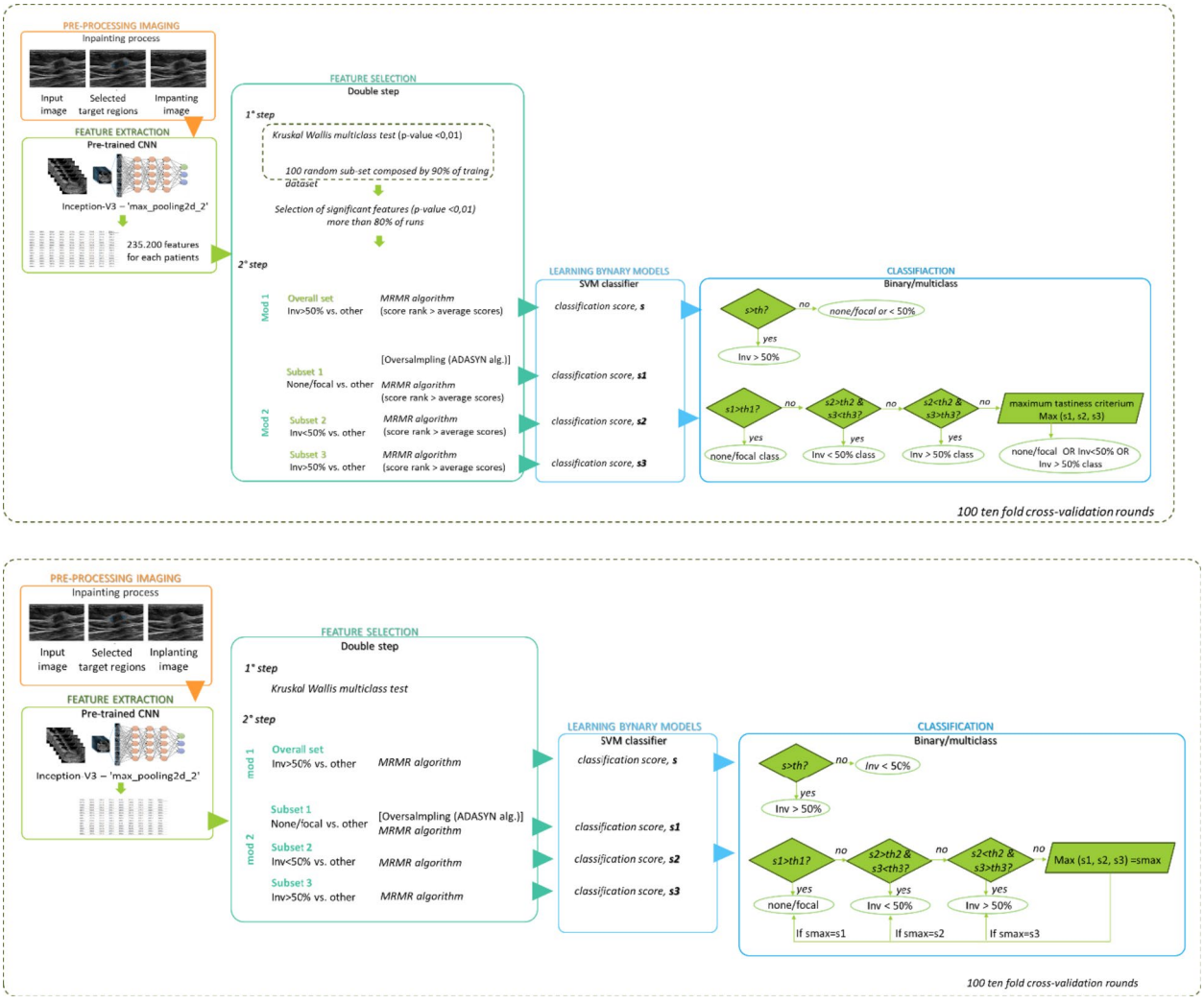

Fig. 2. Analysis pipeline. Two models for predicting the depth of myometrial infiltration have been proposed. For both the analysis pipeline develops in the following phases: (i) pre-processing imaging: an inpainting technique based on coherence transport was adopted; (ii) feature extraction: from the pre-processed image thus obtained, radiomics features were extracted using a transfer learning approach; (iii) feature selection: we implemented a double step of feature selection; (iv) learning model: the final feature subset was used to train a binary Support Vector Machine (SVM); (v) classification models: a binary classification model and a multiclass model were trained; for the multiclass prediction model, a cascade approach has been developed. The two proposed models *mod1* and *mod2* were validated in 100 ten-fold cross-validation rounds.

the group of young women who wish to preserve their fertility since, according to guidelines, the exclusion of myometrial invasion through imaging in low grade disease would make the patient eligible for fertility-sparing treatment[29].

For both proposed models, *mod1* and *mod2*, in order to have a measure of performance variability, we applied a 10-fold cross-validation scheme, iterating 100 times.the pipeline detailed below. Both models include a feature selection phase. Due to the strong imbalance between the number of radiomic features extracted using CNN and the sample size, we implemented a double step of feature selection.

