https://doi.org/10.1007/s44352-025-00019-w

Discover Imaging

RESEARCH Open Access

Automated endometrial segmentation, thickness measurement and pattern prediction on uterine ultrasound images

Hannah E. Pierson[1†], Zachary Shand[2†], Jesse Invik[1], Yang Wang[2], Vishwajeet Ohal[2], Kane Smith[2], Devanshi Patel[2] and Roger A. Pierson[1,3*]

†Hannah E. Pierson and Zachary Shand are co-first authors.

*Correspondence:
Roger A. Pierson
roger.pierson@usask.ca
[1]Synergyne Imaging Technology, Inc., Canmore, AB, Canada
[2]Cybera Inc., Calgary, AB, Canada
[3]Obstetrics and Gynecology, College of Medicine, University of Saskatchewan, Saskatoon, SK, Canada

Abstract

Purpose To develop an AI-based model for automated identification, thickness measurement, and pattern classification of endometrial tissues in uterine ultrasound images.

Methods We utilized a multi-stage model building and training approach with different methods for segmentation, thickness measurement, and pattern assignment. 5985 unique annotated ultrasound images were utilized in model training and validation. The annotated images were randomly divided into three groups (training data, validation data, and test data) at a 70%/10%/20% ratio. 4787 images were used to develop the model; 1198 images were used to evaluate performance. Identification of the endometrium was done utilizing direct object detection and instance segmentation systems with single stage object detectors (YOLO v8). A principal component analysis approach was adapted to determine endometrial thickness. We also trained computational models to assign an endometrial pattern of either 'trilaminar' or 'homogenous'. Intersection over union (IoU), confusion matrices, and error calculations were conducted to assess the model's proficiency and accuracy.

Results The endometrial segmentation model performed at 98% accuracy on the confusion matrix and 92.8% of test data had an intersection over union (IoU) value > 0.75. The endometrial thickness measurement achieved mean absolute error rates of: ± 0.89 mm in length, ± 2.81 mm in position (along the perpendicular-to-lumen axis), and 5.5-degrees in orientation relative to the lumen. Automatic assignment of pattern as either 'trilaminar' or 'homogenous' achieved a 92% accuracy rate.

Conclusions A novel automated method for routine endometrial thickness and pattern assessment is demonstrated. We report a first-of-its kind method for automated endometrial pattern assignment (homogenous / triple-line) on uterine ultrasound images.

1 Introduction

Ultrasonographic measurement of endometrial thickness (ET) and echotextural pattern has been broadly adopted as the standard of care assessment of the endometrium prior to embryo transfer in medically assisted reproduction (MAR) cycles. There are two important considerations that must be acknowledged regarding routine ET and pattern assessments. First, although ET and pattern assessment is commonplace for in vitro fertilization (IVF) therapy, measurement and pattern assignment are subject to broad interpretation and significant error. It is important to acknowledge that variability in the method of imaging, caliper placement, image interpretation, and potential human bias around target thickness cutoffs can be problematic [1, 2]. Second, the clinical utility of ET and pattern assessments is the topic of significant debate [2]. Irrespectively, ET and pattern assignments are widely utilized and are unlikely to be abandoned as a standard of care in MAR cycles. The question then becomes, how do we improve accuracy, reliability, and expediency of these assessments? Standardization and automation of image interpretation could broadly improve measurement objectivity and accuracy, as well as reducing the workload for clinics.

Artificial intelligence (AI) and machine learning (ML) approaches to various aspects of reproductive care are actively being integrated into clinical assessments [3]. Several novel approaches to medical image interpretation are available. Automatic image interpretation has been developed for morphologic oocyte and embryo quality assessments [4–6] and methods for automated ovarian follicle identification and measurement are well established [7, 8]. At present, the common theme among the AI/ML applications to medical image interpretation is that they rely upon imaging methods with high contrast ratios. In the case of automatic follicle detection, follicular fluid is visualized as black polygons in a field of lighter grey tones on the ultrasound images. The echotextural properties of follicular fluid enhances the capacity to execute high efficiency boundary detection between the pixels representing follicular fluid and ovarian stroma. Automatic oocyte and embryo image analytics are similar; the isolated target biomass is readily distinguishable from translucent media in digital microscopy images. The advantages of high contrast imaging lend themselves well to the current capacity of computer vision systems for efficient code development.

Automated approaches to assessing the endometrium are not as straightforward as those for oocytes and embryo quality. [9]. Endometrial tissues are in a constant state of development and differentiation in response to circulating reproductively active hormones. Endometrial tissues cycle through distinct morphological changes over the menstrual cycle: proliferation (follicular phase), differentiation (early luteal phase), functional (secretory phase), and degradation (menses). In addition, the endometrium is composed of two distinct tissue layers, stratum basalis and stratum functionalis, which can be difficult to distinguish from one another and surrounding myometrial tissue depending on cycle stage. The endometrial stratum basalis lines the uterine cavity, interfaces with the myometrium and remains stable in morphology and ultrasonographic appearance over the menstrual cycle. The stratum functionalis interfaces with the stratum basalis, faces the uterine lumen and undergoes significant changes over the menstrual cycle. The stratum basalis is rich in stem / progenitor cells that primarily function to cyclically replenish the stratum functionalis. The s. functionalis is thinnest after menses and becomes thickened through estrogen induced proliferation of endometrial epithelium during the follicular phase. The follicular phase s. functionalis appears hypoechoic relative to s. basalis on ultrasound imaging, creating a tri-laminar pattern on ultrasound images. Maturation / differentiation of the functionalis layer occurs in luteal phase and progressively shifts the ultrasonographic texture of the tissues to a homogenous hyperechoic appearance as branching and coiling of the endometrial glands under the influence of progestins progresses. The fully differentiated s. functionalis imaged in secretory phase is indistinguishable from stratum basalis on ultrasound imaging, creating a homogenous ultrasonographic appearance. During menses, the s. functionalis degrades and is shed while the basalis layer remains and re-initiates the cyclic endometrial changes as the cycle repeats if pregnancy is not established [10].

The significant microanatomic changes that occur over the menstrual cycle pose a challenge to automation of endometrial assessment. Standardization of the timeframe at which imaging is conducted is an obligate requirement with current technologies. In addition, the signal-to-noise ratio is high on ultrasound images of the uterine body and endometrium. Boundaries between tissue layers can be difficult to discern for computer vision systems and the challenge may be amplified when anatomic (such as fibroids or scar tissue) or pathologic features cast acoustic shadows on the endometrium. Few reports have been published with methods for automatic segmentation and measurement of endometrium on ultrasound images [11–13]. All are based on small data sets. Other AI-based models for assessing endometrium have been proposed, however, one extends beyond endometrial assessment and requires manual measurements and input of patient chart data [14] and the other shows lackluster performance on image segmentation [15]. To our knowledge, there are no prior reports of automated methods for endometrial pattern assignment (homogenous / triple-line).

