



HAL
open science

Artificial intelligence-based radiomics models in endometrial cancer: A systematic review

Lise Lecointre

► **To cite this version:**

Lise Lecointre. Artificial intelligence-based radiomics models in endometrial cancer: A systematic review. *Journal of Surgical Oncology*, 2021, 24;S0748-7983(21)005, 10.1016/j.ejso.2021.06.023 . hal-03398184

HAL Id: hal-03398184

<https://hal.science/hal-03398184v1>

Submitted on 5 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1 **Title: Artificial Intelligence-based radiomics models in endometrial cancer: A systematic**
2 **review**

3

4 **Journal:** European Journal of Surgical Oncology

5 **Article type:** Systematic review

6

7 **Authors:** Lise Lecointre^{1,2,3}, Jérémy Dana^{3,4}, Massimo Lodi¹, Chérif Akladios¹, Benoît
8 Gallix^{2,3,5}

9

10 **Affiliations:**

11 ¹ Department of Gynecologic Surgery, Hôpitaux Universitaires de Strasbourg, Strasbourg,
12 France.

13 ² I-Cube UMR 7357 - Laboratoire des sciences de l'ingénieur, de l'informatique et de
14 l'imagerie, Université de Strasbourg, Strasbourg, France

15 ³ Institut hospitalo-universitaire (IHU), Institute for Minimally Invasive Hybrid Image-Guided
16 Surgery, Université de Strasbourg, Strasbourg, France

17 ⁴ Inserm U1110, Institut de Recherche sur les Maladies Virales et Hépatiques, Strasbourg,
18 France

19 ⁵ Department of Diagnostic Radiology, McGill University, Montreal, Canada

20

21 Corresponding author

22 Lise Lecointre MD

23 Work address: 1 avenue Molière, 67200 Strasbourg

24 Email: lise.lecointre@chru-strasbourg.fr

26 **Abstract**

27 *Background.* Radiological preoperative assessment of endometrial cancer (EC) is in some
28 cases not precise enough and its performances improvement could lead to a clinical benefit.

29 Radiomics is a recent field of application of artificial intelligence (AI) in radiology.

30 *Aims.* To investigate the contribution of radiomics on the radiological preoperative
31 assessment of patients with EC; and to establish a simple and reproducible AI Quality Score
32 applicable to Machine Learning and Deep Learning studies.

33 *Methods.* We conducted a systematic review of current literature including original articles
34 that studied EC through imaging-based AI techniques. Then, we developed a novel Simplified
35 and Reproducible AI Quality score (SRQS) based on 10 items which ranged to 0 to 20 points
36 in total which focused on clinical relevance, data collection, model design and statistical
37 analysis. SRQS cut-off was defined at 10/20.

38 *Results.* We included 17 articles which studied different radiological parameters such as
39 deep myometrial invasion, lympho-vascular space invasion, lymph nodes involvement, etc.
40 One article was prospective, and the others were retrospective. The predominant technique
41 was magnetic resonance imaging. Two studies developed Deep Learning models, while the
42 others machine learning ones. We evaluated each article with SRQS by 2 independent
43 readers. Finally, we kept only 7 high-quality articles with clinical impact. SRQS was highly
44 reproducible (Kappa = 0.95 IC 95% [0.907-0.988]).

45 *Conclusion.* There is currently insufficient evidence on the benefit of radiomics in EC.
46 Nevertheless, this field is promising for future clinical practice. Quality should be a priority
47 when developing these new technologies.

48 **Keywords:** endometrial carcinoma, imaging, radiomics, machine learning, deep learning,
49 artificial intelligence.

50 **Introduction**

51 Endometrial cancer (EC) is the first diagnosed gynecological cancer in the United
52 States, accounting approximately 61,880 new cases and 12,160 deaths in 2019 [1]. Incidence
53 appears to be increasing due to the population aging and the rise in obesity rates, two main
54 risk factors of EC [1]. The radiological preoperative assessment (pelvic magnetic resonance
55 imaging and endometrial biopsy) aims to stage the disease and evaluate the risk of
56 recurrence, and therefore to determine the appropriate therapeutic management, in
57 particular for lymph node staging during initial surgery.

58 However, this preoperative radiological assessment is not precise enough.
59 Radiological staging may underestimate disease extension and provide inaccurate
60 assessment of lymph node status. Moreover, most of the prognostic elements of the
61 International Federation of Gynecology and Obstetrics (FIGO) classification can only be
62 obtained after final histological examination of surgical specimens.