First, filtering was applied by selecting only those features individually associated with the outcome of interest. This selection scheme is called maximum relevance selection. On the other hand, features can be selected so that they are mutually distant from each other while still having a "high" correlation with the classification variable. Therefore, a second feature selection step was applied that also considers the relationship with other features. To evaluate the significant association of the deep myometrial invasion (not/focal infiltration, infiltration less than 50%, and infiltration greater than 50%), a first step involves the use of the Kruskal Wallis test iterated 100 times on a random subset equal to 90% of the training sample. At the end of the iterative procedure, only the features that were significant (p-value <0.01) in at least 80% of the runs were selected. The choice of this last selection parameter is the result of experimental evaluations which were not reported in the paper so as not to burden the discussion of the work.

A second feature selection step was applied randomly on the subset of features thus identified. Specifically, the Maximum Relevance — Minimum Redundancy (MRMR) algorithm was applied to identify the features having the most correlation with the outcome and the least correlation between themselves[30]. The MRMR algorithm determines the importance of the single feature by quantifying pairwise mutual information of features and mutual information of a feature and the response. Specifically, the algorithm obtains a measure of relevance and a measure of redundancy, then assigns an importance score to each feature based on the relationship between relevance and redundancy. In this framework, features with a strong relationship with the target (high relevance) but a weak relationship with other predictive features (low redundancy) are favored and then selected. Thus, only those features whose importance score exceeded the average score of all features in the subset determined by the previous feature selection step were selected. The feature selection criterion implemented allows for the identification of a subset of features that jointly contain the greatest predictive power.

The same feature selection procedure was implemented for both the binary outcome predictive model (mod1) and the multiclass outcome predictive model (mod2). However, with reference to the multi-class problem (mod2), it should be noted that the cascade model, as described below, from individual binary models. Therefore, the second feature selection step was implemented by considering in each sub-model the outcome of interest the relevant binary outcome. A first model (*mod1*) was trained to discriminate cases with myometrial infiltration less than 50% (negative class) from those infiltration greater than 50% (positive class). Specifically, the final feature subset was used to train a binary Support Vector Machine (SVM)[31] classifier. The SVM (Support Vector Machine) classifier is a supervised machine learning algorithm that identifies the optimal hyperplane to distinguish between two classes using a kernel function. In this paper, we used the linear function. The default values of all SVM classifier parameters defined by the data analysis software used were retained.

The multiclass classification model (*mod2*) was developed using a cascade approach. We started from the hypothesis that single binary problems are more performing than a multiclass problem, especially when, as in our case study, the sub-samples of each class are not particularly numerous. Therefore, we preferred to use a One Versus All strategy, often called One Versus The Rest, which involves the use of as many binary classifiers as there are classes to predict, each trained to recognize a specific class compared to the rest of the sample. Three SVM binary classifiers with linear kernel were trained independently to solve three different tasks: (*c1*) absent or focal myometrial infiltration (positive class) vs. other (including myometrial infiltration less than 50% or greater than 50%, (*c2*) myometrial infiltration less than 50% (positive class) vs. other (including absent or focal infiltration or greater than 50%), and (*c3*) major infiltration of 50% (positive class) vs. other (including absent or focal infiltration or less than 50%). Task c3 of model mod2 is the same as mod1. In the multiclass classification model, the classifier trained on task c3 is used together with the classifier on task c2 to discriminate patients with $>50\%$ infiltration from those with $<50\%$ infiltration, after first selecting only those patients with absent/focal myometrial infiltration (task c1). Each of the three binary classifiers was trained starting from the subset of features identified resulting from the double feature selection step described above. For the *mod2* model, the Kruskal Wallis test was used in its multiclass version to filter the features with respect to their discriminating power on the original problem, while the MRMR algorithm was instead applied on the sub-samples defined above. The dataset related to the classifier *c1*was imbalanced with respect of the interest outcome. Therefore, before of feature selection and classification algorithms, we carried out an over-sampling technique. Specifically, the balancing was performed through the synthetization of sample for the less-represented class using the Adaptive Synthetic Sampling (ADASYN) approach[32,33].

For each of the three classifiers, the optimal threshold was identified using the Youden index on the ROC curves (*t1*, *t2*, and *t3*, respectively)[34]. Three classification scores were then generated for each patient (*s1*, *s2*, and *s3*, respectively). The multiclass classification criterion of the designed cascade model is structured as follows: if the *s1* classification score was greater than the *t1* classification threshold, therefore leaning towards the positive class, then the absent or focal myometrial infiltration class was assigned; otherwise, given the propensity of the first classifier towards other categories, if the classification score *s2* was greater than the threshold value *t2* and at the same time the classification score *s3* was lower than the threshold value *t3* (i.e. it also suggests, like the first classifier *c1*, a different class), the myometrial infiltration class less than 50% was assigned; if, however, the classification score *s3* was greater than the threshold value t3 and at the same time the classification score *s2* was less than the threshold value *t2*, then the assigned class was infiltration greater than 50% (i.e. it also suggests, like the first classifier *c1*, a different class); finally, in the case in which all the classification scores were lower than

