Deep Learning for Endometrium Segmentation in Transvaginal Ultrasound: A Systematic Review Towards Receptivity Assessment

Asma Amirah Nazarudin[1], Siti Salasiah Mokri[2]*, Noraishikin Zulkarnain[3], Aqilah Baseri Huddin[4], Mohd Faizal Ahmad[5], Ashrani Aizzuddin Abd Rani[6], Seri Mastura Mustaza[7], Huiwen Lim[8]
Department of Electrical, Electronic and Systems, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia[1, 2, 3, 4, 6, 7]
Hospital Canselor Tunku Muhriz UKM, Jalan Yaacob Latiff, Bandar Tun Razak, 56000 Cheras, Kuala Lumpur, Malaysia[5]
Advanced Reproductive Center-Hospital Canselor Tunku Mukhriz, Jalan Yaacob Latiff, Bandar Tun Razak, 56000 Cheras, Kuala Lumpur, Malaysia[8]

***Abstract*—Deep learning (DL) has become a transformative approach in medical image analysis, offering superior accuracy and automation in image segmentation tasks. In reproductive imaging, transvaginal ultrasound (TVUS) serves as a crucial modality for evaluating the endometrial condition, which plays a critical role in assessing ovarian health. Although many studies have applied deep learning to the segmentation of pathological endometrial conditions, research focusing on non-pathological endometrium segmentation remains critically limited. This study presents a comprehensive review of deep learning methods for endometrium segmentation in TVUS, with a focus on non-pathological conditions, including endometrial thickness measurement, morphology analysis, and endometrium receptivity assessment. Following PRISMA guidelines, research articles published between 2015 and 2025 were identified from major scientific databases. The selected studies were analyzed in terms of image processing methods, deep learning architectures, and performance metrics, such as Dice coefficient, Jaccard index, precision, recall, and Hausdorff distance. Although foundational architectures, such as U-Net and its variants, achieve impressive Dice coefficients (up to 0.977), the results often rely on small and single-center datasets, proving limited generalizability across imaging settings. Recent advancements demonstrate the efficacy of hybrid architectures, such as the Deep Learned Snake algorithm and Transformer-based models like SAIM, in optimizing segmentation precision within noisy transvaginal ultrasound images. This review highlights the lack of attention to non-pathological endometrium segmentation and guides future research directions in self-supervised learning, transformer-based architectures, and interpretable deep learning to achieve robust and clinically applicable models for enhancing endometrium receptivity assessment and supporting ovarian health in assisted reproduction technology.**

Keywords*—*Endometrium segmentation; deep learning; image segmentation; image processing; endometrium receptivity assessment; ovarian health

I. Introduction

Fertility is a crucial aspect of reproductive health, with individuals seeking medical assistance to conceive. Due to this, in vitro fertilization (IVF) has emerged as one of the most effective assisted reproductive technologies (ART). Research showed that the success rate of IVF remains below optimal, due to multiple limiting factors such as embryo quality and endometrial receptivity [1]. Transvaginal ultrasound (TVUS) is the primary imaging modality in reproductive medicine to measure endometrial thickness and structure because of its real-time, non-invasive nature, cost-effectiveness, and portable data management [2].

In the context of IVF, accurate segmentation of the endometrium in TVUS images is essential for reliable measurement of endometrial thickness, which is a key biomarker for predicting endometrial receptivity and IVF outcomes [3], [4], [5]. Clinicians often rely on TVUS to measure endometrial thickness and detect abnormalities that may impact implantation. However, manual segmentation of the endometrium thickness from ultrasound images is highly subjective, time-consuming, and prone to inter-observer variability. This variability can affect clinical decision-making and IVF outcomes. Therefore, there is a growing need for automated and accurate segmentation of ultrasound images to assist in clinical decision-making [6].

Current efforts towards automation are primarily focused on automating pathological endometrial conditions, such as cancers and endometriosis. Meanwhile, the assessment of endometrial receptivity in healthy women undergoing IVF remains largely overlooked [7]. This disparity highlights a critical research gap in the field. Although accurate endometrium segmentation for pathological diagnosis is essential, it is equally vital to evaluate endometrial receptivity to improve IVF success [8], [9].

Deep learning models have demonstrated remarkable abilities in tackling a diverse array of medical imaging tasks, spanning classification, detection, segmentation, and registration [6], [10]. While deep learning has shown remarkable success in detecting endometrial pathology, current models largely neglect the detailed analysis for assessing healthy endometrium receptivity. Moreover, the limited generalizability across TVUS devices and protocols is due to

*Corresponding author.

small, single-center datasets lacking in diversity in acquisition settings [11].

Other than that, existing models predominantly focus on anatomical features, such as thickness, while overlooking functional biomarkers that correlate with implantation potential [12]. These challenges escalated because researchers haven't agreed on the best deep learning model for endometrium segmentation and a standard method to assess these tools in real-time clinical use. Although experimental studies report that segmentation accuracy meets the required clinical standard, the results have not been translated into clinical application. The gap between research achievements and practical implementation highlights the need for comprehensive solutions that address both technical and clinical requirements for endometrial receptivity assessment.

Therefore, to systematically evaluate these challenges and opportunities, this literature review addresses the following central research question: How can deep learning segmentation of TVUS images improve objective assessment of endometrial receptivity? To accurately address this question, there are four sub-questions:

- What are the predominant deep learning model architectures applied to TVUS endometrium segmentation?
- Which deep learning architecture achieves a high accuracy in endometrium segmentation?
- Which evaluation metrics best validate segmentation quality for clinical receptivity assessment?
- What technical and clinical gaps persist in translating these models to the ART workflow?

By addressing these questions through synthesized evidence, this study aims to establish methodological best practices while identifying priority areas for future research at the intersection of deep learning and reproductive medicine. Fig. 1 illustrates the central research question and sub-questions guiding the systematic review. The novelty of this review bridges the gap by focusing specifically on how deep learning can segment healthy endometrium for receptivity assessment, which is a clinical need often overlooked in prior reviews of gynecological imaging.

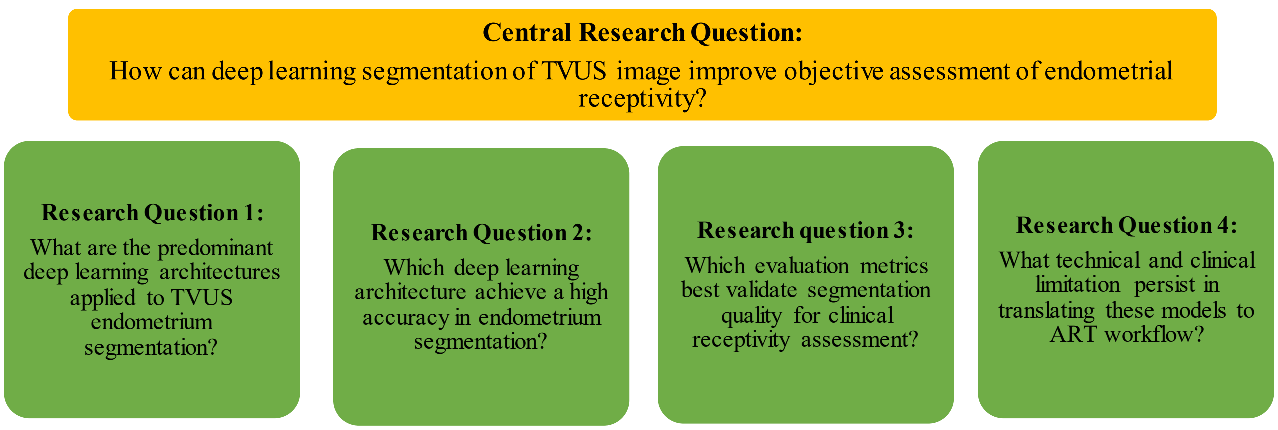

Fig. 1. Overview of the central research question, with research questions that guide this literature review.

This study is organized first to provide a brief overview of related reviews on deep learning for medical, specifically for women's reproductive health system. Later, this study provides brief information on endometrial receptivity and the role of transvaginal ultrasound for assessing and monitoring the endometrium. Next, the study will provide a brief explanation about a deep learning model for image segmentation in medical image analysis. The methodology section outlines the rigorous selection process used to identify relevant studies included in this review. Following this section, the results section synthesizes and presents the selected studies to address the defined research questions. In the future direction section, this study discusses the potential for integrations and clinical applications to improve receptivity assessment. Lastly, the conclusion summarized key insights, identified current gaps in the literature, and outlined actionable future research trajectories. The list of abbreviations used in this study is listed in Table I.

TABLE I. LIST OF ABBREVIATIONS

Word	Abbreviations
ART	Assisted Reproductive Technology
CNN	Convolutional Neural Network
DDRNet	Deep Dual-Resolution Networks
DLS	Deep Learned Snake
ERA	Endometrium Receptivity Assessment
ET	Endometrial Thickness
ET	Endometrium Thickness
FCN	Fully Convolutional Network
IoU	Intersection of Union
IVF	In Vitro Fertilization
MAE	Mean Absolute Error
NLP	Natural Language Processing
ResNet	Residual Network
RSME	Root Square Mean Error
SAIM	Segment Anything with Inception Module
SAM	Segment Anything Module
TVUS	Transvaginal Ultrasound
U-Net	U-Shaped Convolutional Neural Network
VGG	Visual Geometry Group
WOI	Window of Implantation

II. Related Works

Several review articles have surveyed the application of deep learning in medical imaging, including specific analyses of gynecological structures. However, a critical examination of these reviews reveals a consistent focus on pathological diagnosis and has rarely addressed the segmentation of the non-pathological endometrium for receptivity assessment.

Broad systematic reviews in medical imaging, such as those by Litjens et al [10] and Gao et al.[13], highlighted that while deep learning has transformed tasks like classification and segmentation across various anatomical regions, gynecological applications are often overlooked by more extensively researched regions, like the brain, chest, and musculoskeletal system. Furthermore, comprehensive segmentation reviews primarily utilize pathological benchmarks to evaluate model performance. This leaves a significant gap in the synthesized knowledge regarding the automated mapping of healthy anatomical structures.

Within the gynecological field, the existing reviews focused on pathological conditions, particularly oncological diagnostics. For instance, Swarnkar et al. conducted a systematic review of deep learning for the diagnosis of cervical, ovarian, and endometrial malignancies. The review identified 50 relevant studies, including 16 focused specifically on endometrial cancer [14].

Similarly, Aparna and Libish concentrated on research specifically aimed at identifying abnormal and malignant cells in the uterus to discover endometrial cancer [15]. Meanwhile, Piedimonte focused on using machine learning to incorporate clinical and radiologic parameters to pre-operatively stratify high-risk cancer patients [16]. These reviews evaluate the efficacy of models in discriminating between benign and malignant masses or predicting the depth of myometrial invasion, which are tasks inherently different from delineating a healthy endometrial lining [13].

This pathological emphasis persists even in modality-specific reviews. Jiang synthesized the research on deep learning-based imaging for endometrial cancer management across ultrasound, MRI, and hysteroscopy, with a focus on tumor morphology and molecular typing [17]. Zhang evaluated models for classifying common endometrial lesions, including hyperplasia and polyps, but did not address the segmentation of baseline healthy structures [18]. In a comprehensive review, Meiburger highlighted that while automated localization and segmentation techniques represent a developing 'hot topic', gynecological applications remain largely focused on uterine fibroids or follicular monitoring for disease identification [19].

This technical focus on disease is particularly evident in the assessment of female reproductive function. Chen highlighted that while ultrasound is essential for evaluating ovarian reserve and endometrial receptivity (ER), systematic reviews specifically focusing on AI-aided ultrasound for these functional assessments are absent from the literature [7]. Current studies on reproductive segmentation focus primarily on ovarian follicles to diagnose conditions like polycystic ovarian syndrome (PCOS), while the foundational task of segmenting healthy endometrial remains overlooked.

