Scientific Impact Paper No. XX

Peer review draft – August 2024

Artificial Intelligence in Gynaecology Oncology

Saladin Sawan, Noushin Eftekhari, Kristofer Linton-Reid, Nicholas Wood, Tricia A. Numan, Eric O. Aboagye, Claudio Angione, on behalf of the Royal College of Obstetricians and Gynaecologists

Correspondence: Royal College of Obstetricians and Gynaecologists, 10-18 Union Street, London SE1 1SZ
Email: clinicaleffectiveness@rcog.org.uk.

Plain language summary

Artificial Intelligence (AI) is an emerging powerful technology that differs from traditional computer programs in its ability to learn from its results and enhance performance, mimicking human intelligence; hence the name. AI is already an important part of most computer-based tasks in our daily lives. Everyday examples include internet search engines, and products that provide face recognition or predict the outbreak of diseases.

Research interests in AI appear to be subjected to available preexisting information and datasets rather than addressing patients' priorities and clinical needs. The National Institute for Health and Care Excellence (NICE) in England noted that current medical technologies using AI lack robust research backing and NHS patient involvement.

While some AI-based products are currently in clinical use – for example, in identifying abnormal cells in cervical smears - AI remains largely in the research phase in gynaecology oncology. Researchers have reported good results of its performance in fields such as prediction of lymph node involvement in cervical, endometrial, and ovarian cancers, which are important for treatment planning, distinguishing benign from malignant pelvic masses, and cervical cancer screening in low and high-income countries.

There are ethical concerns surrounding the use of AI in health care. Many of these concerns relate to the quality of data used in training AI systems, i.e data should be inclusive so that results can be applicable in the future irrespective of race, ethnicity, socio-economic background or place of residence. It is also not clear who should take responsibility for clinical recommendation made by AI systems: is it the doctor using it, the hospital employing the doctor, or the creators of the AI product. Concerns have also been raised regarding how the roll out of AI might affect jobs for doctors, nurses and administrator staff and their families.

AI is expected to contribute to health care in many positive ways. This can be achieved with good scrutiny and appropriate legislations to protect patients' health and privacy in addition to identifying important research and implementation areas through a collaborative partnership among investors, investigators, clinicians, and patients.

This guidance is for healthcare professionals who care for women, non-binary and trans people. Within this document we use the terms woman and women's health. However, it is important to acknowledge that it is not only women for whom it is necessary to access women's health and reproductive services in order to maintain their gynaecological health and reproductive wellbeing.

Gynaecological and obstetric services and delivery of care must therefore be appropriate, inclusive and sensitive to the needs of those individuals whose gender identity does not align with the sex recorded at birth.

1 Introduction

The term artificial intelligence (AI) is believed to have been coined by John McCarthy et al at the Dartmouth Summer Research Project in 1956, when it was proposed that a machine can be made to simulate 'every aspect of learning or any other features of intelligence.[1] AI is a rapidly evolving field with expanding potentials that is increasingly becoming an integral part our daily lives. Every day examples include internet search engines, recommended posts on social media, financial sector forecast, disease outbreak modelling, defence and weaponry, and even the editing of medical articles.[2-4]

While there is no a universally agreed definition, AI can refer to a branch of informatics that engineer computer systems capable of performing tasks that typically require human intelligence such as reasoning, adaptation, and learning via feedback processes.[5,6] The National Institute for Health and Care Excellence (NICE) has noted that the exact definition of AI in healthcare could be context-dependent and that the extent of AI incorporation into digital health technologies could vary widely[7,8].

In AI, computer systems are built using algorithms, which are sets of mathematical instructions constructed by coding engineers, to uncover patterns and relationships among variables by mining and mapping data and then selecting the best model for a specified purpose.[6,9,10] Algorithms in AI are designed so they can learn and, hence, refine their own performance, unlike conventional algorithms used in traditional computing, which are engineered to follow predefined strict instructions and rules with no inherent capability for learning or performance improvement.[11] Generally, AI algorithms are trained on a dataset (called training data) then are tested to assess performance on another unseen dataset (testing data) prior to implementation on external or validation data. Typically, both training data and testing data are obtained from the same dataset, which is usually divided according to a specified ratio and allocation method.[6]

1.1 Machine learning

Machine learning (ML) is a subfield of AI that facilitates computer systems to enhance their performance in a given task without being programmed explicitly. Machine learning research aims to design algorithms-based models that can learn more efficiently from large and various datasets and examine their applications in multifarious domains.[3,12]

Data used in ML can be labelled or unlabelled. Labelled data comprise input variables (predictor features) which are associated with known outcome values (target values or labels). On the other hand, unlabelled data contains only input variables with no stated outcome values. For example, a dataset for ovarian cancer patients with patients' demographics and cancer characteristics (input variables) would be unlabelled data, unless the dataset also includes survival outcome (outcome value) where it would be labelled data.[6]

There are two types of ML; supervised learning and unsupervised learning. Supervised ML involves training an ML model on labelled data, it aims to learn a function to predict the accurate target value. Supervised ML has succeeded considerably in tasks, such as image recognition, speech recognition, natural language processing and autonomous driving, that would be challenging or unattainable with traditional programming techniques. However, some of the challenges of supervised ML include the need for large amounts of labelled data which could be time and expertise consuming, and the difficulty of handling noisy or ambiguous labels.[6,9,13] Unsupervised ML utilises unlabelled datasets for training to uncover interactions and relationships within the data to identify patterns underlying the data structure. Clustering, dimensionality reduction, and generative modelling are examples of

unsupervised ML. Unsupervised learning can be used to learn abstract and general data representations, and to map the data into compressed representations called embeddings, which retain most of the information of the original data. Autoencoders, generative adversarial networks, and self-organising maps are some of the primary methods for unsupervised learning.[6,14]

1.2 Deep learning

Deep learning (DL) is a subset of ML that uses multi-layered neural networks (NN) to generate complex data representations. DL models perform exceptionally well in domains with high-dimensional input data, such as images, videos, and texts. DL models can autonomously extract hierarchical features from input data and learn to classify, generate, or transform them by layering multiple levels of artificial neurons. DL research focuses on developing new architectures and optimisation techniques for NN and investigating their applications in computer vision, speech recognition, natural language processing, and robotics. Recent advances in DL, such as transformer models, generative adversarial networks (GANs) and diffusion models, have cultivated new ways of human-machine interaction leading to significant AI research breakthroughs.[15-17]

1.3 AI in Cancer research

In recent years, ML and DL have advanced healthcare research including cancer diagnosis, classification and prognosis. These technologies have provided researchers and clinicians with novel tools to further our understanding of the complex mechanisms involved in cancer development and to identify more effective targeted therapeutic options. ML algorithms, for example, can analyse extensive medical records, genetics, and other datasets to unveil patterns that human analysis finds difficult or impossible to recognise.

1.3.1 AI in medical imaging

One significant application of ML in cancer research is image analysis. DL algorithms can analyse medical images, such as X-rays, computerised tomography (CT) scans, and magnetic resonance (MR) images, to detect patterns and anomalies that may demonstrate the presence of cancer. This technology has shown promising results in improving the accuracy of cancer detection, which could lead to cancer diagnosis at early stages, with the opportunity for curative treatment and improved prognosis. DL models can also be employed to analyse medical images to track tumour growth and response to treatment over time.[16] The use of a cloud-based DL system in one institution resulted in significant savings in clinicians' time required for contouring volumes of interest (VOIs) of various organs even when taking into account the time required for correction. The median (range) time for manual VOIs delineation, DL-based segmentation, and subsequent manual corrections were 25.0 (8.0-115.0), 2.3 (1.2-8) and 10.0 minutes (0.3-46.3), respectively in images from 111 patients with various cancer, including female pelvis.[18]

1.3.2 AI in drug discovery

Another area where ML is making significant advances in cancer research in drug discovery. Traditional drug discovery processes are expensive, time-consuming, and often unsuccessful. ML algorithms can scrutinise large datasets of chemical compounds and their interactions with biological systems to identify potential drug candidates. As a result, this technology can substantially expedite the drug discovery process, leading to more effective cancer treatments in a shorter time. Recently, FDA has issued an investigational new drug (IND) clearance for the first time for an AI-generated drug: ISM3091, a ubiquitin-specific protease 1 (USP1) inhibitor.[19]

2 AI: Supporting evidence in health care

The contribution of AI in healthcare is widely celebrated on social and traditional media platforms. It is regarded as an example of good use and a positive role for AI in the face of growing concerns among AI experts regarding its governance in some other fields.[20] While the number of published peer-reviewed articles pertaining to AI in healthcare has increased exponentially in recent years[21], there has been limited robust evidence supporting the implementation of AI or AI-based devices in healthcare.

There are four Cochrane reviews addressing AI to date, all of which were in fields other than gynaecology.[22] NICE has produced Medtech Innovation Briefings (MIB) to advise NHS and social care commissioners when considering new medical technologies. NICE has issued eight MIBs addressing AI systems for all of which there were limited prospective studies and/or a lack of involvement of NHS patients.[23-30] It is worth noting that none of these MIBs were related to gynaecological cancers.

On the other hand, the United States Food and Drug Administration (FDA) has approved or cleared 692 AI-enabled medical devices: 547 (79%) devices were radiology based, while only one system was listed in the obstetrics and gynecology panel (KIDScore D3 for embryo selection).[31-2] The Conformité Européene (CE) mark is not centralised, unlike FDA, and hence there is no readily accessible list of CE-marked AI systems or AI-based devices. Muehlematter et al, have identified 240 AI/ML-based devices approved in Europe between 2015-20, of which only 124 were also approved by FDA. Furthermore, the authors concluded that the majority of CE-marked AI products were not supported by any peer-reviewed publications.[33]

Two systematic reviews have found no clinical trials investigating AI models in gynaecology oncology[34-5]. In addition, there seems to be a paucity in literature of reports into the role of AI in cancers of vulva, vagina and gestational trophoblastic disease[36-7].

2.1 AI reporting standards

Generally, there has been insufficient scrutiny of reporting standards in AI studies in terms of design, methodology, and outcomes.[34,38] Pre-existing reporting guidelines were found to be limited and inadequate to assess AI reporting aricles.[34] Hence, several reporting guidelines have been updated to accommodate specifics pertaining to AI studies. Generally these are referred to as AI extensions such as CONSORT-AI (Consolidated Standards of Reporting Trials-AI); SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials–AI); STARD-AI (Standards for Reporting of Diagnostic Accuracy Studies-AI; TRIPOD-AI (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis-AI); PROBAST-AI (Prediction model Risk Of Bias ASsessment Tool - AI); QUADAS-AI (Quality Assessment of Diagnostic Accuracy Studies-AI); and DECIDE-AI (Developmental and Exploratory Clinical Investigations of DEcision-support systems driven by Artificial Intelligence). [21,34,38-49].