	Subjective assessment	Radiomic-based model	Subjective assessment + Radiomic-based model
AUC	-	90.76 (88.63–92.89)	-
Accuracy	80.52	89.22 (87.61–90.83)*	85.71
Recall	83.33	86.00 (83.68–88.72)*	90.91
Specificity	80.85	91.28 (89.00–93.56)*	85.10
Precision	73.53	86.54 (83.54–89.54)*	86.67*
F1-score	78.12	86.16 (84.18–88.14)*	88.74*

Table 2. Performance evaluation of *depth myometrial invasion* prediction model for binary classification. The metrics are reported in percentage terms. The classification performances of the Radiomic-based model were evaluated in terms of mean and confidence interval 95% (CI95%). *p-value T-test < 0.05.

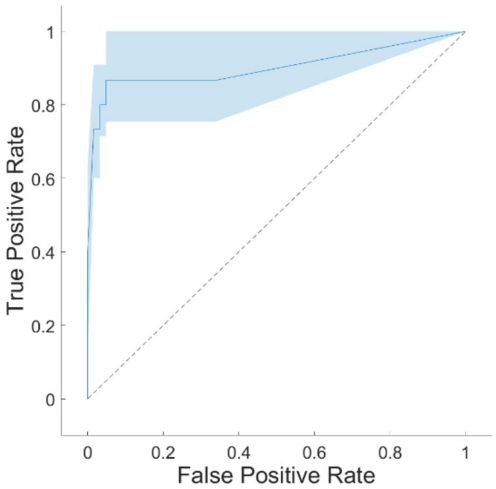

Fig. 3. ROC curve of binary model. The AUC value of the binary radiomic model wa evaluated in terms of mean and confidence interval 95%.

the respective classification thresholds, therefore not showing a particular agreement between them, the class associated with the highest of the three scores was assigned.

Statistical analysis and performance evaluation

The performance of the radiomic-based model was assessed using the mean and 95% confidence intervals. For the binary classification model (mod1), we evaluated performance based on the Area Under the Receiver Operating Characteristic Curve (AUC), as well as accuracy, recall, specificity, precision, and F1-score. These metrics were calculated by determining the optimal threshold using Youden's index on the ROC curves[34].

For the multiclass classification model (mod2), we assessed overall accuracy, per-class F1-score, recall, precision, and F1-score. Per-class recall, precision, and F1-score were computed for each class compared to all other classes. The macro-F1 score, which is the arithmetic mean of per-class F1 scores, was used to account for the imbalance among the three classes, giving equal weight to each class.

We compared the performance of the subjective assessment with that of the radiomic models and the subjective assessment supported by the developed support system using a Student's t-test. Results were deemed statistically significant if the p-value was less than 0.05.

Results

Performance evaluation of depth myometrial invasion prediction model: binary classification

The radiomic-based classification model trained to solve the same binary task was evaluated in 100 ten-fold validation rounds (Table 2). It was highly performed, thus significantly outperforming the subjective assessment. The model reached an AUC value of 90.76 (88.63–92.89 CI95%) (Fig. 3). Accuracy, sensitivity, and specificity were 86.00 (83.28–88.72 CI95%), 85.00% (83.33–86.67 CI95%), and 91.28 (89.00–93.56 CI95%), respectively (Table 2). Moreover, the radiomic-based model shown an optimal precision that was equal to 86.54% (83.54–89.54 CI95%).

On the collected sample, the real-life classification performances of the operators (hereinafter called subjective assessment) were evaluated. Concerning the binary task, such as myometrial infiltration less than 50% (including cases of absent or focal infiltration) vs. infiltration greater than 50%, the overall accuracy of the operators reached 80.52% with a good balancing between two classes: 80.85% and 83.33% for myometrial infiltration less than 50% and myometrial infiltration greater than 50%, respectively.

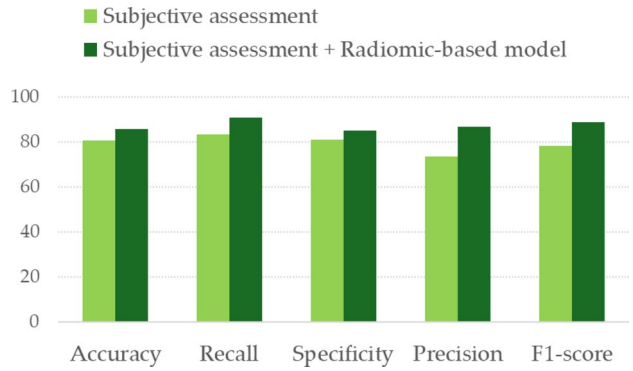

Fig. 4. Comparison of operator performance with and without support system assistance. Binary deep myometrial invasion assessment of clinician before and after being assisted by the radiomic-model.