Consequently, a clear gap exists between the technical literature and clinical application. In addition to cancer diagnosis studies from Swarnkar, a few more studies have explored deep learning segmentation for endometrial hyperplasia and endometriosis [20], [21], [22], [23], [24], [25], [26].

Fig. 2 presents a stacked bar chart visualizing the focus of endometrial image analysis research over the past decade. The data reveals a significant disparity. While deep learning applications are extensively utilized for diagnosing pathological conditions, their application in evaluating healthy endometrial receptivity (healthy endometrium) remains notably overlooked.

The absence of a dedicated synthesis for non-pathological endometrium segmentation represents a significant limitation in translating deep learning from research to clinical practice in assisted reproduction. Therefore, this review systematically addresses this gap by examining deep learning architectures, performance metrics, and clinical translation challenges specifically in the context of endometrial receptivity assessment, thereby providing a focused foundation for the methodological analysis that follows.

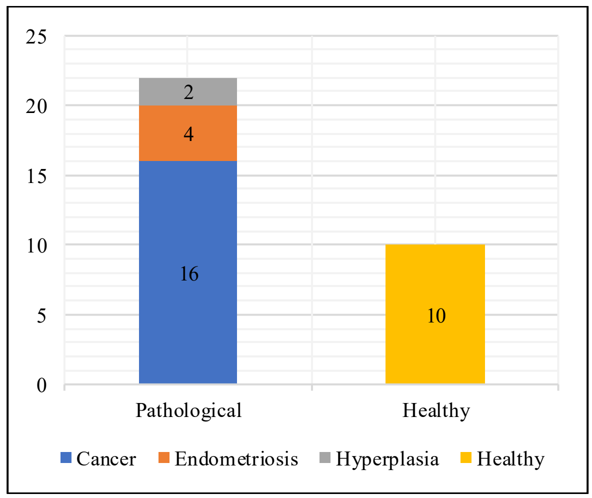

Fig. 2. Comparative analysis of research volume identifying the disparity between endometrial disease diagnosis and healthy assessment.

III. Endometrium Receptivity and Transvaginal Ultrasound

The endometrium, which forms in the uterine lining, is a dynamic tissue essential for female fertility and reproductive health. Throughout the menstrual cycle, the endometrial lining undergoes a series of dynamic changes. The changes include shedding during menstruation, followed by thickening during the proliferative phase and eventually transitioning into the secretory phase to prepare for potential embryo implantation. The condition when the endometrium is ready for embryo implantation is called endometrial receptivity (ET).

Endometrial receptivity refers to a specific period during the menstrual cycle when the endometrium is optimally prepared to facilitate embryo attachment and subsequent pregnancy [27], [28]. This specific period is known as the window of implantation (WOI). An accurate assessment of endometrial thickness and structure is crucial during WOI for improving IVF outcomes [29]. This is because inadequate endometrial development could lead to implantation failure and pregnancy loss, even with high-quality embryos [30], [31].

During in-vitro fertilization (IVF) procedures, ET plays a crucial role in embryo implantation. A receptive endometrial environment is essential for successful implantation and a healthy pregnancy [12].

Endometrial receptivity occurs in a brief period known as the window of implantation (WOI). This period typically occurs between 20 and 24 days of a regular 28-day menstrual cycle. During this critical period, within 3-5 days, the endometrium undergoes changes in molecular and structural modifications to support embryo implantation [32]. Thus, proper timing of embryo transfer within this window is crucial to ensure the success of implantation. Failure synchronization, displaced implantation window, of embryo transfer during this period could lead to implantation failure. The displaced implantation window is one of the causes of recurring implantation failure [33]. The percentage of women with recurrent implantation failure showing displaced implantation windows ranges from 17.7% to 80% [32], [34], [35]. Therefore, the need for an accurate assessment of endometrial receptivity is evident, and transvaginal ultrasound (TVUS) serves as a reliable tool in evaluating the endometrial receptivity for implantation.

The key parameters evaluated through TVUS include endometrial thickness, endometrial pattern, and vascular characteristics [36], [37]. These parameters provide valuable insights into the endometrium's readiness for implantation [38], [39]. A general rule suggests that a thin endometrium reduces implantation success, while an overly thick endometrium may also lower pregnancy rates. However, the optimal endometrial thickness for successful implantation remains inconclusive within the medical community [40]. A study demonstrated that embryo implantation, clinical, and ongoing pregnancy rates were significantly higher in patients with an endometrial thickness greater than 9 mm compared with those less than 9 mm [41]. A few other studies have concluded that the critical threshold of 7-14 mm is the optimal endometrial thickness for embryo transfer [5], [28], [42], [43]. A systematic review found that implantation rates tend to decrease when endometrial thickness is less than 7 mm, whereas a thickness greater than 14 mm does not significantly impact pregnancy outcomes compared to a medium endometrial thickness of 7–14 mm [44]. These inconsistencies underscore the need for additional research to establish a definitive range of endometrial thickness for successful implantation.

In addition to endometrial thickness, the endometrial pattern is a critical parameter for determining endometrial receptivity. Endometrial pattern assessment involves evaluating the endometrial echo pattern using ultrasound imaging. Several studies have demonstrated that the triple-line endometrial pattern is associated with higher implantation and pregnancy rates [12], [44], [45], [46]. Conversely, when the functional layer is non-uniform (heterogeneous), and the central line echo is unclear, receptivity is low [12]. While the endometrial pattern is not entirely overlooked, it tends to receive less attention than other parameters, such as endometrial thickness, especially in studies involving deep learning-based image analysis.

Despite TVUS being the gold standard tool to help medical practitioners monitor and assess endometrial receptivity, it has some limitations. TVUS images often suffer from noise and low resolution, particularly due to speckle noise. Due to this reason, images need to go through a pre-processing step called despeckling of ultrasound images [47]. Additionally, manual interpretation of ultrasound scans is highly operator-dependent, leading to variations in endometrial thickness measurements [4]. Using TVUS imaging, medical practitioners would freeze the screen and measure the endometrial thickness while monitoring the endometrial pattern. This manual segmentation process is tedious, laborious, and time-consuming. In addition, manual segmentation is influenced by the experience and knowledge of medical practitioners. To improve accuracy and consistency, computer-aided segmentation models are crucial for precisely analyzing the endometrium in TVUS images and assessing key endometrial receptivity biomarkers, including thickness and pattern.

IV. Image Processing for Endometrium Segmentation

In 1995, Pierson and Adams concluded that the development of computer-aided image analysis was a significant achievement for improving clinical management of ovarian health [48]. This promise was particularly compelling for the challenging domain of TVUS, where manual segmentation of the endometrium is operator-dependent, tedious, and time-consuming. In the years since, researchers have been exploring and developing image processing techniques to automate this critical task. This is because image analysis is essential for diagnosing pathologies and planning treatments. Primarily, the image processing techniques rely on pixel-level intensity information, spatial relationships, and mathematical morphological operations to delineate anatomical structures. Researchers have categorized these fundamental techniques into several groups:

- Thresholding and region-based methods, which segment based on pixel intensity and spatial homogeneity;
- Edge and contour-based approaches, such as active contour, which delineate anatomical boundaries;
- Morphological operations, used to refine segmented structures, and
- Classical machine learning classifiers, which leverage hand-crafted features.

Together, these techniques laid the foundational toolkit for automated analysis, establishing the feasibility of computer-assisted analysis. These techniques dominated the field until deep learning became superior. While computationally efficient and interpretable, traditional methods face significant challenges, including sensitivity to initialization, manual parameter tuning requirements, limited generalization across imaging conditions, and difficulty handling the inherent noise and ambiguous boundaries characteristic of ultrasound images.

A. Thresholding

Thresholding techniques, while computationally efficient and straightforward, operate by classifying pixels into foreground and background based on a predefined intensity value [49], [50]. This technique is particularly useful for tasks like pre-segmentation or initial region of interest generation, but often requires further refinement due to noise inclusion and an inability to adapt to a variety of image qualities. For instance, segmenting follicles in ultrasound images frequently utilizes thresholding, but the method often deals with noise, thereby needing additional refinement to achieve acceptable accuracy [51]. To address these issues, various thresholding methods were developed, and for medical imaging applications, multi-level thresholding methods were developed to handle complex anatomical structures from medical imaging [52], [53], [54], [55], [56]. Nevertheless, their application to complex anatomical structures, such as the endometrium in TVUS images that are often affected by high noise, low resolution, and unclear boundaries, thresholding techniques remain inadequate for achieving robust and clinically reliable delineation [57], [58].

B. Region-Based Techniques

These techniques focus on grouping neighboring pixels with similar intensity or texture characteristics to overcome the limitation of edge detectors in a noisy environment. Techniques like region growing, for instance, initiate from a seed point and iteratively expand by incorporating neighboring pixels that meet specific homogeneity criteria to construct a comprehensive endometrial region [57]. However, this technique often requires manual seed-point initialization, making it less suitable for fully automated pipelines, especially given the inherent variability in ultrasound image quality and anatomical presentation [59]. Consequently, the reliance on manual seed-point selection introduces subjectivity and can lead to inconsistencies across different medical practitioners, highlighting the need for automated or semi-automated initialization strategies [60].

C. Edge and Contour-Based Techniques

Edge-based and contour-cased techniques have been widely explored for medical image segmentation due to their ability to delineate structural boundaries. Edge detection methods identify discontinuities in image intensity that correspond to tissue interfaces [61], [62], [63]. In contrast, contour-based methods, such as active contours, evolve deformable curves to conform to these boundaries under the influence of internal and external forces [64], [65], [66], [67], [68]. The active contour model, often called snakes, utilizes energy functional minimization to guide the contour towards desired boundaries. This approach offers improvements in adaptability and robustness for complex anatomical structures. However, these techniques require manual initialization and remain highly sensitive to blurred or weak edges and noise, particularly prevalent in ultrasound images, where variations in lesion morphology can further complicate accurate segmentation [19].

D. Morphological Operations

Morphological operations use structuring elements to refine segmented masks, remove small artifacts, and enforce shape constraints. Typically, morphological operation is used to improve defects that are recognized during segmentation [69]. Two basic morphology operations are dilation and erosion. Dilation expands the boundaries of the foreground objects, while erosion shrinks them. Often, the combination of these operations, such as opening and closing, can effectively smooth contours, fill small holes, and remove isolated pixels, thereby enhancing the quality of the endometrial segmentation [70]. Despite their utility in post-processing, morphological operations are primarily refinement tools and do not independently perform segmentation, often requiring other methods to produce an initial segmentation mask.

While these traditional segmentation techniques have laid the foundation for computer-aided image analysis for ovarian health, their performance varies depending on image quality and anatomical complexity. As imaging complexity increased and clinical demand for reliability grew, the limitations of these techniques became apparent. Table II provides a summary of each technique with its strengths and weaknesses.

TABLE II. SUMMARY OF TRADITIONAL SEGMENTATION METHODS

Techniques	Strengths	Weaknesses
Thresholding [49], [50]	• Simple to implement • Fast preprocessing step	• Fails with low contrast image • Sensitive to noise • Fixed threshold
Region-Based [71]	• Capture spatial continuity • Adapts to the region's shape	• Manual seed initialization • Leakage in noisy regions
Edge & Contour Based	• Precise boundary localization • Can handle complex shapes	• Requires a good edge contrast image • Manual initialization •
Morphology Operation [69], [70], [72]	• Capable of removing small artifacts • Capable to smooth and filling holes in the image	• Tendency to over-smooth fine structures in the image • Dependent on element size

The early attempts of automation in endometrium segmentation relied on basic algorithms, such as thresholding, active contours, and region-growing methods [6], [10]. Despite their usefulness, these approaches often struggle to address the fundamental challenges of ultrasound imaging, such as low contrast, speckle noise, and anatomical variability [58]. While these traditional methods provided the foundation, their limited robustness and generalizability eventually drove a shift toward more advanced solutions. Over the past decades, as this article was written, researchers have extensively explored and investigated the application of automated segmentation of ultrasound images, evolving into the application of deep learning-based approaches.