Shahzad et al and Plana et al independently conducted systematic reviews looking into reporting standards in randomised controlled trails (RCT) investigating AI-based interventions until 2021; they found 42 and 41 RCTs respectively with poor adherence to CONSORT-AI guidelines.[34,35]

2.2 AI and statistics

AI has offered exciting new opportunities for exploring and mining big data and uncovering patterns and relationships, including when these are complex or non-linear.[15,39,50-52] In this regard, AI complements traditional computing and inference statistics, offering evidence to inform medical practice and health care delivery.[4] For instance, while supervised ML has provided a complementary approach to regression statistics and survival analysis; unsupervised ML, with the ability to identify

nonlinear relationships among variables, could be used as an alternative technique to correlation statistics where distinct sub-grouping can be recognised.[53-55]

In inference statistics, data are assessed using data models based on specific assumptions, which vary according to the test used. These could include assumptions such as normal distribution, linear relationship, homoscedasticity (equal variances) of errors, and independence of variables.[56,57] The validity of the tests used is then assessed to judge if the findings of the data model are applicable to the data being explored. For instance, in regression analysis, goodness-of-fit tests and residual analysis tests are used while correlation coefficient in correlation can be estimated.[4,53,56]

Most AI algorithms in medicine are designed to estimate the risk (prediction) of a patient having an event presently (diagnosis) or developing one in the future (prognosis).[38,40,58] The way AI and ML carry out a prediction is not always explainable (except for some explainable ML tools such as decision trees). One reason is that the exact structure of algorithms in AI and ML are not known, or not disclosed, unlike in conventional statistical tools. Hence there would be no tests similar to those of goodness-of-fit however instead the prediction function can be validated using test accuracy methods (sensitivity, specificity, and receiver operating characteristic curve (ROC curve) and Area Under the ROC Curve (AUC or AUROC)[41,59] by comparing the prediction results to the observed outcome (such as death or cancer recurrence), or to existing gold standard (assessment ovarian cancer burden on CT by an expert radiologist for example)[4]. In general, an AUC of 0.5 suggests that the test lacks the ability to differentiate (for example between patients who might and those might not develop cancer recurrence), 0.7 to 0.8 is acceptable, 0.8 to 0.9 is considered excellent, while > 0.9 suggests an outstanding performance of the algorithm[60]. It is worthy of noting that metrics alone do not always reflect the quality of ML prediction which has spurred recent research into their interpretability.

3 Cervical cancer

Cervical cancer is caused by persistent infection with high-risk strains of human papilloma virus (HPV)[61-2] which are found in 99.7% of cervical cancers globally.[63] The discovery of the causation role of HPV in cervical cancer has led to two practice-changing developments; HPV-based screening and HPV vaccination[64-66]. Typically, cervical cancer is diagnosed on histological examination of cervical biopsy and often radiological assessment is used when available to predict parametrial invasion, lymph nodes involvement and any distant metastases. The treatment of cervical cancer is largely surgical resection in early stages and chemoradiotherapy in advanced and recurrent disease[67-69].

Cervical cancer is the fourth most common cancer in women globally and the commonest of the gynaecological cancers.[70] The relatively low incidence of cervical cancer in high-income countries such as the UK can be attributed to the success of universal screening programmes and the introduction of the HPV vaccine, which demonstrated the preventable nature of this disease.[71-2] Cervical cancer reflects a profound socioeconomic variation[73], it burdens mostly low- and middle-income countries where 90% of cervical cancer deaths occur.[74] It disproportionately affects young women, and can have a devastating effect on their families and young children.[74] The World Health Organization (WHO) has therefore launched its global strategy to accelerate the elimination of cervical cancer by 2030 by offering more screening and vaccination to all women and young girls globally.[75]

3.1 AI perspectives in cervical cancer

Cervical cancer has been a prime focus for AI research, we have found that the majority of published articles investigating AI in gynaecological cancers are in the cervical cancer domain; these appear to focus on screening, staging and radiotherapy.[76]

Perhaps it is not surprising that one of the earliest attempts to investigate AI in medicine was in cervical cancer screening.[77] In fact the role of AI in cervical cancer screening research can illustrate how AI is transforming medical practice.[78] Automation of cervical cancer screening has been an urgent need, since cytology-based assessment was widely introduced, given the large number of smears performed globally, with up to 200,000 cell per slide.[79,80] In the 1950s, Cytoanalyzer was one of the early attempts in this field, however this was by using traditional computing.[81] The clinical field then was dominated by automation using conventional algorithms such as ThinPrep Imaging System and the Becton Dickinson Focal Point GS Imaging System.[82-84]. Interestingly, in 1995, PapNet received FDA approval[85] and it was one of the earliest AI-enabled medical devices where it used NN to identify abnormal smears based on malignant and premalignant morphologic criteria.[86,87]

3.2 AI: Prognostication in cervical cancer

One systematic review addressing ML research in cervical cancer prediction (screening, detection survival and recurrence rates) has identified 50 articles, 33 of which were published in Asia with only seven articles in Europe and seven studies in America. This systematic review also found that AI models performed differently with CNN achieving the highest positive predictive value (PPV) of 99.5, while KNN had a modest PPV of 80.7.[88] A more recent systematic review looked at the use of ML in survival predictions for cervical cancer patients found 13 suitable articles which used a variety of AI models most commonly RF. It also reported a wide range of AUC: 0.40 – 0.99. The authors also recognised that of interpretability, explainability, and imbalanced datasets remained one of the biggest challenges facing AI research in cervical cancer[89].

3.3 AI in cervical cancer screening

3.3.1 AI in cervical smear screening

Shen et al investigated the cost-effectiveness of three screening methods: HPV testing, manual liquid-based cytology (LBC) and AI-assisted LBC testing with six different frequencies for each (18 screening strategies) in a cohort of 100,000 women. They concluded that the most cost-effective method would be AI-assisted LBC every 5 years.[90] Assessment of cloud-based DL system to analyse digitalised cervical smear slides (using portable whole-slide microscope scanner and uploaded with mobile network in rural Kenya) when samples from a small (740) high-risk women (infected with human immunodeficiency virus (HIV)) used to train and test the system to achieve detection of atypia sensitivity 100% and specificity 78.4% (cytologist assessing physical slides was the gold standard).[91] A cohort study of more than 700,000 women showed an concordance rate of 94.7%, Kappa 0.92 between AI and manual cytology. When considering histologically confirmed cervical intraepithelial neoplasia grade 2 or worse (CIN2+) (also known as high-grade squamous intraepithelial lesion [HSIL]) the sensitivity was (90.1% vs 84.3%) and specificity was (94.8% vs 95.2%) of AI compared with manual cytology respectively.[92]

3.3.2 AI in cervical cancer clinical screening

Clinical inspection can be used In low resource settings where there is a limited access to smear-based screening for cervical cancer.[93] A systematic review evaluating AI-based cervical cancer screening using images taken during visual inspection with acetic acid (VIA) identified 11 suitable articles with sensitivity and specificity, ranging from 0.22 to 0.93 and 0.67 to 0.95, respectively. It was noted that these studies used highly selected images which would not necessarily represent routine practice.[94] A

frequently faced challenge in image capture is the movement of the cervix during acetic acid assessment due to the patient or camera moving. Guo et al have developed a self-supervised RGB-colored DL-based image registration method to automatically align the images, which does not require manual input. This has improved the Dice score by an average of 12.6% [95].

3.4 AI in cervical cancer histopathology

Whole slide imaging (WSI) segmentation and analysis have the potential to predict survival and develop improved treatment plans for patients. A potential association between histological image and cervical cancer prognosis was investigated using a deep neural network (DNN) to extract potential risk factors from WSI to predict overall survival and disease-free survival with AUC of 0.80.[96] A fully automated cervical lesion analysis of conventional cervical smear samples - using WSI - was performed for the first time. Each image is converted into a tile-based pyramid format to handle gigapixel data efficiently and then fed into a multi-layer DL architecture. This system uses a coarse-to-fine strategy for semantic segmentation and tissue detection, making it ideal for rapidly identifying CIN2+/HSIL lesions. At the coarse level, the goal is to quickly identify tissues of interest for further screening, whereas, at the satisfactory level, HSILs are discovered using the findings of the first screening. The proposed system is capable of segmenting HSIL or higher lesions with PPV of 0.93 and sensitivity of 0.90.[97]

3.5 AI in cervical cancer radiology

3.5.1 AI prediction of lymphadenopathy and parametrial invasion in cervical cancer

The diagnosis of lymph node metastasis or parametrial involvement in cervical cancer patients is clinically relevant as it could identify patients whose cancer is too advance to recommend surgical treatment.[98] One systematic review and meta-analysis study investigating AI use for preoperative prediction of lymph node metastasis in abdominopelvic malignancies identified 17 studies of sufficient reporting quality, five of which were in cervical cancer patients. It found that in gynaecology cancers, the pooled AUC was 0.893, (95 %CI, 0.847−0.939) for AI which outperformed radiologist-pooled AUC of 0.749 (95 %CI, 0.656−0.842) when histology was used as the diagnostic endpoint.[99]

Systematic review by Charoenkwan et al have used RF model in retrospective data to predict parametrial involvement with cancer in patients who had surgical resection. Interestingly they used histological and clinical data rather than radiology, some of this data would not usually be available prior to surgical resection of cancer which would raise questions about its usefulness in clinical practice even in low-resource countries with limited access to cross sectional radiology[100].

3.5.2 AI in cervical cancer radiotherapy planning

A systematic review of DL in CT image segmentation for radiotherapy in cervical cancer patients identified 14 articles reporting Dice Similarity Coefficient (DSC) for clinical target volume (CTV) or organ at risk (OAR), which ranged between 0.83 and 0.92. This lead the authors to conclude that DL has good accuracy in automatic segmentation of CT images of cervical cancer.[101]

4 Uterine malignancy

4.1 Endometrial cancer

Endometrial cancer is the most common gynaecological malignancy in high-income countries, with increasing incidence and mortality due to, at least in part, ageing population and prevalent obesity.[102-103] Generally, endometrial cancer patients have a relatively good prognosis since most present with postmenopausal bleeding (PMB) which leads to early diagnosis and treatment. Typically, endometrial cancer is diagnosed on histological examination of endometrial biopsy, and staged with radiology assessing particularly the depth of myometrial invasion (MI), the involvement of lymph nodes and any distant metastases. Treatment of endometrial cancer patients could include surgical resection, radiotherapy and chemotherapy according to the cancer stage and characteristics.[104,105]

4.1.1 AI in endometrial cancer histopathology

Levine et al, from the TCGA research network, proposed a four-category classification for endometrial cancer based on integrated genomics, transcriptomics and proteomics. These are polymerase epsilon (POLE) ultramutated, microsatellite instability hypermutated, copy-number high, and copy-number low groups.[106] Subsequently, surrogate markers were shown to distinguish these four groups into POLEmut, mismatch repair deficient (MMRd), p53abn and non-specific molecular profile (NSMP) respectively.[107] This classification system has recently been adopted to stratify cancer risk for mortality and recurrence; has formed the basis for an international trial investigating targeted management[108], and has been incorporated into the International Federation of Gynecology and Obstetrics' (FIGO) most recent staging system.[109-112]

AI has emerged as a promising tool in endometrial cancer research, potentially improving diagnostic accuracy, risk stratification and treatment planning. Fremond et al investigated interpretable DL pipeline for WSI-based prediction of the endometrial cancer four molecular groups using H&E slides obtained from the Post-Operative Radiation Therapy for Endometrial Carcinoma (PORTEC) trials. This model was able to allocate patients into these groups with AUROC of 0·849, 0·844, 0·928 and 0·883 for POLEmut, MMRd, p53abn and NSMP respectively. This study can be seen as a good example of collaboration among pathologists, clinicians, and clinical and AI scientists to address important clinical issues relevant to patients' care.[113]