Recent advances in computer vision modelling provides the opportunity to push the leading edge and develop standardized / automated approaches to assist in the assessment of the ET and pattern. Our objective was to develop an AI/ML approach for automation of three key features of routine endometrial assessment: 1) identification and segmentation of endometrial echoes from ultrasound images of the uterus; 2) measurement of ET; and, 3) assignment of endometrial pattern as either trilaminar or homogenous.

2 Materials and methods

2.1 Ultrasound image acquisition and patient demographics

Ultrasound imaging was conducted in partnership with Canadian fertility clinics as their patients proceeded through standard of care IVF/MAR therapy. Patient records dated between 2014 and 2023 were accessed retrospectively. Ultrasound images and patient case information were de-identified for this project. The criterion for patient case selection was limited to availability of the required transverse ultrasound images; no other inclusion or exclusion criteria were considered. Transvaginal ultrasound images were captured according to the following parameters: 1) the probe was positioned to identify the thickest aspect of the endometrial echoes ~ 5–10 mm from the myometrial / endometrial junction at the uterine fundus in mid-sagittal plane of section; 2) the probe was then rotated by 90-degrees to visualize the region of interest in transverse plane; and, 3) transverse images of the uterus were acquired at the thickest aspect of the fundus/body at the level of the uterine cornu. Descriptive statistics were prepared to provide baseline patient demographics.

2.2 Data set

Five thousand nine hundred and eighty-five (n = 5985) unique ultrasound images from 5985 corresponding individual patient cases were utilized in computational model training and validation. Each image was annotated by a team of skilled medical research personnel using purpose-built software developed in our laboratory. Image polygon and line annotations were utilized to demarcate the endometrial / myometrial interface, measure ET, and assign the endometrial pattern on each image. All images were successfully annotated. Annotations were conducted by a team of 3 individuals. To minimize inter-observer biases or inconsistencies in annotation each annotator received one-on-one training and was provided with a detailed standard operating procedure manual on label creation and pattern assignment. Additionally, each image file annotated by one individual was reviewed and validated by one of the other two individuals. Cross-validation of the annotations by two individuals was designed to reduce errors and biases. Any case where a consensus could not be reached regarding appropriate placement of the annotation markers or pattern assignment was reviewed and confirmed by the most experienced senior individual.

Image preprocessing was limited to cropping-out patient identifiers from the top ~ 10% of the ultrasound images for de-identification purposes. In addition, the standard Yolo-V8 model enables several image augmentations during training. The default image resizing (640px × 640px), mosaic augmentations, image scaling, translations, and horizontal flipping were utilized. Ad-hoc testing showed shear, HSV augmentation and vertical flipping either did not improve or slightly worsened the training results. The annotated image database was randomly divided into standard train/validation/test splits of 70%/10%/20%, corresponding to 4189 training images, 598 validation images, and 1198 testing images. These splits were selected as the standard machine learning splits. Follow up of the training size versus accuracy metrics plateau between 500 to 1000 images. With a training size of 4189 images and no strong class imbalances, no cross validation was deemed necessary. The training and validation sets were jointly used for fitting and hyperparameter tuning while the 1198 test images were used for reporting accuracy and other performance metrics.

2.3 Model building: endometrial segmentation and thickness measurement

Image segmentation is a process that partitions an image into multiple regions based on visual or statistical characteristics [16]. Segmentation was used to delineate the endometrium in uterine ultrasound images. We employed a ML approach based on the You Only Look Once (YOLO) version 8 framework [17], a modern single-stage object detection model designed for real-time performance. The primary rationale for choosing YOLOv8 was its ability to produce high accuracy while also having high inference speeds. The YOLOv8 model consists of a so-called backbone, neck, and head stages. The backbone is a typical convolutional neural network, the neck enhances features from the backbone, and the head is trained to predict bounding boxes, object classes, and segmentation masks. Specifically related to our clinical context, it allowed us to train and predict on ultrasound images in an efficient manner. A light-weight model with high inference speed is also an important consideration for deployment on non-specialized hardware, making it more accessible for clinical use. U-Net based architectures may provide a slight increase in segmentation performance, however, its introductory paper quotes an IoU value of 0.92 [18], and newer versions have accuracies ranging from 70 to 95% [19] which were comparable to the performance in our model. Transformer models were not considered due to the larger dataset sizes needed for these types of architectures. The at the time available SAM (Segment Anything Model) from Meta was also briefly tested; however, it required substantial manual selection of the middle of objects and failed to provide valuable segmentation results in the grayscale and "fuzzy" ultrasound images.

The model was trained to identify anatomical features from ultrasound images and learned to detect the presence of the endometrium and generate corresponding segmentation masks that defined its boundaries. To support accurate segmentation, the model incorporated a dual-path architecture that operated in parallel. One path generated a set of general-purpose image-wide masks that captured different spatial features while the other predicted how to combine the masks to isolate individual anatomical structures. These components were then combined linearly to produce a detailed mask for each detected region. This approach eliminated the need for slower sequential processing and preserved spatial accuracy by avoiding re-sampling of localized features. Additionally, the use of multi-resolution feature maps allowed the model to effectively capture both broad and fine-grained details—an essential advantage when segmenting the endometrium where precise boundary definition is critical for downstream measurements like ET.

Once the endometrial tissue region of interest (ROI) boundaries were defined using segmentation, a principal component analysis (PCA) approach was adapted to determine ET perpendicular to the uterine lumen at the thickest aspect of the endometrial echoes. Principal component analysis was examined as a robust way to determine the shape of an ellipse [20]. We utilized PCA to find the axis centered at the means of the points along the boundaries that maximize the variances along the founded axis. The shape of the endometrium is a slightly deformed ellipse which is transformed by dents or bulges on either one side or two sides of a standard ellipse around the minor axis. In the first step, an interpolation to the predicted boundaries was applied to smooth the curve and emphasize the interval where curvature was significant. Next, the PCA process fetched major and minor axes of the deformed ellipse after which the means of the points outlining the ellipse were determined. The equation of the minor axis was generated (mathematical proof provided in Supplemental Materials, Appendix). Finally, the intersection of the minor axis and the predicted boundaries were found by identifying the point closest to the minor axis after interpolation. This method was effective in finding the thickest part of a slightly deformed ellipse (Fig. 1).