V. DEEP LEARNING MODELS FOR ENDOMETRIUM SEGMENTATION

In recent years, extensive literature reviews have been conducted on deep learning approaches to medical image analysis, highlighting methodologies, technical progress, and clinical translation limitations [10], [73], [74], [75], [76], [77]. These efforts have shown deep learning as a transformative paradigm, demonstrating superior performance in image processing across diverse medical imaging modalities. Despite the various studies and reviews in deep learning for medical images, only a few studies have focused on deep learning segmentation of the endometrium in TVUS images for endometrium receptivity assessment (ERA). This highlights the need to explore alternative deep learning models for ERA to ensure repeatability and high segmentation accuracy in clinical applications. To address this issue, recent research has applied a range of deep learning architectures aiming to improve segmentation accuracy, robustness, and clinical applicability specifically for endometrium analysis.

A. Foundation: Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is the baseline for deep learning in medical image analysis due to its ability to capture spatial features and patterns effectively [78], [79], [80]. These networks leverage convolutional layers to automatically learn hierarchical representations from raw image data, eliminating the need for manual feature identification and extraction. This ability is crucial in medical imaging for tasks such as identifying anatomical structures of the endometrium, where subtle textural and morphological cues are vital for accurate segmentation and subsequent analysis [81]. The strength of CNNs lies in their capacity to capture local spatial features, such as edges, textures, and patterns, through convolutional filters. In the context of endometrium segmentation, CNNs have shown potential in learning discriminative features from TVUS. However, a fundamental limitation of CNNs is their reliance on local context, which constrains their ability to model long-range dependencies and global context within an image [82]. Fig. 3 below shows the basic architecture of a convolutional neural network.

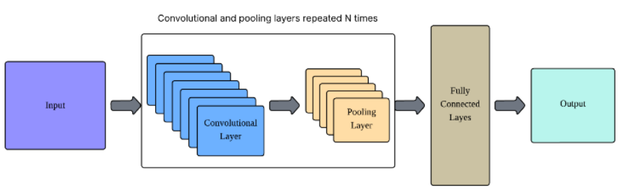

Fig. 3. Basic architecture of a convolutional neural network.

1) U-NET and its variant: Among CNN-based architectures, the U-Net architecture is the most prominent in medical image segmentation. This is due to its iconic encoder-decoder structure and skip connections. The architecture effectively combines multi-scale feature extraction with precise spatial localization [83].

From this fundamental design, numerous adaptations have emerged for medical image segmentation, in which annotated datasets are limited. In endometrium segmentation, U-Net-based architectures have been widely used due to their efficiency in learning from small TVUS datasets. Variants such as VGG-based U-Net [3], ResNet50-U-Net [84], and 3D U-Net [57] have been developed to enhance feature extraction and volumetric analysis. Fig. 4 is the architecture of the ResNet50-UNet model.

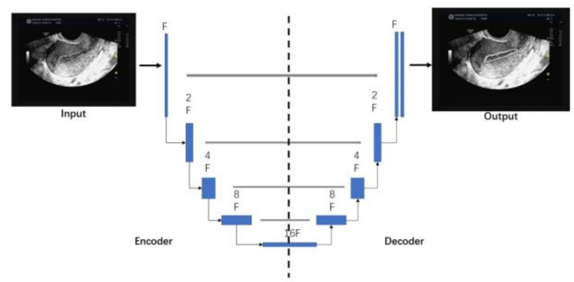

Fig. 4. Schematic diagram of the architecture ResNet50-UNet model from the study [79].

For instance, the VCG-based U-Net achieved a Dice score ranging from 0.83 to 0.9, with a mean absolute error of 1.23 mm and 1.38 mm across two datasets based on [3]. Liu et al. (2022) proposed a ResNet50-SegNet deep learning model for endometrial segmentation, achieving a Dice coefficient of 0.82 with thickness errors within ±2 mm. However, a major limitation of this method was its low segmentation accuracy for endometrial linings ≤3 mm, with an accuracy rate of only 55.3% [4]. In comparison to the previous architecture, Fig. 5 below shows the architecture of ResNet50-SegNet.

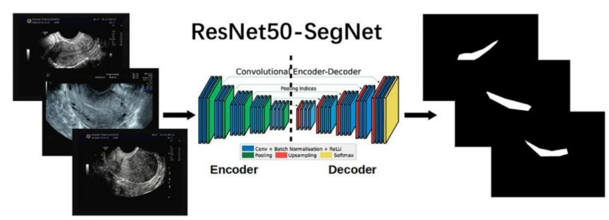

Fig. 5. ResNet50-SegNet architecture [4].

Meanwhile, the 3D U-Net has been successfully applied to volumetric TVUS data, capturing inter-slice contextual information, achieving a 0.91 Dice score with 94.20% of automatic thickness measurements falling within clinically accepted error margins [57]. These variants collectively demonstrate the adaptability and effectiveness of the U-Net models. Despite these achievements, these architectures are still facing challenges in capturing long-range dependencies and global contextual information.

B. Transformer-Based Application

In recent years, transformer-based architectures have gained attention in medical image segmentation due to their ability to overcome the key limitation of CNNs, modeling long-range dependencies and global contextual relationships. Originally, this architecture was introduced in natural language processing (NLP) [85]. Its self-attention mechanism allows the network to capture inter-pixel relationships across the entire image, building a holistic understanding of image structure [86]. This capability is particularly beneficial in medical imaging, where global context is often critical for accurate segmentation of complex anatomical structures like the endometrium [1]. Fig. 6 shows the transformer model overview.

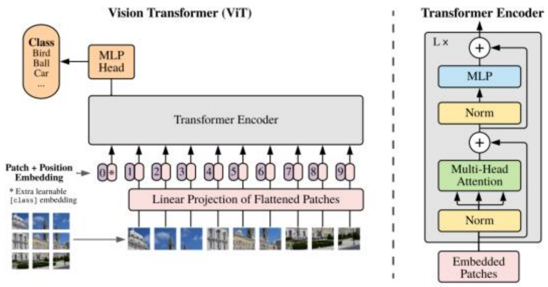

Fig. 6. Vision Transformer (ViT) overview [86].

C. Hybrid and Adversarial Architectures

To overcome the limitations of CNNs, researchers have developed hybrid and adversarial learning frameworks to enhance the segmentation performance. Hybrid models integrate the feature extraction capabilities with other computational approaches or architectural innovations. A prominent example is the Deep Learned Snake model proposed by Singhal [87], which incorporated an active contour model with a deep learning model for curve evolution. From this approach, Park et al. 2019 introduced a key-point discriminator that used endometrial boundary detections from an FCN to guide the active contour model and achieved a Dice score of 82.76%.[58]. These frameworks effectively mimic human experts' analysis, allowing the segmentation to learn more precise delineations. While these approaches demonstrated a significant enhancement in boundary detection, challenges persist in managing the architectural complexity and dependence on large annotated datasets for optimal performance.

D. Emergence Models and Foundational Architectures

The latest phase of advancement in endometrium segmentation research has witnessed the emergence of innovative architectures that emphasize generalizability, computational efficiency, and reduced reliance on large annotated datasets. Deep Dual-Resolution Networks (DDRNet) represent an approach employing parallel branches to efficiently process both high- and low-resolution features, while combining contextual and spatial information [60]. This design allows for a more nuanced understanding of complex endometrial morphology, combining fine-grained details with broader anatomical context.

Building on the evolution of foundation models, recent studies have explored adaptation of the segment anything model (SAM) for medical image segmentation. For example, Qiu introduced a segment anything with inception module (SAIM), which integrates inception-based encoders and point prompts to improve segmentation precision in the noisy ultrasound data [88]. The study utilizes an open-source SAM and introduces enhancements to the image encoder structure. Fig. 7 shows the SAIM architecture.

Another notable method, MultiStudentNet, leverages the weights of multiple models to facilitate feature sharing. This method employs multiple student models to integrate labeled and unlabeled data to improve reliability and model robustness [89]. Fig. 8 shows the framework for MultiStudentNet.

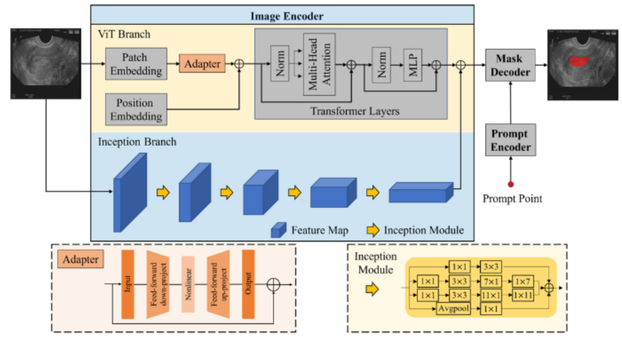

Fig. 7. SAIM architecture [88].

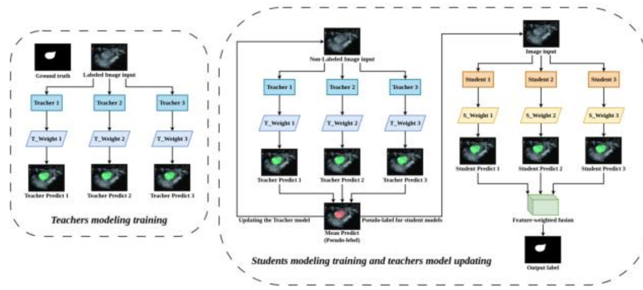

Fig. 8. MultiStudentNet architecture [89].

VI. Methodology

Following established methodological guidelines for a systematic literature review [90], this review was conducted in accordance with the Preferred Reporting Items for Systematic Review and Meta-Analysis (86). The review process encompasses five key stages: 1) search strategy development, 2) study selection and screening, 3) data extraction, 4) quality assessment, and 5) narrative synthesis. Each of these stages is explained in the subsections.

A. Search Strategy and Information Resources

The literature search began in April 2025. This search employed Google Scholar as the primary search database to identify and extract relevant primary research studies. The primary database is complemented by manual cross-checking in IEEE Xplore, Semantic Scholar, Springer Link, and ScienceDirect to ensure coverage of the study. The search strategy for this study used a Boolean operator and a combination of keywords related to endometrium segmentation, deep learning, and transvaginal ultrasound. The Boolean query string was (“endometrium segment” OR “endometrial segmentation”) AND (“deep learning”) AND (“TVUS” OR “Transvaginal Ultrasound”).

B. Eligibility Criteria

In the initial search, 453 papers were found in Google Scholar, 139 in Springer Link, 36 in Semantic Scholar, 24 in ScienceDirect, 4 in PubMed, and 3 in IEEE Xplore, totaling 659 papers. The inclusion and exclusion criteria for the paper selection in this review are as follows:

Inclusion Criteria:

- Paper published in English between 2015 and 2025.

- Application related to endometrium receptivity assessment, thickness measurement.
- Studies with mixed cohorts (pathological and healthy) were included if receptivity-related outcomes were reported.

Exclusion Criteria:

- Duplicate papers between the databases
- Paper published before 2015
- Inaccessible paper
- Non-English paper
- Studies focusing exclusively on pathological segmentation

The exclusion criteria narrowed down the papers to a total of 210 papers. Then, a total of 190 papers were discarded from the keywords, title, and abstract screening. After that, the exclusion criteria for non-pathological disease of segmentation were selected, and a total of 10 papers were analyzed. Fig. 9 below shows the simplified visual of the search methods conducted for the literature review based on the PRISMA guideline [91].

C. Data Extraction

Data extraction for this review was guided by the four research questions raised in the introduction. For each included study, relevant information was systematically extracted to address these questions. This structured approach ensured all extracted data directly supported the analytical aim of this review.