4.1.2 AI in endometrial cancer imaging

4.1.2.1 AI prediction of myometrial invasion in endometrial cancer

The depth of myometrial invasion (MI) in endometrial cancer is an important clinical criterion; not only does it determine the cancer stage and thus guide treatment options, it is also used in the NHS to triage patients for secondary or tertiary care facility for surgical treatment.[105] Several studies have developed ML and DL tools to detect MI. The efficacy of DL using T2-weighted imaging (T2WI)-based MR was assessed in 530 patients with pathologically confirmed endometrial cancer. DL-based detection and classification algorithms were developed to automatically locate the cancer area and calculate the MI depth. This model achieved an average accuracy of 77.14% in sagittal images and 86.67% in coronal images for lesion identification and reported accuracy of 84.78% detecting deep MI. Combining the knowledge of radiologists with a trained network model improved accuracy to 86.2%.[114] The same research group later developed a technique which first used the U-net to segment tumour and uterus on MR images, and then analysed the segmentation pictures for MI depth using

three AI models (rapid thinning, fit-ellipse, and area ratio), they reported accuracy of 87.1%, 90.3% and 85.8% respectively.[115] A pilot study evaluating radiomics-powered ML to detect deep MI in 54 endometrial cancer patients, 17 of whom had deep MI. This was a multistep model, radiologists performed lesion segmentation, features were extracted, and an RF wrapper was then used to select the most informative features - followed by an ensemble of J48 decision trees. This model achieved accuracy of 91% in testing data, which also appeared to improve radiologists' performance when using ML.[116]

4.1.2.2 AI prediction of lymphadenopathy in endometrial cancer

AI models have been evaluated for the prediction of lymph node metastasis in endometrial cancer. A recent systematic review of the role of ML in preoperative identification of lymph node involvement found 50 studies with 103,752 patients, including 12,579 with positive lymph node on histopathology. The best performing model was that constructed by combining radiomics and clinical features with pooled sensitivity and specificity of 0.81(95%CI: 0.70-0.89) and 0.84(95%CI: 0.76-0.89) respectively, which outperformed clinical decisions using Mayo criteria[117] in its specificity 0.59(95%CI: 0.38-0.77) while maintained the sensitivity rate 0.81(95%CI: 0.66-0.90).[118] Similarly, Yan et al used MR radiomics aided with an AI model (MRMR) to predict lymph node involvement in patients who had lymphadenectomy for confirmed endometrial cancer. Their model achieved AUC of 0.91 compared with 0.81 and 0.84 for two radiologists.[119] In a systematic review by Lecointre et al in 2021, 17 articles were identified that used AI-based radiomics in endometrial cancer for the prediction of MI and lymph node and lymphovascular space involvement. The authors concluded that while this was a promising field, there was insufficient evidence on the advantages of AI-based radiomics in endometrial cancer.[120]

4.2 AI in uterine smooth muscle neoplasms

The differentiation between uterine leiomyosarcoma and leiomyoma is a clinically challenging one, particularly in women who wish to preserve their fertility. A systematic review in 2021 found six studies that predominantly used AI and radiomics on MR images. The authors of the review concluded that there was insufficient evidence to support radiomics in clinical leiomyosarcoma diagnosis.[121] A more recent study, which included 200 leiomyoma patients and 63 leiomyosarcoma patients showed that DNN model had a comparable accuracy diagnosing sarcoma to experience radiologist (91.3% and 88.3% respectively) but superior to that of less experienced radiologist (accuracy 80.1%).[122]

5 Ovarian cancer

Ovarian cancer is a heterogeneous disease at anatomical, cellular and molecular pathway aspects.[123,124] Ovarian cancers can be epithelial or non-epithelial. Non-epithelial ovarian cancers are germ cell tumours (such as immature teratoma) or sex cord stromal cancers (e.g. granulosa cell tumour)[125]. Epithelial ovarian cancers include high-grade serous carcinoma (HGSC) and low-grade serous carcinoma, which are currently viewed as two distinct diseases rather than one malignancy with two grades.[126] The most common ovarian cancer, and one with a poor prognosis, is HGSC.[124] It is now well accepted that the majority of HGSC arise from the fallopian tube precursor lesions, while rare cases may arise from the peritoneum in addition to the ovarian origin. Thus it is referred to as HGSC of tubo-ovarian or primary peritoneal origin. The term 'ovarian cancer' is often used as an umbrella term to refer to these groups of cancers.[125,127-129] Treatment for ovarian cancer broadly consists of maximum cytoreductive surgery which aims to achieve complete cytoreduction (also

known as no macroscopic residual disease [NMRD]), and systemic anticancer therapy (SACT), which include chemotherapy (platinum-based generally) and targeted therapies such as poly adenosine diphosphate ribose polymerase inhibitors (PARPi) and anti-angiogenetic agents.[124,127]

5.1 AI Perspectives in Ovarian Cancer

The recent growing appreciation of the heterogeneity of ovarian cancer has paved the way for more targeted and personalised treatment options.[123] In addition, the availability of multiple data sources such as electronic patient records, radiology, digital histopathology images, and biomarkers, has also offered new opportunities for utilising AI models to address existing clinical challenges as well as to explore new ones. AI has shown great promise in ovarian cancer research, with numerous studies exploring its potential to improve diagnosis, treatment, and prognosis. In recent years, there has been growing interest in integrating multiple data types, such as radiogenomics, multi-omics, and fluxomics data, to improve our understanding of ovarian cancer and develop more effective diagnostic and treatment strategies.

AI-based research in ovarian cancer appears to have focused on diagnosis, prognosis, prediction of surgical resectability and the response to chemotherapy. A systematic review that identified 39 studies investigating ovarian cancer diagnosis and prognosis, found that the majority (19 studies) used high-throughput omics data, while 13 utilised serum biomarkers and electronic patient records, with 7 studies using histopathology or radiology images.[130] This is interesting, since in the current clinical practice, imaging and biomarkers are dominantly used for clinical decision making. While this might reflect the availability and suitability of omics data for AI-based research it could also indicate the direction for future research in ovarian cancer. Importantly, this review found that the quality of the studies was not entirely satisfactory, with wide gaps in the predictive performances of AI models. This review also pointed out the importance of AI model selection to suite the type of investigated data. For example, support vector machine (SVM) appeared to be suitable for ovarian cancer diagnosis using ultrasound scan imaging, while deep convolutional neural networks (DCNN) algorithm reached a modest accuracy of 78.20% in Haemotoxylin and Eosin (H&E) histology slide images.[130,131]

5.2 AI: Treatment Planning in Ovarian Cancer

5.2.1 AI in Pelvic Mass Stratification

Several studies have investigated the performance of AI models in determining the nature of ovarian mass (malignant, benign, or borderline), which is a relevant and common clinical encounter. In addition, malignancy risk prediction of pelvic masses is currently to triage patients to surveillance, secondary treatment or cancer centre surgery.[132,133] One systematic review and meta-analysis of literature in the English and Chinese languages identified 11 studies that investigated the use of AI technology using radiology images in ovarian cancer diagnosis. It found a pooled AUROC of 0.94 (95% CI 0.88-1.00), 0.82 (95% CI 0.71–0.93) and 0.82 (95% CI 0.78–0.86) for ultrasound, MR and CT respectively.[134] Another systematic review evaluating AI in ultrasound imaging has also suggested a better performance for AI utilising ultrasound compared with MR and CT, with a pooled AUC of 0.95 (0.93−0.97), 0.90 (0.87−0.92), and 0.82 (0.78−0.85) respectively. When compared with human clinicians the pooled AUC was 0.91 (0.88−0.93) for AI and 0.85 (0.81−0.88) for human clinicians. This systematic review did not find a significant difference in the performance of ML and DL with pooled sensitivity and specificity of 89% (85−92%) and 88% (82−92%) for ML and 88% (84−91%) and 84% (80−87%) for DL, respectively[135]. A systematic review specifically looking at AI in ultrasound diagnosis

of ovarian cancer identified 14 studies with a wide range of sensitivity and specificity rates, 40%-99% and 76%-99%, respectively. The identified studies used varying AI models such as SVM, DCNN, K-nearest number classifier (KNN), decision tree (DT), DNN and probabilistic neural network (PNN). However, it was challenging to compare AI modality performance given the heterogeneity in methodology including feature extraction and segmentation techniques.[136]

One study using four AI classifiers KNN, SVM, random forest (RF) and logistic regression (LR) on CT images has found that an ensemble model (combined radiomics, DL, and clinical data) outperformed each model individually with a test accuracy of 82% in cases with confirmed histological diagnosis. This was comparable to senior radiologists (> 10 years' experience) but outperformed radiologists with less than 10 years' experience (respective accuracy 83% and 66%).[137] Another study investigated MR based single-and-multiparameter (MP) ML model to distinguish borderline ovarian tumours from early stage ovarian cancers, as confirmed by histology, achieved AUC of 0.920 compared to AUC 0.797 for radiologists.[138] Concordant conclusions were reached by another group, which constructed a late multiparametric (LMP) model based on multiple instance convolutional neural network (MICNN) to distinguish borderline from malignant ovarian tumours as confirmed by histology, achieving AUC of 0.884 (95%CI 0.831-0.938) compared to pooled AUC of 0.797 for radiologists.[139] Similarly, Wang et al have shown that DL outperformed radiologist in distinguishing borderline from malignant tumours with AUCs of 0.87 and 0.75 resepectively.[140] This remains an area of active research, particularly with new work highlighting end-to-end radiomics-based model capable of adnexal mass segmentation and classification, with a comparable predictive performance (AUC 0.90) to the published performance of expert subjective assessment (gold standard), and current risk models. The false discovery and false positive rate levels of the best models currently in the field encourages use of these AI tools in a two-step approach: to initially identify the 'high-risk' adnexal mass that warrant further evaluation by an expert ultrasound examiner in a second step, thus reducing clinical workload.[141]

Another study sought to use the ML Minimum Redundancy - Maximum Relevance (MRMR) feature selection method applied to biochemical markers, and achieved sensitivity and specificity of 1.00 and 0.90 (compared to 0.92 and 0.97 respectively when the risk of ovarian malignancy algorithm (ROMA) was used).[142] Reilly et al have developed an ovarian cancer risk assessment tool in women with known pelvic masses. They called it multivariate index assay (MIA3G), which is a deep feedforward neural network model using features of patient age, menopausal status and seven biomarkers: cancer antigen 125 (CA125), human epididymis protein 4 (HE4), beta-2 microglobulin, apolipoprotein A-1, transferrin, transthyretin, and follicle-stimulating hormone. They used over 3,000 patients to train, test and validate this tool, with an impressive negative predictive value (NPV) of 99.38% in a population with a prevalence of 4.9% however this was at the cost of a reduced PPV of 22.45% and low sensitivity in early stage cancer (76.92%).[143] Ahmad et al investigated several biomarkers-based ML models including RF, SVM, decision tree (DT), extreme gradient boost (XGBoost), LR, Gradient Boosting Machine (GBM) and Light Gradient Boosting Machine (LGBM) with accuracy ranging between 0.59% and 91% distinguishing malignant from benign cases.[144]