		Subjective assessment	Radiomic-based model	Subjective assessment + Radiomic-based model
Overall accuracy		74.02	84.93 (82.72–87.15)*	82.82*
Per-class recall	Absent or focal infiltration	93.33	76.67 (69.84–83.48)*	93.33
	Infiltration less than 50%	56.25	85.63 (83.01–88.24)*	71.88*
	Infiltration greater than 50%	83.33	88.33 (84.79–91.88)*	90.00
Per-class precision	Absent or focal infiltration	70.00	78.22 (71.57–84.87)*	82.35
	Infiltration less than 50%	78.26	86.57 (83.09–90.06)*	85.19
	Infiltration greater than 50%	73.53	87.34 (85.42–89.26)*	78.79
Per-class F1-score	Absent or focal infiltration	80.00	76.86 (71.73–82.02)*	87.50
	Infiltration less than 50%	65.45	86.06 (83.22–88.91)*	77.97
	Infiltration greater than 50%	78.12	87.69 (86.07–89.31)*	84.40
Macro F1-score		74.52	84.48 (82.10–86.86)*	83.29

Table 3. Performance evaluation of *depth myometrial invasion* prediction model for multiclass classification. The metrics are reported in percentage terms. The classification performances of the Radiomic-based model were evaluated in terms of mean and confidence interval 95% (CI95%). *p-value T-test < 0.05.

The performances of the operator assisted by the radiomic-based decision-making system were statistically superior in terms of precision, rising to 86.67% (Fig. 4). The other metrics are still greater than performances without the aid of the operator although they are not statistically significant.

Performance evaluation of depth myometrial invasion prediction model: multiclass classification

For the multiclass radiomic-based model, all the calculated classification metrics outperformed the subjective assessment ones, except for the recall of the absent or focal infiltration class (Table 3). Radiomic-based model reached an average overall accuracy of 84.93% (82.72–87.15 CI95%) and a Macro F1-score of 84.48% (82.10–86.86 CI95%) with a good balance for all per-class metrics. Specifically, the per-classes recall was 93.33%, 71.88% and 90.00%, for focal infiltration, infiltration less than 50%, and infiltration greater than 50% class, respectively; whereas per -class precision were 78.22%, 86.57%, and 87.34%, respectively.

The operator's overall accuracy dropped to 74.02% respect to the binary task (Fig. 5). Specifically, greater uncertainty emerges in the identification of the infiltration less than 50% class whose per-class recall was equal to 56.25% compared to 93.33% and 83.33% of the absent or focal infiltration and infiltration greater than 50% classes, respectively.

The performances of the operator assisted by the radiomic-based decision-making system were statistically superior both in terms of overall accuracy and in terms of per-class recall in identifying the intermediate class (rising to 82.82% and 71.88% compared to 71.88% and 56.25%, respectively). Also in this case, the other metrics are still greater than performances without the aid of the operator although they are not statistically significant.

Discussions and conclusions

In a patient with EC, the evaluation of myometrial infiltration suspected on imaging at the time of staging of a patient with EC influences the choice of the most appropriate management[9]. In women who underwent radical surgery, in case of absence of myometrial invasion, systematic lymphadenectomy is not recommended, and sentinel lymph node biopsy can be omitted. Sentinel lymph node biopsy can be considered for staging purposes in patients with low-risk/intermediate-risk disease. Systematic lymph node staging should be performed in patients with high–intermediate-risk/high-risk disease, if sentinel lymph node is not detected

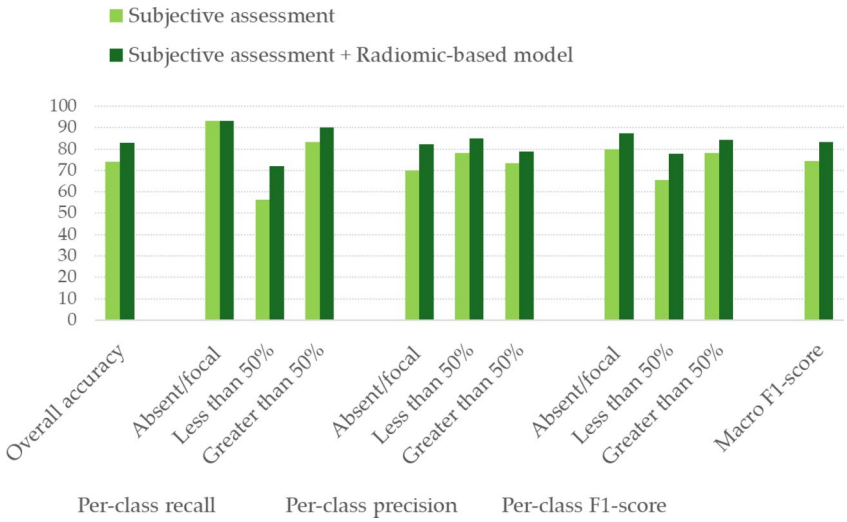

Fig. 5. Comparison of operator performance with and without support system assistance. Binary deep myometrial invasion assessment of clinician before and after being assisted by the radiomic-model.

