

Artificial Intelligence for Ovarian Health Assessment: A Survey of Ultrasound, Multimodal Learning, Explainable AI and Longitudinal Approaches

Prajnyashree Giri
Department of Computer
Engineering
Smt. Indira Gandhi College
of Engineering
(Autonomous & NAAC A
Grade Accredited)

Samruddhi Gopalkar
Department of Computer
Engineering
Smt. Indira Gandhi College
of Engineering
(Autonomous & NAAC A
Grade Accredited)

Sara Pawar
Department of Computer
Engineering
Smt. Indira Gandhi College
of Engineering
(Autonomous & NAAC A
Grade Accredited)

Dr. Kajol Chaudhari
Assistant Professor
Smt. Indira Gandhi College
of Engineering
(Autonomous & NAAC A
Grade Accredited)

Abstract - Evaluating the health of ovaries demands corroboration across symptoms, menstrual and reproductive history, lab values, ultrasound imagery, and prior clinical encounters. We review recent artificial-intelligence work surrounding ovarian health across ultrasound analysis, clinical natural-language processing, multimodal learning, explainable AI, uncertainty estimation, and longitudinal modelling. Our work included researching recent indexed, peerreviewed literature with targeted searches, screening the results, and analyzing two datasets based on modality, task, algorithm, dataset attributes, validation strategies, explainability, and temporal modelling. Current literature is advancing rapidly in image-based and multimodal ovarian diagnosis but also exhibits several key consistent limitations including retrospective datasets with little to no external validation, heterogeneous labels, incomplete patient-level multimodal pairings, and reliance on cross-sectional prediction. Although the volume of longitudinal-specific ovarian AI is sparse in comparison, recent research illustrates that longitudinal modelling can be done explicitly. Through our analysis we identify an opportunity for a connected, doctor-facing system that leverages clinical text, sonography reports and ultrasound images, other medical or lab reports, visit-level structured evidence, and patient-level longitudinal reasoning. From here we motivate OvaCare, whose output for each encounter can be broken down into two parts: OUTPUT 1, a structured clinical insight contributed to a patient timeline, and OUTPUT 2, a longitudinal clinical insight generated from analyzing subsequent visits. Ultimately, the end result of this review is not only a current summary of work in ovarian-focused AI, but also provides research-gapdriven justification for OvaCare's multimodal and temporal methodologies.

Index Terms - artificial intelligence, clinical decision support, literature review, multimodal learning, ovarian health, longitudinal modelling, explainable AI, ultrasound.

INTRODUCTION

A. Background and Motivation

Ovarian disorders are looked at using types of information instead of just one test. Clinical notes can include symptoms, menstrual history, past health previous problems with the

ovaries and family history. Lab results and other medical reports give context-based information while ultrasound gives both written reports and pictures. The research is starting to show this way of looking at patient data but many artificial-intelligence systems still focus on a narrower problem that looks at one point in time.

Ultrasound is a way to look at the ovaries and nearby areas because it is safe available everywhere and can show the shape of the area. Recent studies have used learning, transferring models from other studies analyzing images in detail dividing parts of the image and using new types of models to look at PCOS ovarian cysts, tumors and cancer [3]–[5] [10]–[18] [23]–[29]. Some studies that use than one type of data show that health information can add to the details from images [8] [9] [17]–[19] [23]–[27].

Clinical natural-language processing is another part because symptoms and health history are often written in free text. Early work showed that NLP can find PCOS-related information in records [1] and more recent work has used language models to find signs of ovarian cancer before it is diagnosed from health records [32]. These results support taking out health information in a structured way instead of just looking at the text as a whole.

The main problem that is still not solved in the research is dealing with time. Most systems look at the situation of a problem or a patient. But health care also depends on whether a finding's new has stayed the same is getting better has come back is getting bigger is more complicated or is changing along with symptoms and lab results. A recent review clearly separates prediction looking at a series of results models that know about time and AI that helps make decisions [7]. OvarXNet shows how to model the ovaries over time using repeated PET/CT scans and health data [31]. This is different from the OvaCare setup in the type of data used the way it works and the goal of the work.

B. Objective of the Review

The goal of this review is to look at what current ovarian AI can do find the problems that're still important for patients and, for technology and find the research gap that leads to OvaCare. Special attention is given to how the type of data used the model chosen, being able to explain the results, testing and looking at time changesre connected. The result is a way to move from what has been done in the past to choosing the OvaCare research method.

II. LITERATURE REVIEW

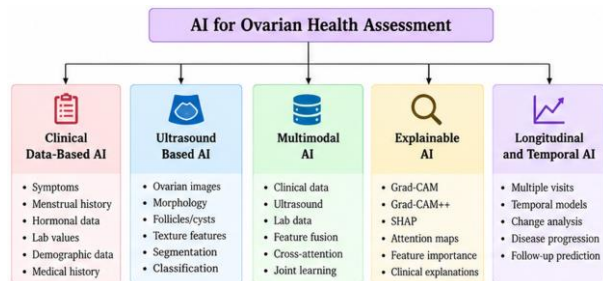


Fig. 1. Taxonomy of AI approaches relevant to ovarian-health assessment.

A. Ultrasound-Based Ovarian AI

PCOS research has moved from machine-learning methods to transfer learning, segmentation and explainable ultrasound models.

Recently researchers have combined ultrasound with clinical and biochemical data while other studies have concentrated on image-only classification or segmentation [14] [17] [18] [21] [22].

These studies show that ultrasound holds shape information but they also reveal large differences in how images are taken, marked and how classes are defined. For cysts and benign lesions deep learning has been used for segmentation and for telling different diagnoses apart. AdaResU-Net segmentation and ultrasound radiomics have been studied for cysts and for separating endometrioma from dermoid cyst [12] [13].

Ovarian torsion has also been examined with image-based learning [29]. These studies show task-specific usefulness. They usually focus on a single decision at a single moment. Ovarian-tumor research has become more multiclass and multimodal. OvaMTA uses a multi-task ultrasound pipeline for detecting masses for segmentation and for classification [10]. Multicentre studies have looked at deep-learning help, for tumors and ovarian cancer and international validation has shown that broad external datasets are vital [16] [23]. Systematic reviews report overall performance yet they also reveal great heterogeneity and limits in how well the results generalise outside the study setting [3] [4].