5.2.2 AI Prediction of peritoneal metastasis in ovarian cancer

AI has been applied in ovarian cancer research in radiomics analysis. Quantitative imaging features were extracted from preoperative MR images. Feature screening was performed using a minimum redundancy maximum correlation (MrMc) and least absolute shrinkage selection operator (LASSO) methods. Four radiomics models were constructed based on three MR sequences. Then, combined

with radiomics characteristics and clinicopathological risk factors, a multi-factor logistic regression method was used to build a radiomics nomogram. The radiomics nomogram based on the combined multiparametric MR (MP-MR) sequence showed good predictive accuracy for peritoneal carcinomatosis in patients with ovarian cancer (AUC 0f 0.90), allowing for identifying PC in ovarian cancer patients before surgery.[145] The association between protein abundance and various CT image traits and texture features in patients with HGSC was investigated using the Kendall tau rank correlation coefficient and the Mann-Whitney U test.[146] A potential connection between CT-based tumour heterogeneity metrics and protein abundance was revealed for the first time. The connections with argininosuccinate synthase 1 (ASS1) were the most intriguing.[146] The protein abundance of cysteine-rich protein two was inversely linked with tumour involvement of the mesentery, a known major limiting factor for primary debulking surgery (CRIP2). Even after controlling for multiple testing, this connection remained statistically significant. CRIP2 is a tumour suppressor and a regulator of cell proliferation.[147,148] In addition, supradiaphragmatic lymphadenopathy was positively linked with the protein abundance of MAGE family member A4 (MAGE4). Increased MAGE4 expression in ovarian cancer cells is an independent predictor of mortality related to reduced overall survival. Similar studies of CT radiomics have provided linkage to ovarian cancer phenotypes or integration of phenotypic information to improve prediction.[149-152]

5.2.3 AI Prediction of lymphadenopathy in ovarian cancer

In an interesting attempt, Yao et al used residual neural network (RNN) and SMV models on Fludeoxyglucose F18 (18F-FDG) positron emission computed tomography (PET) in apparently early-stage ovarian cancer patients to evaluate lymph node metastasis. They reported an impressive performance of their model with AUC of 0.93 (95% CI 0.84-0.99), sensitivity of 81% and specificity of 100% when compared with final H&E histology assessed by human histopathologists. Unfortunately, this study did not provide details of the surgical procedure, the extent of surgical lymph node dissection or the number of lymph nodes excised.[153]

5.2.4 AI prediction of cytoreductive surgery outcome in ovarian cancer

One ultimate clinical goal in the management of ovarian cancer patients is to offer cytoreductive surgery for only those who are likely to benefit from this extensive surgery.[154] One review has looked at the role of AI in predicting NVRD in ovarian cancer patients, it identified only 2 studies with a modest accuracy of 77.7% and 65.8%.[58,155,156] Laios et al used XGBoost to construct an intraoperative scoring system in patients undergoing cytoreductive surgery for advanced ovarian cancer which was found to predict NMRD with AUC 0.88 (95% CI 0.85-0.91).This was found to be superior to Peritoneal Carcinomatosis Index (PCI) and the Intra-operative Mapping for Ovarian Cancer (IMO) scoring systems, which had AUC of 0.73 and 0.67 respectively.[157] Maubert et al have shown - using intraoperative findings in patients undergoing cytoreductive surgery and hyperthermic intraperitoneal chemotherapy (HIPEC) of whom 153 patients (49%) had gynaecological cancers - that RF model surpassed, with an accuracy of 98%, other classification algorithms, which included simple classification, conditional tree (CT) and SVM, in predicting resectability of peritoneal carcinomatosis.[158] In another study, using preoperative data which included radiology, age, CA-125, performance status, *BRCA* status, and surgical complexity scores, it was reported that an RF model can successfully predict complete cytoreduction (residual disease 0 cm/NVRD) and optimal cytoreduction (residual disease ≤ 1 cm), with AUC of 89.0% and 84.0% respectively.[159]

5.3 AI in ovarian cancer histopathology

Histopathologic diagnosis is one area where AI has been applied in ovarian cancer research. A DL - based approach was applied to evaluate histopathologic patterns in ovarian cancer. The first step was to segment ovarian cancer regions from WSI. Then, a deep interactive learning approach was used to efficiently train the ovarian segmentation model, achieving an intersection-over-union (IoU), sensitivity and PPV of 0.74, 0.86 and 0.84 respectively; and automatically extracting HGSC patches. After segmentation, a *BRCA* classification model processed cancer patches to produce patch-level scores indicating the likelihood of a *BRCA* mutation, AUC for *BRCA* classification ranged between 0.49 and 0.67 on the validation dataset.[160] Another study applied an attention-based NN to predict somatic *BRCA1/2* gene status and survival data. The model was tested on a cohort of 664 ovarian cancer patients, of whom 233 (35.1%) had a somatic *BRCA1/2* mutation. The training and testing sets achieved an area AUC of 0.7 and 0.55, respectively.[161]

The identification of tubal intraepithelial carcinoma (STIC), which is a precursor for HGSC, and tubal intraepithelial lesion (STIL) has been explored by Boaerts et al. They investigated a DL algorithm (U-Net with resnet50 backbone) to distinguish STIC/STIL from benign tissues on WSI from 682 patients. They achieved AURC 0.95 (95% CI: 0.90–0.99) on the external test data when compared to panel review of specialist gynaecology pathologists.[162] Another group used digital H&E WSI to predict the effectiveness of treatment with bevacizumab in ovarian cancer patients. They used a two-step hybrid DL framework which included efficient weekly supervised cascaded DL for rapid identification of regions of interest (ROIs) followed by DL based classifier to predict treatment effectiveness. This precluded the need for human pathologist input and achieved a high accuracy of 0.882 and sensitivity of 0.912.[163] Ma et al have constructed an ovarian cancer-specific predictive framework to inform clinical use in terms of platinum response and prognosis. They utilised multiple biomarkers including circulating tumour cells (CTCs) to investigate the performance of eight ML classifiers: RF, SVM, Gradient Boosting Machine (GBM), Conditional RF, NN, Naive Bayes, Elastic Net, and LR. RF model came on top in predicting platinum-resistance with AUC of 0.81.[164]

5.4 AI in precision medicine for ovarian cancer

AI has been used in biomarker discovery and to explore mechanisms underlying ovarian cancer. An ML algorithm was applied to analyse the proteomic dataset from ovarian cancer patients, TOP1, PDIA4, and OGN was identified as candidate biomarkers and potential mechanisms underlying the disease. This approach improves the understanding of ovarian cancer and guides the development of new treatments.[165]

A potential capability of ML models is to help predict the effectiveness of pharmacological therapy based on the individualised genetic profiles of patient tumours, an important goal of contemporary cancer medicine.[166] Since several alternative biochemical pathways can contribute to the development of the same cancer type, the responses of different individuals to the same chemotherapeutic agent might vary considerably. Therefore, the transcriptomics data were analysed using SVM to enhance the predictability of patients' responses to therapy. Using gene expression profiles of 152 cancer patients obtained from the TCGA database, the response of individual patients treated with gemcitabine or 5-FU was predicted with >81% accuracy.[166]

Utilising data from the cancer genome atlas (TCGA), Chen et al used gradient boosting decision tree (GBDT) algorithm to analyse genetic interactions related to chemoresistance in ovarian cancer. They

identified 24 signature gene pairs and 10 individual signature genes with AUC for chemoresistance prediction of 0.97 and 0.68 respectively. The authors concluded that these findings could improve clinical practice and inform decision-making and treatment choices for patients and their clinicians.[167] Another study also used gene expression data, indicating genes such as *TLR4*, *BSCL2, CDH1, ERBB2, SCGB2A1*, and *BRCA2* as critical prognostic indicators.[168]

6 Ethical considerations

AI implementation in gynaecologic oncology, in line with other health care domains, raises several controversial issues which should be carefully addressed to ensure a safe, effective, and equitable use of this technology. These considerations include cybersecurity, data protection, bias and equity, accountability, validity, and reliability. The impact on patients' experience and health workers' skills and job security is a real concern too.

Cybersecurity and data protection in AI is of paramount importance. AI relies on access to large amounts of patient data, including sensitive information such as medical history and genetic information. These data must be protected from cyber threats such as hacking, data breaches, and ransomware attacks. Health organisations and researchers must take the appropriate measures to ensure the privacy and security of patients' data. Another significant concern is the potential inequity in AI algorithms where there is the potential for AI reinforcing existing biases in healthcare, particularly concerning race, ethnicity, and socioeconomic status compromising further equal access to medical care. AI outcome is driven by the quality of training data used, if the data are incomplete or not inclusive, this could lead to wrong results or inappropriate treatment recommendation. This could affect some patients more than others according to the representativeness of training data (race, ethnicity, socioeconomic class, or place of residence). Therefore, health organisations and researchers must carefully consider issues of equity and bias in developing and implementing AI algorithms to ensure that they are fair and accurate for all patients.[169]

Additionally, the lack of interpretability of most AI models could hinder incorporating AI results into clinical decision-making. While AI system may produce accurate results, it can be difficult for clinicians to understand how the algorithm arrived at its conclusions, making it hard to support their implementation. One other challenge is the transparency in AI systems in healthcare and subsequently with the liability for AI-based clinical outcomes. In addition to jurisdiction consideration, the 'black box' nature, where the exact final structure of the constructed algorithms is unknown or cannot be known, which could form a major obstacle.[170,171] Accountability is another critical ethical consideration when implementing AI in health care, as questions are raised, i.e: who is responsible for unintended consequences if they occur? Would that be the clinician in direct contact with patients, the hospital employing that clinician, or the company marketing the used AI system? Health organisations and researchers must ensure that AI systems are transparent, explainable, and accountable. Patients must be helped to understand how AI is used in their care, and health workers must be trained to interpret and act on AI-generated results appropriately. The validity and reliability of AI algorithms are also important considerations. Health organisations and researchers must ensure that AI algorithms are validated and tested rigorously to provide accurate and reliable results. AI should not replace clinical judgment or patient input but rather be used to augment and inform clinical decision-making.[172,173]

Finally, the implementation of AI models in gynaecologic oncology may impact patients' experience and health workers' skills and job security. Patients may feel uncomfortable or sceptical about the use

of AI in their care, and health workers may feel threatened by the potential for AI to replace or reduce their role with negative impact on career satisfaction, and financial constraints affecting families. Therefore, health organisations should ensure that patients are informed about the use of AI in their care and that health workers are trained to use AI appropriately and to understand its limitations.[174,175]

7 Discussion

While AI has shown promise in gynaecological oncology, there are still limitations to its implementation in clinical practice. AI research in gynaecology oncology appears to be more concerned with discovering the best AI model fitting available data and identifying algorithms with the highest AUC rate, rather than addressing the patients' priorities and investigating clinical needs. The developers of this paper have found a few precious examples of productive collaboration among AI scientist, biology scientist and clinicians.[113] AI could be a powerful tool in areas of pressing need for academic and clinical progress, such as symptom-based early diagnosis of ovarian cancer, endometrial cancer stratification, chemotherapy resistance prediction and cervical cancer screening in low- and middle-income countries. In fact, there have been several publications setting priorities and goals as seen by patients and their clinicians that we recommend AI investors and investigators can consult for future guidance.[176-180]

It is possible that the reason underpinning this phenomenon is that AI scientists are limited with their research to the data they have access to. However, this could be compromising AI research results in that these data are not AI specific, they were collected selectively to suit existing tools for which AI algorithms might not be able to exercise their full intelligence given that the 'missing' uncollected data might be important predictive features. Another challenge with existing data is the need to make them AI-compatible. This is called data curation, a process which includes filtering, cleansing, integration, alteration and reduction. On some occasions, this can hinder the data, which become less representative, too ideal.[39] This ultimately could affect the performance of AI models trained and tested in noise-reduced datasets, leading to difficulty maintaining performance when implanted in real-world data (overfitting).[9] There are several other challenges faced when implementing AI in the healthcare system, other than where it was trained, which should be taken into account when considering generalisability. These include differences in clinical practice according to health system type and settings, to jurisdiction, or as they evolve over time; patients' demographics, social and cultural characteristics, and genotypic and phenotypic specifics. In addition to the wide range of hardware and software used to capture data and the type of data collected. [59,181] Some obstacles can be practice-specific, for example the IBM Watson for Oncology, trained by specialists in Memorial Sloan Kettering Cancer Center (MSK), has some of its recommended management plans ignored in health systems with practices dissimilar to that where it was trained. [5,182-185] Perhaps, a crucial obstacle for AI implementation is the lack of clinical trials demonstrating and evidencing AI benefits to patients with gynaecological cancers.