on either pelvic side (Table 2). Ovarian preservation can be considered in pre-menopausal patients aged < 45 years with well-defined histological and staging criteria, and among these with presumed myometrial invasion < 50%[29]. Moreover, fertility-sparing treatment should be considered only in patients with atypical hyperplasia (AH), endometrial intraepithelial neoplasia (EIN) or grade 1 endometrioid EC without myometrial invasion[29]. Therefore, in case of diagnosis of EC confirmed with hysteroscopic biopsy, an accurate radiological assessment of myometrial invasion is fundamental. The pre- operative mandatory work-up of EC patient includes expert transvaginal or transrectal ultrasound[35,36]. An expert ultrasound examination can substitute pelvic resonance imaging (MRI) scan[36]. Indeed, numerous studies demonstrate that the diagnostic performance of transvaginal ultrasound and MRI for detecting myometrial invasion in EC are quite similar[37,38].

It is important to note that to define myometrial infiltration, the expert sonographer relies on subjective evaluation with a higher degree of agreement with histopathology[22]. Nevertheless, there are numerous confounding factors that can complicate the ultrasound assessment of myometrial invasion in endometrial cancer. One such factor may be the presence of coexisting uterine pathologies, such as uterine fibroids or adenomyosis[39,40]. These conditions can distort the normal uterine anatomy and make it challenging to accurately determine the extent of myometrial invasion[39,40].

Another confounding factor is the MELF pattern of myometrial infiltration. MELF, which stands for microcystic, elongated, and fragmented, is a distinct histologic pattern of invasion characterized by the presence of small, irregular clusters of cancer cells within the myometrium. These clusters can be difficult to detect on ultrasound, as they may not form a distinct mass or may be mistaken for normal myometrial tissue. The MELF pattern of invasion can lead to an underestimation of the true extent of myometrial involvement[41]. In addition to these factors, other variables such as the position of the uterus and patient body habitus can also influence the accuracy of ultrasound assessment. For example, a retroverted uterus or a large amount of abdominal fat can make it more challenging to obtain optimal ultrasound images and visualize the uterine wall clearly[42]. Moreover, the accuracy of ultrasound evaluation in assessing myometrial infiltration in EC is subjective, as it relies heavily on the radiologist's expertise, experience, and interpretive skills. While ultrasound is a valuable tool for evaluating the extent of myometrial invasion, the reliability of the results can vary depending on the individual radiologist's proficiency and familiarity with the specific imaging characteristics of endometrial cancer[43]. Furthermore, a gynecologist with expertise in gynecological diagnostics, specifically dedicated to the study of gynecological oncology, is not a professional figure available in all medical centers, but only in tertiary care centers[38]. To overcome these limitations, radiomics, which extracts quantitative image features from medical imaging, has been proposed to predict myometrial infiltration[43].

In this work, we propose two different machine learning models based on the extraction of radiomics features from diagnostic ultrasound images using pre-trained CNNs [45–46]. Following the approach widely proposed in the literature on magnetic resonance images, the first model aims to solve the broader problem of classification of myometrial depth, i.e. infiltration less than 50% vs. infiltration greater than 50%[14–19]. However, as previously anticipated, the therapeutic approach is also influenced by the infiltration depth of the infiltration class less than 50%, i.e. whether the infiltration is absent or contained (focal) or is more extensive within the 50% limit. In this context, this work proposes a second multiclass classification model, which provides the third class with focal or absent infiltration. The binary classification model predictive of the depth of myometrial infiltration proved to be highly performing with an AUC of 90.76 (88.63–92.89 CI95%). In particular, the proposed model exceeds the operator's classification performance, especially with reference to precision (86.54% vs. 73.53%), thus demonstrating greater reliability of the prediction provided.

In our study, we have evaluated the subjective assessment supported by automated prediction model to evaluate the real benefit of using a support decision system in the clinal practice. We have verified that the

Paper	Image	Sample size	Methods	Performance
Xiong, L. et al. (2023)[14]	MRI	154	Ellipse fitting algorithm	AUC: 89% Accuracy: 87%
Stanzione, A. (2021)[15]	MRI	54	PyRadiomics – Random forest	AUC: 89% Accuracy: 87%
Chen, X. (2020)[16]	MRI	530	DL network	AUC: 89% Accuracy: 85%
Zhun, X. (2021)[17]	MRI	79	Geometric and texture features – SVM classifier	AUC: 92% Accuracy: 94%
Ueno, Y. (2017)[18]	MRI	137	Texture feauters – RF classifier	AUC: 84% Accuracy: 81%
Liu, Xiaoling, et al. (2024)[19]	US	604	EfficientNet-B6	AUC: 81% Accuracy: 80%
Our proposal (*mod1*)	US	77	Pre-trained Inception-V3 – SVM classifier	AUC: 91% Accuracy: 89%

Table 4. The state-of-the-art performances achieved in previous works about DMI prediction by radiomic-based binary models.

operator's classification performance increased significantly both in terms of overall accuracy and in terms of per-class precision in identifying the intermediate class of the multiclass task. An accurate prediction of myometrial invasion in patients with EC ensures the best modulation of patient treatment, allowing the evaluation of whether fertility-sparing or ovarian conservation treatments are feasible in younger women and avoiding overtreatment in patients undergoing radical surgery[29]. Although histological examination remains the absolute truth, myometrial invasion should not be assessed by frozen section because of poor reproducibility and agreement with definitive paraffin Sect[29]. Therefore, during the staging phase, this challenging role is entrusted to imaging techniques[38].