To address RQ1, the types of deep learning models, backbones, and any hybrid or custom modifications were documented and categorized. This included recording architectures such as U-Net variants, Transformers, and hybrid frameworks like SAIM to identify the most frequently employed architecture in TVUS endometrium segmentation.

For RQ2 and RQ3, quantitative performance metrics and evaluation approaches were extracted. Segmentation metrics and thickness-measurement errors were recorded to facilitate accuracy comparisons across deep learning models. All reported evaluation metrics and assessments were further examined for their alignment with clinical standards, such as the acceptance rate of thickness-measurement within +-2mm, to determine which metrics are regarded as gold-standard for receptivity assessment in ART.

Finally, to answer RQ4, author-reported limitations and clinical translation hindrances were recorded. This included documenting issues related to dataset size, single-center bias, model generalizability, and clinical-integration challenges. These insights were synthesized to identify the gaps and to guide future research directions in the translation of deep learning models into ART workflows.

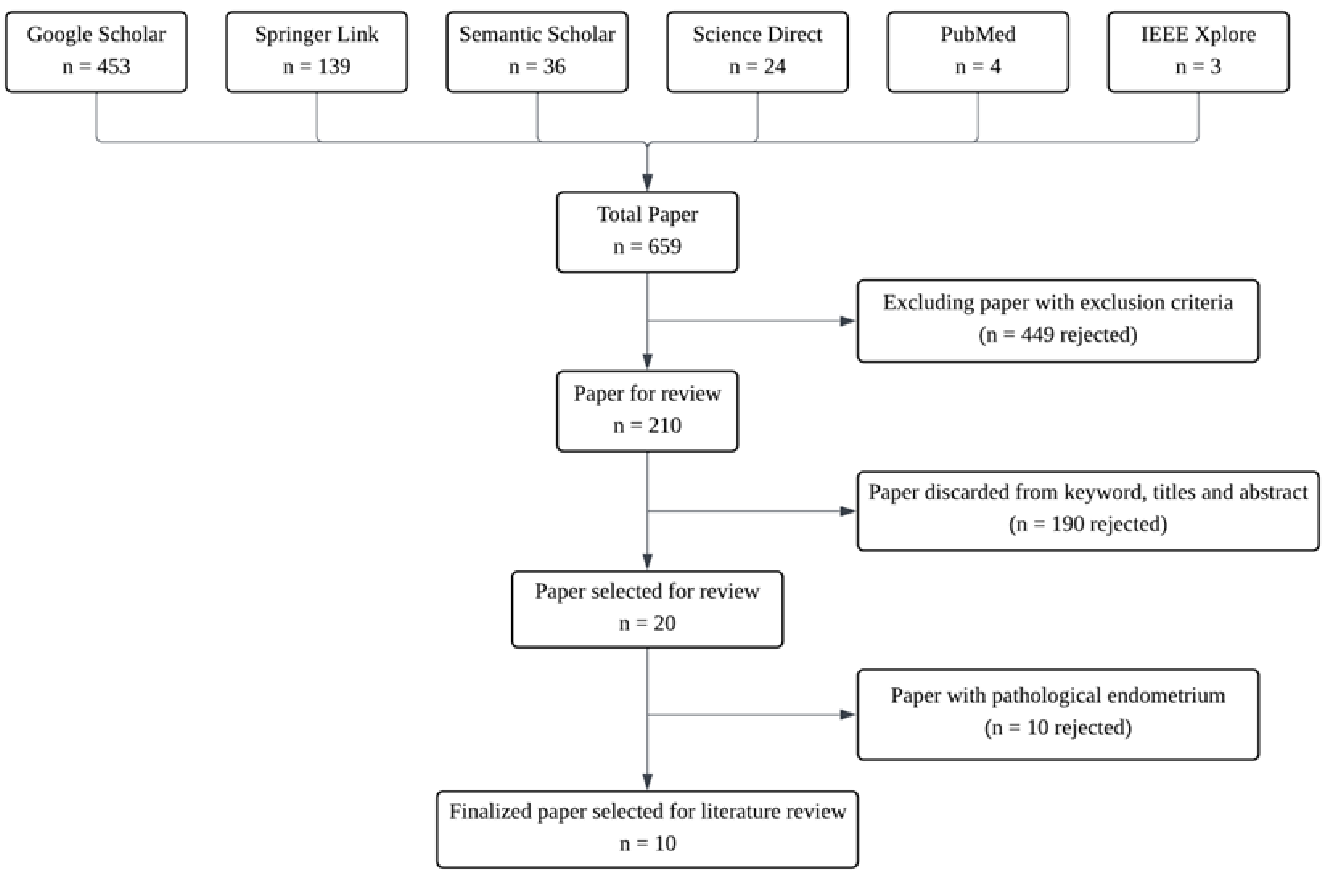

Fig. 9. PRISMA flow diagram of included studies.

VII. Results

A. RQ1: Predominant Deep Learning Architectures for TVUS Endometrium Segmentation

Endometrium segmentation in medical imaging is a critical yet challenging task for quantitative diagnostic analysis. Based on the literature, the most common approaches for endometrial image segmentation have increasingly relied on deep learning (DL) techniques, particularly in Convolutional Neural Networks (CNNs) as the foundation architecture. This is due to the proven effectiveness of the models in medical image analysis [3]. Accordingly, researchers have adapted and modified core CNNs to develop various segmentation network architectures specifically aimed at delineating the boundaries of the endometrium lining.

Built upon CNN architectures, the U-Net architecture and its numerous variants are most frequently cited as the benchmark for this specific segmentation task. U-Net variants have been established as the backbone of segmentation methodologies [92], [93], [94]. The classical U-Net architecture has been widely adapted into diverse variants, such as the VGG-based U-Net, which utilizes the VGG16 network as its encoder for hierarchical feature extraction to enhance spatial representation [3]. In comparative evaluations, a pure U-Net model was tested for endometrium segmentation [57], [58], [84]. Furthermore, the use of ResNet-based U-Net and SegNet models shows the importance of deep residual networks, pre-trained on large-scale datasets such as ImageNet, in mitigating vanishing gradient problems and improving feature depth [4], [84], [95]. For volumetric analysis, the 3D U-Net extension is employed to process three-dimensional transvaginal ultrasound (3D TVUS) images, thereby capturing crucial inter-slice structural information [57].

Beyond U-Net, other general frameworks such as Deep Dual-Resolution Networks (DDRNet) and SegNet are also commonly utilized for this purpose. One comparative study evaluated six different models based on these architectures, specialized combinations such as ResNet50 U-Net, ResNet50 SegNet, U-Net mini, VGG SegNet, alongside pure U-Net and DDRNet. The result from this comparison found that DDRNets showed the best endometrium segmentation with a Dice score of 0.895 [84] using 1050 images. Each architecture, while built on CNNs, employs a different architecture to optimize performance for medical image segmentation.

Innovative approaches have also emerged through the development of hybrid variational models. The hybrid models combine data-driven and energy-based techniques. One notable example is the Deep Learned Snake (DLS) model proposed in 2019 [87]. In the research, the hybrid model integrates the robustness of a Fully Convolutional Network (FCN) with the mathematical rigor of a variational level-set method for curve propagation. In this framework, the FCN generates a deep-learned endometrium probability map that acts as a soft, adaptive shape into the level-set functional for curve propagation. This model combined the robustness of learned features with the interpretability of deformable models. The efficacy of this hybrid method is demonstrated by its performance, yielding approximately 30% improvement over standalone supervised learning methods and producing endometrial thickness measurements within a clinical tolerance of 2mm in 87% of cases.

Additionally, recent research explores adversarial learning frameworks to further refine segmentation precision. In 2019, Park et al. proposed a key-point discriminator, informed by anatomical landmarks and spatial features, to strategically guide the segmentation network towards more accurate boundary delineation in the presence of ultrasound artifacts [58]. The discriminator trained the segmentation network to predict accurate boundaries, to ensure robust segmentation for unclear edges and heterogeneous texture. This method achieved a Dice score of 82.67% and a Jaccard index of 70.46%.

More recently, the architectural evolution in endometrial segmentation is shifting into utilizing Transformers, which are uniquely capable of capturing long-range, global image context in comparison to pure CNNs. A leading example is the Segment Anything with Inception Module (SAIM) [88]. This method utilizes a dual-branch encoder that comprises of Vision Transformer (ViT) for global context and an Inception-based CNN for multi-scale feature extraction. This architecture is critically dependent on clinician-guided point prompts to isolate the endometrium. This approach achieved a segmentation result of Dice Score 76.31% and Intersection over Union of 63.71% using a single-point prompt. Furthermore, its performance improved with additional prompts, reaching a Dice score of 81.30% with five-point prompts [88].

While U-Net-based models form a common foundation, the formation of their variants shows that the U-Net model requires significant task-specific modifications. As the literature affirms, an effective solution must be customized to overcome specific limitations, especially in TVUS. Thus, the shift towards current and advanced hybrid systems, such as SAIM, shows that the most performed models are hybrid systems, which integrate multiple network architectures.

Table III synthesizes the architectural landscape and key characteristics of each model reviewed. The table illustrated the predominance of U-Net variants, while documenting the emergence of hybrid and transformer-based architectures. The domination of UNet variants is driven by three core factors: the ability to preserve critical boundaries, the capacity to capture global and local, and reliable performance on small datasets.

U-Net variants can preserve critical boundaries through skip connections that transfer high-resolution features from the encoder directly to the decoder. This mechanism endures that local spatial information, which typically lost during the down sampling process, is retained to facilitate precise localization of the endometrial lining. Consequently, the architecture effectively integrates spatial and contextual information across its symmetrical path to achieve accurate pixel-level predictions. This precision is vital for identifying the thin, echogenic line of endometrium, especially in cases where boundaries are blurred or highly irregular.

Furthermore, U-Net is celebrated for its reliable performance and robustness when trained on small datasets, a common limitation in clinical reproductive research. This is because medical datasets often consist of only a few hundred images and often are not shareable. U-Net's lightweight and simple structure makes it inherently less prone to overfitting. These characteristics make U-Net particularly suited to the challenges of endometrium segmentation in TVUS images, where boundary clarity is low, anatomical information is essential, and annotated data are limited.

TABLE III. SUMMARY OF DEEP LEARNING ARCHITECTURES IN INCLUDED STUDIES

No	Study (Author, Year)	Architecture Family	Specific Model / Variant	Backbone / Key Components	Dataset Size	Key Architectural Notes
1	Hu et al., 2019 [3]	CNN-based (U-Net)	VGG-based U-Net	VGG16 encoder	73	Skip connections, encoder-decoder
2	Park et al., 2019 [58]	Hybrid	Key-point Discriminator	FCN + adversarial guidance	320	Adversarial learning for boundary refinement
3	Singhal et al., 2017 [87]	Hybrid	Deep Learned Snake	FCN + active contour	110	Combines DL with variational methods
4	Wang et al., 2022 [57]	CNN-based (U-Net)	3D U-Net	3D convolutional blocks	113	Volumetric segmentation
5	Liu et al., 2022 [4]	CNN-based	ResNet50-SegNet	ResNet50 encoder	1050	Encoder-decoder with residual connections
6	Liu et al., 2022 [84]	CNN-based (U-Net & SegNet)	U-Net SegNet	ResNet50, Vanilla CNN, Vanilla Mini, VGG 16, DDR-Net	840	U-Net with residual backbone
7	Qiu et al., 2024 [88]	Transformer-based	SAIM	ViT + Inception-CNN	180	Hybrid transformer-CNN with prompting
8	Peng et al., 2024 [89]	Advanced / Semi-supervised	MultiStudentNet	Ensemble of CNNs	215	Semi-supervised, multi-model
9	Ithani et al., 2024 [96]	CNN-based (U-Net)	Classic U-Net	Basic encoder-decoder	25	Small-scale feasibility study
10	Yan et al., 2024 [95]	CNN-based (U-Net)	Custom U-Net variant	Not specified	180	Clinical receptivity-focused

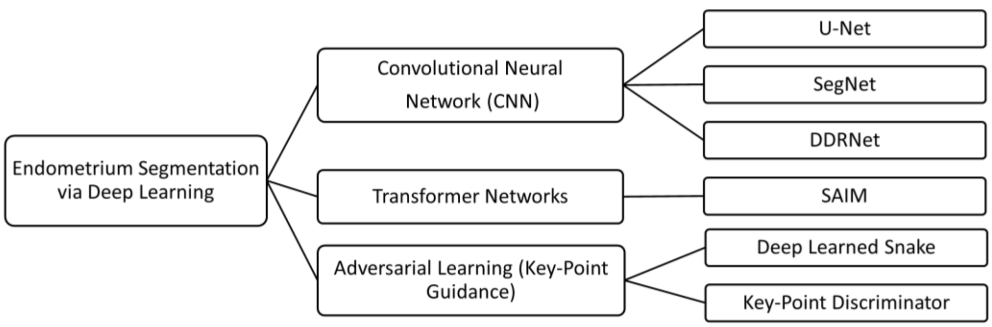

Fig. 10. Taxonomy of current deep learning architectures for healthy endometrium segmentation.