B. Multimodal Learning

Multimodal ovarian AI mixes ultrasound with biochemical, radiological or textual data.

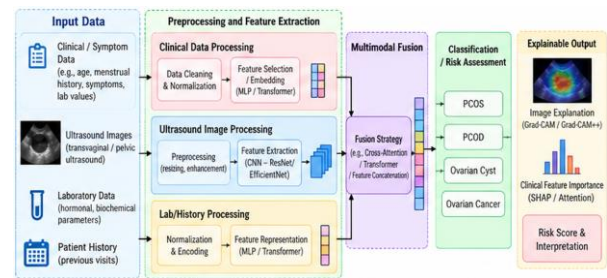
Wang and colleagues reported a multimodality model that uses imaging with clinical information [8].

OvcaFinder blends deep learning, O-RADS and routine clinical variables and it also includes external validation [9].

UMORSS adds uncertainty estimation and a workflow designed to aid ovarian cancer risk assessment [19].

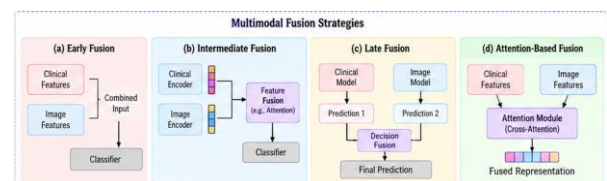
Other recent studies link ultrasound with variables or clinical text [23]–[27].

Fig. 2. Conceptual multimodal ovarian-health processing pipeline. Adapted from the earlier OvaCare survey draft.



Across the studies that I looked at multimodal design means trying to show how different types of evidence fit together. Ultrasound gives shape and details of lesions; clinical records give symptoms, history and exam context; and lab or other reports give chemical and other background info. The main problem is not just gathering these sources. Making sure they line up for the same patient and the same visit without mixing up data or matching wrong things. This is especially crucial for OvaCare because a system that follows patients over time can only be trustworthy if the data for each visit is accurate. The research backs the idea that using kinds of data can make risk assessment better but having many types of data does not automatically mean the system looks at the long term. A model might combine pieces of information at one visit and still only stay at that single point in time. For OvaCare this difference is key: the system must first pull together the evidence, in a single visit and then think about how things change across multiple visits.

Fig. 3. Comparison of early, intermediate, late and attention-based multimodal fusion strategies.

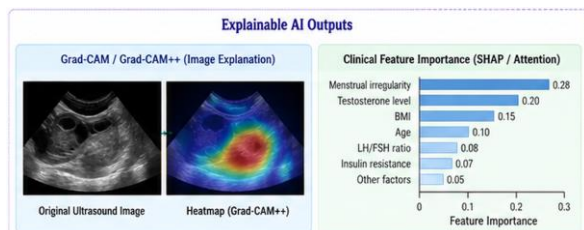


The fusion strategies in Fig. 3 illustrate why the review treats architecture selection as a question. Early fusion is simple and computationally attractive. It can require careful normalization of heterogeneous features. Late fusion preserves modality-specific predictions. Can remain robust when one modality is unavailable but it may fail to learn useful interactions. Intermediate fusion permits modality-specific encoders before joint representation learning. Attention-based fusion can model relevance between modalities but it adds parameters and should therefore be compared against simpler baselines. For OvaCare the final choice should depend on size, modality completeness and patient-level pairing.

C. Explainability and Uncertainty

Grad-CAM is widely used to highlight image regions contributing to deep-network predictions [36] while SHAP provides attribution for structured variables [37]. Ovarian AI studies increasingly include heatmaps, feature attribution or comparison with clinical scoring systems [9] [17] [19] [26]. However visual explanations should not automatically be interpreted as proof of validity. Uncertainty estimation is also important because a doctor-facing decision-support system must indicate when evidence is weak, conflicting or, outside the model's experience [19].

Fig. 4. Illustrative explainability outputs: image-region attribution and structured-feature importance.

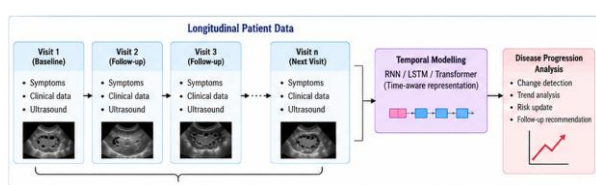


Explainability should be seen as showing evidence than proving that something caused an outcome. A heatmap can show which part of an image affected the model's decision and a feature-attribution method can list the variables in order of importance. Neither of these things automatically proves that the highlighted part is actually causing the result in a way. The job of XAI in OvaCare is to help the doctor look at the evidence that supports OUTPUT 1 and to show when there is disagreement or uncertainty if the different ways of getting information do not agree with each other.

D. Clinical NLP and Longitudinal AI

Clinical NLP can take notes. It turns them into information such as whether a symptom is present how long it has been going on how bad it is, menstrual history, past diseases, family history and what the doctor thinks. The PCOS EHR research shows that NLP works with coded data and recent studies show that big language models can help find features related to ovarian cancer. Clinical language models like ClinicalBERT offer a base, for learning how to represent information. Longitudinal ovarian AI is not as advanced as current-visit prediction models. OvarXNet used repeated PET/CT scans and clinical data with a CNN and Bi-GRU to model how the disease progresses and returns in advanced ovarian cancer [31]. This shows that modeling time-based changes is possible.. It does not fully address the challenge of combining repeated clinical text, ultrasound reports, ultrasound images and lab results into a consistent patient timeline. This gap highlights an opportunity, for OvaCare.

Fig. 5. Conceptual longitudinal patient representation across repeated visits.



E. Interpretation of Clinical Evidence

According to the literature clinical text must be interpreted as evidence that requires context. A useful extraction layer should separate findings from prior findings negated symptoms family history and uncertain statements. For OvaCare the proposed structured schema encapsulates demographics menstrual and reproductive history symptoms PCOS-related features reproductive history past conditions family history symptom details context and clinician impression. The aim is to prepare narrative clinical evidence for fusion while retaining source and uncertainty. This is consistent with work in EHR/NLP and LLM-based ovarian-cancer feature extraction [1] [32].

III. METHODOLOGY: REVIEW OF LITERATURE

A. Review Design

The review concentrated on intelligence for health assessment with recent peer-reviewed and indexed studies providing the evidence basis informing the methods as needed. The search and screening process was designed to advance the decision-making of the project than estimate a pooled clinical effect.