63 Therefore, surgery remains a key element in the management of endometrial cancer.
64 Its two main objectives are on one hand to perform ablation of the primitive tumor, and on
65 the other to precisely stage the disease and evaluate its prognosis factors. While the first
66 objective may be obtained with a “simple” hysterectomy, the latter requires much more
67 extensive intervention, including a complete omentectomy, pelvic lymphadenectomy and
68 lumbo-aortic lymphadenectomy [2, 3], whose therapeutic benefit remains controversial.
69 Patients with endometrial carcinoma are fragile and performing aggressive surgery may lead
70 to severe complications of critical concern in obese, older and fragile patients. There is
71 therefore a need of maximizing pre-operative diagnostic performances. A better selection of
72 patients undergoing surgery would result in reducing overtreatment risk, morbidity and
73 mortality through personalized care.

74 Radiological assessment may benefit from a tumor characterization using radiomics.
75 They are reproducible and quantitative image features allowing the non-invasive
76 characterization of the tumor heterogeneity [4]. Artificial intelligence in radiology is a recent
77 discipline allowing high-throughput extraction of digital medical imaging data to obtain
78 predictive and/or prognostic information about patients and their pathologies by describing
79 tumor heterogeneity and indirectly molecular and genetic characteristics of the tumor. It
80 could allow prediction of diagnosis, treatment response and prognosis. This new field of
81 research is gaining momentum in oncology as its applications are wide and promising,
82 especially for clinical decision-making and personalized treatment [5]. There is a strong
83 correlation between radiomic data and clinical outcomes. This concept has already proven
84 its effectiveness in the preoperative prediction of different solid tumors [6], including lung,
85 breast and colorectal cancers [7, 8].

86 The approach of AI differs from the conventional radiological method as it is an
87 automatic, reproducible and quantitative analysis of images that can go beyond the human
88 eye. AI algorithms can be trained to analyze either pre-determined parameters (*i.e.*, machine
89 learning), such as tumor size, tumor shape, lymph nodes, etc., or without human supervision
90 (*i.e.*, deep learning) with a free analysis chain that may not be clearly explainable for human
91 intelligence. An example of free analysis chain is the artificial neural network, which is
92 basically circuits of functions that take images as input and produce analysis on the output.
93 Neural network can be more or less complex, according to the objective and the input type.
94 A neural network is called “deep” when it is formed of “hidden layers” in which the
95 information flows. The more hidden layers are present, the deeper the network is, the more
96 complex it is. Such a model is powerful enough to adjust perfectly to a specific training
97 dataset with a risk of resulting in poor performances on new information, which is called

98 “overfitting”. Therefore, different techniques of internal and external validation aim to avoid
99 this problem which tampers applicability of the algorithm. “Deep learning” refers to the use
100 of these deep neural networks.

101 Artificial Intelligence in Radiology is a promising field that may lead to a better
102 preoperative radiological assessment of patients with EC, and some literature have already
103 been published. However, because of its novelty, complexity and rapid evolution, assessing
104 the quality of this literature can be challenging. Therefore, our main objective was to study
105 the contribution of radiomics on the preoperative radiological assessment of patients with
106 EC through a systematic review. Our second objective was to establish a simple and
107 reproducible AI Quality Score applicable to Radiomics and Deep Learning studies.

108

109

110 **Methods**

111 *Literature Search Approach*

112 Medline (Pubmed) was searched on October 10, 2020, by two independent reviewers (LL,
113 JD) using the following keywords: endometrial carcinoma, imaging, radiomics, machine
114 learning, deep learning, artificial intelligence. Original articles were considered if they
115 studied endometrial carcinoma through imaging-based AI techniques. Non-English articles,
116 editorials or review articles were not included. Inclusion of studies was made in consensus.

117

118 *Data extraction*

119 Data extracted included first author name, year of publication, main objective, study design
120 (retrospective or prospective, mono or multicentric), imaging technique used, sample size
121 (training, validation and testing), model type (Machine learning or Deep Learning), classifier
122 and diagnostic performance metrics.

123

124 *Quality assessment using a Simplified and Reproducible AI Quality Score (SRQS)*

125 Assessing radiology research on artificial intelligence (AI) may be complex. Scores have been
126 proposed to evaluate the quality of AI studies. The Radiomics Quality Score (RQS) proposed
127 by Lambin *et al.* intended to provide guidance and evaluation criteria to readers [5].
128 However, this score presents some limitations. It is complex to be used in routine, doesn't
129 apply to Deep Learning Research, and doesn't include clinical relevance. Based on published
130 guidelines for radiomics to ensure standardization, reproducibility and high-quality of
131 research articles [9, 10], we modified the RQS into a simplified and reproducible score fitted
132 to Radiomics and Deep Learning studies. This score (SRQS), based on 10 items and ranging
133 from 0 to 20 points, was designed to focus on essential steps and discriminate robust AI