Regulatory and ethical issues must be addressed before AI can be widely adopted in gynaecological oncology. These include issues related to data privacy and security, as well as the potential for AI to replace human expertise and decision-making. Despite these limitations, AI has the potential to significantly improve the accuracy and efficiency of gynaecological oncology diagnosis and treatment. Ongoing research and development will be critical to addressing these challenges and realising the full potential of AI in this field.

8 Conclusion

AI is a collective set of self-teaching algorithms used by multiple computer programs in our daily lives. AI has emerged as a powerful tool in gynaecology oncology which is likely to shape future clinical practice.

It is currently in clinical use in automated cytology in cervical smears and has shown good results in the fields of cervical cancer screening, staging and radiotherapy planning. AI models are investigated in endometrial cancer staging and prediction of malignant potential in uterine tumours. In ovarian cancer, AI has been shown to aid triaging of pelvic masses, predict cancer stage and resectability. Although several ML and DL models have been proposed for the integration of multi-omics and image data for gynaecological cancers, several challenges remain to deploy and improve these methods, such as the lack of single-cell RNA-seq data with different available data types and treatment information, the simulation of intra-omics interactions, and incorporating multimodal data into a machine learning model that can be interpreted biologically. Moreover, further research and validation of these methods are needed to ensure their effectiveness and safety in clinical settings.

To date, AI remains largely in the research phase in gynaecological cancer domains. Significant efforts addressing practical, ethical and legal concerns must be made to allow safe, efficient, and accountable implementation of AI. An effective collaborative partnership among stakeholders, AI, biology and clinical scientists, clinicians, policymakers, investors, and patients, is of paramount importance if the full potential of AI to be realised.

9 Opinion

- AI is a broad spectrum of emerging and evolving tools utilising computational algorithms, which offer exciting opportunities with potential significant challenges.
- AI research has been utilising data collected for other purposes that might be also biased and not inclusive which could limit its ability and mask important discoveries.
- AI research is largely focused on discovering AL algorithms and models and identifying the ones best performing in training data.
- AI research focus in gynaecology oncology requires urgent readjustment to address the crucial issues of clinical needs and patients' priorities.
- There have been some examples of joint efforts of AI scientists with biologists, clinical scientists, and clinicians to produce meaningful and applicable research. This could form a model to guide future efforts via partnerships among investigators, investors, clinicians, policymakers, and patients.
- AI implementation could be hindered unless serious issues with ethical, legal and security implications are addressed and acted upon.
- The RCOG is advised that preparedness for AI-based technology in time is crucial. We recommend the RCOG educates members and fellows for an AI future, and incorporate AI into the training curriculum.
- The RCOG could consider, in line with other medical colleges, establishing a dedicated committee or task-specific group overseeing AI research, progress and implementation.

10 Artificial intelligence abbreviations

AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
RF	Random Forest
NN	Neural Network
RNN	Residual Neural Network
PNN	Probabilistic Neural Network
DNN	Deep Neural Network
DCNN	Deep Convolutional Neural Network
MICNN	Multiple Instance Convolutional Neural Network
GAN	Generative Adversarial Network
SVM	Support Vector Machine
KNN	K-nearest Number classifier
DT	Decision Tree
CT	Conditional Tree
LR	Logistic Regression
LMP	Late Multiparametric
MRMR	Minimum Redundancy - Maximum Relevance
XGBoost	Extreme Gradient Boost
GBDT	Gradient Boosting Decision Tree
GBM	Gradient Boosting Machine
MrMc	Minimum Redundancy Maximum Correlation
LASSO	Least Absolute Shrinkage Selection Operator

References

1. McCarthy J, Minsky ML, Rochester N, Shannon CE. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. AI Magazine. 2006; *27*: 12-5. 10.1609/aimag.v27i4.1904.
2. Smith GC, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. BMJ. 2003; *327*: 1459-61.
3. Russell SJ, Norvig P (eds). Artificial Intelligence a Modern Approach. Pearson Education: 2010.
4. Breiman L. Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical Sci. 2001; *16*: 199-231.
5. Alami H, Lehoux P, Auclair Y, de Guise M, Gagnon MP, Shaw J et al (2020) Artificial intelligence and health technology assessment: anticipating a new level of complexity. J Med Internet Res. 2020; *22*: e17707.
6. Lidströmer N, Aresu F, Ashrafian H. Basic Concepts of Artificial Intelligence: Primed for Clinicians. In: N. Lidströmer, H. Ashrafian, eds. Artificial Intelligence in Medicine. Springer International Publishing. 2022; 3-20.
7. Unsworth H, Wolfram V, Dillon B, Salmon M, Greaves F, Liu X et al. Building an evidence standards framework for artificial intelligence-enabled digital health technologies. The Lancet Digital Health. 2022; *4*: e216-e217.
8. National Institute for Health and Clinical Excellence (NICE). Evidence standards framework for digital health technologies: user guide. 2019; NICE, London. https://www.nice.org.uk/corporate/ecd7/resources/evidence-standards-framework-for-digital-health-technologies-user-guide-pdf-11696158815685.
9. Ertel W (ed). Introduction to Artificial Intelligence. Springer International Publishing. 2017; 1-21. 10.1007/978-3-319-58487-4_1.
10. Bhardwaj H, Tomar P, Sakalle A, Sharma U. Principles and Foundations of Artificial Intelligence and Internet of Things Technology (chapter 20). In: Kaur G, Tomar P, Tanque M (eds). Artificial Intelligence to Solve Pervasive Internet of Things Issues. Academic Press. 2021; 377-92. https://doi.org/10.1016/B978-0-12-818576-6.00020-4.
11. Ford L. Artificial intelligence and software engineering: a tutorial introduction to their relationship. Artificial Intelligence Review. 1987; *1*: 255-73.
12. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015; *521*: 436-44.
13. ter Haar Romeny BM. Introduction to Artificial Intelligence in Medicine. In: Lidströmer N, Ashrafian H (eds). Artificial Intelligence in Medicine. Springer International Publishing. 2022; 75-97. 10.1007/978-3-030-64573-1_27.
14. Hira MT, Razzaque MA, Angione C, Scrivens J, Sawan S, Sarker M. Integrated multi-omics analysis of ovarian cancer using variational autoencoders. Sci Rep. 2021; *11:* 6265.
15. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019; *25*: 44-56.
16. Munir K, Elahi H, Ayub A, Frezza F, Rizzi A. Cancer diagnosis using deep learning: a bibliographic review. Cancers (Basel). 2019; *11*: 1235
17. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015: *521*; 436-44.
18. Strolin S, Santoro M, Paolani G, Ammendolia I, Arcelli A, Benini A et al. How smart is artificial intelligence in organs delineation? Testing a CE and FDA-approved Deep-Learning tool using multiple expert contours delineated on planning CT images. Front Oncol. 2023; *13*.
19. Li Y, Wu J, Liu J, Qin L, Cai X, Qiao J et al. Abstract 502: ISM3091, a novel selective USP1 inhibitor as a targeted anticancer therapy. Cancer Research. 2023; *83(7 suppl):* 502.
20. Bengio Y, Russell S, Musk E, Wozniak S, Harari YN (2023) Pause Giant AI Experiments: An Open Letter. https://futureoflife.org/open-letter/pause-giant-ai-experiments/Future of Life. [Accessed 1 July 2024].

21. Vasey B, Novak A, Ather S, Ibrahim M, McCulloch P (2023) DECIDE-AI: a new reporting guideline and its relevance to artificial intelligence studies in radiology. Clin Radiol. 2023; *78*: 130-36.
22. Review, C. (2024). Cochrane Review. https://www.cochranelibrary.com/search.[Awaiting reference update]
23. NICE. Paige Prostate for prostate cancer. Medtech innovation briefing. NICE: 2021; https://www.nice.org.uk/advice/mib280. [Accessed 1 July 2024]
24. NICE. Immunoscore for predicting risk of colon cancer relapse. Medtech innovation briefing [MIB269]. NICE: 2021; https://www.nice.org.uk/advice/mib269. [Accessed 1 July 2024].
25. NICE. Digital technologies for the detection of melanoma. Medtech innovation briefing. NICE: 2022; https://www.nice.org.uk/advice/mib311/ [Accessed 1 July 2024].
26. NICE. Butterfly iQ+ for diagnostic ultrasound imaging. Medtech innovation briefing. NICE: 2021; https://www.nice.org.uk/advice/mib254. [Accessed 1 July 2024].
27. NICE. Artificial intelligence in mammography. Medtech innovation briefing. NICE: 2021; https://www.nice.org.uk/advice/mib242. [Accessed 1 July 2024].
28. NICE. Artificial intelligence for analysing chest X-ray images. Medtech innovation briefing. NICE: 2022; https://www.nice.org.uk/advice/mib292. [Accessed 1 July 2024].
29. NICE. Artificial intelligence for analysing chest CT images. NICE: 2021; https://www.nice.org.uk/advice/mib243. [Accessed 1 July 2024].
30. NICE. AMBLor for identifying low-risk non-ulcerated early-stage cutaneous melanomas. NICE: 2022; https://www.nice.org.uk/advice/mib294. [Accessed 1 July 2024].
31. Zhu S, Gilbert M, Chetty I, Siddiqui F. The 2021 landscape of FDA-approved artificial intelligence/machine learning-enabled medical devices: An analysis of the characteristics and intended use. Int J Med Inform. 2022; *165*, 104828.
32. FDA. Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. 2022; https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices. [Accessed 1 July 2024].
33. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. The Lancet Digital Health. 2021; *3*, e195-e203. https://doi.org/10.1016/S2589-7500(20)30292-2.
34. Shahzad R, Ayub B, Siddiqui MAR. Quality of reporting of randomised controlled trials of artificial intelligence in healthcare: a systematic review. BMJ Open. 2022; *12*: e061519.
35. Plana D, Shung DL, Grimshaw AA, Saraf A, Sung JJY, Kann BH. Randomized clinical trials of machine learning interventions in health care: a systematic review. JAMA Netw Open. 2022; *5*: e2233946-e2233946.
36. Inzani F, Santoro A, Angelico G, Mastrosimini MG, Masciullo V, Salutari V et al. (2022). Neuroendocrine tumor (NET) of the vagina in the light of WHO 2020 2-tiered grading system: clinicopathological report of the first described case. Virchows Archiv. 2022; *480*: 687-91.
37. Choschzick M, Alyahiaoui M, Ciritsis A, Rossi C, Gut A, Hejduk P et al. Deep learning for the standardized classification of Ki-67 in vulva carcinoma: A feasibility study. Heliyon. 2021; *7*: e07577.
38. Collins GS, Dhiman P, Navarro CLA, Ma J, Hooft L, Reitsma JB et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open. 2021; *11*: e048008. 10.1136/bmjopen-2020-048008.
39. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nature Medicine. 2022; *28*: 31-8. 10.1038/s41591-021-01614-0.
40. Vasey B, Collins GS. Transparent reporting of artificial intelligence models development and evaluation in surgery: The TRIPOD and DECIDE-AI checklists. Surgery. 2023; 174:727-29