B. Search Strategy

The review employed sets of the following concepts: health; ovarian cancer; ovarian tumour; ovarian cyst; PCOS; ovarian torsion; ultrasound; clinical text; natural-language processing; multimodal learning with artificial intelligence; uncertainty; longitudinal modelling; and temporal artificial intelligence. The initial searches retrieved indexed records from the publisher databases and subsequent screening resulted in the inclusion of reports if they directly addressed a methodological gap or contributed a pertinent ovarian-AI capability.

C. Inclusion and Exclusion Criteria

The following criteria guided the selection of studies: proximity to the project's clinical tasks; employment of machine learning, deep learning, radiomics, natural-language processing or multimodal artificial intelligence; description of the validation procedures; specification of the dataset characteristics, model architecture and evaluation metrics. Additionally the review included studies that provided an methodological basis for industry trends and excluded those with no direct relevance to gynecological assessment as well as those that lacked a specific task or were redundant.

D. Data Extraction and Critical Appraisal

For each included study the review extracted the task, scope of disease, input type, dataset size, model families, multimodal design, methods for implementing evidence/provenance traceability, validation design and temporal capabilities. The review then organised findings around five themes: artificial intelligence; multimodal artificial intelligence for clinical evidence fusion; methods for implementing explainability or uncertainty traceability in natural-language processing; and

finally, temporal and longitudinal artificial intelligence for capturing a patient's evolving health state. By focusing on these themes the review aimed to characterise the state-of-the-art capabilities identify gaps in methods and articulate recommendations for OvaCare.

IV. RESULTS AND DISCUSSION

A. Findings from the Reviewed Literature

The reviewed clinical evidence highlights three trends. First there is a shift from single-modality image classification to multimodal and contextual intelligence for health assessment. Ultrasound-based artificial intelligence now tackles PCOS, ovarian cysts, ovarian tumours, cancer and torsion with transfer learning, segmentation-based radiomics and transformer-based models being common [10]–[18] [21]–[29]. Multimodal methods consistently show that adding evidence beyond the image can improve performance [8] [9] [17]–[19] [23]–[27].

Second the literature suggests that the validation design is as important as the performance metric. Multicentre and external-validation studies demonstrate the value of institutional, imaging-protocol and population diversity [9] [16] [23] [26] [27]. Systematic reviews highlight the prevalence of single-centre, retrospective designs and the resulting inconsistencies in performance estimates [3] [4]. As a result OvaCare should report sensitivity, specificity, F1, AUROC/AUPRC, calibration and external test results than accuracy as the sole metric.

Third there is a growing emphasis on model transparency with intelligence and uncertainty estimation enabling clinicians to trust model outputs. Grad-CAM and SHAP tools show which image regions and structured variables contribute to the diagnosis [36] [37]. Likewise uncertainty-aware systems allow clinicians to receive warnings when the model is not confident signalling that the result requires oversight [19]. Taken together these findings suggest that OvaCare should adopt an evidence-centred 'white box' design.

The literature can be broadly grouped along a spectrum starting with single-modality image classification and moving toward multimodal approaches. Across this progression the focus has gradually shifted from predicting a diagnosis to providing decision support that also takes uncertainty into account. The literature also shows a shift from cross-sectional to modelling with several recent papers addressing the ovarian-cancer trajectory for each patient. The literature does not yet show that these advances have been combined into a workflow that supports ovarian-health assessment for individual patients. This observation has led to the following recommendations.

B. Research Gaps Identified

Gap 1. Integration of modalities. Many studies combine ultrasound with a set of clinical variables but a complete workflow involving clinical text, ultrasound image, sonography report and other medical or laboratory reports is uncommon. This motivates an evidence schema in OvaCare.

Gap 2. Static prediction. Reviewed systems answer a current-visit question. When follow-up exists it is often used for labelling rather than explicit modelling of change. OvaCare therefore requires patient-level linkage and temporal reasoning.

Gap 3. Clinical-text integration. Clinical notes contain symptoms, history, negation and context that're difficult to represent using image-only models. Ovarian-specific clinical NLP evidence exists,. It is not commonly integrated with repeated ultrasound and other reports in one decision-support workflow [1] [32].

Gap 4. Validation and data limitations. Many studies are retrospective, single-centre or based on specialised datasets. Multimodal development also requires matched patient-level data; unrelated datasets cannot be safely combined simply because they contain similar variables. This limits both training and longitudinal evaluation.

Gap 5. Limited decision-support orientation. High-performing classifiers do not automatically provide an useful explanation of what changed why it matters and whether uncertainty remains. A doctor-facing system should present evidence, changes and limitations than only a disease probability.

C. Comparative Interpretation of the Reviewed Approaches

I have observed that models are useful when the available variables are already structured. Models cannot represent morphology that is visible in ultrasound. Image-only systems capture patterns. These systems can miss symptoms, history and biochemical context. Multimodal models address part of this limitation. Most recent systems still operate at the level of a diagnostic task or a single encounter. Uncertainty-aware systems improve the safety of the workflow by indicating when the prediction should be treated cautiously while longitudinal models demonstrate that repeated observations can provide information. These capabilities should therefore be viewed as layers than competing alternatives.

A second comparison looks at the level at which information is represented. Some systems work directly on images; other systems first cut out a lesion. Pull out radiomic features thereby handling information in a different way. Clinical models may rely on variables that humans define by hand while language models can turn records into ideas and thereby change information. For OvaCare this means that an intermediate evidence representation is helpful: each modality should produce findings that're meaningful, to clinicians and tied back to the original source before the final fusion. That evidence representation can then be checked on its own. Reused by the layer.

D. Path from Research Gaps to OvaCare Approach

The research gaps we identified shape the design of the OvaCare system. Since ultrasound models have shown performance and transfer learning is well proven in medical image analysis [34] OvaCare can begin with a pretrained image encoder and then fine-tune it for specific ovarian tasks. This

avoids training an image model from scratch saving time and resources. Clinical text holds details and contextual clues so a clinical-language encoder or instruction-tuned language model works well for extracting information. The main goal at this stage is to pull out evidence rather than generate broad diagnostic statements [1] [32] [42] [43].