2.4 Model building: endometrial pattern assignment

Computational object detection models were trained to assess for the presence or absence of a stratum functionalis layer using segmentation. Presence of the s. functionalis was used as an indication of tri-laminar layering of the endometrium thus allowing the ability to assign a pattern value of either 'trilaminar' or 'homogenous'. A small subset of the annotated dataset was prepared with 110 images displaying ROI polygons around the s. functionalis tissue layer. We prepared an object detection model utilizing

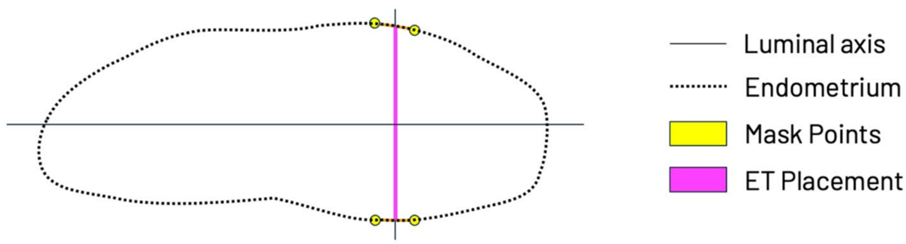

Fig. 1 Determination of the position at which to measure ET. Following segmentation of the endometrial tissues (dashed-line polygon) from myometrium the thickest aspect of the endometrial echoes, perpendicular to the luminal axis was identified (yellow mask points) and selected for thickness measurement (magenta line). The method was developed to minimize the impact of potential distortions in the uterine cavity caused by contractile waves, fibroids, and/or polyps on the accuracy of the final thickness measurement

the 110-image subset to detect the presence of a visually distinct s. functionalis tissue layer. We also prepared a synthetic calculated stratum functionalis data set to implement in model training. This was done by calculating the difference in thicknesses between the s. functionalis and s. basalis layers (training based on skilled human annotations) and shrinking the endometrial polygon by a uniform thickness to generate an approximate stratum functionalis polygon. Visual inspection and comparison to a limited number of fully annotated s. functionalis polygons indicated relatively accurate synthetic s. functionalis polygons were generated. Logic was written to process the output of the model to classify endometrial pattern as either homogeneous or tri-laminar [21]. If both an endometrial segmentation and a distinct functionalis segmentation were detected by the model, the classification label 'tri-laminar' was applied. Homogenous pattern classification was assigned based upon the absence of a distinct s. functionalis tissue layer within the endometrial echoes.

2.5 Model parameters

The model parameters were default for YOLO v8. The learning rate was 0.01. Batch size for the training set was set to 16. The loss function was a combination of multiple components, including box loss, class loss, and objectness loss, with respective values of 7.5, 0.5, and 1.0. The number of epochs for training was set to 200. The following HSV augmentation parameters for HSV adjustments were utilized: 'hsv_h': 0.015; 'hsv_s': 0.7; 'hsv_v': 0.4; 'translate': 0.1; 'scale': 0.5; 'fliplr': 0.5; 'mosaic': 0.5.

2.6 Model performance analysis

Intersection over union (IoU) was utilized to assess the model's proficiency in segmentation of the endometrium. In brief, the area of overlap between the predicted mask and the actual endometrium is divided by the area of union between the predicted and actual endometrium. The closer the IoU is to 1.0, the better the match. A confusion matrix is a tool used to evaluate the predictive performance of an algorithm; usually on a binary question, such as "Did the algorithm correctly identify the endometrium?" (with a IoU of 0.50 or more). Within the confusion matrix, there is a tally of true positives (the algorithm correctly identified endometrium where endometrium was present), false positives (the algorithm identified endometrium where it was absent), false negatives (the algorithm failed to identify endometrium where it was present) and true negatives (the algorithm did not identify endometrium where endometrium was absent). We did not provide any images that did not contain endometrium, so there were no true negatives in our dataset. This tally allows us to calculate accuracy, precision (positive predictive value), specificity (true negative rate) and sensitivity (true positive rate) of the algorithm [21] as well as calculate the Matthews Correlation Coefficient (MCC) and Bookmaker informedness [22]; two other measures of the efficacy of the segmentation algorithm. The MCC provides a single value, the correlation between correct classifications and the predicted classifications (range poor, − 1 to good, 1). Confusion matrices and confidence curves were utilized to assess ET predictions and pattern.

3 Results

3.1 Patient demographics & ultrasound characteristics

Mean patient age (n = 5985) was 36.49 years ± 4.36 STDEV (range: 23.47 years to 55.42 years), age distribution is provided in Supplemental Fig. 1. Age data were normally distributed (174 records did not have age listed). Frozen-thaw embryo transfer cycles (FET) represented 85% of the cohort (n = 5087); 15% of cycles were fresh embryo transfers (n = 898). Mean endometrial thickness was 9.22 mm with a standard deviation of 2.74 and right skew of 0.57. Homogenous endometrial patterns made up 48% of the image collection (n = 2873) while triple-line endometrial patterns made up the remaining 52% (n = 3112).

3.2 Automated endometrial segmentation

The segmentation models trained on the uterine ultrasound images were standard object detection models that generated labels and image masks. The masks were then converted to polygon outlines. The resultant model has high accuracy and tended to identify one and only one endometrium (for the 1-class model) and zero or one s. functionalis, but in some cases too many or zero segmentations can occur. The main endometrium segmentation model achieved impressive results and on its test set and scored 97.7% accuracy, 98.4% precision and 97.3% sensitivity (recall) for its metrics[1] calculated at the standard IoU (50% mask overlap). Performance was categorized based upon standardized IoU cut-offs. High performance accuracy on annotated images after endometrial segmentation was defined as an IoU greater than or equal to 0.75 (IoU ≥ 75%); moderate performance was defined as IoU between 0.5 and 0.749 (IoU ≥ 50% to < 75%); and low performance was defined as IoU below 0.5 (IoU < 50%). Additionally, 92.8% of test data had an intersection over union (IoU) value greater than the 'high performance' 0.75 threshold (Fig. 2). YOLO v8 provides two metrics to help determine how a model is functioning during the training phase and validation phase, mAP50 and mAP50-95. mAP stands for mean Average Precision with 50 referring to a 50% IoU threshold. mAP50 checks the performance of the model only at the IoU 50% threshold, providing a fast, easily calculated score, while mAP.5:0.95 provides an evaluation averaged over multiple thresholds of IoU from 50 to 95%. A good mAP50 score, anything over 50% indicates that the model works well with easier cases. If the mAP50-95 score is significantly lower than the mAP50 score, this can indicate the model is not working as well for more complex images [23, 24]. In this study, the primary endometrial segmentation mAP50 score and mAP50:95 scores are similar, and both considered excellent [23]. Mean IoU value was 0.91 ± 0.058 STDEV. Sample images with both manual annotation

[1] Additionally, because there are no "true negative" for the object detection for "background" specificity is an undefined value, and similarly the NPV or PPV values would not be appropriate.