Fig. 10 shows the simplified taxonomy of the deep learning architecture models for endometrium segmentation, illustrating the evolution from CNN-based foundations towards hybrid and transformer-integrated architectures. The following section, which answers RQ2, examines whether this architectural prevalence correlates with superior segmentation accuracy or if emerging models offer a competitive advantage under specific clinical or dataset conditions.

B. RQ2: High Accuracy Deep Learning Architectures

Determining a singular superior method for endometrial segmentation is challenging because segmentation performance highly depends on various conditions, including dataset size, image quality, and evaluation metrics. Nevertheless, recent studies highlight several deep learning architectures that have demonstrated notably high accuracy, though these results must be interpreted considering dataset characteristics and clinical applicability.

For example, the foundational U-Net architecture achieved a remarkable Dice coefficient of 0.977 in one study. However, this result was achieved on a limited dataset of only 20 training and 5 test images [96]. This highlights an important point where simpler models like U-Net can perform well due to a small dataset available. Consequently, while this result is numerically high, its generalizability using larger and more diverse clinical datasets remains uncertain.

For volumetric analysis, the extension to 3D U-Net has proven significantly more effective than its 2D counterpart for processing three-dimensional transvaginal ultrasound (3D TVUS) images [57]. In a dedicated study, the 3D U-Net achieved a Dice coefficient of 90.83%, a metric value higher than the results obtained from 2D segmentation approaches on the same volumetric data. This performance [57] is attributed to the model's capacity to capture crucial inter-slice contextual information and spatial relationships within the full volume. Therefore, for applications requiring holistic anatomical assessment, 3D convolutional approaches represent a highly accurate solution.

Beyond standard supervised learning, semi-supervised frameworks such as Multi-StudentNet are designed to enhance accuracy by leveraging both labeled and unlabeled data. Thus, it reduces the heavy reliance on manual annotations. This architecture reported an overall Dice score of 0.81, with its performance further increasing to 92.33% for polyp cases and

90.20% for cancer cases [89]. This indicates a particular robustness in segmenting pathologically altered endometria, which often present more complex imaging characteristics. The model's design utilizes an ensemble of student models, which has effectively improved segmentation reliability and performance across varied clinical presentations.

Among architectures designed for efficient and contextual feature extraction, Deep Dual-Resolution Networks (DDRNets) have demonstrated leading performance in comparative analyses with an average Dice score of 0.895 [84]. This high accuracy is facilitated by its design of two parallel branches operating at different resolutions, which enables effective multi-scale information fusion. The architecture's efficiency also suggests strong potential for real-time semantic segmentation applications in clinical settings.

Other deep learning models have also achieved competitive results by using different design improvements. The VGG16-based U-Net, employed in an automated measurement pipeline, achieved a Dice score of 0.83, performing better at 0.853 in the proliferative phase compared to 0.796 in the secretory phase. Similarly, a key-point discriminator framework significantly improved results compared to the standard U-Net of 58.69%, achieving a Dice score of 82.67% [58]. Furthermore, the Segment Anything with Inception Module (SAIM) adaptation recorded a Dice score of 76.31% and increased to 81.30% with multiple point prompts [88]. The comparative performance of the reviewed architecture is summarized in Table III. The table also highlights the influence of dataset size and clinical applicability.

TABLE IV. LIST OF METHODS AND RESULTS FOR ENDOMETRIUM SEGMENTATION

No.	Method	Result
1	DLS	DLS 85%-87% tolerance limit compared to U-Net with 60%-70% tolerance limit.
2	VGG-Based U-Net	DSC 0.83 Thickness measurement:- MAE/RMSE: 1.23/1.79 on 27 DL test cases- MAE/RMSE: 1.38/1.85 on 46 Thickness test cases
3	Key-discriminator	DSC 82.67%, Jaccard 70.46%
4	6 Different U-Net	DDRNets is the highest average of DSC, Recall, Precision, and Specificity with 0.895, 0.884, 0.910, and 0.998.
5	ResNet50-based SegNet	Average DSC: 0.82
6	3D U-Net	DSC 90.83%, Jaccard 83.35%, Sensitivity 90.85
7	Multi StudentNet	DSC 0.81, specificity 0.99, sensitivity 0.87
8	SAIM	DSC 76.31% and IoU 63.71%, a higher multipoint prompt receives a higher result in metrics DSC and IoU.
9	Transformer - CNN	U-Net DSC 0.977, Transformer DSC 0.956
10	ResNet-50	AUC 0.853

In conclusion, while a classic U-Net excels on minimal data, the most accurate and generalizable models for broader clinical application appear to be the 3D U-Net, Multi-StudentNet, and DDRNet, as evidenced by their consistently high Dice scores across more substantial datasets. However, to draw a definitive conclusion about which network is highly accurate is challenging and potentially misleading because all these studies used different datasets with varying characteristics and sizes. This is due to the limited availability of common datasets for healthy endometrium TVUS images. For example, the study using the 3D U-Net had a total of 113 images [89], in comparison to the study comparing six networks that had a total of 1050 images [84]. To address this limitation, the MultiStudentNet specifically uses a semi-supervised approach [89].

C. RQ3: Metrics and Validation for Deep Learning in Endometrium Segmentation

The evaluation of deep learning models for endometrium segmentation requires a comprehensive approach that reflects both technical segmentation accuracy and clinical measurement reliability. To achieve this, researchers commonly employ a range of quantitative and statistical metrics to assess segmentation accuracy and endometrial thickness measurement performance. Table IV presents the list of methods and results for endometrium segmentation.

1) Segmentation evaluation metrics: For endometrium segmentation, the objective is to measure how close the automated segmentation overlaps with the ground-truth annotation. One of the most used metrics is the Dice Similarity Coefficient (DSC). The values range from 0 to 1, with 1 as a perfect overlap between the segmented region and the ground truth. In different literatures, DSC is also known as Dice Score, Dice coefficient, and F1 Score. The formula for measuring DSC is defined as:

$$DSC = \frac{2|A \cap B|}{|A|+|B|} \tag{1}$$

where, A is the segmented region to be assessed, and B is the corresponding ground truth.

The recoded DSC values were reported in most of the main literature for this study. 9 out of 10 papers used DSC values as a metric to evaluate segmentation performance. The recorded DSC values for healthy endometrium segmentation using deep learning range from 0.76 to 0.98. The highest DSC was observed in the U-Net with 0.977 on a limited dataset of 25 images [96]. This indicates that while DSC is a useful benchmark for general overlap, it can be sensitive to dataset size, and a high DSC does not automatically mean an excellent segmentation result.

A recent deep learning model also measured a high DSC at 0.895 for a Deep Dual-Resolution Network (DDRNet) in a comparative analysis [84]. However, while proposing the key-point discriminator method, the DSC value recorded was 82.67%, outperforming conventional methods like U-Net (58.69%) and FCN (78.39%) [58]. This shows that a little tuning in the model architecture can improve segmentation results, especially if the model is designed to better identify boundaries.

Another commonly used metric for segmentation evaluation is the Jaccard Index, also known as Intersection over Union (IoU). The DSC and IoU are proportional to the number of spatial overlaps between the segmented and ground truth images. The value ranges from 0 to 1, with 1 as the perfect matching [97]. The formula of IoU is defined as:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

However, the Jaccard index was reported far less frequently than DSC in the included papers. For this review, out of 10 main literature papers, only 3 studies use the Jaccard index as a performance metric. Values reported include 70.46% using the key-point discriminator method [58] and 83.35% using the 3D U-Net method [57]. This low number of adoptions in evaluation may be attributed to the Jaccard index's mathematical property of generally resulting in lower numerical values compared to Dice for the same segmentation quality. However, for clinical applications such as endometrial thickness measurement, Jaccard index sensitivity to boundary intersection may be helpful in a realistic assessment.

This pattern of selective metric reporting extends to a few other complementary metrics for segmentation evaluation. For example, another commonly used metric is accuracy. The accuracy value is calculated through precision and recall, which quantify the model's ability to correctly identify endometrial pixels while minimizing false positives and negatives. In certain papers, recall is also known as sensitivity, which measures the proportion of relevant instances of true positive pixels that are retrieved by the model and focuses on how well the model avoids false negatives. The formulas for accuracy, precision, and recall are defined below:

$$Accuracy = \frac{TP+FN}{TP+FP+TN+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall\ (Sensitivity) = \frac{TP}{TP+FN} \quad (5)$$

where, TP- True Positive, FP - False Positive, TN - True Negative, FN – False Negative.

In a comparative study, a DDRNet demonstrated superior performance, achieving a precision of 0.910 and a recall of 0.884, the highest among the evaluated deep learning models. This review revealed that these complementary metrics are infrequently documented, with only 3 out of 10 papers reporting precision and 2 out of 10 papers reporting recall (sensitivity). This indicates that while precision and recall provide distinct and complementary insights in quantifying false positive and false negative rates, the Dice score continues to be the favored standard metric for overall segmentation evaluation.

Beyond the mentioned metrics, specificity is another segmentation metric that measures the proportion of correctly identified negative pixels out of all actual negative pixels. This metric is also known as the true negative rate, as it indicates the model's ability to minimize false positives [84], [89]. The formula for specificity is written as:

$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

These metrics were reported in a study comparing variations of CNN architectures and evaluation in the developing Multi-StudentNet model. The reported specificity value is generally 0.99 [84], which is high, indicating robust performance in excluding non-endometrial regions. In medical image segmentation, specificity is important because the background makes up most of the image. High specificity shows the model is good at not mistakenly labelling normal tissue as endometrium. This helps prevent false alarms and builds trust in the tool for clinical application. Although specificity provides valuable information about background exclusion, it is less frequently reported because the aim is to detect the endometrium boundary, the foreground region of interest, and not to verify the background classification.

Another underrated metric is the 95th percentile Hausdorff Distance (HD95). This metric measures the maximum boundary between a predicted segmentation and the ground truth. While Dice scores calculate overall overlap pixels, the HD95 specifically targets the worst-case alignment errors by using the 95th percentile to ignore extreme outliers. In this study, this metric was explicitly used in only one paper [57]. The 3D U-Net model achieved the best HD95 score of 12.75mm when using the Enhanced Augmented Data (EAD) method.

From this information, a boundary-focused metric like HD95 is more clinically informative than a general overlap metric like DSC because it directly relates to measurement accuracy. Since clinical decisions are based on measurements like endometrial thickness, a minor boundary error can lead to a significant measurement error. HD95 assesses this specific risk. Thus, the HD95 provides a more direct assessment of a model's utility in a clinical setting, where the ±2 mm tolerance is critical.