Combining types of evidence requires multimodal fusion. Of producing one single hard-to-interpret probability right away each data source should first provide its own structured input. This includes findings from text, sonographic measurements and descriptions from the report features extracted from images and relevant lab results or report data. These individual inputs are then combined to create OUTPUT 1. A visit-level insight that reflects the full picture at one point in time.

The gap in prediction leads directly to the second stage of the system. When OUTPUT 1 is stored with an identifier and visit date the system builds a timeline of structured observations over time. OUTPUT 2 must analyze this sequence. Determine what has changed. It should flag what is new what remains the same what has resolved, improved, worsened, recurred, contradicted findings or remains uncertain. For example a lesion might grow from 3.2 cm in January to 3.8 cm in March. 4.5 Cm in June, with increasing complexity. The useful output isn't just listing the numbers. It's interpreting the trend giving insight into how the condition is changing.

TABLE I. COMPARISON OF REPRESENTATIVE RECENT OVARIAN AI APPROACHES

Study	Year	Task	Input	Multimodal	XAI/UQ	External validation	Temporal
Wang et al. [8]	2024	Benign vs malignant	US + menopause + serum	Yes	Limited	No	No
Xiang et al. [9]	2024	Ovarian cancer	US + O-RADS + clinical	Yes	SHAP/heatmaps	Yes	No
Dai et al. [10]	2024	Ovarian masses	US images/videos	No	Limited	Yes	No
Du et al. [15]	2024	Benign/borderline/malignant	US radio-mics + DL + clinical	Yes	Grad-CAM	No	No
Christiansen et al. [16]	2025	Ovarian cancer	US	No	Limited	Multicentre	No
UMORS [19]	2025	Ovarian risk	US + clinical	Yes	Uncertainty	Yes	Follow-up

Study	Year	Task	Input	Multimodal	XAI/UQ	External validation	Temporal
Agirioy et al. [17]	2025	PCOS	Clinical + biochemical + US	Yes	SHAP	Yes	No
Lu/Chiu et al. [25],[30]	2025-26	Ovarian tumors/cancer	US + clinical text	Yes	Limited	Yes	No
Jeong et al. [27]	2026	Benign vs malignant	US + age + CA-125	Yes	Limited	Multicentre	No
Lai et al. [26]	2026	Tumor malignancy	US + Doppler + clinical	Yes	O-RADS comparison	Yes	No
Sadeghi et al. [31]	2026	Ovarian cancer relapse	PET/CT + clinical	Yes	Attention/calibration	No	Yes
OvaCare (proposed)	—	Longitudinal ovarian assessment	Clinical text + US + reports/labs	Yes	Evidence + uncertainty	Future	Yes

Research gaps can become design requirements. First research gaps tell us that heterogeneous current-visit evidence must be linked for each patient and each encounter. Second research gaps say that each visit should create an OUTPUT 1 that stays stable enough to compare later. Third research gaps require that source attribution and uncertainty stay in the record so that conflicting evidence shows up. Fourth research gaps mean the timeline must keep visit order and dates of treating follow-up data as a jumbled pile. Fifth research gaps urge us to evaluate multimodal and longitudinal versions so that any improvement can be seen as coming from added capability, not just from a bigger model.

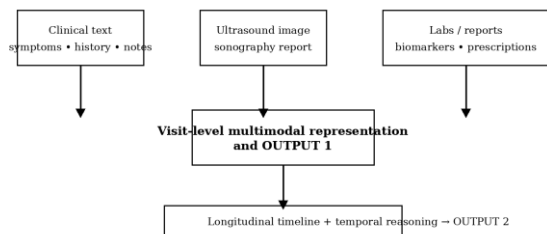
E. Implications for the Research Project

The review outcome gives requirements for the research project. First the system must be designed for doctors to use directly. It should support decision-making without taking over the doctors judgment. Second the three main inputs from a visit should be seen as working clinical text the sonography report along with the ultrasound image and other medical or laboratory reports. All three should be used in combination. Third the system needs to turn each encounter into an OUTPUT 1. This format should make it easy to store information and go back to it later. Fourth the timeline feature must work at the level. That means repeated visits, by the patient should be connected properly. Fifth the longitudinal part of the system should do more than just summarize past data. It should reason through time. Understanding how events unfold over weeks, months or years. Finally evaluation must cover areas. We need to check performance for each input modality. We also need to assess

how well the system handles modalities together. Calibration is important too. Making sure predictions are reliable. Explainability matters so users know why the system gives results. We must test everything on patient-level data that was not used during training.

F. Proposed OvaCare Workflow

Fig. 7. OvaCare concept: current-visit multimodal evidence produces OUTPUT 1, which is retained for longitudinal reasoning toward OUTPUT 2.



OvaCare AI: Visit-Level Multimodal and Longitudinal Workflow

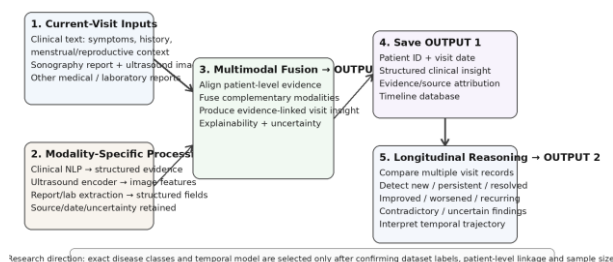
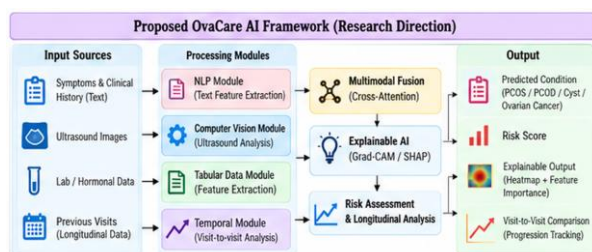


Fig. 8. OvaCare AI workflow from current-visit multimodal evidence to longitudinal clinical insight.

The proposed workflow is therefore: current-visit data acquisition → modality-specific preprocessing and feature extraction → multimodal fusion → OUTPUT 1 → save → patient timeline storage → longitudinal temporal analysis → OUTPUT 2. I think the workflow should only lock the disease classes and final model family after we confirm label availability, sample size and patient-level linkage, in the selected datasets. This workflow prevents the project from forcing a disease taxonomy that the available evidence cannot support.

G. Rationale for the Selected Research Direction

Fig. 9. Earlier conceptual OvaCare framework showing modality-specific modules, multimodal fusion, explainability and longitudinal analysis.