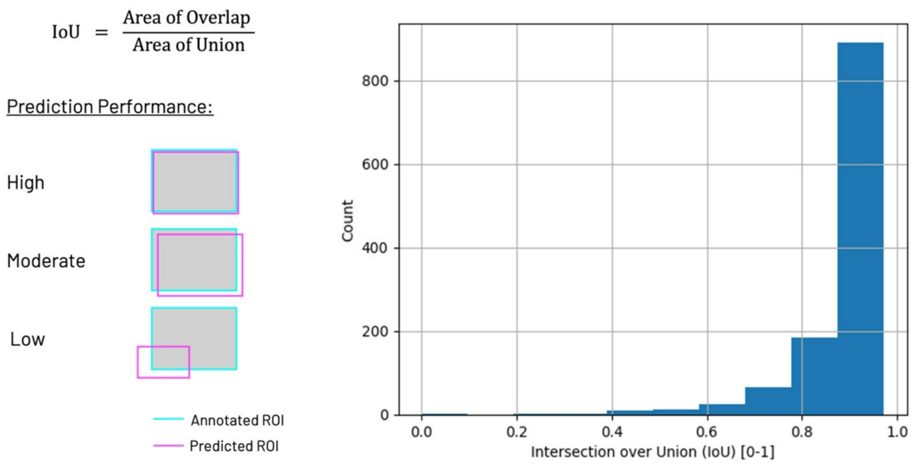

Fig. 2 Intersection over union for endometrial segmentation performance. The IoU calculation is shown (left) along with graphic representations of theoretical high, low, and moderate performance predictions for identification of the region of interest (ROI). The distribution of IoU values for the test data (right) demonstrates a strong right-shifted distribution representing > 92.8% of test predictions with an IoU value ≥ 0.75

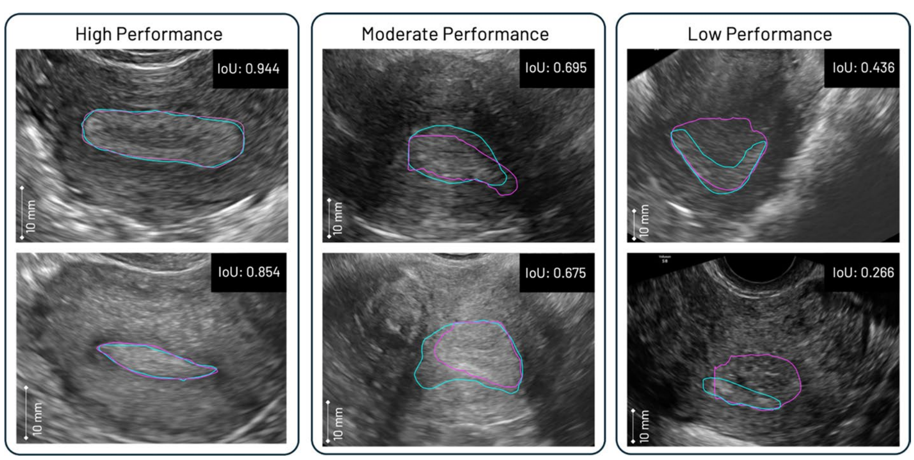

Fig. 3 Endometrial segmentation performance. Sample images with annotations of high performance (IoU ≥ 75%), moderate performance (IoU ≥ 50% to < 75%), and low performance (IoU < 50%) accuracies on annotated images after endometrial segmentation. The cyan polygons represent human annotation. Magenta polygons represent computational model predictions. The IoU value for each sample image is identified at the top right-hand corner of each panel

of the endometrial tissues and computationally modeled annotations are shown in Fig. 3. Confusion matrix provided values for accuracy, 0.98, precision, 0.99, specificity or true negative rate (Supplemental Figs. 3 and 4), 1.0, sensitivity or true negative rate, 0.99, Matthews Correlation coefficient 0.98, and Bookmakers informedness, 0.99.

3.3 Automated endometrial thickness measurement

Endometrial thickness was predicted by our model through placement of a linear annotation at the thickest aspect of the endometrial echoes in the transverse plane. Accuracy of the predicted ET measurement on each image in test data set was assessed based on three parameters: 1) length (predicted thickness measurement compared to human annotated measurement), 2) position (placement of the predicted point at which the ET was measured), and 3) orientation (degrees off-axis relative to the human placed annotation). These error conditions were not mutually exclusive. The ET measurement achieved mean absolute error rates of: ± 0.89 mm in length; ± 2.81 mm in position; and ± 5.5-degrees in orientation relative to the lumen (Fig. 4). Absolute error rates were calculated in pixels and converted to mm (adjusted for scale / magnification on each individual image).

3.4 Automated endometrial pattern assignment

Automated prediction of the presence of a visually appreciated distinction between the s. functionalis and the s. basalis endometrial tissue layers was utilized as an indicator of the pattern (homogenous versus triple-line) (Fig. 5 / Supplemental Fig. 2). The s. functionalis prediction model trained on both the ground truth endometrial segmentations and synthetic s. functionalis polygons had an overall accuracy of 83.1%, a precision of 82.0%, and a recall of 80.5%; however as a basic classifier (sorting endometrial pattern by ‘homogenous’ versus ‘triple-line’) it performed at 91.9% accuracy on test the data set. This classifier strategy accurately predicted endometrial pattern on 1101 images of the 1198 test data set and failed to predict endometrial pattern on 97 images. A detailed breakdown of the standard YOLO object detection and segmentation (box and mask metrics, respectively) is presented in Table 1. The proportion of ultrasound images representing homogenous versus triple-line patterns in the train / validation / test data sets is shown in Table 2.

4 Discussion

Our primary objective was to develop an automated method for assessing ET and pattern on ultrasonographic images. The signal-to-noise ratio is high on ultrasound images relative to other medical imaging modalities. Existing AI/ML applications for medical image interpretation tend to take advantage of high contrast anatomical features that

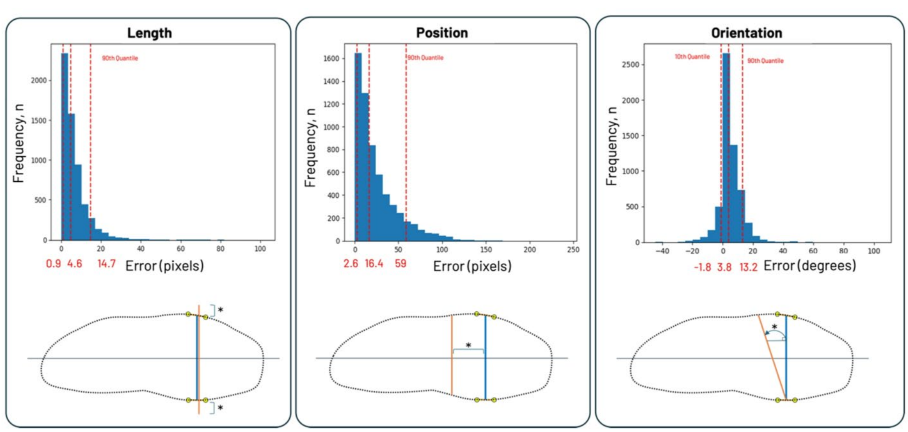

Fig. 4 Endometrial thickness prediction performance analysis. The model prediction performance in endometrial line length (left), position at which the ET is assessed (center), and orientation relative to the luminal axis (right) is demonstrated. Histograms for each error / performance analysis are shown (above). Diagrammatic representations of the error being assessed on the endometrial line annotations are shown (below), the dashed polygon represents the endometrial polygon, the black horizontal line represents the luminal axis, the dark blue line represents a theoretical human placed annotation, the orange line represents a theoretical predicted annotation, the yellow spots represent the mask points, and the ‘*’ brackets indicate the error being assessed. Dashed-lines and numbers represent percentiles on the histograms, eg: for the length histogram, 90% of the images had a 14.7 pixel error or less in the length of the line compared to the original annotation of the line. The line on the left represents the 10th percentile, the middle line the 50th percentile, and the line on the right, the 90th percentile