41. Sounderajah V, Normahani P, Aggarwal R, Jayakumar, S, Markar SR, Ashrafian H. Reporting Standards and Quality Assessment Tools in Artificial Intelligence–Centered Healthcare Research. In: Lidströmer N, Ashrafian H (eds.) Artificial Intelligence in Medicine. Springer International Publishing. 2022; 385-395
42. Sachdeva G, Han D, Keane PA, Denniston AK, Liu X. Demonstrating Clinical Impact for AI Interventions. In: Sachdeva G, Han D, Keane PA, Denniston AK, Liu X (eds). AI in Clinical Medicine. Wiley. 2023: 459-468.
43. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, Chan AW et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nature Medicine. 2020; *26*: 1364-74.
44. Ibrahim H, Liu X, Rivera SC, Moher D, Chan A-W, Sydes MR et al. (2021). Reporting guidelines for clinical trials of artificial intelligence interventions: the SPIRIT-AI and CONSORT-AI guidelines. Trials. 2021; *22*: 11
45. Cruz Rivera S, Liu X, Chan A-W, Denniston AK, Calvert MJ, Darzi A et al (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nature Medicine. 2020; *26*: 1351-63.
46. Cruz Rivera S, Liu X, Chan A-W, Denniston AK, Calvert MJ, Ashrafian H et al. (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. The Lancet Digital Health. 2020; *2*: e549-e560.
47. Sounderajah V, Ashrafian H, Aggarwal R, De Fauw J, Denniston AK, Greaves F et al. (2020). Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: The STARD-AI Steering Group. Nat Med. 2020; *26*: 807-8.
48. Lee J, Mulder F, Leeflang M, Wolff R, Whiting P, Bossuyt PM. QUAPAS: An Adaptation of the QUADAS-2 Tool to Assess Prognostic Accuracy Studies. Ann Intern Med. 2022; *175*, 1010-1018. 10.7326/m22-0276.
49. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022; *28*: 924-33.
50. Davids J, Ashrafian H. AIM in Nanomedicine. In: Lidströmer N, Ashrafian H (eds.) Artificial Intelligence in Medicine. Springer International Publishing. 2022: 1169-1185.
51. Gospic KAM, Passmore G. Importance of AI in Medicine. In: Lidströmer N, Ashrafian H (eds.) Artificial Intelligence in Medicine. Springer International Publishing. 2022: 99-114.
52. Mistry P. The New Frontiers of AI in Medicine. In: Lidströmer N, Ashrafian H (eds.) Artificial Intelligence in Medicine. Springer International Publishing. 2022: 115-27.
53. Ley C, Martin RK, Pareek A, Groll A, Seil R, Tischer T. Machine learning and conventional statistics: making sense of the differences. Knee Surg Sports Traumatol Arthrosc. 2022; *30*: 753-7.
54. Valiente Fernández M, Lesmes González de Aledo A, Martín Badía I, Delgado Moya FdP. Comparing traditional regression and machine learning models in predicting acute respiratory distress syndrome mortality. Crit Care Med. 2024; *52: 104-6*
55. Villar J, González-Martín JM, Hernández-González J, Armengol MA, Fernández C, Martín-Rodríguez C et al. Predicting ICU Mortality in Acute Respiratory Distress Syndrome Patients Using Machine Learning: The Predicting Outcome and STratifiCation of severity in ARDS (POSTCARDS) Study. Crit Care Med. 2023; *51*: 1638-49.
56. Bewick V, Cheek L, Ball J. Statistics review 7: Correlation and regression. Crit Care. 2003; *7*, 451-9.
57. Bewick V, Cheek L, Ball J. Statistics review 14: Logistic regression. Crit Care. 2005; *9*: 112-18.
58. Enshaei A, Robson CN, Edmondson RJ. Artificial intelligence systems as prognostic and predictive tools in ovarian cancer. Ann Surg Oncol. 2015; *22*: 3970-75.
59. Park SH, Han K, Jang HY, Park JE, Lee J-G, Kim DW et al. Methods for Clinical Evaluation of Artificial Intelligence Algorithms for Medical Diagnosis. Radiology. 2023; *306*: 20-31.

60. Mandrekar JN. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. J Thorac Oncol. 2010; *5*: 1315-16.
61. Dürst M, Gissmann L, Ikenberg H, zur Hausen H (1983). A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. Proc Natl Acad Sci USA, 1983; *80*: 3812-15.
62. zur Hausen H. Papillomaviruses in the causation of human cancers — a brief historical account. Virology. 2009; *384*: 260-265.
63. Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol. 1999; *189*, 12-19.
64. Kitchener HC, Almonte M, Thomson C, Wheeler P, Sargent A, Stoykov, B et al. HPV testing in combination with liquid-based cytology in primary cervical screening (ARTISTIC): a randomised controlled trial. The Lancet Oncology. 2009; *10*: 672-82.
65. Kitchener HC, Almonte M, Gilham C, Dowie R. Stoykova B, Sargent A et al. ARTISTIC: A randomised trial of human papillomavirus (HPV) testing in primary cervical screening. Health Technol Assess. 2009; *13*, 1-126.
66. Gilham C, Sargent A, Kitchener HC, Peto J. HPV testing compared with routine cytology in cervical screening: Long-term follow-up of ARTISTIC RCT. Health Technol Assess. 2019; *23*: 1-43.
67. Boon SS, Luk HY, Xiao C, Chen Z, Chan PKS. Review of the Standard and Advanced Screening, Staging Systems and Treatment Modalities for Cervical Cancer. Cancers (Basel). 2022; *14*: 2913.
68. Crafton SM, Venkat PS, Salani R. A review of the state of cervical cancer: updates from prevention to recurrent disease. Curr Opin Obstet Gynecol. 2024; *36*, 28-33.
69. Montgomery A, Durden A, Sundararajan S, Al-Booz H, Newton C. Review of invasive cervical cancer. Obstetrics, Gynaecology & Reproductive Medicine. 2023; *33*: 281-85.
70. Globocan. Absolute numbers, Incidence, Females, in 2022. 2024; Globocan: https://gco.iarc.fr/today/en/dataviz/pie?mode=cancer&group_populations=1&sexes=2&populations=900. [Accessed 2 July 2024]
71. Falcaro M, Castañon A, Ndlela B, Checchi M, Soldan K, Lopez-Bernal J et al. The effects of the national HPV vaccination programme in England, UK, on cervical cancer and grade 3 cervical intraepithelial neoplasia incidence: a register-based observational study. Lancet. 2021; *398*: 2084-92.
72. Quinn M. Babb P, Jones J, Allen E. Effect of screening on incidence of and mortality from cancer of cervix in England: evaluation based on routinely collected statistics. BMJ. 1999; *318*: 904-08.
73. Markovina S, Rendle KA, Cohen AC, Kuroki LM, Grover S, Schwarz JK. Improving cervical cancer survival–A multifaceted strategy to sustain progress for this global problem. Cancer. 2022; *128*: 4074-84.
74. World Health Organization. Cervical Cancer. WHO. 2024: https://www.who.int/health-topics/cervical-cancer#tab=tab_1. [Accessed 2 July 2024].
75. World Health Organization. Global strategy to accelerate the elimination of cervical cancer as a public health problem. WHO. 2020: https://www.who.int/publications/i/item/9789240014107. [Accessed 2 July 2024].
76. Idlahcen F, Idri A, Goceri E. Exploring data mining and machine learning in gynecologic oncology. Artificial Intelligence Review. 2024; *57*: 20.
77. McKenna SJ, Ricketts, IW, Cairns AY, Hussein KA (1993). A comparison of neural network architectures for cervical cell classification. Dundee University. 1993; https://staff.computing.dundee.ac.uk/stephen/cervicalcellclassification.pdf. [Accessed 2 July 2024].

78. Rahaman MM, Li C, Wu X, Yao Y, Hu Z, Jiang T et al. A Survey for Cervical Cytopathology Image Analysis Using Deep Learning. IEEE Access. 2020; *99: 1.*
79. Jiang P, Li X, Shen H, Chen Y, Wang L, Chen H et al. A systematic review of deep learning-based cervical cytology screening: from cell identification to whole slide image analysis. Artificial Intelligence Review. 2023; *56*: 2687-758.
80. Bengtsson E, Malm P. Screening for cervical cancer using automated analysis of PAP-smears. Comput Math Methods Med. *2014*; 842037.
81. Tolles WE. Section of biology: the cytoanalyzer—an example of physics in medical research*. Transactions of the New York Academy of Sciences. 1955; *17*: 250-6.
82. Wilbur DC, Black-Schaffer WS, Luff RD, Abraham KP, Kemper C, Molina JT et al. (2009). The Becton Dickinson FocalPoint GS Imaging System: Clinical Trials Demonstrate Significantly Improved Sensitivity for the Detection of Important Cervical Lesions. Am J Clin Pathol. 2009; *132*: 767-75.
83. Biscotti CV, Dawson AE, Dziura B, Galup L, Darragh T, Rahemtulla A et al. Assisted Primary Screening Using the Automated ThinPrep Imaging System. Am J Clin Pathol. 2005; *123*: 281-87.
84. Kitchener HC, Blanks R, Dunn G, Gunn L, Desai M, Albrow R et al. (2011). Automation-assisted versus manual reading of cervical cytology (MAVARIC): a randomised controlled trial. Lancet Oncol. 2011; *12*:
85. Mango LJ. The FDA Review Process: Obtaining Premarket Approval for the PAPNET Testing System. Acta Cytol. 1996; *40*: 138-40.
86. Valente PT, Schantz HD. Cytology automation: An overview. Laboratory Medicine. 2001; *32*: 686-90.
87. Luck R, Tjon-Fo-Sang R, Mango L, Recht J, Lin E, Knapp J. PAPNET TM: an automated cytology screener using image processing and neural networks. SPIE Digital Library. 1992; https://www.spiedigitallibrary.org/conference-proceedings-of-spie/1623/0000/PAPNET-TM--an-automated-cytology-screener-using-image-processing/10.1117/12.58066.short. [Accessed 2 July 2024]
88. Gutierrez-Espinoza S, Cabanillas-Carbonell M. Machine Learning Analysis for Cervical Cancer Prediction, a Systematic Review of the Literature. *2021 International Conference on e-Health and Bioengineering (EHB)*, Iasi, Romania. 2021; 1-6
89. Rahimi M, Akbari A, Asadi F, Emami H. Cervical cancer survival prediction by machine learning algorithms: a systematic review. BMC Cancer. 2023; *23*, 341.
90. Shen M, Zou Z, Bao H, Fairley CK, Canfell K, Ong JJ et al. Cost-effectiveness of artificial intelligence-assisted liquid-based cytology testing for cervical cancer screening in China. Lancet Reg Health West Pac. 2023; *34*: 100726.
91. Holmström O, Linder N, Kaingu H, Mbuuko N, Mbete J, Kinyua F et al. (2021). Point-of-care digital cytology with artificial intelligence for cervical cancer screening in a resource-limited setting. JAMA Network Open. 2021; *4*: e211740
92. Bao H, Sun X, Zhang Y, Pang B, Li H, Zhou L et al. The artificial intelligence-assisted cytology diagnostic system in large-scale cervical cancer screening: A population-based cohort study of 0.7 million women. Cancer Med. 2020; *9*, 6896-906.
93. World Health Organization. WHO guideline for screening and treatment of cervical pre-cancer lesions for cervical cancer prevention. WHO. 2021; https://www.who.int/publications/i/item/9789240030824 [Accessed 2 July 2024].
94. Viñals R, Jonnalagedda M, Petignat P, Thiran J-P, Vassilakos P(2023). Artificial Intelligence-Based Cervical Cancer Screening on Images Taken during Visual Inspection with Acetic Acid: A Systematic Review. Diagnostics (Basel). *13*, 836.
95. Guo P, Xue Z, Angara S, Antani SK. Unsupervised Deep Learning Registration of Uterine Cervix Sequence Images. Cancers. 2022; *14*: 2401.