The literature does not support locking in one algorithm before the available datasets and labels are audited. The literature does however support a staged approach. A pretrained ultrasound encoder makes sense because transfer learning works with medical-image datasets. Clinical-language representations also

offer a foundation for pulling information from narrative records. Multimodal fusion should be tested against baselines. Attention-based fusion can be explored as an option. Attention-based fusion should not be assumed to be the choice. The longitudinal component should be chosen based on how visits there are, how apart they are the pattern of missing data and how much training data is available. Recurrent models, temporal transformers or simpler change-detection methods should all be compared. There is no reason to assume that the complex model will always perform best. This step-by-step selection comes directly from the gaps identified in the review.

This staged plan works well for a final-year project. A reliable baseline can be built before tackling modeling. If the data is limited and the full three-modality system cannot be used the project can still share results from modalities. The project can clearly point out which parts need data in the future. On the hand if enough matched data becomes available the same architecture can be extended to validation and prospective testing. The literature supports this approach of pushing to implement every component at once.

The final stage is evaluating the output. The system should deliver OUTPUT 1 as a evidence-based insight into a visit. OUTPUT 2 should provide a view showing what has changed across visits. Evaluation needs to check whether the output matches the stored evidence, whether any important changes are left out whether conclusions are made without support and whether uncertainty is properly communicated. This is especially crucial for a doctor-facing system because even a fluent explanation can mislead if it creates relationships that're not actually present in the records.

The fourth stage is the engine. Each validated visit-level representation becomes an OUTPUT 1 record. The first baseline for time-based analysis can be a comparison: measurements, symptoms, imaging descriptions and lab values between visits. Then a learned temporal model can be tested against this baseline. The goal is to see whether the learned model picks up patterns that simple change rules miss. Examples include lesions, growth over time complexity recurrence after healing or a mismatch between symptom trends and imaging findings.

The third stage is visit-level multimodal fusion. The experiment should compare models using text, images, only reports and progressively combined inputs. This setup allows for an ablation study showing how much each input contributes. If the multimodal model performs better the improvement must be tested within subgroups—not reported as a single overall score. Missing-modality tests should also be included to see how well the model handles cases where one modalities absent.

The second stage is modality-modeling. The clinical-text model should first be tested on information extraction before being connected to the classifier. The ultrasound branch should establish a transfer-learning baseline. If segmentation is needed a lesion-localization baseline should be set. The sonography-report branch should extract descriptors and measurements. Laboratory and other medical reports should be normalized while keeping units and reference context. At this stage the aim is not to get the score but to confirm that each branch produces consistent and reproducible evidence.

The implementation should happen in stages so that each part can be validated before being combined with the next. The first stage is. Schema mapping. Every candidate dataset should be reviewed for identifiers, encounter dates, which modalities are disease labels, image quality, missing values and annotation quality. Variables should then be mapped to the common OvaCare schema. This step is essential because similar terms, across datasets do not mean the meaning. For instance a measurement might be recorded in units a diagnosis label might reflect suspicion than confirmation or an ultrasound image might not match the report's encounter.

H. Disease. Clinical Task Decomposition

I note that the reviewed literature covers ovarian and related conditions but the clinical questions differ for each disease. PCOS is often treated as a syndrome-level classification or detection problem while an ovarian cyst or tumor is naturally seen as a lesion-level assessment problem. Ovarian torsion requires a decision, whereas endometrioma and other benign lesions often need to be set apart from other possible diagnoses. Ovarian cancer and malignant ovarian tumors also require reasoning because cancer can be viewed at many levels: benign versus non-benign, benign versus borderline versus malignant, tumor type and finally a clinically relevant cancer category. A single flat classifier would therefore create a structure for varied tasks.

For OvaCare the literature shows that it is better to split the problem into subtasks rather than force every condition into a single label. A visit-level representation can hold details about disease or lesion hypotheses, imaging, symptoms, lab results and uncertainty. The downstream insight can then display the chosen interpretation with any remaining options. This is especially important when the same patient has a lesion at one visit and a specific diagnosis after further imaging or follow-up. The longitudinal layer should record these changes replacing the state.

The disease scope also shapes how datasets are built. Public datasets may carry labels for one disease. Provide little information, about another. Therefore the final training scope should be chosen after mapping the datasets to the OvaCare schema. The system should not claim that the system covers a disease merely because that disease appears on the list. This review sees disease coverage as a decision that depends on data. This review suggests reporting which diseases and tasks were included during the experiment.

I. Representation of the Three Input Modalities

The main design problem is how to show kinds of evidence before we combine them. Clinical text reads like a story. It often says words such as "not present" "before" or "maybe". Sonography reports are a bit more ordered than notes. They still use words, numbers and impressions. Ultrasound pictures give data that you cannot get fully from the written report. Other medical and laboratory reports give numbers or categories that may come at times and use different reference ranges. Because of this we should not just paste all of them together as data without a step.

The proposed OvaCare representation treats each visit as one synchronized evidence unit. The clinical text part pulls out facts

such as age, reproductive and menstrual history, symptoms, past problems, family history, how the symptoms last and the doctor's impression. The sonography report part pulls out lesion size, shape, side, labels and the reported impression. The ultrasound image part gives features. When needed the ultrasound image part marks where the lesion's. The laboratory and other medical report part adds any blood test, blood cell count or other relevant facts. Each piece that is pulled out should keep its source. When possible each piece records the visit date and how certain it's

This design has two benefits. First it makes the combination of ways of showing evidence clear because the final conclusion can be traced back to the pieces of evidence. Second it creates a step for the long-term engine. If the system only stored a disease probability, for each visit the time model would lose what actually changed. By storing findings reasoning can look at measurements, symptoms, lesion shape and lab trends.

J. Clinical Text Processing and Information Extraction

Clinical natural-language processing should first be seen as a job of finding evidence not as a job of creating reasoning. Research shows that information about the ovaries can be found in health records. This information includes symptoms, medical history and features that appear before a diagnosis [1] [32]. Previous work in natural-language processing also highlights the need to understand negation and context. For example a phrase like "no pain" must be understood differently from "pain for three months." Likewise a family history of a condition should not be treated the same as a diagnosis.