134 research articles (Figure 1). Points were awarded based on the presence of each of following
135 criteria. First, this score acknowledges the clinical impact and relevance of the study (2
136 points). In a second step, specific details of data acquisition are evaluated: 5 points are
137 granted if it was prospective, 2 points if image acquisition was performed according to an
138 up-to-date and standardized imaging protocol, 2 points if the segmentation was performed
139 automatically, and 2 points if the authors considered the pathological diagnosis as the gold
140 standard. Thirdly, the score focuses on the AI model design, i.e., training and testing details.
141 Specific criteria in the training section are proposed to echo each other between Radiomics
142 and Deep Learning with an allocation of 3 points if the following criteria were present.
143 Regarding Radiomics, the model should be IBSI-compliant (Image Biomarker Standardization
144 Initiative [11]). Extraction and selection (correlation and redundance) of features and
145 classifiers should be detailed. Regarding Deep Learning, neural network architecture
146 (number of hidden layers, nodes, activation functions...), initialization of model parameters
147 (random or transfer learning from a pre-trained network) and training approach (data
148 augmentation, hyperparameters) should be detailed. As mentioned, testing a model,
149 preferably using an independent and external dataset, is mandatory to ensure its
150 generalization. The absence of a test cohort was sanctioned in our score with the withdrawal
151 of 5 points because it is the major step in the development of an AI model. Diagnostic
152 performance of the model should be evaluated using reliable performance metrics such as
153 the area under the curve, sensitivity or specificity to be granted 2 points.

154 Each article was evaluated by two independent readers using this score. The threshold of the
155 SRQS for sufficient quality was set at 10/20. We also evaluated the reliability of our
156 simplified RQS by performing a Cohen's Kappa test.

157 **Results**

158 **Articles selected**

159 Seventeen articles [12-28] were included in this review study (Table 1). The clinical
160 objectives of these articles were to classify cancer severity according to: deep myometrial
161 invasion (n=5) [14, 16, 19, 22, 25], lympho-vascular space invasion (n=3) [12, 16, 24], lymph
162 nodes involvement (n=4) [18, 19, 23, 27], “high-risk” endometrial carcinoma (n=2) [20, 28],
163 histologic grade (n=5) [12, 13, 16, 17, 21], high-risk histological subtype (n = 1) [19];
164 endometrial precancerous lesion versus early-stage carcinoma (n=1) [26], benign versus
165 malignant endometrial lesion (n=1) [15], recurrence-free survival (n=1) [17], cervical stroma
166 invasion (n=1) [19].

167 Of the 17 studies, only 2 were multicentric [18, 28]. None was prospective. The predominant
168 technique was Magnetic Resonance Imaging (n=14) [12-19, 21, 22, 24, 25, 27, 28]. Positron
169 emission tomography computed tomography (n=2) [23, 26] and contrast-enhanced CT (n=1)
170 [20] were the two other modalities studied. Only two studies developed a Deep Learning-
171 based model [14, 22]. The other 15 studies employed radiomics-based models [12, 13, 15-
172 21, 23-28].

173

174 **Quality Assessment**

175 Quality assessment is shown in Table 2. Mean SRQS ranged from 0 to 14. The median score
176 was 4.

177 The highest SRQS (14/20) was for the article studying high risk group classification which
178 included different clinical and radiological parameters [28]. Absence of testing dataset and
179 retrospective design of the study were the most decisive criteria. Based on this analysis, we
180 have selected seven articles [22-28] that demonstrated sufficient quality (> 10/20):

181 myometrial invasion (n = 2) [22, 25], lympho-vascular space invasion (n = 1) [24], lymph
182 nodes involvement (n = 2) [23, 27], high-risk endometrial cancer (n = 1) [28]and on
183 differential diagnosis between endometrial precancerous lesion and early-stage carcinoma
184 (n = 1) [26].

185

186 Article analysis

187 Only the results of the 7 articles of sufficient quality were analyzed [22-28]. Performance
188 metrics of these seven articles are summarized in Table 3.

189 Chen [22] and Stanzione's [25] articles, both published in 2020, studied myometrial
190 invasion by endometrial carcinoma. Both were retrospective studies based on MRI images.
191 The study by Chen *et al.* included 530 patients including 138 in the testing dataset and aimed
192 to determine the diagnostic performance of a deep learning model in assessing myometrial
193 invasion. This model demonstrated higher accuracy than radiologists (84.8% *versus* 78.3%)
194 based on pathology examination, but the best results were observed when both were
195 associated (86.2%). However, this study had limitations, including an unbalanced database
196 with only 18 cases of deep invasion ($\geq 50\%$) and 120 cases of shallow invasion ($< 50\%$) in the
197 testing dataset. The authors concluded that their model was effective and could be used in
198 clinical practice by radiologists [22].

199 The article by Stanzione *et al.* aimed to measure the performance of a learning
200 machine model to detect deep myometrial invasion. They included 54 patients (of which
201 20% were in the testing dataset and 80% in the training dataset). They found an accuracy of
202 92% *versus* 82% for unaided radiologists, and an accuracy of up to 100% (n = 11) when the
203 radiologist was assisted by the algorithm. The main limitation of these results was the low
204 number of patients. Nevertheless, this study demonstrated the feasibility of this model and

205 its usefulness in improving radiologists' performance and reducing inter-individual variability
206 [25].