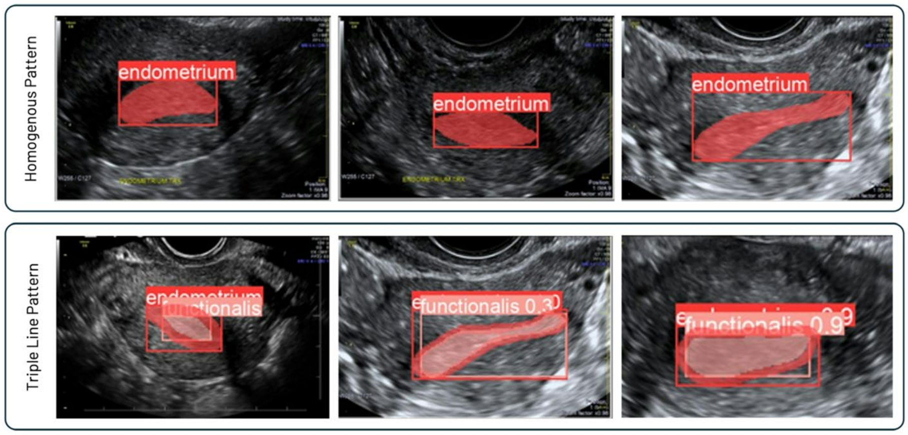

Fig. 5 Examples of visually distinct functionalis and basalis layers as a discriminator for endometrial pattern. Top Row: Representative samples of automated detection of a single homogenous endometrial tissue pattern (red bounding box—labeled 'endometrium'). Bottom Row: representative samples of automated detection of a distinct functionalis layer (pink bounding box—labeled 'functionalis') within the endometrial echoes (red bounding box) in a triple-line patterned endometrium

Table 1 Standard YOLO computer vision metrics

Model	Class	Box				Mask or segmentation			
		Prec	Rec	mAP50	mAP50-95	Prec	Rec	mAP50	mAP50-95
Main Segmentation	Endometrium	0.989	0.978	0.994	0.855	0.984	0.973	0.99	0.831
Synthetic Functionalis	All	0.904	0.887	0.918	0.706	0.802	0.804	0.804	0.465
	Endometrium	0.978	0.978	0.99	0.854	0.816	0.816	0.775	0.438
	s. Functionalis	0.831	0.796	0.847	0.558	0.825	0.791	0.834	0.492

Evaluated on the test data set (n = 1198). Box metrics are included for completeness and represent the success at finding a simple bounding box around the detected objects and are obviousnullly higher than the mask metrics which represent the per-pixel labelling of the images used to generate outlining polygons. The precision anext (COCO) evaluations. For all metrics values closer to 1.0 are better and can be interpreted as percentage-like fractions.d recall are computed at a standard IoU of 0.5, the mAP50 score is the mean average precision calculated by integrating the precision-recall curve at an IoU of 0.5 and the mAP[50::95] is the average over IoU values 0.50 through 0.95 as a more stringent evaluation metric as used in Common Objects in Cont ext (COCO) evaluations. For all metrics values closer to 1.0 are better and can be interpreted as percentage-like fractions

Table 2 Proportion of endometrial pattern

	Homogenous pattern	Triple line pattern
Total (N = 5985)	2811	3174 (53%)
Train (N = 4189)	1935	2254 (54%)
Validation (N = 598)	286	312 (52%)
Test (N = 1198)	590	608 (51%)

Homogenous versus triple-line patterns are represented at comparable levels in the training / validation /test datasets

present clear boundaries upon imaging. The endometrial / myometrial junction is a clear anatomic boundary. However, the density of endometrial and myometrial tissues is similar, which results in an acoustic image where the two tissue layers can be resolved, but at low contrast. We chose to segment the endometrium from the myometrium in our model building to: 1) constrain the ET measurement and pattern assignment to the precise region of interest (the endometrium); and 2) structure the logic around identification of the widest aspect of the endometrial echoes. This approach was designed to improve the accuracy with which the ET measurements and pattern assignments were made.

Secondarily, we aimed to introduce precise standardization of the method by which the thickness and pattern assessments are conducted. To standardize, we first needed to identify and acknowledge sources of variability in standard of care endometrial assessments. Two key variability points were identified: 1) the location at which ET was measured; and, 2) the subjectivity based on individual clinical experience. Therefore, we standardized ET measurement using a line drawn from the endometrial–myometrial interface at the visually thickest superior and inferior aspects of the endometrial cavity, within 5 to 10 mm of the fundal aspect of the endometrium (baseline measurement location). Changes in ET over the ovarian cycle in a clinically typical population are demonstrable using this clearly defined baseline measurement location [10]. Rotation of the probe by ~ 90 degrees from the baseline mid-sagittal to transverse plane of visualization achieves two goals: 1) it locks the imaging in the 5–10 mm zone of interest as endometrial echoes outside of this zone will not be visualized in transverse plane; and, 2) it increases the number of pixels in the final captured image representing endometrial tissue. Importantly, the ET measurement remains unchanged between the mid-sagittal imaging plane and transverse plane upon rotation of the probe. Using the transverse imaging plane effectively reduces the potential for error by standardizing image capture in the target measurement location, while simultaneously providing more pixels for analysis for AI/ML training and validation.

The novel AI/ML-based method for assessing ET and pattern presented herein performs three distinct tasks at levels exceeding 90% accuracy. First, the model accurately identifies and segments endometrial tissues from myometrium on 2D ultrasound images. Our automatic segmentation of endometrial tissues performs well, and the majority of predicted endometrial segmentations fell into the high-performance prediction category with IoU values higher than 0.75. In addition to the standard metrics, the average IoU or overlap/agreement on the actual segmentation was high with a mean IoU of 0.916 and a mAP score of 0.995; both of which indicate the actual outlines generated have very high agreement beyond the typical 50% standards for reporting IoU thresholds. This high-acuity performance is in stark contrast to a previously published method that showed generally poor performance in segmentation of endometrial echoes [15]. The pool of ML based ET estimators to compare and contrast with the method we describe is very limited. The absolute error rate in ET measurement we report is less than 1 mm when compared to human annotation which represents a significantly tighter error margin than previously published reports [11–13], which showed measurement errors of up to 4 mm. The work of Hu et al. [25] uses segmentation with a medial axis transformation. The sample size of the work described is < 300 images with a test set of 27 images, leading to the typical concerns with implementation of these types of approaches on small data sets. The absolute error in ET measurement is also more than double the method we propose. Many clinics utilize an ET cut-off of 7 mm to advance to embryo transfer, therefore measurement errors of up to 4 mm (meaning > 50% of the thickness cut off) are not acceptable. The low error (< 1 mm) our model achieved is not only a significant improvement from previous attempts at automated ET measurements; it is also further evidence that the upstream endometrial segmentation masks, on which the ET computations rely, are modeled with greater accuracy. The placement of the measurement line in terms of position, axis and overall measurement error are all minimal and within error clinical tolerances.