2) Thickness measurement performance metrics: As established, endometrial thickness serves as a critical biomarker for assessing reproductive health, guiding clinical decisions in areas such as fertility treatments and screening. Consequently, evaluating the performance of thickness measurement moves beyond overlap metrics. The performance of this measurement requires rigorous evaluation against the clinical gold standard of manual expert measurement. This is to ensure that the automated results are reliable and clinically acceptable. Among the relevant metrics are Acceptance Rate, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE).

The reliability of the evaluation is benchmarked against a well-defined clinical tolerance. Clinical tolerance represents the acceptable error range, which, for endometrial thickness, is consistently set at ±2mm based on clinical guidelines [3], [87]. Based on the clinical tolerance, the acceptance rate measures the percentage of a model's measurements that fall within the range. Thus, model efficacy is primarily reported through Acceptance Rate. The definition of Acceptance Rate in mathematical expression is:

$$Acceptance\ Rate = \frac{\sum_{i=1}^{n} |y_i - x_i| < 2mm}{n} \times 100\% \quad (7)$$

where, y is the endometrial thickness measured by deep learning methods, x is the ground truth, and n is the total number of validation images or cases.

By achieving a high acceptance rate, such as 89.3% [4] or 94.20% [57], the deep learning methods demonstrate their potential to provide a reliable measurement that aligns with clinical requirements. Notably, 4 out of 10 papers measure the acceptance rate of the thickness measurements.

While the Acceptance Rate indicates clinical applicability, the MAE and RSME provide deeper insight of the measurement errors. The MAE calculates the average absolute difference, a straightforward measure of typical error magnitude [4], [57]. In contrast, RSME is mathematically structured to evaluate error magnitude, serves as a sensitive indicator for inconsistent measurement, and the presence of significant outliers [57], [98]. The formulas of MAE and RSME are written as follows:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{8}$$

$$RSME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \tag{9}$$

Reported MAE values demonstrate model accuracy, with methods like VGG-based U-Net pipeline achieving MAEs of 1.23mm for the deep learning test set and 1.38mm for the thickness test set [3]. The results are considered as it falls within the clinical tolerance of ±2mm. The 3D U-Net segmentation method reported an even lower MAE of 0.75 mm with a corresponding RSME of 1.07mm [57]. For comparison, the ResNet50-SegNet model reported an overall MAE of 2.3mm across all validation data, although performance improved to an MAE of 2.0mm when analyzing cases where endometrial thickness is more than 3mm [4].

The evaluation of endometrial thickness measurement remains overlooked in the current included papers. The most critical metric, the acceptance rate, was reported on 40%. The fact that 60% of the included papers did not report this value may suggest that most studies are technically focused rather than clinically validated. This shows a great hindrance for clinical translation.

Table V presents the performance metrics used for both segmentation and thickness measurement evaluation from the literature. Current metric reporting in endometrium segmentation prioritizes technical overlap, especially DSC value, abundantly, but does not sufficiently validate clinical applicability. This focus overlooks critical metrics, such as acceptance rate, which directly assess whether automated thickness measurements fall within clinically permissible rate margins. Consequently, this condition limits the assessment for practicality in real-life assessment, specifically for clinical integration. In conclusion, to ensure clinical readiness, future work must standardize metric reporting that prioritizes measurement accuracy and reliability for an effective integration into ART workflows. The inconsistency and underutilization of clinical decision metrics are direct symptoms of broader limitations to clinical applications, which are explored in RQ4.

D. RQ4: Technical and Clinical Gaps to Translate Deep Learning Model into Art

Translating promising deep learning segmentation and measurement performance into reliable and adoptable tools for ART requires addressing significant technical and clinical gaps that compromise reliability and clinical utility. Although most models demonstrate high accuracy on internal test sets, there are limitations regarding data diversity, integration of complex clinical features, and performance on diagnostically challenging cases.

One major limitation is ensuring that models perform reliably across diverse patient populations and clinical settings, addressing issues of data scarcity and single-source dependency [95]. Many studies rely on retrospective data. The models' ability to generalize can be compromised if the original dataset is not sufficiently large or diverse [4], [88], [89]. Furthermore, the results from small sample sizes cannot be guaranteed to generalize to other segmentation tasks [4].

In addition, models are often trained and tested on data collected from a single manufacturer or institution. This fact may affect the robustness of the model when deployed elsewhere [95]. External validation using datasets outside the source institution has not always been carried out. Consequently, the performance of some deep learning models, such as those predicting WOI, is acknowledged to require greater diversification, specifically needing a greater representation of cases with thin endometrial thickness (<7mm).

TABLE V. DISTRIBUTION OF EVALUATION METRICS IN ENDOMETRIUM SEGMENTATION LITERATURE

Method	Segmentation Performance Metrics							Thickness Measurement Performance Metrics			
	DSC	Jaccard	Sensitivity	Specificity	Accuracy	Precision	HD95	MAE	RMSE	Acceptance Rate	SDE
DLS										/	
VGG-Based U-Net	/							/	/	/	/
Key-discriminator	/	/									
6 Different U-Net	/		/	/		/					
ResNet50-based SegNet	/							/	/	/	/
3D U-Net	/	/	/				/	/	/	/	/
Multi StudentNet	/		/	/		/					
SAIM	/	/									
Transformer - CNN	/										
ResNet-50	/		/		/	/					

While endometrial thickness is a crucial biomarker, relying solely on it is incomplete to predict ART outcomes. Future models must be designed to assess a broader range of complex physiological variations of other critical features of endometrial receptivity, such as endometrial pattern. The evolution towards multiple parameter assessment requires a more sophisticated approach for validation. Furthermore, adequate differentiation across various pathological conditions, such as cancer and hyperplasia, is needed, as these conditions are known to impact echo pattern and segmentation [3], [4], [95].

Another setback for deep learning systems is the poor performance on thin endometria. There's one study that showed that for cases with endometrial thickness less than 3 mm, the acceptance rate was significantly lower at only 54.5%, compared to nearly 98.3% for thicker endometria (ET>10 mm) [4]. This inaccuracy in thin endometria could lead to false positive diagnoses, leading to unnecessary invasive examinations for patients.

As discussed in the introduction, the technical limitations of ultrasound remain a fundamental obstacle, with issues such as low contrast and speckle noise complicating the precise delineation of the endometrial boundary to measure the endometrial thickness effectively [89]. Most models fail in cases where the endometrium boundary is blurred or has a slightly irregular shape [3], [4].

To translate these technically advanced models into a clinical application, these models should be integrated into user-friendly interfaces for seamless routine ART screening protocols [95]. To this end, researchers are actively exploring and developing novel architectures, such as semi-supervised and key-point guided adversarial networks, to enhance segmentation robustness. This enhancement is critical to ensure a reliable endometrial receptivity assessment for clinical application.

The identified limitations and gaps reveal that clinical translation is hindered by clinical validation. To progress, research must shift from proof-of-concept or technical evaluation studies to rigorous development. To bridge the identified gaps, future research should adopt three key actions. First, the establishment of large, multi-center, and prospectively collected datasets is essential to ensure models are robust across diverse populations and critically challenging subgroups, such as thin endometria. Second, to adapt clinical validation, evaluation must adopt a mandatory and standardized set of metrics that are highly reliable in medical practice. Third, model development should evolve beyond endometrium segmentation to thickness measurement and other ultrasound biomarkers, such as echo pattern and volume. By addressing these actions, it can transform deep learning beyond technical research into a reliable clinical translation to aid in improving ART in the future.

VIII. Conclusion

This review systematically investigated the role of deep learning in TVUS image segmentation for the objective assessment of endometrial receptivity. The analysis provides clear answers for the research questions, revealing both the significant progress and the critical path forward.

In finding the predominant architectures for the deep learning segmentation (RQ1), it is found that the field is not dominated by a single model but is built upon a U-Net-based foundation, with a clear evolution towards sophisticated hybrid systems, such as SAIM, and tailored variants like 3D U-Net. When evaluating which of these models achieve high accuracy (RQ2), models such as the 3D U-Net, DDRNet, and Multi-StudentNet demonstrated superior performance. However, a definitive ranking is impossible due to a major problem, which is the lack of a common, large-scale dataset. This problem prevents fair comparison and highlights a fundamental need for the community.

Subsequently, this review also identifies the metrics used to validate segmentation quality (RQ3). It is revealed that not all metrics are equally meaningful for clinical translation. While the Dice score is most used, a hierarchy of clinical utility exists. The most critical metrics are the Acceptance Rate, supported by error magnitude metrics, which are MAE and RMSE. These metrics work together best to validate a model's readiness for clinical receptivity assessment.

Despite these advancements, significant technical and clinical limitations (RQ4) were revealed to prevent the integration of the models into ART workflows. These persistent limitations include poor model performance on thin endometria, inadequate validation across the menstrual cycle and various pathological conditions, and a narrow focus on thickness over a holistic, multi-parametric receptivity assessment.

In conclusion, while deep learning offers a powerful pathway to standardize and objectify endometrial evaluation, its full potential for clinical application has not yet been utilized. The transition from a promising algorithm to a trusted clinical tool hinges on future work that prioritizes robust, multi-parametric models, rigorous and stratified clinical validation, and the development of standardized benchmarks and datasets. By focusing on these challenges, the field can finally translate computational promise into enhanced diagnostic precision and improved outcomes in assisted reproduction.

Acknowledgment

The authors would like to express their gratitude to the Ministry of Higher Education (MOHE) of Malaysia and Universiti Kebangsaan Malaysia for their financial support under the grant FRGS/1/2024/TK07/UKM/02/6 for the resources that made this review possible.

References

[1] Z. Liu, Z. Zhang, and P. Xie, "Global research trends in endometrial receptivity from 2000 to 2024: bibliometric analysis," 2024, Frontiers Media SA. doi: 10.3389/fmed.2024.1465893.

[2] A. A. Nazarudin, N. Zulkarnain, A. Hussain, S. S. Mokri, and I. N. A. M. Nordin, "Review on automated follicle identification for polycystic ovarian syndrome," Bulletin of Electrical Engineering and Informatics, vol. 9, no. 2, pp. 588–593, 2020, doi: 10.11591/eei.v9i2.2089.

[3] S.-Y. Hu, H. Xu, Q. Li, B. A. Tefler, L. J. Brattain, and A. E. Samir, "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), IEEE, 2019, pp. 993–997.

[4] Y. Liu et al., "Automatic Measurement of Endometrial Thickness From Transvaginal Ultrasound Images," Front. Bioeng. Biotechnol., vol. 10, Mar. 2022, doi: 10.3389/fbioe.2022.853845.

[5] S. Archana, K. M. Kundavi, V. Geetha, and N. Naaram, "Enhancing IVF Success: The Role of Endometrial Receptivity and Uterine Scoring in Predicting Pregnancy Outcomes," International Journal of Science and Research (IJSR), vol. 12, no. 12, pp. 1319–1325, Dec. 2023, doi: 10.21275/sr231214102717.

[6] S. Liu et al., "Deep Learning in Medical Ultrasound Analysis: A Review," Apr. 01, 2019, Elsevier Ltd. doi: 10.1016/j.eng.2018.11.020.

[7] Z. Chen, Z. Wang, M. Du, and Z. Liu, "Artificial Intelligence in the Assessment of Female Reproductive Function Using Ultrasound: A Review," Jun. 01, 2022, John Wiley and Sons Ltd. doi: 10.1002/jum.15827.

[8] H. Achache and A. Revel, "Endometrial receptivity markers, the journey to successful embryo implantation," Nov. 2006. doi: 10.1093/humupd/dml004.

[9] J. Bellver and C. Simón, "Implantation failure of endometrial origin: What is new?," Aug. 01, 2018, Lippincott Williams and Wilkins. doi: 10.1097/GCO.0000000000000468.