207 The article by Luo *et al.*, published in 2020, was interested in evaluating lympho-
208 vascular space invasion in preoperative imaging, as it is an important prognostic parameter
209 currently only accessible after post-operative pathological examination. The authors
210 performed a retrospective study based on MRI and developed a nomogram based on clinical
211 features and a radiomics score. They included 144 patients of which 43 were in the testing
212 dataset. The nomogram had a sensitivity of 94%, a specificity of 78.6%, and an area under
213 the curve (AUC) of 0.807. However, this study did not have a dataset validation, thus
214 exposing itself to the risk of overfitting and poorer diagnostic performance on the testing
215 dataset. These results could be used to predict preoperative lympho-vascular space invasion
216 and thus possibly adapt the therapeutic sequence for endometrial cancer [24].

217 Crivellaro *et al.* and Xu *et al.* studied lymph node involvement in endometrial cancer.
218 Crivellaro *et al.* published a retrospective study of 167 patients with early-stage endometrial
219 cancer in 2020. They evaluated the role of the 18FDG PET-scanner in predicting lymph node
220 involvement. They showed a moderate contribution in this configuration, especially since it
221 was based on a morphological and not a morphometabolic parameter, thus eventually
222 suggesting a better efficiency of MRI [23].

223 Xu *et al.* published in 2019 a retrospective study evaluating MRI for preoperative
224 prediction of metastatic lymph node involvement in 200 patients with endometrial cancer.
225 The algorithm was developed on 70% of the patients and tested on the remaining 30%. The
226 authors found an accuracy of about 85%, especially for intermediate sized nodes [27].

227 Yan *et al.* published in 2020 a retrospective multicenter study on a nomogram based
228 on clinical and MRI-based radiomic data to predict preoperatively high-risk endometrial

229 cancers (defined as the presence of at least one of deep myometrial invasion, grade 3 tumor,
230 lympho-vascular space invasion, cervical stroma invasion, lymph node metastasis, extra-
231 uterine invasion, or non-endometrioid adenocarcinoma) and therefore their optimal surgical
232 management. They included 717 patients and compared the therapeutic proposals based on
233 the nomogram with those taken for these patients. They found on one hand good
234 preoperative prediction rates (AUC between 0.896 and 0.919) and on the other hand a
235 benefit in terms of correction of clinical decisions, thus suggesting a possible application in
236 clinical practice for the choice of surgical treatment [28].

237 Finally, Wang *et al.* investigated the differential diagnosis between atypical
238 hyperplasia with field cancerization and early-stage endometrial cancer (IA). They used
239 radiomics data from PET-scans of 170 patients. Maximum SUV and peak SUV were the best
240 predictors of diagnosis, especially between field cancerization and invasive cancer [26].

241

242 Reliability of SRQS

243 The two independent readers were an oncology surgeon and a radiologist. There was an
244 excellent agreement between the two reviewer's assessment (Kappa = 0.95 IC 95% [0.907-
245 0.988]). There was almost no variability between the two readers, except for the item
246 "clinical outcome improving patient management" in 4 articles [17, 19, 21, 24].

247

248

249 **Discussion**

250 We conducted a systematic review of the currently available data on the impact of
251 radiomics on preoperative imaging assessment in endometrial cancer. There is a need for a
252 non-invasive, efficient and reproducible method of stratification of the risk of recurrence to
253 adapt the therapeutic management of EC. Most of research studies were MRI-based, the
254 imaging modality of reference, and focusing on relevant clinical issues [14, 16-18, 22-24, 27,
255 28]. They demonstrated strong diagnostic performance in various objectives such as deep
256 myometrial invasion, lympho-vascular space invasion, lymph nodes involvement or
257 histologic grade. However, only few articles matched quality requirements [22-28], and none
258 provided a tool with a short-term applicability in clinical practice. We are therefore
259 compelled to consider that to date, we are only at the beginning of the development of this
260 field. Quality requirements should be of critical concern for researchers and clinicians in this
261 domain. For this reason, we developed a novel quality assessment score which avoids the
262 pitfalls of those already available. Nonetheless, radiomics is a very promising field and its
263 application in endometrial cancer could change our future practices.

264

265 **Quality issues**

266 First, we noted a certain degree of heterogeneity among methodologies. This
267 heterogeneity is due to a lack of standardization, and this seems to be true not only for
268 mathematical methodology, but also for analyzed data.

269 Secondly, most of the included articles did not provide an independent testing of the
270 developed model preventing them from generalizability and external applicability. Indeed,
271 performances of AI algorithms can be impressive on training datasets due to overfitting.

272 While multicentric and prospective studies are necessary for accurately assessing impact on
273 clinical outcomes, none have been published.