Our model measures thickness with an absolute error of 0.89 mm, which means that any endometria with a thickness greater than 7.9 mm would move forward to transfer without concern of error impacting the decision. Likewise, any endometria with thickness lower than 6.11 mm would be screened out and transfer deferred. The described error rates will not impact clinical decision making for these cases as they would be clearly noted either above or below the 7 mm threshold of endometrial thickness widely considered to be a candidate for embryo transfer. The cycles with endometrial thickness between 6.11 and 7.9 mm could potentially be assigned for transfer, or deferral, erroneously with the degree of error present in our model. However, in our experience less than 20% of cycles fall into this thickness range. While this is a significant number of cases, it is a minority of cases, which could potentially be flagged for human review. Alternatively individual clinics could choose to set their cut-off threshold to ~ 6 mm to offset this error potential, since ET of 6 and above have also been shown to be positive predictors of outcome [26]. It is important to note that this same error potential exists with human measurements, and it is known that human error in measurement is in the range of 10–20% of the total thickness [27]. The 0.89 mm absolute error our model exhibits is a lower absolute error than is tolerated by human image assessment [28–30].

Finally, our model performs endometrial pattern assignment. To our knowledge, this is the first report of automated endometrial pattern assignment. The performance metrics of our pattern recognition model appear moderate when compared to the endometrial segmentation and measurement model. This is likely attributed to the use of the approximate 'synthetic' polygons for training the computer vision system to identify the endometrial s. functionalis layer. Alternatively, it is possible that distinguishing s. functionalis from s. basalis is a more difficult computer vision task, especially when one endometrial tissue layer is significantly thinner than the other (depending on the state of glandular development in response to circulating hormones), the underlying microanatomic structures are intermingled and the relative echotextural image attributes vary in response to circulating reproductively active hormones. When training the model, however, the performance was improved by the addition of the full endometrial segmentation polygons (perhaps from learning from the higher quality outlines for identifying boundaries in the ultrasound images). Evaluation metrics for the prediction / identification of s. functionalis had an overall accuracy of 83.1% (precision 82.0% / recall 80.5%). These metrics were calculated at the IoU overlap of 50% meaning the model generated poorer quality s. functionalis polygons than our over-all endometrial segmentation model. However, it was still able to identify at high accuracy when there was a visually appreciable differentiation between the endometrial tissue layers correctly labelling portions of the endometrium as 'functionalis'. This is more obvious if we consider the higher "box" metrics in Table 1. When we used this object detection model to predict only if an s.unctionalis layer is visually appreciable (without evaluating segmentation accuracy/IoU) we increased performance by 8.8% and achieved a classifier[2] with 91.9% accuracy.

We acknowledge that the utility of this method presupposes the utility of standard of care ET and pattern assessments which are the subject of significant debate within the literature. We also acknowledge the ultrasound image data set utilized for this research program were acquired outside of the standard imaging timing in many clinical practices

[2] When evaluating methods for classification, we tested a more traditional image classification model, the VGG image classifier, which performed worse with a 75% as a classification success rate.

(~48 h prior to day-5 / day-6 embryo transfer), which may limit its implementation in the field. Further, we utilized transverse plane of section ultrasound images, acquired in the 5–10 mm fundal zone to train and validate our AI/ML models. While the rationale for this decision is discussed in detail, we acknowledge that standard clinic practice typically relies upon mid-sagittal ET measurement. However, transverse plane of section images may be acquired as described and the method may be implemented with minimal adjustment to standard clinical imaging practices. Our initial predictive modeling results also indicated positive performance of this method on mid-sagittal images and future work may be undertaken to adapt this strategy to mid-sagittal images which are routinely utilized in clinical practice. Full analysis of the performance and clinical significance will be explored and communicated in future studies.

Automation of these routine endometrial assessments could support clinical workflow and enhance patient care through time savings and standardization. Reduction of clinical time load for patients, ultrasonographers, and clinicians is meaningful for all parties. The inference time of this method is ~2.2 s per image, which constitutes significant time savings over annotation by human personnel. Conceivably, a full series of ultrasound images could be collected and the code could run post-hoc during down-time between patients, saving time for patients and clinical staff. Standardization of key processes in IVF therapies has also been demonstrated to improve clinical workflows and outcomes [31]. As such, standardization through automation of these measurements has the potential to reduce inter-observer variability and enhance patient care. In the future it is plausible that such an automated measurement / pattern assignment system could be integrated into existing ultrasound machine software or photo archiving systems to aide in clinical workflows.

Constraints to deployment of this automated approach to ET/pattern assessment method in clinical settings are primarily related to image quality. Excessive motion artifact, imbalanced image gains, over/undersaturated pixel densities, and anatomic shadowing from fibroids or fluid may significantly impact the performance of these computer vision tools. However, it is noteworthy that 388 of the 5985 (~6.5%) of the images utilized in development of this model had notations for the presence of leiomyomas (uterine fibroids). Many fibroids cast shadows and/or caused deformation of the uterine cavity, which further demonstrates the utility of the method described herein in the real-world clinical setting. Utilization of modern ultrasound instrumentation by trained personnel can successfully eliminate the imaging concerns, however, anatomic structures which cast significant shadows or generate specular reflections would need to be addressed through more extensive training and validation of AI/ML systems on larger datasets in future work.

We understand that the utilization of one segmentation technique (YOLO v8) is a limitation to the work. Exploration and/or integration of other modern segmentation techniques, such as U-Net, DeepLabv3+, or SAM will be the subject of future work.

The use of AI in the medical field is not without controversy. Utilizing computational models in medical decision making can have serious implications. The black box modelling often employed in AI/ML methods can limit clinician and/or patient confidence in decision making based on these types of tools. For this reason, explainable AI has been developed in some medical applications, to provide a greater transparency and improved confidence in the outputs. Some AI models used in medicine provide enhancements to imaging that allow for more consistent diagnoses, without taking the human element completely out of the decision making process, giving those at the receiving end a greater connection to the end result [32]. In addition, ongoing assessment of the quality of image segmentation provided by AI/ML becomes a challenge once the training phase is complete. Once launched, the absence of human annotations of ground truth comparator data for routine quality and improvement checks creates a need for more inventive methods of assessment about the reliability of decision making based on computational models [33].

We report an automated method to enhance, streamline, and standardize routine endometrial assessment. This method is intended to provide a tool that can reduce clinical workload if implemented but ultimately relies upon the clinicians utilizing the technology to make the final decision of whether or not their patient(s) proceed with an embryo transfer procedure.

5 Conclusions

A novel computer vision approach to automation of routine ET and pattern assessment has been developed. Our results were over 90 percent accurate in routine use, provide a new standardized tool for endometrial assessment, and pave the way for deeper inquiry into expanding the role of AI/ML in clinical assessment of endometrial function.