Table 4summarizes the main works proposed in the literature. To our knowledge, the literature is poor in evaluating the predictive power of radiomics on ultrasound images in predicting deep myometrial invasion. In[19], the authors have validated the three pretrained DL models (EfficientNet-B6, EfficientNet-B0, and ResNet-50) to determine the deep myometrial invasion (infiltration less than 50% vs. greater than 50%) on the US image of the EC. In their experimental study, EfficientNet-B6 showed the best performance in the testing set in terms of area under the curve (AUC) and accuracy reached 81.40% (74.60–88.20 95%CI) and 80.20 (73.30–85.50 95%CI). Although the performances declared by the authors are 9% points lower than those achieved by our model, it should be underlined that the case study in[19] is multicentric and more numerous than ours.

. Most of the studies proposed in the literature on the same research task have used MRI. Numerous studies have shown that the two modalities have comparable diagnostic efficacy; however, compared with MRI, ultrasound is cheaper and takes less time to perform. In addition, ultrasound is a first-level examination, whereas MRI is a second- or third-level in-depth examination. The preliminary results shown in this article are comparable to those obtained in state-of-the-art works on MR images. Clearly given the different nature of the images used and the models implemented to solve the task of interest, the comparison is purely qualitative.

Further effort has been made to provide a model that can predict the depth of myometrial infiltration with a greater level of detail. We have developed a multiclass model that allows us to accurately identify cases with absent or focal infiltration already during the diagnosis phase. To the best of our knowledge, our work represents the first attempt to evaluate the predictive power of radiomics on ultrasound images in predicting multiclass deep myometrial invasion.

The proposed multiclass model is highly performing with an overall accuracy of value of 84.93% and a good balance of all per-class metrics, outperforming the operator's performance especially in terms of recall. Although the results obtained from the study are highly encouraging, the study has limitations related to the sample size. In fact, further validation studies and model optimizations are needed to improve performance and imagine real clinical applicability. In particular, evaluation studies should be performed on a larger dataset to ensure generalizability. In addition, the inclusion of other operators for the evaluation of the same case was not considered in this preliminary work. Hovewer, US scans may vary depending on the operator, the device used, and the acquisition settings, affecting the reliability of the model; therefore, specific studies on the stability of performance by varying not only the operator but also the device are needed. The proposed models can represent a valid support tool for the clinicians for an accurate estimate of the deep myometrial infiltration already in the diagnosis phase with an easily accessible, low-cost examination that causes no discomfort for the patient such as ultrasound.

Moreover, in this paper we use Inception-V3. It has an architecture with Inception modules, which reduce the number of parameters compared to deeper networks such as ResNet-152, DenseNet, VGG-19. It offers a good balance between accuracy and inference speed, which is useful in clinical scenarios where rapid diagnosis is crucial. However, future studies include the evaluation of other architectures, such as EfficientNet or similary, that for the same performance could offer an even lower computational cost.

Furthermore, although these models have the potential to standardize examinations based on subjective evaluations such as ultrasound, to be applied correctly, it remains essential to properly perform the imaging examination, have an adequate knowledge of the most appropriate methodology of execution by the operators, and possess the ability to use the instrument, in this case, the ultrasound machine. Automated artificial

intelligence models can improve the reproducibility of diagnostic examinations over subjective assessments, ensuring more reliable analysis based on standardized data. In addition, they can be decision support tools for less experienced clinicians, accelerating the learning curve and increasing diagnostic confidence. This is to ensure that the analyzed images are truly representative of the carcinoma, and it is not overestimated or underestimated with subsequent repercussions on management [45–46].

Data availability

The data presented in this study are available on request from the corresponding author. The data are not publicly available because are proper of the Azienda Ospedaliera Policlinico Consorziale-University of Bari.