[10] G. Litjens et al., "A survey on deep learning in medical image analysis," Dec. 01, 2017, Elsevier B.V. doi: 10.1016/j.media.2017.07.005.

[11] J. Wu, J. Sheng, X. Wu, and Q. Wu, "Ultrasound assessed endometrial receptivity measures for the prediction of in vitro fertilization embryo transfer clinical pregnancy outcomes: A meta analysis and systematic review," Exp. Ther. Med., vol. 26, no. 3, Aug. 2023, doi: 10.3892/etm.2023.12152.

[12] C. H. Zhang et al., "An endometrial receptivity scoring system basing on the endometrial thickness, volume, echo, peristalsis, and blood flow evaluated by ultrasonography," Front. Endocrinol. (Lausanne)., vol. 13, Aug. 2022, doi: 10.3389/fendo.2022.907874.

[13] X. Jiang et al., "Enhancing clinical decision-making in endometrial cancer through deep learning technology: A review of current research," Jan. 01, 2024, SAGE Publications Inc. doi: 10.1177/20552076241297053.

[14] B. Swarnkar, N. Khare, and M. Gyanchandani, "Systematic Review of Deep Learning Techniques for Gynecological Cancer Diagnosis," 2025, Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ACCESS.2025.3603458.

[15] P. R. Aparna and T. M. Libish, "Image processing and machine learning approaches for the automatic diagnosis of endometrial cancer," Int. J. Health Sci. (Qassim)., pp. 2827–2847, Jun. 2022, doi: 10.53730/ijhs.v6ns5.9255.

[16] S. Piedimonte et al., "Evaluating the use of machine learning in endometrial cancer: a systematic review," International Journal of Gynecological Cancer, vol. 33, no. 9, pp. 1383–1393, Sep. 2023, doi: 10.1136/ijgc-2023-004622.

[17] Y. Gao, Y. Jiang, Y. Peng, F. Yuan, X. Zhang, and J. Wang, "Medical Image Segmentation: A Comprehensive Review of Deep Learning-Based Methods," May 01, 2025, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/tomography11050052.

[18] Y. Z. Zhang et al., "Deep learning model for classifying endometrial lesions," J. Transl. Med., vol. 19, no. 1, Dec. 2021, doi: 10.1186/s12967-020-02660-x.

[19] K. M. Meiburger, U. R. Acharya, and F. Molinari, "Automated localization and segmentation techniques for B-mode ultrasound images: A review," Comput. Biol. Med., vol. 92, pp. 210–235, Jan. 2018, doi: 10.1016/j.compbiomed.2017.11.018.

[20] S. Visalaxi and T. Sudalaimuthu, "Automated segmentation of endometriosis using transfer learning technique," F1000Res., vol. 11, p. 360, Oct. 2022, doi: 10.12688/f1000research.110283.2.

[21] W. K. R. Figueredo, A. C. Silva, A. C. de Paiva, J. O. B. Diniz, A. Brandão, and M. A. P. Oliveira, "Automatic segmentation of deep endometriosis in the rectosigmoid using deep learning," Image Vis. Comput., vol. 151, p. 105261, Nov. 2024, doi: 10.1016/j.imavis.2024.105261.

[22] S. Visalaxi and T. Sudalai Muthu, "Automated prediction of endometriosis using deep learning," International Journal of Nonlinear Analysis and Applications, vol. 12, no. 2, pp. 2403–2416, 2021, doi: 10.22075/ijnaa.2021.5383.

[23] S. Visalaxi, D. Punnoose, and T. S. Muthu, "An Analogy of Endometriosis Recognition Using Machine Learning Techniques," in 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), IEEE, Feb. 2021, pp. 739–746. doi: 10.1109/ICICV50876.2021.9388403.

[24] S. Visalaxi, D. Punnoose, and T. S. Muthu, "Lesion Extraction of Endometriotic images using Open Computer Vision," in 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), IEEE, Mar. 2021, pp. 747–751. doi: 10.1109/ICAIS50930.2021.9395822.

[25] P. Hu, Y. Gao, Y. Zhang, and K. Sun, "Ultrasound image-based deep learning to differentiate tubal-ovarian abscess from ovarian endometriosis cyst," Front. Physiol., vol. 14, Feb. 2023, doi: 10.3389/fphys.2023.1101810.

[26] B. Dungate, D. R. Tucker, E. Goodwin, and P. J. Yong, "Assessing the Utility of artificial intelligence in endometriosis: Promises and pitfalls," Jan. 01, 2024, SAGE Publications Ltd. doi: 10.1177/17455057241248121.

[27] M. F. Blanco-Breindel, M. Singh, and J. Kahn, Endometrial Receptivity. 2025.

[28] E. A. Devyatova, K. A. Tsaturova, and E. V. Vartanyan, "Predicting of successful implantation at IVF cycles," Gynecological Endocrinology, vol. 32, pp. 27–29, Sep. 2016, doi: 10.1080/09513590.2016.1232803.

[29] S. Xu, H. Diao, Y. Xiong, C. Zhang, Y. Zhang, and Y. Zhang, "The study on the clinical efficacy of endometrial receptivity analysis and influence factors of displaced window of implantation," Sci. Rep., vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-91745-y.

[30] S. M. Dahiphale et al., "A Comprehensive Review of the Endometrial Receptivity Array in Embryo Transfer: Advancements, Applications, and Clinical Outcomes," Cureus, Aug. 2024, doi: 10.7759/cureus.67866.

[31] S. Xu, H. Diao, Y. Xiong, C. Zhang, Y. Zhang, and Y. Zhang, "The study on the clinical efficacy of endometrial receptivity analysis and influence factors of displaced window of implantation," Sci. Rep., vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-91745-y.

[32] A. López-Luna, D. Hernández-Melchor, L. Ramírez-Martínez, and E. López-Bayghen, "The Genetic and Biochemical Blueprint of Endometrial Receptivity: Past, Present, and Future Factors Involved in Embryo Implantation Success," in Modern Medical Genetics and Genomics, IntechOpen, 2019. doi: 10.5772/intechopen.80452.

[33] N. Ahmed Bahgat and W. Syed, "Displacement of Window of Implantation in Cases of Unexplained Recurrent Implantation Failure as Detected by Endometrial Receptivity Array Testing," Journal of Gynecology and Obstetrics, vol. 10, no. 2, p. 69, 2022, doi: 10.11648/j.jgo.20221002.13.

[34] P. Edimiris, C. Doehmen, D. M. Baston-Buest, J. S. Kruessel, and A. P. Bielfeld, "One center experience with a personalized frozen-thawed embryo transfer in patients with recurrent implantation failure," J. Assist. Reprod. Genet., vol. 40, no. 7, pp. 1639–1647, Jul. 2023, doi: 10.1007/s10815-023-02835-7.

[35] Y. Zheng et al., "Novel endometrial receptivity test increases clinical pregnancy and live birth rates in patients with recurrent implantation failure: Secondary analysis of a prospective clinical trial," International Journal of Gynecology & Obstetrics, Apr. 2025, doi: 10.1002/ijgo.70157.

[36] P. Kovacs, S. Matyas, K. Boda, and S. G. Kaali, "The effect of endometrial thickness on IVF/ICSI outcome," Human Reproduction, vol. 18, no. 11, pp. 2337–2341, 2003, doi: 10.1093/humrep/deg461.

[37] A. Stanziano et al., "The use of real time strain endometrial elastosonography plus endometrial thickness and vascularization flow index to predict endometrial receptivity in IVF treatments: a pilot study," BMC Med. Imaging, vol. 23, no. 1, Dec. 2023, doi: 10.1186/s12880-023-01071-w.

[38] Y. Wang et al., "Ideal embryo transfer position and endometrial thickness in IVF embryo transfer treatment," International Journal of Gynecology and Obstetrics, vol. 143, no. 3, pp. 282–288, Dec. 2018, doi: 10.1002/ijgo.12681.

[39] X. Cheng et al., "Endometrial echo patterns of embryo transfer day affect pregnancy outcomes in frozen embryo transfer cycles: a retrospective clinical study," BMC Pregnancy Childbirth, vol. 25, no. 1, Dec. 2025, doi: 10.1186/s12884-025-07501-7.

[40] E. Boys and M. Chapman, "Review: The Influence Of Endometrial Thickness On IVF Outcomes," J. Fertili. In Vitro, vol. 02, no. 03, 2012, doi: 10.4172/2165-7491.1000107.

[41] N. Noyes, H.-C. Liu, K. Sultan, G. Schattman, and Z. Rosenwaks, "Endometrial thickness appears to be a significant factor in embryo implantation in in-vitro fertilization," 1995.

[42] A. Kasius et al., "Endometrial thickness and pregnancy rates after IVF: A systematic review and meta-analysis," Hum. Reprod. Update, vol. 20, no. 4, pp. 530–541, 2014, doi: 10.1093/humupd/dmu011.

[43] J. Zhao, Q. Zhang, and Y. Li, "The effect of endometrial thickness and pattern measured by ultrasonography on pregnancy outcomes during IVF-ET cycles," 2012. [Online]. Available: http://www.rbej.com/content/10/1/100

[44] Z. Liao et al., "The Effect of Endometrial Thickness on Pregnancy, Maternal, and Perinatal Outcomes of Women in Fresh Cycles After IVF/ICSI: A Systematic Review and Meta-Analysis," Feb. 11, 2022, Frontiers Media S.A. doi: 10.3389/fendo.2021.814648.

[45] F. Ahmadi, F. Akhbari, M. Zamani, F. Ramezanali, and R. Cheraghi, "Value of Endometrial Echopattern at HCG Administration Day in Predicting IVF Outcome," 2017.

[46] H. Elhassan, J. G. Thornton, W. Malouf, and N. Raine-Fenning, "EP32.15: The prognostic accuracy of endometrial patterns on ultrasound to predict implantation," Ultrasound in Obstetrics & Gynecology, vol. 62, no. S1, pp. 306–307, Oct. 2023, doi: 10.1002/uog.27230.

[47] A. M. Moubark, M. H. Mohd Zaman, M. A. Zulkifley, S. H. Md Ali, L. Nie, and S. Freear, "Despeckling Ultrasound B-Mode Images Denoised by a New Unsharp Masking Method for Improved Cyst Segmentation," in Proceedings - 2020 IEEE EMBS Conference on Biomedical Engineering and Sciences, IECBES 2020, Institute of Electrical and Electronics Engineers Inc., Mar. 2021, pp. 178–182. doi: 10.1109/IECBES48179.2021.9398839.

[48] R. A. Pierson and G. P. Adams, "Computer-assisted image analysis, diagnostic ultrasonography and ovulation induction: Strange bedfellows," Theriogenology, vol. 43, no. 1, pp. 105–112, 1995, doi: 10.1016/0093-691X(94)00014-L.

[49] A. A. Nazarudin et al., "Performance Analysis of a Novel Hybrid Segmentation Method for Polycystic Ovarian Syndrome Monitoring," Diagnostics, vol. 13, no. 4, p. 750, Feb. 2023, doi: 10.3390/diagnostics13040750.

[50] P. M. M. Pereira, L. M. N. Tavora, R. Fonseca-Pinto, R. P. Paiva, P. A. A. Assuncao, and S. M. M. De Faria, "Image segmentation using gradient-based histogram thresholding for skin lesion delineation," in BIOIMAGING 2019 - 6th International Conference on Bioimaging, Proceedings; Part of 12th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2019, SCITEPRESS, 2019, pp. 84–91. doi: 10.5220/0007354100840091.

[51] A. A. Nazarudin, N. Zulkarnain, S. S. Mokri, W. M. D. W. Zaki, and A. Hussain, "An implementation of Otsu thresholding and the Chan-Vese method on the PCO segmentation of ultrasound images," ICECIE 2020 - 2020 2nd International Conference on Electrical, Control and Instrumentation Engineering, Proceedings, 2020, doi: 10.1109/ICECIE50279.2020.9309595.