For the OvaCare clinical-text branch a clear and organized extraction schema is the chosen approach. The schema proposed by natural-language processing includes demographics; history; symptoms; features linked to PCOS; reproductive history; past diseases or surgeries; family history; details about when symptoms began how long they lasted how severe they were and how they changed over time; context; and the clinician's impression. Labels for context must clearly separate conditions, past conditions, family history negated statements and uncertain information. When the original text is unclear the system should preserve the uncertainty of turning it into a positive conclusion.

A practical way to build this system is to start with a model that has already been trained on language or with a medical language model tuned with instructions. Next fine-tune that model for the job of extracting information. If computing power is limited or labeled data is scarce methods such as LoRA or QLoRA can be employed. These methods allow adaptation of the model without requiring resources.

The first step in evaluation is to check how well the model extracts information. Focus on precision, recall and F1 score for entities and their attributes. After this step decide whether the text-based extracted information should be used in any decisions. Keeping this separation helps figure out whether a mistake originates from misunderstanding language or, from combining information across data types.

K. Ultrasound Image Analysis and Sonography Report Processing

Ultrasound gives both shape details and measurements which makes it very important for checking problems. Research has models that perform segmentation use transfer learning apply radiomics and employ neural networks to study ovarian cysts, PCOS, tumors and to determine if something is cancerous [10]–[18] [23]–[29]. These studies show that using pretrained encoders and then fine-tuning them for specific tasks works better than starting from scratch each time.

For the image part a good starting point is to check whether a standard transfer-learning model such as ResNet or EfficientNet provides a way to represent the data. A transformer-based visual encoder can be tried if there is data and computing power to handle the extra complexity. Segmentation can be added when the model needs to locate the lesion and using UNet methods gives an easy-to-understand baseline. Tools such as Grad-CAM can show which parts of the image help make a prediction. These images should only be used as help not as proof that the result is correct [36] [44].

The sonography report needs to be handled not with the image. Important items to note include the size of the lesion, which side it's, on its appearance whether it is solid or contains fluid, any internal walls, any bumps, mentions of blood flow and the doctor's opinion. Keeping the report separate lets the model learn whether the words add anything beyond the image. It also allows the system to work even if there is no image. The final design that uses both parts should work even if some information is missing.

L. Multimodal Fusion and Missing-Modality Handling

The reviewed literature describes fusion strategies: early intermediate, late and attention-based. Early fusion is simple. It requires that different types of features be brought into the format through normalization. Late fusion keeps the predictions from each modality separate. This works well when some data is missing. It may miss deeper connections between the different types of information. Intermediate fusion allows each modality to develop its specialized representation before combining them in a shared learning step. Attention-based fusion can focus on the parts of the data that matter most for a case. It adds complexity. Needs more data to work properly.

For OvaCare an intermediate or attention-based approach looks promising. The three main inputs— text, sonography reports and ultrasound images—have very different forms.. The literature does not show that more complex fusion methods always perform better than simpler ones. A fair test must include least one basic baseline model along with the new multimodal system. Ablation studies should remove one modality at a time. The goal is to see how performance changes. This would help answer a real-world question: do text, sonography reports, ultrasound images and other reports each provide unique information that isn't already covered by the others?

Handling different modalities is just as important. In clinical practice a patient might have a report without the original image, an image, without lab results or incomplete clinical notes. A reliable system must. Work with missing inputs or be clear that its confidence is lower. The model should never make

up missing data. This rule also affects how the dataset is built. Artificially filling in all missing inputs may not reflect real-world conditions.

M. Longitudinal Representation and Temporal Reasoning

The unique requirement of OvaCare is to move beyond a visit prediction to a long-term clinical insight. Recent studies look at steps such as prediction, serial summarisation, explicit time-aware modelling and decision-linked temporal AI. OvarXNet shows how ovarian changes can be mapped over time using repeated PET/CT scans and clinical data. That study is not the same as OvaCare because OvaCare brings together text, sonography reports, ultrasound images and other laboratory evidence at visits.

The OvaCare longitudinal engine should use stored OUTPUT 1 records that contain raw multimodal inputs. Each OUTPUT 1 can hold the visit date, structured findings, disease or lesion hypotheses, measurements, symptoms, laboratory evidence, uncertainty and supporting sources. The time engine can then compare visits that're not consecutive to spot new persistent resolved improved, worsened, recurring or contradictory findings. This method gives a picture that is more complete than simply averaging predictions across visits.

A useful time design must separate a change in measurement from a change in interpretation. For example an ovarian cyst may grow while staying benign. It may grow while gaining features. These two paths should not look the same. Likewise symptoms may ease while an imaging abnormality stays the same. OvaCare long-term insight should therefore describe evidence streams and clearly state when they agree or disagree.

The time model can be built in steps. A rule-based change detector can give an easy-to-understand baseline for measurement and categorical changes. A sequence model, transformer or recurrent network can then be tested for trajectory representation. Finally the OvaCare project can check whether a time-aware decision layer helps doctors more, than a summary. This staged approach keeps the risk low when claiming time intelligence without showing that the model really uses time.

N. Evaluation, Validation and Clinical Reliability

The evaluation of OvaCare must match the tasks it is meant to perform. For clinical-text extraction, precision, recall and F1 score are used to measure how well the system identifies entities and their attributes. When classifying diseases or lesions sensitivity, specificity, precision, recall, F1 score, AUROC and AUPRC should all be taken into account. Calibration matters a lot when the system gives risk or confidence estimates because these numbers need to be trusted. Confusion matrices can show where errors are happening often— in critical disease categories. For reasoning tasks the evaluation must also check whether the system accurately detects the direction and type of change between visits.

It is important to split data at the level. If data from one patient appear in both training and testing sets the model might learn patterns to that person. This leads to an optimistic estimate of performance. The same rule applies to visits—each visit belonging to a patient must stay in one partition. Whenever possible use a dataset or data from another institution to test

how well the model generalizes. Many ovarian AI studies stress the importance of multicentre or external validation [9] [16] [23].

Beyond performance clinical reliability must also be assessed. A system built for doctors needs to display the evidence behind every insight. It should make a difference, between what is observed and what has been concluded by the system. When evidence is missing or conflicting uncertainty must be shown. Tools like feature attribution and image heatmaps can help clinicians review results. These tools do not prove that the system truly understands relationships. Therefore the final evaluation of the system should not focus on technical metrics. It must also consider how useful and consistent the presentation of evidence really is.