Appendix

Mathematical proof of using principal components analysis to find the thickest part of a slightly deformed elipse. A brief proof of using PCA to find the major and minor axis of a deformed ellipse is presented.

First, we use a mathematical expression to define a deformed ellipse using parameterized equations, which can be imagined as an endometrial centered at the (0, 0) on a Cartesian plane with its major and minor axis sitting on y-axis and x-axis, respectively. To do so, we need to add transformation based on the standard ellipse in parametric equations:

$$\boldsymbol{x = a \cdot cos(\theta)},$$

$$\boldsymbol{y = b \cdot sin(\theta)}$$

Now we add some dents or bulges to the ellipse:

$$\boldsymbol{M(c, k, \theta, h) = -c \cdot e^{-k(1 - sin\theta - h)}} \tag{1}$$

$$\boldsymbol{y = b \cdot sin\theta \cdot (1 + M)} \tag{2}$$

$$\boldsymbol{x = a \cdot cos\theta} \tag{3}$$

Here M adds the dents or bulges to the standard ellipses. The parameter c controls whether the deformation should be a “dent” or a “bulge”; k is the parameter to control if the deformation is more localized or less localized; h determines where the deformation happens, and if h = 0, the deformation happens around the y-axis. Here we discuss the situation where h is fixed at 0. The Eq. (1) turns into:

$$\boldsymbol{M(c, k, \theta, h) = -c \cdot e^{-k(1 - sin(\theta))}} \tag{4}$$

By the process of PCA (Trevor et al 2009), we can get the PCA axis from applying Singular Value Decomposition on the covariance matrix of x and y data points and find the eigenvectors:

$$\begin{bmatrix} Var(x) & Cov(x,y) \\ Cov(x,y) & Var(y) \end{bmatrix}$$

And when h = 0, we have:

$$\begin{aligned} Cov(x,y) &= E[xy] - E[x]E[y] \\ &= \int_{-\infty}^{\infty} a \cdot cos\theta \cdot b \cdot sin\theta \cdot (1+M) d\theta \\ &= \int_{-\infty}^{\infty} a \cdot cos\theta \cdot b \cdot sin\theta \cdot \left(1 + -c \cdot e^{-k(1-sin(\theta))}\right) d\theta = 0 \end{aligned}$$

Since:

$$E[x] = 0$$

And

$$a \cdot cos\theta \cdot b \cdot sin\theta \cdot \left(1 + -c \cdot e^{-k(1-sin(\theta))}\right)$$

is an odd function, and thus its integral will be zero if the range of the integral is symmetric around (0, 0).

So the Covariance matrix of x and y becomes:

$$\mathbf{P} = \begin{bmatrix} Var(x) & 0 \\ 0 & Var(y) \end{bmatrix}$$

And let $v_1 = (1,0)^T$ and $v_2 = (1,0)^T$. Since:

$$\mathbf{P}v_1 = Var(x)\, v_1$$

$$\mathbf{P}v_2 = Var(y)\, v_2$$

We have the directions of the principal components, which is $(1,0)^T$ and $(1,0)^T$ which are the directions of x-axis and y-axis. To accommodate the shift of the means when the deformations are not around (0, 0), we use interpolation to address the intervals on the deformed ellipses where curvature is non-zero in our study. So instead of getting the axis centered at (0, 0), the means are now shifted to the region where the deformation is more significant, and the axis will be rotated.

A further improvement of this method can be found by utilizing mathematical models that emphasize on local curvatures, and advanced machine learning algorithms such as kernel principal components analysis, known as kernel PCA, which can capture the curvatures without interpolation, and suitable for more deformed ellipse, such as a concave one.

Abbreviations

AI	Artificial intelligence
COCO	Common objects in context
ET	Endometrial thickness
FET	Frozen embryo transfer

IoU	Intersection over union
IVF	In vitro fertilization
MAR	Medically assisted reproduction
ML	Machine learning
PCA	Principal component analysis
ROI	Region of interest
2D	Two-dimensional
YOLO	You only look once

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s44352-025-00019-w.

Supplementary Material 1.

Acknowledgements

We thank Byron Chu of Cybera Inc for his dedication and assistance in project management. We are grateful to John Deptuch of Synergyne ART Analytics, Inc. for his support in database curation and manuscript development.

Author contributions

HP led / designed the study and prepared the manuscript for publication with support from JI. ZS led the data science. ZS, VO, YW, KS, and DP contributed to the coding and development of the AI/ML tools described. RAP, ZS, VO, YW, and KS contributed to manuscript development and revisions. YW provided mathematical proofs. RAP led the scientific aspects of the study. All authors have reviewed and approved the manuscript.

Funding

No funding was received to assist with preparation of this manuscript.

Data availability

The data that support the findings of this study are available from Synergyne Imaging Technology Inc. however restrictions apply to the availability of the data, which were used under licence for the current study, and so are not publicly available. Data are however available from the corresponding author upon reasonable request and with permission of Synergyne Imaging Technology Inc.

Declarations

Ethics approval and consent to participate

This study involving de-identified patient health information was conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The Human Ethics—Biomedical Research Ethics Board of the University of Saskatchewan approved this study (BIO-5251). Waiver of consent was granted by the Human Ethics—Biomedical Research Ethics Board of the University of Saskatchewan (BIO-5251).

Consent for publication

Not applicable.

Competing interests

Financial interests: RAP is CSO of Synergyne and a shareholder within the company. HP and JI are employees of Synergyne. ZS is an employee of Cybera Inc. YW, VO, KS, and DP were employed by Cybera Inc. during development of the computational models. No other financial interests to disclose. Non-financial interests: The authors declare they have no non-financial interests to declare.

Received: 14 May 2025 / Accepted: 4 November 2025

Published online: 12 November 2025

References

1. De Geyter C, Schmitter M, De Geyter M, Nieschlag E, Holzgreve W, Schneider HP. Prospective evaluation of the ultrasound appearance of the endometrium in a cohort of 1,186 infertile women. Fertil Steril. 2000;73(1):106–13.
2. Mathyk B, Schwartz A, DeCherney A, Ata B. A critical appraisal of studies on endometrial thickness and embryo transfer outcome. Reprod Biomed Online. 2023;47(4):103259.
3. Fernandez EI, Ferreira AS, Cecílio MHM, Chéles DS, de Souza RCM, Nogueira MFG, et al. Artificial intelligence in the IVF laboratory: overview through the application of different types of algorithms for the classification of reproductive data. J Assist Reprod Genet. 2020;37(10):2359–76.
4. Fjeldstad J, Qi W, Mercuri N, Siddique N, Meriano J, Krivoi A, et al. An artificial intelligence tool predicts blastocyst development from static images of fresh mature oocytes. Reprod Biomed Online. 2024;48(6):103842.
5. Rubio I, Galán A, Larreategui Z, Ayerdi F, Bellver J, Herrero J, et al. Clinical validation of embryo culture and selection by morphokinetic analysis: a randomized, controlled trial of the EmbryoScope. Fertil Steril. 2014;102(5):1287-1294.e5.
6. Storr A, Venetis C, Cooke S, Susetio D, Kilani S, Ledger W. Morphokinetic parameters using time-lapse technology and day 5 embryo quality: a prospective cohort study. J Assist Reprod Genet. 2015;32(7):1151–60.