Received: 10 December 2024; Accepted: 2 May 2025
Published online: 07 May 2025

References

1. World Health Organization. : Estimated cancer incidence, mortality and prevalence worldwide in 2012. (2015).GLOBOCAN (2012).
2. Colombo, N. et al. ESMO-ESGO-ESTRO consensus conference on endometrial cancer: diagnosis, treatment and follow-up. *International Journal of Gynecologic Cancer* 26.1 (2016).
3. Creasman, W. T. et al. Surgical pathologic spread patterns of endometrial cancer: a gynecologic oncology group study. *Cancer* **60. S8**, 2035–2041 (1987).
4. Chattopadhyay, S. et al. Tumour-free distance from Serosa is a better prognostic indicator than depth of invasion and percentage myometrial invasion in endometrioid endometrial cancer. *BJOG: Int. J. Obstet. Gynecol.* **119** (10), 1162–1170 (2012).
5. Ozbilen, O. et al. Comparison of myometrial invasion and tumor free distance from uterine Serosa in endometrial cancer. *Asian Pac. J. Cancer Prev.* **16** (2), 519–522 (2015).
6. Berek, J. S. et al. FIGO staging of endometrial cancer: 2023. *Int. J. Gynecol. Obstet.* **162** (2), 383–394 (2023).
7. de Biase, D. et al. Integrated clinicopathologic and molecular analysis of endometrial carcinoma: prognostic impact of the new ESGO-ESTRO-ESP endometrial cancer risk classification and proposal of histopathologic algorithm for its implementation in clinical practice. *Front. Med.* **10**, 1146499 (2023).
8. Cho, K. R. et al. International society of gynecological pathologists (ISGyP) endometrial cancer project: guidelines from the special techniques and ancillary studies group. *Int. J. Gynecol. Pathol.* **38**, S114–S122 (2019).
9. Cubo-Abert, M. et al. Diagnostic performance of transvaginal ultrasound and magnetic resonance imaging for preoperative evaluation of low-grade endometrioid endometrial carcinoma: prospective comparative study. *Ultrasound Obstet. Gynecol.* **58** (3), 469–475 (2021).
10. Antonsen, S. et al. MRI, PET/CT and ultrasound in the preoperative staging of endometrial cancer—a multicenter prospective comparative study. *Gynecol. Oncol.* **128** (2), 300–308 (2013).
11. Singh, N. et al. Pathologic prognostic factors in endometrial carcinoma (other than tumor type and grade). *Int. J. Gynecol. Pathol.* **38**, S93–S113 (2019).
12. Shrestha, P. et al. A systematic review on the use of artificial intelligence in gynecologic imaging–background, state of the art, and future directions. *Gynecol. Oncol.* **166** (3), 596–605 (2022).
13. Lecointre, L. et al. Artificial intelligence-based radiomics models in endometrial cancer: A systematic review. *Eur. J. Surg. Oncol.* **47** (11), 2734–2741 (2021).
14. Xiong, L. et al. A computer-aided determining method for the myometrial infiltration depth of early endometrial cancer on MRI images. *Biomed. Eng. Online.* **22** (1), 103 (2023).
15. Stanzione, A. et al. Deep myometrial infiltration of endometrial cancer on MRI: a radiomics-powered machine learning pilot study. *Acad. Radiol.* **28** (5), 737–744 (2021).
16. Chen, X. et al. Deep learning for the determination of myometrial invasion depth and automatic lesion identification in endometrial cancer MR imaging: a preliminary study in a single institution. *Eur. Radiol.* **30**, 4985–4994 (2020).
17. Zhu, X. et al. Detection of deep myometrial invasion in endometrial cancer MR imaging based on multi-feature fusion and probabilistic support vector machine ensemble. *Comput. Biol. Med.* **134**, 104487 (2021).
18. Ueno, Y. et al. Endometrial carcinoma: MR imaging–based texture model for preoperative risk stratification—a preliminary analysis. *Radiology* 284.3 : 748–757. (2017).
19. Liu, X. et al. A transvaginal ultrasound-based deep learning model for the noninvasive diagnosis of myometrial invasion in patients with endometrial cancer: comparison with radiologists. *Acad. Radiol.* **31** (7), 2818–2826 (2024).
20. Leone, F. P. G. et al. Terms, definitions and measurements to describe the sonographic features of the endometrium and intrauterine lesions: a consensus opinion from the international endometrial tumor analysis (IETA) group. *Ultrasound Obstet. Gynecology: Official J. Int. Soc. Ultrasound Obstet. Gynecol.* **35** (1), 103–112 (2010).
21. Education and Practical Standards Committee. Minimum training recommendations for the practice of medical ultrasound. *Ultraschall in der Medizin (Stuttgart, Germany*: 27.1 (2006): 79–105. (1980).
22. Pineda, L. et al. Agreement between preoperative transvaginal ultrasound and intraoperative macroscopic examination for assessing myometrial infiltration in low-risk endometrioid carcinoma. *Ultrasound Obstet. Gynecol.* **47** (3), 369–373 (2016).
23. Criminisi, A., Pérez, P. & Toyama, K. Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Process.* **13**, 1200–1212 (2004).
24. Bornemann, F., März, T. & Folkmar, and Fast image inpainting based on coherence transport. *J. Math. Imaging Vis.* **28**, 259–278 (2007).
25. LeCun, Y., Bengio, Y. & Hinton, G. *Deep Learn. Nat.* **521**.7553 : 436–444. (2015).
26. Russakovsky, O. et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vision*. **115**, 211–252 (2015).
27. Bera, K. et al. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nat. Reviews Clin. Oncol.* **19** (2), 132–146 (2022).
28. Szegedy, C. et al. Rethinking the inception architecture for computer vision. *Proceedings of the IEEE conference on computer vision and pattern recognition*. (2016).
29. Concin, N. et al. ESGO/ESTRO/ESP guidelines for the management of patients with endometrial carcinoma. *Int. J. Gynecologic Cancer* **31**.1. (2021).
30. Ding, C. Minimum redundancy feature selection from microarray gene expression data. *J. Bioinform. Comput. Biol.* **3** (02), 185–205 (2005).
31. Burger, C. J. C. *A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery* (WORKSHOP ON DATA MINING AND KNOWLEDGE DISCOVERY, 1998).
32. He, H. et al. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*. Ieee, (2008).