96. Chen C, Cao Y, Li W, Liu Z, Liu P, Tian X et al. The pathological risk score: A new deep learning-based signature for predicting survival in cervical cancer. Cancer Med. 2023; *12*: 1051-63.
97. Wang CW, Liou YA, Lin YJ, Chang CC, Chu PH, Lee YC et al. Artificial intelligence-assisted fast screening cervical high grade squamous intraepithelial lesion and squamous cell carcinoma diagnosis and treatment planning. Sci Rep. 2021; *11*: 16244.
98. Reed N, Balega J, Barwick T, Buckley L, Burton K, Eminowicz G et al. British Gynaecological Cancer Society (BGCS) cervical cancer guidelines: Recommendations for practice. Eur J Obstet Gynecol Reprod Biol. 2021; *256*: 433-65.
99. Bedrikovetski S, Dudi-Venkata NN, Maicas G, Kroon HM, Seow W, Carneiro G et al. Artificial intelligence for the diagnosis of lymph node metastases in patients with abdominopelvic malignancy: A systematic review and meta-analysis. Artif Intell Med. 2021; *113*: 102022.
100. Charoenkwan P, Shoombuatong W, Nantasupha C, Muangmool T, Suprasert P, Charoenkwan K. iPMI: Machine Learning-Aided Identification of Parametrial Invasion in Women with Early-Stage Cervical Cancer. Diagnostics (Basel). 2021; *11*: 1454.
101. Yang C, Qin L-H, Xie Y-E, Liao J-Y. Deep learning in CT image segmentation of cervical cancer: a systematic review and meta-analysis. Radiat Oncol. 2022; *17*: 175.
102. Crosbie EJ, Kitson SJ, McAlpine JN, Mukhopadhyay A, Powell ME, Singh N. Endometrial cancer. Lancet. 2022; *399*: 1412-28.
103. Agnew HJ, Kitson SJ, Crosbie EJ. Gynecological malignancies and obesity. Best Pract Res Clin Obstet Gynaecol. 2023; *88*: 102337.
104. Kuhn TM, Dhanani S, Ahmad S. An Overview of Endometrial Cancer with Novel Therapeutic Strategies. Curr Oncol. 2023; *30*: 7904-19.
105. Morrison J, Balega J, Buckley L, Clamp A, Crosbie E, Drew Y et al. British Gynaecological Cancer Society (BGCS) uterine cancer guidelines: Recommendations for practice. Eur J Obstet Gynecol Reprod Biol. 2022; *270*: 50-89.
106. Levine DA, Getz G, Gabriel SB, Cibulskis K, Lander E, Sivachenko A et al. Integrated genomic characterization of endometrial carcinoma. Nature. 2013; *497*, 67-73.
107. Vermij L, Smit V, Nout R, Bosse T. Incorporation of molecular characteristics into endometrial cancer management. Histopathology. 2020; *76*: 52-63.
108. Consortium RR. Refining adjuvant treatment in endometrial cancer based on molecular features: the RAINBO clinical trial program. Int J Gynecol Cancer. 2022; 33:109-17.
109. Berek JS, Matias-Guiu X, Creutzberg C, Fotopoulou C, Gaffney D, Kehoe S et al. (2023). FIGO staging of endometrial cancer: 2023. J Gynecol Oncol. 2023; *34: e85*
110. Jumaah AS, Al-Haddad HS, McAllister KA, Yasseen AA. The clinicopathology and survival characteristics of patients with POLE proofreading mutations in endometrial carcinoma: A systematic review and meta-analysis. PLoS One. 2022; *17*: e0263585.
111. Yusof MNM, Chew KT, Hafizz AMHA, Azman SHA, Razak WSA, Hamizan MRu et al (2023). Efficacy and Safety of PD-1/PD-L1 Inhibitor as Single-Agent Immunotherapy in Endometrial Cancer: A Systematic Review and Meta-Analysis. Cancers (Basel). 2023; *15*: 4032.
112. Casanova J, Duarte GS, da Costa AG, Catarino A, Nave M, Antunes T et al. Prognosis of polymerase epsilon (POLE) mutation in high-grade endometrioid endometrial cancer: Systematic review and meta-analysis. Gynecol Oncol. 2024; *182*, 99-107.
113. Fremond S, Andani S, Barkey Wolf J, Dijkstra J, Melsbach S, Jobsen JJ et al. Interpretable deep learning model to predict the molecular classification of endometrial cancer from haematoxylin and eosin-stained whole-slide images: a combined analysis of the PORTEC randomised trials and clinical cohorts. Lancet Digit Health. 2023; *5*: e71-e82.
114. Chen X, Wang Y, Shen M, Yang B, Zhou Q, Yi Y 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. 2020; *30*, 4985-94.

115. Mao W, Chen C, Gao H, Xiong L, Lin Y. Quantitative evaluation of myometrial infiltration depth ratio for early endometrial cancer based on deep learning. Biomedical Signal Processing and Control. 2023; *84*: 104685
116. Stanzione A, Cuocolo R, Del Grosso R, Nardiello A, Romeo V, Travaglino A et al. (2021). Deep Myometrial Infiltration of Endometrial Cancer on MRI: A Radiomics-Powered Machine Learning Pilot Study. Acad Radiol. 2021; *28*: 737-44.
117. Mariani A, Webb MJ, Keeney GL, Haddock MG, Calori G, Podratz KC. Low-risk corpus cancer: Is lymphadenectomy or radiotherapy necessary? Am J Obstet Gynecol. 2000; *182*: 1506-19.
118. Ren Z, Chen B, Hong C, Yuan J, Deng J, Chen Y et al (2023). The value of machine learning in preoperative identification of lymph node metastasis status in endometrial cancer: a systematic review and meta-analysis. Front Oncol. 2023; *13*: 1289050.
119. Yan BC, Li Y, Ma FH, Zhang GF, Feng F, Sun MH et al. Radiologists with MRI-based radiomics aids to predict the pelvic lymph node metastasis in endometrial cancer: a multicenter study. Eur Radiol. 2021; *31*: 411-22.
120. Lecointre L, Dana J, Lodi M, Akladios C, Gallix B. Artificial intelligence-based radiomics models in endometrial cancer: A systematic review. Eur J Surg Oncol. 2021; *47*: 2734-41.
121. Ravegnini G, Ferioli M, Morganti AG, Strigari L, Pantaleo MA, Nannini M et al. Radiomics and Artificial Intelligence in Uterine Sarcomas: A Systematic Review. J Pers Med. 2021; *11*: 1179.
122. Toyohara Y, Sone K, Noda K, Yoshida K, Kurokawa R, Tanishima T et al. Development of a deep learning method for improving diagnostic accuracy for uterine sarcoma cases. Sci Rep. 2022; *12*, 19612.
123. Veneziani AC, Gonzalez-Ochoa E, Alqaisi H, Madariaga A, Bhat G, Rouzbahman M et al. Heterogeneity and treatment landscape of ovarian carcinoma. Nat Rev Clin Oncol. 2023; *20*, 820-42.
124. Hollis RL. Molecular characteristics and clinical behaviour of epithelial ovarian cancers. Cancer Lett. 2023; *555*: 216057.
125. WHO Classification of Tumours Editorial Board. WHO Classification of Tumours, 5th edn, vol 4: Female Gential Tumours. WHO. 2020; IARC Publications.
126. Grisham RN, Manning-Geist BL, Chui MH. The highs and lows of serous ovarian cancer. Cancer. 2023; *129*: 2613-20.
127. Berek JS, Renz M, Kehoe S, Kumar L, Friedlander M. Cancer of the ovary, fallopian tube, and peritoneum: 2021 update. Int J Gynecol Obstet. 2021; *155*: 61-85.
128. McCluggage WG, Singh N, Gilks CB. Key changes to the World Health Organization (WHO) classification of female genital tumours introduced in the 5th edition (2020). Histopathology. 2022; *80*: 762-78.
129. Burling MJ, Gamet K, Eva L, Tan AL. Referral patterns for genetic counselling of women diagnosed with tubo-ovarian or peritoneal high-grade serous carcinoma (HGSC) within the Auckland Gynaecological Oncology Centre. Aust N Z J Obstet Gynaecol. 2019; *59*, 444-49.
130. Zhou J, Cao W, Wang L, Pan Z, Fu Y. Application of artificial intelligence in the diagnosis and prognostic prediction of ovarian cancer. Comput Biol Med. 2022; *146*: 105608.
131. Wu M, Yan C, Liu H, Liu Q. Automatic classification of ovarian cancer types from cytological images using deep convolutional neural networks. Biosci Rep. 2018; *38*: BSR20180289.
132. Nelson-Piercy C, Dean C, Shehmar M, Gadsby R, O'Hara M, Hodson K, et al; the Royal College of Obstetricians and Gynaecologists. The Management of Nausea and Vomiting in Pregnancy and Hyperemesis Gravidarum (Green-top Guideline No. 69). BJOG. 2024; 131(7): e1–e30
133. Kaloo PD, Louden KA, Khazzali S, Hoy D, Sadoon SL the Royal College of Obstetricians and Gynaecologists. The Management of Suspected Ovarian Masses in Premenopausal Women. Green-top Guideline No.62. 2011; https://www.rcog.org.uk/media/yhujmdvr/gtg_62-1.pdf [Accessed 2 July 2024].