274 Quality assessment using a Simplified AI Quality Score

275 Assessing radiology research on artificial intelligence (AI) may be complex. Scores
276 have been proposed to evaluate the quality of AI studies. The Checklist for Artificial
277 Intelligence in Medical Imaging (CLAIM) [9] was inspired from validated ones such as the
278 STARD-AI [29] and CONSORT-AI [30] checklists. The CLAIM intended to provide a writing and
279 reading map to ensure standardization, reproducibility and high-quality of research articles.
280 Similarly, Bluemke *et al.* had previously proposed the key considerations for authors,
281 reviewers, and readers, from the definition of image datasets to the explainability of AI
282 algorithms [10]. Moreover, the Radiomics Quality Score proposed by Lambin *et al.* intended
283 to provide guidance and evaluation criteria to readers [5]. However, this score presents
284 some limitations. It is complex to be used in routine, does not apply to Deep Learning
285 Research and does not include clinical relevance.

286 To fill with the described requirements, we modified the Radiomics Quality Score into
287 a novel Simplified and Reproducible AI Quality Score (SRQS) applicable to Radiomics and
288 Deep Learning studies based on 10 items and ranging from 0 to 20 points. This score was
289 designed to focus on essential steps and distinguish robust AI research articles. Its main
290 advantages are to be easy to use, reproducible and discriminant. First, we considered that
291 no study could be justified without a clinical substrate. Therefore, any study that did not
292 attempt to address a clinical issue was considered irrelevant. Secondly, we highlighted
293 details of dataset acquisition as selection biases are of critical concern, especially for AI
294 algorithms which can perfectly fit to a specific dataset. It is an understatement to say that
295 the prospective aspect of a study shields from severe bias. Thirdly, model development is

296 defined by the training and testing details. Training an AI model requires consistency and
297 robustness. Every detail should be provided to the reader to ensure reproducibility by
298 another research team. In a matter of reproducibility and standardization in the field of
299 “traditional” radiomics, the Image Biomarker Standardization Initiative, an independent
300 international collaboration, gave guidelines and definitions for the extraction of high-
301 throughput image biomarkers. Such definitions of image biomarkers cannot apply to the
302 field of Deep Learning as they are free of human intervention. As mentioned, testing a
303 model, preferably using an independent and external dataset, is mandatory to assure its
304 generalization. We stated in this score that testing a model was the major step in the
305 development of an AI model. Finally, statistical analysis – including performance metrics –
306 should be detailed and were therefore included in our score.

307

308 **Conclusion**

309 Although there is insufficient evidence of a benefit of radiomics in endometrial
310 cancer, preliminary data suggest that these new technologies – applied in conjunction with
311 human intelligence – can address some of the clinical issues. We found that some models
312 are already under development and show good performances. These models need to follow
313 high-quality standards and undergo external validation through prospective clinical trials. In
314 view of the rapid expansion of this domain, we expect to see in a short-term future new and
315 evidence-based models with applicability in clinical practice.

316

317

318

319

320

321 **Declaration of interests**

322 None

323 **Acknowledgements**

324 This work was supported by French state funds managed within the “Plan Investissements

325 d’Avenir” and by the ANR (reference ANR-10-IAHU-02)