7. Deutch TD, Joergner I, Matson DO, Oehninger S, Bocca S, Hoenigmann D, et al. Automated assessment of ovarian follicles using a novel three-dimensional ultrasound software. Fertil Steril. 2009;92(5):1562–8.
8. Schnorr J, McLaughlin J, Cook H, Slowey M, Schnorr J. Automation of follicular measurements during ovarian stimulation using a convolutional neural network. J Fertil Reprod Health. 2023;1(1):3–8.
9. Pierson HE. Non-invasive methodologies for endometrial evaluation in medically assisted reproduction. J Reprod. 2024;12:173–81.
10. Baerwald AR, Pierson RA. Endometrial development in association with ovarian follicular waves during the menstrual cycle. Ultrasound Obstet Gynecol. 2004;24(4):453–60.
11. Liu Y, Zhou Q, Peng B, Jiang J, Fang L, Weng W, et al. Automatic measurement of endometrial thickness from transvaginal ultrasound images. Front Bioeng Biotechnol. 2022;10:853845. https://doi.org/10.3389/fbioe.2022.853845/full.
12. Park H, Lee HJ, Kim HG, Ro YM, Shin D, Lee SR, et al. Endometrium segmentation on transvaginal ultrasound image using key-point discriminator. Med Phys. 2019;46(9):3974–84.
13. Wang X, Bao N, Xin X, Tan J, Li H, Zhou S, et al. Automatic evaluation of endometrial receptivity in three-dimensional transvaginal ultrasound images based on 3D U-net segmentation. Quant Imaging Med Surg. 2022;12(8):4095–108.
14. Liang X, He J, He L, Lin Y, Li Y, Cai K, et al. An ultrasound-based deep learning radiomic model combined with clinical data to predict clinical pregnancy after frozen embryo transfer: a pilot cohort study. Reprod Biomed Online. 2023;47(2):103204.
15. Asch Schuff R, Suarez J, Laugas N, Ramirez MLZ, Alkon T. Artificial intelligence model utilizing endometrial analysis to contribute as a predictor of assisted reproductive technology success. J IVF-Worldw. 2024;2(2):1–8.
16. Vlăsceanu GV, Tarbă N, Voncilă ML, Boiangiu CA. Selecting the right metric: a detailed study on image segmentation evaluation. BRAIN Broad Res Artif Intell Neurosci. 2024;15(4):295–318.
17. Bolya D, Zhou C, Xiao F, Lee YJ. YOLACT: real-time instance segmentation [Internet]. arXiv; 2019 [cited 2025 Apr 30]. Available from: http://arxiv.org/abs/1904.02689
18. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation [Internet]. arXiv; 2015 [cited 2025 Sep 20]. Available from: http://arxiv.org/abs/1505.04597
19. Jiangtao W, Ruhaiyem NIR, Panpan F. A comprehensive review of U-Net and its variants: advances and applications in medical image segmentation [Internet]. arXiv; 2025 [cited 2025 Sep 20]. Available from: http://arxiv.org/abs/2502.06895
20. Wijewickrema S, Paplinski A. Principal component analysis for the approximation of a fruit as an ellipse. In 2004 [cited 2025 Apr 30]. Available from: https://www.semanticscholar.org/paper/Principal-Component-Analysis-for-the-Approximation-Wijewickrema-Paplinski/d29d5531fa666cecc0a80ea662bd870c1245758f
21. Vanacore A, Pellegrino MS, Ciardiello A. Fair evaluation of classifier predictive performance based on binary confusion matrix. Comput Stat. 2024;39(1):363–83.
22. Chicco D, Tötsch N, Jurman G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. Biodata Min. 2021;14(1):13.
23. How to Interpret YOLOv8 Results: A Complete Guide [Internet]. 2024 [cited 2025 Sep 20]. Available from: https://yolov8.org/how-to-interpret-yolov8-results/
24. What Is mAP50 in YOLOv8? A Simple Explanation [Internet]. 2025 [cited 2025 Sep 20]. Available from: https://yolov8.org/what-is-map50-in-yolov8/
25. Hu SY, Xu H, Li Q, Telfer BA, Brattain LJ, Samir AE. Deep learning-based automatic endometrium segmentation and thickness measurement for 2D transvaginal ultrasound. In: 2019 41st Annual international conference of the IEEE engineering in medicine and biology society (EMBC) [Internet]. 2019 [cited 2025 Sep 20]. p. 993–7. Available from: https://ieeexplore.ieee.org/document/8856367
26. Mahutte N, Hartman M, Meng L, Lanes A, Luo ZC, Liu KE. Optimal endometrial thickness in fresh and frozen-thaw *in vitro* fertilization cycles: an analysis of live birth rates from 96,000 autologous embryo transfers. Fertil Steril. 2022;117(4):792–800.
27. Pierson H, Cadesky K, Meriano J, Invik J, Laskin C, Pierson R. Ultrasound based endometrial receptivitiy scoring accurately identifies cycles with low probability of pregnancy. In: ASRM Abstracts [Internet]. 2021. Available from: https://matrisart.com/wp-content/uploads/2022/05/Matris-ASRM-2021.pdf
28. Epstein E, Valentin L. Intraobserver and interobserver reproducibility of ultrasound measurements of endometrial thickness in postmenopausal women. Ultrasound Obstet Gynecol. 2002;20(5):486–91.
29. Saha TK, Amer SAKS, Biss J, Thakare H, Williams S, Farrell CT, et al. The validity of transvaginal ultrasound measurement of endometrial thickness: a comparison of ultrasound measurement with direct anatomical measurement. BJOG Int J Obstet Gynaecol. 2004;111(12):1419–24.
30. Spandorfer SD, Arrendondo-Soberon F, Loret de Mola JR, Feinberg RF. Reliability of intraobserver and interobserver sonographic endometrial stripe thickness measurements. Fertil Steril. 1998;70(1):152–4.
31. Alikani M, Campbell A. Shaping the future of the IVF laboratory: standardization for more predictable outcomes. Reprod Biomed Online. 2025;50(4):104854.
32. Houssein E, Gamal A, Younis E, Mohamed E. Explainable artificial intelligence for medical imaging systems using deep learning: a comprehensive review. Cluster Comput. 2025;28:469.
33. Sikha OK, Riera-Marín M, Galdran A, López JG, Rodríguez-Comas J, Piella G, et al. Uncertainty-aware segmentation quality prediction via deep learning Bayesian modeling: comprehensive evaluation and interpretation on skin cancer and liver segmentation. Comput Med Imaging Graph. 2025;123:102547.
34. Trevor H, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2009. Available from: https://hastie.su.domains/ElemStatLearn/

Publisher's Note