134. Ma L, Huang L, Chen Y, Zhang L, Nie D, He W et al. AI diagnostic performance based on multiple imaging modalities for ovarian tumor: A systematic review and meta-analysis. Front Oncol. 2023; *13*: 1133491.
135. Xu HL, Gong TT, Liu FH, Chen HY, Xiao Q, Hou, Y et al. Artificial intelligence performance in image-based ovarian cancer identification: A systematic review and meta-analysis. EClinicalMedicine. 2022; *53*, 101662.
136. Mitchell S, Nikolopoulos M, El-Zarka A, Al-Karawi D, Al-Zaidi S, Ghai A et al. Artificial Intelligence in Ultrasound Diagnoses of Ovarian Cancer: A Systematic Review and Meta-Analysis. Cancers. 2024; *16*: 422.
137. Jan Y-T, Tsai P-S, Huang W-H, Chou L-Y, Huang S-C, Wang J-Z et al. Machine learning combined with radiomics and deep learning features extracted from CT images: a novel AI model to distinguish benign from malignant ovarian tumors. Insights Imaging. 2023; *14*: 68.
138. Li Ya, Jian J, Pickhardt PJ, Ma F, Xia W, Li H et al. MRI-Based Machine Learning for Differentiating Borderline From Malignant Epithelial Ovarian Tumors: A Multicenter Study. J Magn Reson Imaging. 2020; *52*: 897-904.
139. Jian J, Li Ya, Xia W, He Z, Zhang R, Li H et al. MRI-Based Multiple Instance Convolutional Neural Network for Increased Accuracy in the Differentiation of Borderline and Malignant Epithelial Ovarian Tumors. J Magn Reson Imaging. 2022; *56*: 173-81.
140. Wang Y, Zhang H, Wang T, Yao L, Zhang G, Liu X et al. Deep learning for the ovarian lesion localization and discrimination between borderline and malignant ovarian tumors based on routine MR imaging. Sci Rep. 2023; *13*, 2770.
141. Barcroft JF, Linton-Reid K, Landolfo C, Al-Memar M, Parker N et al. Machine learning and radiomics for segmentation and classification of adnexal masses on ultrasound. NPJ Precis Oncol. 2024; *8*: 41.
142. Lu M, Fan Z, Xu B, Chen L, Zheng X, Li J et al. Using machine learning to predict ovarian cancer. Int J Med Inform. 2020; *141*: 104195.
143. Reilly G, Bullock RG, Greenwood J, Ure DR, Stewart E, Davidoff P et al. Analytical Validation of a Deep Neural Network Algorithm for the Detection of Ovarian Cancer. JCO Clin Cancer Inform. 2022; e2100192.
144. Ahamad MM, Aktar S, Uddin MJ, Rahman T, Alyami SA, Al-Ashhab S et al. Early-Stage Detection of Ovarian Cancer Based on Clinical Data Using Machine Learning Approaches. J Pers Med. 2022; *12*: 1211.
145. Yu XY, Ren J, Jia Y, Wu H, Niu G, Liu A et al. Multiparameter MRI Radiomics Model Predicts Preoperative Peritoneal Carcinomatosis in Ovarian Cancer. Front Oncol. 2021; *11*: 765652.
146. Beer L, Sahin H, Bateman NW, Blazic I, Vargas HA, Veeraraghavan H et al. Integration of proteomics with CT-based qualitative and radiomic features in high-grade serous ovarian cancer patients: an exploratory analysis. Eur Radiol. 2020; *30*: 4306-16.
147. Cheung AK, Ko JM, Lung HL, Chan KW, Stanbridge EJ, Zabarovsky E et al. Cysteine-rich intestinal protein 2 (CRIP2) acts as a repressor of NF-kappaB-mediated proangiogenic cytokine transcription to suppress tumorigenesis and angiogenesis. Proc Natl Acad Sci U S A. 2011; *108*: 8390-95.
148. Zhou L, Wang Y, Zhou M, Zhang Y, Wang P, Li X et al. HOXA9 inhibits HIF-1alpha-mediated glycolysis through interacting with CRIP2 to repress cutaneous squamous cell carcinoma development. Nat Commun. 2018; *9*, 1480.
149. Lu H, Lou H, Wengert G, Paudel R, Patel N, Desai S et al. Tumor and local lymphoid tissue interaction determines prognosis in high-grade serous ovarian cancer. Cell Rep Med. 2023; *4*: 101092.
150. Lu H, Arshad M, Thornton A, Avesani G, Cunnea P, Curry E et al. A mathematical-descriptor of tumor-mesoscopic-structure from computed-tomography images annotates prognostic- and molecular-phenotypes of epithelial ovarian cancer. Nat Commun. 2019; *10*: 764.

151. Crispin-Ortuzar M, Woitek R, Reinius MAV, Moore E, Beer L, Bura V et al. Integrated radiogenomics models predict response to neoadjuvant chemotherapy in high grade serous ovarian cancer. Nat Commu. 2023; *14:* 6756.
152. Boehm KM, Aherne EA, Ellenson L, Nikolovski I, Alghamdi M, Vázquez-García I et al. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. Nat Cancer. 2022; *3*: 723-33.
153. Yao H, Zhang X. Prediction model of residual neural network for pathological confirmed lymph node metastasis of ovarian cancer. BioMed Res Int. *2022*: 9646846.
154. Fotopoulou C, Hall M, Cruickshank D, Gabra H, Ganesan R, Hughes C et al. British Gynaecological Cancer Society (BGCS) epithelial ovarian/fallopian tube/primary peritoneal cancer guidelines: recommendations for practice. Eur J Obstet Gynecol Reprod Biol. 2017; *213*: 123-39.
155. Laios A, Gryparis A, DeJong D, Hutson R, Theophilou G, Leach C. Predicting complete cytoreduction for advanced ovarian cancer patients using nearest-neighbor models. J Ovarian Res. 2020; *13*: 117.
156. Parpinel G, Laudani ME, Piovano E, Zola P, Lecuru F. The use of artificial intelligence for complete cytoreduction prediction in epithelial ovarian cancer: A narrative review. Cancer Control. 2023; *30*: 10732748231159553.
157. Laios A, Kalampokis E, Johnson R, Munot S, Thangavelu A, Hutson R et al. Development of a novel intra-operative score to record diseases’ anatomic fingerprints (anafi score) for the prediction of complete cytoreduction in advanced-stage ovarian cancer by using machine learning and explainable artificial intelligence. Cancers (Basel). 2023; *15*: 966.
158. Maubert A, Birtwisle L, Bernard JL, Benizri E, Bereder JM. Can machine learning predict resecability of a peritoneal carcinomatosis? Surgical Oncology. 2019; *29*: 120-25.
159. Piedimonte S, Erdman L, So D, Bernardini MQ, Ferguson SE, Laframboise S et al (2023). Using a machine learning algorithm to predict outcome of primary cytoreductive surgery in advanced ovarian cancer. J Surg Oncol. 2023; *127*: 465-72.
160. Ho DJ, Chui MH, Vanderbilt CM, Jung J, Robson ME, Park CS et al. Deep Interactive Learning-based ovarian cancer segmentation of H&E-stained whole slide images to study morphological patterns of BRCA mutation. J Pathol Inform. 2023; *14*: 100160.
161. Nero C, Boldrini L, Lenkowicz J, Giudice MT, Piermattei A, Inzani F et al. Deep-learning to predict brca mutation and survival from digital h&e slides of epithelial ovarian cancer. Int J Mol Sci. 2022; *23*: 10.3390/ijms231911326.
162. Bogaerts JMA, Bokhorst J-M, Simons M, van Bommel MHD, Steenbeek MP, de Hullu JA et al. Deep learning detects premalignant lesions in the Fallopian tube. npj Women's Health. 2024; *2*: 11.
163. Wang C-W, Chang C-C, Lee Y-C, Lin Y-J, Lo S-C, Hsu P-C et al. Weakly supervised deep learning for prediction of treatment effectiveness on ovarian cancer from histopathology images. Comput Med Imaging Graph. 2022; *99*: 102093.
164. Ma J, Yang J, Jin Y, Cheng S, Huang S, Zhang N et al. Artificial intelligence based on blood biomarkers including ctcs predicts outcomes in epithelial ovarian cancer: A prospective study. Onco Targets Ther. 2021; *14*: 3267-80.
165. Farinella F, Merone M, Bacco L, Capirchio A, Ciccozzi M, Caligiore D. Machine Learning analysis of high-grade serous ovarian cancer proteomic dataset reveals novel candidate biomarkers. Sci Rep. 2022; *12*: 3041.
166. Huang C, Clayton EA, Matyunina LV, McDonald LD, Benigno BB, Vannberg F et al. Machine learning predicts individual cancer patient responses to therapeutic drugs with high accuracy. Sci Rep. 2018; *8*: 16444.
167. Chen K, Xu H, Lei Y, Lio P, Li Y, Guo H et al. Integration and interplay of machine learning and bioinformatics approach to identify genetic interaction related to ovarian cancer chemoresistance. Brief Bioinform. 2021; *22*: bbab100.

168. Hossain MA, Saiful Islam SM, Quinn JMW, Huq F, Moni MA. Machine learning and bioinformatics models to identify gene expression patterns of ovarian cancer associated with disease progression and mortality. J Biomed Inform. 2019; *100*: 103313.
169. WHO. Development of guidance on ethics and governance of artificial intelligence for health. WHO consultation. WHO. 2021: https://iris.who.int/bitstream/handle/10665/340089/9789240012752-eng.pdf?sequence=1. [Accessed 2 July 2024].
170. Lehmann LS. Ethical Challenges of Integrating AI into Healthcare. In: Lidströmer N, Ashrafian H (eds). Artificial Intelligence in Medicine. Springer International Publishing. 2022; 139-44.
171. Ploug T, Holm S. Right to Contest AI Diagnostics. In: Lidströmer N, Ashrafian H (eds). Artificial Intelligence in Medicine. Springer International Publishing. 2022; 227-38.
172. Scott IA, Carter SM, Coiera E. Exploring stakeholder attitudes towards AI in clinical practice. BMJ Health & Care Informatics. 2021; *28*: e100450.
173. Xue P, Tang C, Li Q, Li Y, Shen Y, Zhao Y et al. Development and validation of an artificial intelligence system for grading colposcopic impressions and guiding biopsies. BMC Medicine. 2020; *18*, 406.
174. Chen M, Wang J, Xue P, Li Q, Jiang Y, Qiao Y. Evaluating the feasibility of machine-learning-based predictive models for precancerous cervical lesions in patients referred for colposcopy. Diagnostics (Basel). 2022; *12*: 3066.
175. Chen M, Zhang B, Cai Z, Seery S, Gonzalez MJ, Ali NM. (2022). Acceptance of clinical artificial intelligence among physicians and medical students: A systematic review with cross-sectional survey. Front Med (Lausanne). 2022; *9*: 990604.
176. Wan YL, Beverley-Stevenson R, Carlisle D, Clarke S, Edmondson RJ, Glover S et al. Working together to shape the endometrial cancer research agenda: The top ten unanswered research questions. Gynecol Oncol. 2016; *143*: 287-93.
177. McCormack M, Gaffney D, Tan D, Bennet K, Chavez-Blanco A, Plante M. The Cervical Cancer Research Network (Gynecologic Cancer InterGroup) roadmap to expand research in low- and middle-income countries. Int J Gynecol Cancer. 2021; *31*: 775-78.
178. Bast Jr, RC, Matulonis UA, Sood AK, Ahmed AA, Amobi AE, Balkwill FR et al. Critical questions in ovarian cancer research and treatment: Report of an American Association for Cancer Research Special Conference. Cancer. 2019; *125*: 1963-72. https://doi.org/10.1002/cncr.32004.
179. Vergote I, Gonzalez-Martin A, Lorusso D, Gourley C, Mirza MR, Kurtz J-E et al. Clinical research in ovarian cancer: consensus recommendations from the Gynecologic Cancer InterGroup. Lancet Oncol. 2022; *23*: e374-e384.
180. Virani S, Baiocchi G, Bowtell D, Cabasag CJ, Cho KR, Fortner RT et al. Joint IARC/NCI International Cancer Seminar Series Report: expert consensus on future directions for ovarian carcinoma research. Carcinogenesis. 2021; *42*: 785-93.
181. Futoma J, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research and machine learning in health care. Lancet Digital Health. 2020; *2*: e489-e492.
182. Kumar A, Tejaswini P, Nayak O, Kujur AD, Gupta R, Rajanand A et al. A Survey on IBM Watson and Its Services. Journal of Physics: Conference Series. 2022; *2273*: 012022.
183. Scott RE, Mars M. Principles and Framework for eHealth Strategy Development. J Med Internet Res. 2013; *15*: e155.
184. Emani S, Rui A, Rocha HAL, Rizvi RF, Juaçaba SF, Jackson GP et al. Physicians' Perceptions of and Satisfaction With Artificial Intelligence in Cancer Treatment: A Clinical Decision Support System Experience and Implications for Low-Middle-Income Countries. JMIR Cancer. 2022; *8*: e31461.
185. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface. 2018; *15*, 20170387.