327 **References**

- 328 [1] Colombo N, Creutzberg C, Amant F, Bosse T, Gonzalez-Martin A, Ledermann J, et al.
329 ESMO-ESGO-ESTRO Consensus Conference on Endometrial Cancer: diagnosis, treatment and
330 follow-up. *Ann Oncol.* 2016;27:16-41.
- 331 [2] Kwon JS, Carey MS, Goldie SJ, Kim JJ. Cost-effectiveness analysis of treatment strategies
332 for Stage I and II endometrial cancer. *J Obstet Gynaecol Can.* 2007;29:131-9.
- 333 [3] Berek JS, Hacker NF. *Practical gynecologic oncology.* Philadelphia: Lippincott Williams &
334 Wilkins; 2005.
- 335 [4] Savadjiev P, Chong J, Dohan A, Agnus V, Forghani R, Reinhold C, et al. Image-based
336 biomarkers for solid tumor quantification. *Eur Radiol.* 2019 Oct;29(10):5431-5440.
- 337 [5] Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al.
338 Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin*
339 *Oncol.* 2017;14:749-62.
- 340 [6] Sun R, Limkin EJ, Vakalopoulou M, Derclé L, Champiat S, Han SR, et al. A radiomics
341 approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1
342 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol.*
343 2018;19:1180-91.
- 344 [7] Dohan A, Gallix B, Guiu B, Le Malicot K, Reinhold C, Soyer P, et al. Early evaluation using a
345 radiomic signature of unresectable hepatic metastases to predict outcome in patients with
346 colorectal cancer treated with FOLFIRI and bevacizumab. *Gut.* 2020;69:531-9.
- 347 [8] Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a
348 Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal
349 Cancer. *J Clin Oncol.* 2016;34:2157-64.
- 350 [9] Mongan J., Moy L., Kahn C.E. Checklist for Artificial Intelligence in Medical Imaging
351 (CLAIM): A Guide for Authors and Reviewers. *Radiology: Artificial Intelligence.*
352 2020;2:e200029.
- 353 [10] Bluemke DA, Moy L, Bredella MA, Ertl-Wagner BB, Fowler KJ, Goh VJ, et al. Assessing
354 Radiology Research on Artificial Intelligence: A Brief Guide for Authors, Reviewers, and
355 Readers-From the Radiology Editorial Board. *Radiology.* 2020;294:487-9.
- 356 [11] Zwanenburg A, Vallieres M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The
357 Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-
358 Throughput Image-based Phenotyping. *Radiology.* 2020;295:328-38.
- 359 [12] Bereby-Kahane M, Dautry R, Matzner-Lober E, Cornelis F, Sebbag-Sfez D, Place V, et al.
360 Prediction of tumor grade and lymphovascular space invasion in endometrial
361 adenocarcinoma with MR imaging-based radiomic analysis. *Diagn Interv Imaging.*
362 2020;101:401-11.
- 363 [13] Chen T, Li Y, Lu SS, Zhang YD, Wang XN, Luo CY, et al. Quantitative evaluation of
364 diffusion-kurtosis imaging for grading endometrial carcinoma: a comparative study with
365 diffusion-weighted imaging. *Clin Radiol.* 2017;72:995 e11- e20.
- 366 [14] Dong HC, Dong HK, Yu MH, Lin YH, Chang CC. Using Deep Learning with Convolutional
367 Neural Network Approach to Identify the Invasion Depth of Endometrial Cancer in
368 Myometrium Using MR Images: A Pilot Study. *Int J Environ Res Public Health.* 2020;17.

369 [15] Kierans AS, Doshi AM, Dunst D, Popiolek D, Blank SV, Rosenkrantz AB. Retrospective
370 Assessment of Histogram-Based Diffusion Metrics for Differentiating Benign and Malignant
371 Endometrial Lesions. *J Comput Assist Tomogr.* 2016;40:723-9.

372 [16] Ueno Y, Forghani B, Forghani R, Dohan A, Zeng XZ, Chamming's F, et al. Endometrial
373 Carcinoma: MR Imaging-based Texture Model for Preoperative Risk Stratification-A
374 Preliminary Analysis. *Radiology.* 2017;284:748-57.

375 [17] Yamada I, Oshima N, Miyasaka N, Wakana K, Wakabayashi A, Sakamoto J, et al. Texture
376 Analysis of Apparent Diffusion Coefficient Maps in Cervical Carcinoma: Correlation with
377 Histopathologic Findings and Prognosis. *Radiology: Imaging Cancer.* 2020;2:e190085.

378 [18] Yan BC, Li Y, Ma FH, Zhang GF, Feng F, Sun MH, et al. Radiologists with MRI-based
379 radiomics aids to predict the pelvic lymph node metastasis in endometrial cancer: a
380 multicenter study. *Eur Radiol.* 2021;31:411-22.

381 [19] Ytre-Hauge S, Dybvik JA, Lundervold A, Salvesen OO, Krakstad C, Fasmer KE, et al.
382 Preoperative tumor texture analysis on MRI predicts high-risk disease and reduced survival
383 in endometrial cancer. *J Magn Reson Imaging.* 2018;48:1637-47.

384 [20] Ytre-Hauge S, Salvesen OO, Krakstad C, Trovik J, Haldorsen IS. Tumour texture features
385 from preoperative CT predict high-risk disease in endometrial cancer. *Clin Radiol.* 2021;76:79
386 e13-79 e20.

387 [21] Yue W, Meng N, Wang J, Liu W, Wang X, Yan M, et al. Comparative analysis of the value
388 of diffusion kurtosis imaging and diffusion-weighted imaging in evaluating the histological
389 features of endometrial cancer. *Cancer Imaging.* 2019;19:9.

390 [22] Chen X, Wang Y, Shen M, Yang B, Zhou Q, Yi Y, et al. Deep learning for the
391 determination of myometrial invasion depth and automatic lesion identification in
392 endometrial cancer MR imaging: a preliminary study in a single institution. *Eur Radiol.*
393 2020;30:4985-94.

394 [23] Crivellaro C, Landoni C, Elisei F, Buda A, Bonacina M, Grassi T, et al. Combining positron
395 emission tomography/computed tomography, radiomics, and sentinel lymph node mapping
396 for nodal staging of endometrial cancer patients. *Int J Gynecol Cancer.* 2020;30:378-82.