O. Research Questions Emerging from the Review

The literature review leads to a set of research questions that can guide the implementation phase. First does combining text, sonography reports and ultrasound images improve visit-level ovarian-health assessment compared with single-modality baselines? Second does the addition of laboratory or other medical reports provide value when the principal imaging and clinical inputs are already available? Third does storing an OUTPUT 1 improve the ability to identify change across repeated visits compared with a conventional sequence of raw predictions? Fourth which fusion strategy provides the balance between performance, robustness to missing modalities and interpretability for the dataset size? Fifth can a temporal model distinguish progression from measurement variation? Sixth can the system communicate uncertainty and evidence in a form that's more useful to clinicians than a disease probability? These questions convert the review from a survey into an agenda. The questions should be answered through controlled comparisons than, by reporting the final proposed model. Baseline models, ablation studies, patient-level validation and external testing should be included where data permit. Such comparisons will allow the project to determine whether each architectural component is justified by evidence.

P. Limitations of the Current Review and Proposed System

This review has some limitations. This review is a targeted narrative review, not an analysis. Because of this this review does not give an effect size for AI. The literature varies widely in how diseases are defined, how datasets are used how imaging protocols are applied how labels are assigned and how evaluation methods are employed. This variation makes direct numerical comparisons difficult. Also recent publications are evolving quickly in multimodal and longitudinal AI. Therefore the conclusions of this review should be seen as design-oriented evidence than proof that one architecture works best in all situations.

The OvaCare system also depends on data availability. Public datasets often do not include all three modalities for a patient at a visit. For example it can be hard to get matched text sonography reports, ultrasound images and laboratory results, at scale. The project should therefore report the pairing and coverage that have been achieved. The project should not suggest data when the different parts come from sources. If any components are trained using modality-specific datasets this limitation must be clearly stated. The final multimodal

evaluation should use matched patient-level cases whenever possible.

Finally OvaCare is meant to support doctors not replace them. The proposed outputs—OUTPUT 1 and OUTPUT 2—should help doctors organize evidence point out changes and show uncertainty. These outputs should not be used as substitutes for judgment. This distinction matters most when the system faces disease types, incomplete records, conflicting information or imaging findings that fall outside the training data.

Q. Overall Synthesis of the Review

Taken together, the literature supports a layered architecture for OvaCare. The first layer should transform heterogeneous inputs into structured, source-linked evidence. The second layer should combine those representations to generate a visit-level clinical insight. The third layer should persist the visit-level representation in a patient timeline. The fourth layer should compare multiple stored encounters to produce a longitudinal clinical insight. This sequence reflects the distinction between multimodal fusion within a visit and temporal reasoning across visits.

The review also indicates that model sophistication should follow evidence availability. A complex transformer or attention mechanism is not justified if the dataset is small, labels are inconsistent or patient-level modality pairing is weak. Conversely, a carefully structured multimodal representation can provide value even with relatively simple baseline models because it preserves clinical context and supports later longitudinal analysis. The recommended research path is therefore incremental: establish reliable modality-specific extraction, establish a strong visit-level baseline, validate multimodal fusion, and only then add explicit temporal modelling.

This synthesis defines the central contribution expected from the OvaCare research project. The novelty is not the isolated use of ultrasound AI, clinical NLP, multimodal learning or temporal modelling, since each has precedent in the literature. Instead, the project investigates how these capabilities can be combined into a doctor-facing ovarian-health workflow in which each encounter becomes structured longitudinal memory and subsequent encounters are interpreted as a trajectory. The value of the proposed system will therefore depend on demonstrating that the combination produces more informative and reliable clinical insight than independent single-visit predictions.

V. CONCLUSION

This review shows that ovarian AI has progressed substantially in ultrasound classification, multimodal learning, explainability, uncertainty estimation and clinical NLP. Strong recent studies demonstrate that ultrasound can support detection and classification of PCOS, cysts and ovarian tumors, while multimodal models show the value of adding clinical information. At the same time, external validation, dataset diversity, uncertainty and clinical interpretability remain important limitations.

The central research gap identified by the review is longitudinal integration. Existing systems generally analyse one encounter at a time, while explicit longitudinal ovarian modelling remains comparatively sparse and is often focused on specialised oncology imaging such as PET/CT. The literature therefore supports a research direction that connects current-visit multimodal assessment with patient-level temporal reasoning.

OvaCare is consequently positioned as a doctor-facing clinical decision-support research framework. Its first stage produces an evidence-linked OUTPUT 1 from clinical text, sonography information and other reports; its second stage stores these findings in a timeline and produces OUTPUT 2 by reasoning over change across visits. The literature review thus provides the research gap and the path to the selected multimodal and longitudinal approach. Clinical deployment should remain outside the scope of the research claim until prospective and external validation demonstrate safety and usefulness.