[52] Erwin, A. Noorfizir, M. N. Rachmatullah, Saparudin, and G. Sulong, "Hybrid multilevel thresholding-otsu and morphology operation for retinal blood vessel segmentation," Engineering Letters, vol. 28, no. 1, pp. 180–191, 2020.

[53] D. Y. Huang, T. W. Lin, and W. C. Hu, "Automatic multilevel thresholding based on two-stage Otsu's method with cluster determination by valley estimation," International Journal of Innovative Computing, Information and Control, vol. 7, no. 10, pp. 5631–5644, 2011.

[54] P. S. Liao, T. S. Chen, and P. C. Chung, "A fast algorithm for multilevel thresholding," Journal of Information Science and Engineering, vol. 17, no. 5, pp. 713–727, 2001.

[55] S. Pare, A. Kumar, G. K. Singh, and V. Bajaj, "Image Segmentation Using Multilevel Thresholding: A Research Review," Iranian Journal of Science and Technology - Transactions of Electrical Engineering, vol. 44, no. 1, 2020, doi: 10.1007/s40998-019-00251-1.

[56] P. Yang, W. Song, X. Zhao, R. Zheng, and L. Qingge, "An improved Otsu threshold segmentation algorithm," International Journal Of Computational Science and Engineering, vol. 22, no. 1, pp. 146–153, 2020.

[57] X. Wang et al., "Automatic evaluation of endometrial receptivity in three-dimensional transvaginal ultrasound images based on 3D U-Net segmentation," Quant. Imaging Med. Surg., vol. 12, no. 8, pp. 4095–4108, Aug. 2022, doi: 10.21037/qims-21-1155.

[58] H. Park et al., "Endometrium segmentation on transvaginal ultrasound image using key-point discriminator," Med. Phys., vol. 46, no. 9, pp. 3974–3984, Sep. 2019, doi: 10.1002/mp.13677.

[59] M. Rahmani, A. Kazemi, M. J. Aziz, E. Najafzadeh, P. Farnia, and A. Ahmadian, "A Systematic Review on Medical Image Segmentation using Deep Learning."

[60] Y. Fang et al., "A self-supervised classification model for endometrial diseases," J. Cancer Res. Clin. Oncol., vol. 149, no. 20, pp. 17855–17863, Dec. 2023, doi: 10.1007/s00432-023-05467-7.

[61] M. Nikolic, E. Tuba, and M. Tuba, "Edge detection in medical ultrasound images using adjusted Canny edge detection algorithm," 24th Telecommunications Forum, TELFOR 2016, pp. 1–4, 2017, doi: 10.1109/TELFOR.2016.7818878.

[62] B. Padmapriya, T. Kesavamurthi, and H. W. Ferose, "Edge based image segmentation technique for detection and estimation of the bladder wall thickness," Procedia Eng., vol. 30, no. 2011, pp. 828–835, 2012, doi: 10.1016/j.proeng.2012.01.934.

[63] P. Hiremath, P. S. Hiremath, and J. R. Tegnoor, "Automatic Detection of Follicles in Ultrasound Images of Ovaries using Edge Based Method Texture Analysis and its applications, 3D face recognition, text localization and detection in natural scene images View project Bacterial Cell Image Analysis Using D," 2010. [Online]. Available: https://www.researchgate.net/publication/45718436

[64] T. F. Chan and L. A. Vese, "Active contours without edges," IEEE Transactions on Image Processing, vol. 10, no. 2, pp. 266–277, 2001, doi: 10.1109/83.902291.

[65] N. I. C. Marzuki, N. H. Mahmood, and M. A. A. Razak, "Segmentation of white blood cell nucleus using active contour," J. Teknol., vol. 74, no. 6, pp. 115–118, 2015, doi: 10.11113/jt.v74.4675.

[66] P. S. Hiremath and J. R. Tegnoor, "Follicle detection in ultrasound images of ovaries using active contours method," in Proceedings of the 2010 International Conference on Signal and Image Processing, ICSIP 2010, 2010, pp. 286–291. doi: 10.1109/ICSIP.2010.5697484.

[67] R. J. Hemalatha, V. Vijaybaskar, and T. R. Thamizhvani, "Performance Evaluation of Contour Based Segmentation Methods for Ultrasound Images," Advances in Multimedia, vol. 2018, 2018, doi: 10.1155/2018/4976372.

[68] Eliyani, S. Hartati, A. Musdholifah, and D. Dasuki, "Active contour without edge and watershed for follicle detection in ultrasound image of ovary," 2020 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2020, pp. 295–300, 2020, doi: 10.1109/ICACSIS51025.2020.9263115.

[69] S. F. Ameer, Z. T. Nayyef, Z. H. Fahad, and I. R. N. AlRubee, "Using morphological operation and watershed techniques for breast cancer detection," International journal of online and biomedical engineering, vol. 16, no. 5, pp. 140–149, 2020, doi: 10.3991/IJOE.V16I05.12999.

[70] N. Zulkarnain, A. A. Nazarudin, A. H. A. N. Al Has, S. S. Mokri, A. Hussain, and I. N. A. M. Nordin, "Ultrasound Image Segmentation For Detecting Follicle In Ovaries Using Morphological Operation And Extraction Methods," J. Pharm. Negat. Results, vol. 13, no. 4, pp. 659–665, Oct. 2022, doi: 10.47750/pnr.2022.13.04.088.

[71] C. Panagiotakis and A. Argyros, "Region-based Fitting of Overlapping Ellipses and its application to cells segmentation," Image Vis. Comput., vol. 93, p. 103810, 2020, doi: 10.1016/j.imavis.2019.09.001.

[72] M. J. Rao and Dr. R. K. Kumar, "Follicle Detection in Digital Ultrasound Images using Bi-Dimensional Empirical Mode Decomposition and Mathematical Morphology," International Journal of Engineering and Management Research, vol. 8, no. 4, 2018, doi: 10.31033/ijemr.8.4.20.

[73] R. Aggarwal et al., "Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis," Dec. 01, 2021, Nature Research. doi: 10.1038/s41746-021-00438-z.

[74] S. Shamshirband, M. Fathi, A. Dehzangi, A. T. Chronopoulos, and H. Alinejad-Rokny, "A review on deep learning approaches in healthcare

systems: Taxonomies, challenges, and open issues," Jan. 01, 2021, Academic Press Inc. doi: 10.1016/j.jbi.2020.103627.

[75] B. Sistaninejhad, H. Rasi, and P. Nayeri, "A Review Paper about Deep Learning for Medical Image Analysis," 2023, Hindawi Limited. doi: 10.1155/2023/7091301.

[76] X. Liu, L. Song, S. Liu, and Y. Zhang, "A review of deep-learning-based medical image segmentation methods," Sustainability (Switzerland), vol. 13, no. 3, pp. 1–29, Feb. 2021, doi: 10.3390/su13031224.

[77] Y. Wang, S. Mastura Mustaza, and M. Syuhaimi Ab-Rahman, "Pulmonary Nodule Segmentation Using Deep Learning: A Review," 2024, Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ACCESS.2024.3447697.

[78] L. Alzubaidi et al., "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," J. Big Data, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00444-8.

[79] Y. Xu, R. Quan, W. Xu, Y. Huang, X. Chen, and F. Liu, "Advances in Medical Image Segmentation: A Comprehensive Review of Traditional, Deep Learning and Hybrid Approaches," Oct. 01, 2024, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/bioengineering11101034.

[80] C. Wang, J. Zhang, and S. Liu, "Medical Ultrasound Image Segmentation With Deep Learning Models," IEEE Access, vol. 11, pp. 10158–10168, 2023, doi: 10.1109/ACCESS.2022.3225101.

[81] J. Suarez, "Advancements in Artificial Intelligence in the study of Endometrium," The Journal of Reproduction, pp. 212–221, Aug. 2024, doi: 10.58779/issn.2954-467x.tjor2024.v2.n4.57.

[82] S. Takahashi et al., "Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review," Dec. 01, 2024, Springer. doi: 10.1007/s10916-024-02105-8.

[83] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," May 2015, [Online]. Available: http://arxiv.org/abs/1505.04597

[84] Y. Liu et al., "Automatic endometrial segmentation in ultrasound images using deep learning," in Proceedings - 2022 IEEE 15th International Symposium on Embedded Multicore/Many-Core Systems-on-Chip, MCSoC 2022, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 67–71. doi: 10.1109/MCSoC57363.2022.00020.

[85] H. Thisanke, C. Deshan, K. Chamith, S. Seneviratne, R. Vidanaarachchi, and D. Herath, "Semantic Segmentation using Vision Transformers: A survey," May 2023, [Online]. Available: http://arxiv.org/abs/2305.03273

[86] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Jun. 2021, [Online]. Available: http://arxiv.org/abs/2010.11929

[87] N. Singhal, S. Mukherjee, and C. Perrey, "Automated Assessment of Endometrium From Transvaginal Ultrasound Using Deep Learned Snake," in 2017 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE, 2017.

[88] Y. Qiu, Z. Xie, Y. Jiang, and J. Ma, "Segment anything with inception module for automated segmentation of endometrium in ultrasound images," Journal of Medical Imaging, vol. 11, no. 03, May 2024, doi: 10.1117/1.jmi.11.3.034504.

[89] B. Peng et al., "Multi-StudentNet: a Semi-supervised Deep-Learning Framework for Endometrial Segmentation in Transvaginal Ultrasound Images," in 2024 IEEE International Conference on E-Health Networking, Application and Services, HealthCom 2024, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/HEALTHCOM60970.2024.10880821.

[90] H. Snyder, "Literature review as a research methodology: An overview and guidelines," J. Bus. Res., vol. 104, pp. 333–339, Nov. 2019, doi: 10.1016/j.jbusres.2019.07.039.

[91] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," Mar. 29, 2021, BMJ Publishing Group. doi: 10.1136/bmj.n71.

[92] S. Visalaxi and T. Sudalaimuthu, "Automated segmentation of endometriosis using transfer learning technique," F1000Res., vol. 11, p. 360, Mar. 2022, doi: 10.12688/f1000research.110283.1.

[93] J. Wang, L. Han, and D. Ran, "Architectures and Applications of U-net in Medical Image Segmentation: A Review," in 2023 9th International Symposium on System Security, Safety, and Reliability (ISSSR), IEEE, Jun. 2023, pp. 84–94. doi: 10.1109/ISSSR58837.2023.00022.

[94] N. Siddique, P. Sidike, C. Elkin, and V. Devabhaktuni, "U-Net and its variants for medical image segmentation: theory and applications," 2019.

[95] S. Yan, F. Xiong, Y. Xin, Z. Zhou, and W. Liu, "Automated assessment of endometrial receptivity for screening recurrent pregnancy loss risk using deep learning-enhanced ultrasound and clinical data," Front. Physiol., vol. 15, 2024, doi: 10.3389/fphys.2024.1404418.

[96] Q. T. Ithani, S. S. Mokri, N. Zulkarnain, and M. F. Bin Ahmad, "Deep Learning Model for Endometrium Segmentation in Transvaginal Ultrasound (TVUS) Images," 2024, pp. 66–80. doi: 10.2991/978-94-6463-566-9_6.

[97] S. N. Kumar, A. Lenin Fred, H. Ajay Kumar, and P. Sebastin Varghese, "Performance Metric Evaluation of Segmentation Algorithms for Gold Standard Medical Images," in Recent Findings in Intelligent Computing Techniques, Springer Singapore, 2018, pp. 457–469. doi: 10.1007/978-981-10-8633-5_45.

[98] U. Sara, M. Akter, and M. S. Uddin, "Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study," Journal of Computer and Communications, vol. 07, no. 03, pp. 8–18, 2019, doi: 10.4236/jcc.2019.73002.