397 [24] Luo Y, Mei D, Gong J, Zuo M, Guo X. Multiparametric MRI-Based Radiomics Nomogram
398 for Predicting Lymphovascular Space Invasion in Endometrial Carcinoma. *J Magn Reson*
399 *Imaging.* 2020;52:1257-62.

400 [25] Stanzione A, Cuocolo R, Del Grosso R, Nardiello A, Romeo V, Travaglino A, et al. Deep
401 Myometrial Infiltration of Endometrial Cancer on MRI: A Radiomics-Powered Machine
402 Learning Pilot Study. *Acad Radiol.* 2020.

403 [26] Wang T, Sun H, Guo Y, Zou L. (18)F-FDG PET/CT Quantitative Parameters and Texture
404 Analysis Effectively Differentiate Endometrial Precancerous Lesion and Early-Stage
405 Carcinoma. *Mol Imaging.* 2019;18:1536012119856965.

406 [27] Xu X, Li H, Wang S, Fang M, Zhong L, Fan W, et al. Multiplanar MRI-Based Predictive
407 Model for Preoperative Assessment of Lymph Node Metastasis in Endometrial Cancer. *Front*
408 *Oncol.* 2019;9:1007.

409 [28] Yan BC, Li Y, Ma FH, Feng F, Sun MH, Lin GW, et al. Preoperative Assessment for High-
410 Risk Endometrial Cancer by Developing an MRI- and Clinical-Based Radiomics Nomogram: A
411 Multicenter Study. *J Magn Reson Imaging.* 2020;52:1872-82.

412 [29] Sounderajah V, Ashrafian H, Aggarwal R, De Fauw J, Denniston AK, Greaves F, et al.
413 Developing specific reporting guidelines for diagnostic accuracy studies assessing AI
414 interventions: The STARD-AI Steering Group. *Nature Medicine.* 2020;26:807-8.

415 [30] Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, Spirit AI, et al. Reporting guidelines
416 for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI
417 Extension. BMJ. 2020;370:m3164.
418

419
420
421
422
423
424
425
426
427
428
429
430
431
432

Clinical Impact and Relevance

- Clinical issue addressed **+2**

Data

- Prospective data **+5**
- Image Acquisition: protocol quality **+1**
- Automatic Segmentation **+1**
- Gold Standard Quality (e.g. pathology) **+2**

Model Design

Training Details **+3**

Machine Learning

- IBSI-Compliant
- Features : extraction and selection
- Details of training approach

Deep Learning

- Neural Network Architecture
- Initialization of model parameters (random or transfer learning)
- Details of training approach

Testing Details

External **+5**

Internal **+3**

Absent **-5**

Statistical Analysis

- Performance Metrics **+1**

Table 1: Characteristics of the studies included in the review

Author	Year of Publication	Objective	Study Design	Imaging technique	Sample size (n)			Model Type
					Training	Validation	Testing	
<i>Chen</i>	2020	Deep myometrial invasion	Retrospective Monocenter	MRI	313	79	138	Deep Learning
<i>Stanzione</i>	2020	Deep myometrial invasion	Retrospective Monocenter	MRI	43	0	11	Machine Learning
<i>Luo</i>	2020	Lympho-vascular space invasion	Retrospective Monocenter	MRI	101	0	43	Machine Learning
<i>Crivellaro</i>	2020	Lymph nodes involvement	Retrospective Monocenter	PET-CT	69	0	28	Machine Learning
<i>Xu</i>	2019	Lymph nodes involvement	Retrospective Monocenter	MRI	140	Cross-validation	60	Machine Learning
<i>Yan</i>	2020	High-risk endometrial carcinoma	Retrospective Multicenter	MRI	394	0	323	Machine Learning
<i>Wang</i>	2019	Endometrial precancerous lesion versus early-stage carcinoma	Retrospective Monocenter	PET-CT	170	0	Details are missing	Machine Learning
<i>Kierans</i>	2016	Benign versus malignant endometrial carcinoma	Retrospective Monocenter	MRI	54	0	0	Machine Learning
<i>Yue</i>	2019	Histologic grade	Retrospective Monocenter	MRI	91	0	0	Machine Learning
<i>Yamada</i>	2019	Histologic grade Recurrence-free survival	Retrospective Monocenter	MRI	121	Cross-validation	0	Machine Learning
<i>Chen</i>	2017	Histologic grade	Retrospective Monocenter	MRI	73	0	0	Machine Learning

Ueno	2017	Deep myometrial invasion Lympho-vascular space invasion Histologic grade	Retrospective Monocenter	MRI	137	Cross-validation	0	Machine Learning
Dong	2020	Deep myometrial invasion	Retrospective Monocenter	MRI	24	6	48 (including the 6 patients of the validation group)	Deep Learning
Bereby-Kahane	2020	Histologic grade Lympho-vascular space invasion	Retrospective Monocenter	MRI	73	Cross-validation	0	Machine Learning
Ytre-Hauge	2020	High-risk endometrial carcinoma	Retrospective Monocenter	CT	155	0	0	Machine Learning
Ytre-Hauge	2018	Deep myometrial invasion Cervical stroma invasion Lymph node involvement High-risk histological subtype	Retrospective	MRI	180	0	0	Machine Learning
Yan	2020	Lymph nodes involvement	Retrospective Multicentric	MRI	622	0	0	Machine Learning