33. Gu, X., Angelov, P. P. & Eduardo, A. Soares. A self-adaptive synthetic over-sampling technique for imbalanced classification. *Int. J. Intell. Syst.* **35** (6), 923–943 (2020).
34. Youden, W. J. Index for rating diagnostic tests. *Cancer* 3.1 : 32–35. (1950).
35. Alcázar, G. et al. Transvaginal ultrasound versus magnetic resonance imaging for preoperative assessment of myometrial infiltration in patients with endometrial cancer: a systematic review and meta-analysis. *Journal of gynecologic oncology* 28.6 (2017).
36. Christensen, J. W. et al. Assessment of myometrial invasion in endometrial cancer using three-dimensional ultrasound and magnetic resonance imaging. *Acta Obstet. Gynecol. Scand.* **95** (1), 55–64 (2016).
37. Alcazar, J. et al. Transvaginal/transrectal ultrasound for assessing myometrial invasion in endometrial cancer: a comparison of six different approaches. *J. Gynecologic Oncol.* **26** (3), 201 (2015).
38. Madár, I. et al. Diagnostic accuracy of transvaginal ultrasound and magnetic resonance imaging for the detection of myometrial infiltration in endometrial cancer: A systematic review and Meta-Analysis. *Cancers* **16** (5), 907 (2024).
39. Van den Bosch, T. et al. Terms, definitions and measurements to describe sonographic features of myometrium and uterine masses: a consensus opinion from the morphological uterus sonographic assessment (MUSA) group. *Ultrasound Obstet. Gynecol.* **46** (3), 284–298 (2015).
40. Van den Bosch, T. et al. Sonographic classification and reporting system for diagnosing adenomyosis. *Ultrasound Obstet. Gynecology: Official J. Int. Soc. Ultrasound Obstet. Gynecol.* **53** (5), 576–582 (2019).
41. Okcu, O. et al. Prognositc significance of microcystic elongated and Fragmanted (MELF) myometrial invasion pattern: A retrospective study. *Medeni Med. J.* **37**, 212–219 (2022).
42. Van den Bosch, T. Ultrasound in the diagnosis of endometrial and intracavitary pathology: an update. *Australasian J. Ultrasound Med.* **15** (1), 7–12 (2012).
43. Dong, H. C. et al. Using deep learning with convolutional neural network approach to identify the invasion depth of endometrial cancer in myometrium using MR images: a pilot study. *International journal of environmental research and public health* 17.16 : 5993. [45] Massafra, Raffaella, et al. Robustness evaluation of a deep learning model on sagittal and axial breast DCE-MRIs to predict pathological complete response to neoadjuvant chemotherapy. *Journal of Personalized Medicine* 12.6 (2022): 953.[46] Bove, Samantha, et al. A CT-based transfer learning approach to predict NSCLC recurrence: The added-value of peritumoral region. *PLoS One* 18.5 (2023): e0285188. (2020).

Author contributions

Conceptualization, F.A., A.F., and R.M.; methodology, A.F., S.B., M.C.C., and R.M.; software, A.F.; validation, A.F., F.A. and R.M.; formal analysis, F.A., A.F., C.Co., R.M., and V.L.; resources, G.Co., R.M., V.L.; data curation, F.A., R.M., M.G., G. L., G.Ca., and E.S.; writing—original draft preparation, F.A., A.F., R.M., S.B., M.C.C., and R.M.; writing—review and editing, F.A., A.F., R.M., E.C., S.B., M.C.C., M.G., G.L., S.L., G.Ca., E.S., El.V., En.V., G.Co., R.M, and V.L.; supervision, F.A, A.F., G.Co., R.M. and V.L. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by funding from the Italian Ministry of Health "5 per 1000" project (Deliberation n. 655/2022).

Declarations

Competing interests

The authors declare no competing interests.

Conflicts of Interest

The authors declare no conflict of interest.

Institutional review board statement

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of the Azienda Ospedaliera Policlinico Consorziale - University of Bari, Italy (protocol code n. 6398/2020).

Informed consent statement

Written informed consent for participation was required for this retrospective observational study.

Author disclaimer

The authors affiliated with IRCCS Istituto Tumori "Giovanni Paolo II", Bari are responsible for the views expressed in this article, which do not necessarily represent the ones of the institute.

Additional information

Correspondence and requests for materials should be addressed to S.B. or M.C.C.