Table 2: Quality assessment of the studies included in the review

Author, year	Reading	Data (/4)			Model Design									Statistical analysis	Clinical outcome improving patient management (0/2)	Total	
		Image protocol quality (0/1)	Automatic segmentation (0/1)	Gold standard (0/2)	Prospective study (0/5)	Testing (-5/3/5)*	If Machine Learning / 3			If Deep Learning / 3			Performance Metrics (0/1)			For each reading	R1/R2 mean
							Package IBSI-compliant (0/1)	Features (extraction, correlation, redundancy, selection...) (0/1)	Details of training approach (classifier) (0/1)	Neural network architecture detailed (0/1)	Initialization of parameters (randomized, transfer learning) (0/1)	Details of training approach (data augmentation, hyperparameters,...) (0/1)					
<i>Chen 2020</i>	R1	1	0	2	0	3	NA	NA	NA	1	1	0	1	2	11	11	
	R2	1	0	2	0	3	NA	NA	NA	1	1	0	1	2	11		
<i>Stanzione 2020</i>	R1	1	0	2	0	3	1	1	1	NA	NA	NA	1	0	10	10	
	R2	1	0	2	0	3	1	1	1	NA	NA	NA	1	0	10		
<i>Luo 2020</i>	R1	1	0	2	0	3	1	1	1	NA	NA	NA	1	0	10	11	
	R2	1	0	2	0	3	1	1	1	NA	NA	NA	1	2	12		
<i>Crivellaro 2020</i>	R1	1	0	2	0	3	1	1	1	NA	NA	NA	1	2	12	12	
	R2	1	0	2	0	3	1	1	1	NA	NA	NA	1	2	12		
<i>Xu 2019</i>	R1	1	0	2	0	3	1	1	1	NA	NA	NA	1	2	12	12	
	R2	1	0	2	0	3	1	1	1	NA	NA	NA	1	2	12		
<i>Yan 2020</i>	R1	1	0	2	0	5	1	1	1	NA	NA	NA	1	2	14	14	
	R2	1	0	2	0	5	1	1	1	NA	NA	NA	1	2	14		
<i>Wang 2019</i>	R1	1	1	2	0	3	1	1	1	NA	NA	NA	1	0	11	11	
	R2	1	1	2	0	3	1	1	1	NA	NA	NA	1	0	11		
<i>Kierans 2016</i>	R1	1	0	2	0	-5	0	0	0	NA	NA	NA	1	0	0	0	
	R2	1	0	2	0	-5	0	0	0	NA	NA	NA	1	0	0		

Yue 2019	R1	1	0	2	0	-5	1	0	0	NA	NA	NA	1	0	0	1
	R2	1	0	2	0	-5	1	0	0	NA	NA	NA	1	2	2	
Yamada 2019	R1	1	0	2	0	-5	0	1	1	NA	NA	NA	1	0	1	2
	R2	1	0	2	0	-5	0	1	1	NA	NA	NA	1	2	3	
Chen 2017	R1	1	0	2	0	-5	0	0	0	NA	NA	NA	1	0	0	0
	R2	1	0	2	0	-5	0	0	0	NA	NA	NA	1	0	0	
Ueno 2017	R1	1	0	2	0	-5	0	1	1	NA	NA	NA	1	2	3	3
	R2	1	0	2	0	-5	0	1	1	NA	NA	NA	1	2	3	
Dong 2020	R1	1	0	2	0	-5	NA	NA	NA	1	1	1	1	2	4	4
	R2	1	0	2	0	-5	NA	NA	NA	1	1	1	1	2	4	
Bereby- Kahane 2020	R1	1	0	2	0	-5	0	1	1	NA	NA	NA	1	0	1	1
	R2	1	0	2	0	-5	0	1	1	NA	NA	NA	1	0	1	
Ytre- Hauge 2020	R1	0	0	2	0	-5	0	0	0	NA	NA	NA	1	0	0	0
	R2	0	0	2	0	-5	0	0	0	NA	NA	NA	1	0	0	
Ytre- Hauge 2018	R1	1	0	2	0	-5	0	0	0	NA	NA	NA	1	0	0	0.5
	R2	1	0	2	0	-5	0	0	0	NA	NA	NA	1	2	1	
Yan 2020	R1	1	0	2	0	-5	1	1	1	NA	NA	NA	1	2	4	4
	R2	1	0	2	0	-5	1	1	1	NA	NA	NA	1	2	4	