REFERENCES

- [1] V. Castro, Y. Shen, S. Yu, S. Finan, C. T. Pau, V. Gainer, C. C. Keefe, G. Savova, S. N. Murphy, T. Cai, and C. K. Welt, "Identification of subjects with polycystic ovary syndrome using electronic health records," *Reproductive Biology and Endocrinology*, vol. 13, art. 116, 2015, doi: 10.1186/s12958-015-0115-z.
- [2] M. J. Page et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, n71, 2021, doi: 10.1136/bmj.n71.
- [3] I. Garcia-Atutxa, F. Villanueva-Flores, E. D. Barrio, J. I. Sanchez-Villamil, J. Martínez-Más, and A. Bueno-Crespo, "Artificial intelligence for ovarian cancer diagnosis via ultrasound: a systematic review and quantitative assessment of model performance," *Frontiers in Artificial Intelligence*, vol. 8, art. 1649746, 2025, doi: 10.3389/frai.2025.1649746.
- [4] R. Li, J. Lei, X. Tang, S. Zheng, J. Qu, Y. Xu, and H. Zheng, "Artificial intelligence based on ultrasound for initial diagnosis of malignant ovarian cancer: a systematic review and meta-analysis," *Frontiers in Oncology*, vol. 15, art. 1626286, 2025, doi: 10.3389/fonc.2025.1626286.
- [5] F. Moro, M. T. Giudice, M. Ciancia, D. Zace, G. Baldassari, M. Vagni, H. E. Tran, G. Scambia, and A. C. Testa, "Application of artificial intelligence to ultrasound imaging for benign gynecological disorders: systematic review," *Ultrasound in Obstetrics & Gynecology*, vol. 65, no. 3, pp. 295–302, 2025, doi: 10.1002/uog.29171.
- [6] H. Yu and L. Peng, "Artificial intelligence in ovarian pathophysiology and management: a systematic review and meta-analysis," *Journal of Ovarian Research*, vol. 19, art. 224, 2026, doi: 10.1186/s13048-026-02083-0.
- [7] Y. Liu, Y. Liu, Y. Zhao, Y. Luo, and X. Hao, "From static snapshots to longitudinal trajectories: artificial intelligence in women's reproductive and ovarian health," *Frontiers in Endocrinology*, vol. 17, art. 1893963, 2026, doi: 10.3389/fendo.2026.1893963.
- [8] Z. Wang et al., "Multi-modality deep learning model reaches high prediction accuracy in the diagnosis of ovarian cancer," *iScience*, vol. 27, no. 4, art. 109403, 2024, doi: 10.1016/j.isci.2024.109403.
- [9] H. Xiang et al., "Development and validation of an interpretable model integrating multimodal information for improving ovarian cancer diagnosis," *Nature Communications*, vol. 15, art. 2681, 2024, doi: 10.1038/s41467-024-46700-2.
- [10] W.-L. Dai et al., "Development and validation of a deep learning pipeline to diagnose ovarian masses using ultrasound screening: a retrospective multicenter study," *eClinicalMedicine*, vol. 78, art. 102923, 2024, doi: 10.1016/j.eclinm.2024.102923.
- [11] M. Xi et al., "Ultrasonographic diagnosis of ovarian tumors through the deep convolutional neural network," *Ginekologia Polska*, vol. 95, no. 3, pp. 181–189, 2024, doi: 10.5603/gpl.94956.
- [12] M. Sha, "Segmentation of ovarian cyst in ultrasound images using AdaResU-net with optimization algorithm and deep learning model," *Scientific Reports*, vol. 14, art. 18868, 2024, doi: 10.1038/s41598-024-69427-y.
- [13] L. Liu, W. Cai, C. Zhou, H. Tian, B. Wu, J. Zhang, G. Yue, and Y. Hao, "Ultrasound radiomics-based artificial intelligence model to assist in the differential diagnosis of ovarian endometrioma and ovarian dermoid cyst," *Frontiers in Medicine*, vol. 11, art. 1362588, 2024, doi: 10.3389/fmed.2024.1362588.
- [14] J. Kermanshahchi, A. J. Reddy, J. Xu, G. K. Mehrook, and F. Nausheen, "Development of a machine learning-based model for accurate detection and classification of polycystic ovary syndrome on pelvic ultrasound," *Cureus*, vol. 16, no. 7, e65134, 2024, doi: 10.7759/cureus.65134.
- [15] Y. Du, W. Guo, Y. Xiao, H. Chen, J. Yao, and J. Wu, "Ultrasound-based deep learning radiomics model for differentiating benign, borderline, and malignant ovarian tumours: a multi-class classification exploratory study," *BMC Medical Imaging*, vol. 24, art. 89, 2024, doi: 10.1186/s12880-024-01251-2.
- [16] F. Christiansen et al., "International multicenter validation of AI-driven ultrasound detection of ovarian cancer," *Nature Medicine*, vol. 31, no. 1, pp. 189–196, 2025, doi: 10.1038/s41591-024-03329-4.
- [17] M. Agirsoy and M. A. Oehlschlaeger, "A machine learning approach for non-invasive PCOS diagnosis from ultrasound and clinical features," *Scientific Reports*, vol. 15, art. 33638, 2025, doi: 10.1038/s41598-025-10453-9.
- [18] M. S. Sundari, N. V. Sailaja, D. Swapna, S. Vikkurty, V. C. Jadala, K. B. K. S. Durga, and P. Thottempudi, "Transfer learning-enhanced CNN model for integrative ultrasound and biomarker-based diagnosis of polycystic ovarian disease," *Scientific Reports*, vol. 15, art. 34519, 2025, doi: 10.1038/s41598-025-17711-w.
- [19] X. Wang et al., "A multimodal uncertainty-aware AI system optimizes ovarian cancer risk assessment workflow," *npj Digital Medicine*, vol. 8, art. 614, 2025, doi: 10.1038/s41746-025-01986-4.
- [20] M. Ghaderzadeh, A. Garavand, and C. Salehnasab, "Artificial intelligence in polycystic ovary syndrome: a systematic review of diagnostic and predictive applications," *BMC Medical Informatics and Decision Making*, vol. 25, art. 427, 2025, doi: 10.1186/s12911-025-03255-6.
- [21] S. Reka, T. S. Praba, M. Prasanna, V. N. N. Reddy, and R. Amirtharajan, "Automated high precision PCOS detection through a segment anything model on super resolution ultrasound ovary images," *Scientific Reports*, vol. 15, art. 16832, 2025, doi: 10.1038/s41598-025-01744-2.
- [22] Y. Natarajan, S. P. K. R., and S. D. M., "Feature fusion context attention gate UNet for detection of polycystic ovary syndrome," *Scientific Reports*, vol. 15, art. 42386, 2025, doi: 10.1038/s41598-025-26314-4.
- [23] Y. Wang et al., "Ultrasound-based deep learning model as an assistant improves the diagnosis of ovarian tumors: a multicenter study," *Insights into Imaging*, vol. 16, art. 221, 2025, doi: 10.1186/s13244-025-02112-4.
- [24] C. Su et al., "Multimodal deep learning based on ultrasound images and clinical data for better ovarian cancer diagnosis," *Journal of Digital Imaging*, vol. 39, no. 2, pp. 1168–1180, 2025, doi: 10.1007/s10278-025-01566-8.

- [25] V. Lakshmi and B. Pushpa, "Explainable multimodal deep learning using cross-attention fusion of ultrasound and clinical features for PCOS classification," *Discover Computing*, vol. 29, art. 7, 2026, doi: 10.1007/s10791-025-09901-x.
- [26] R. Lai et al., "Investigation of multimodal deep learning models for predicting ovarian tumor malignancy based on ultrasound images and clinical information – a comprehensive comparative study against readers and O-RADS," *BMC Medical Imaging*, vol. 26, art. 257, 2026, doi: 10.1186/s12880-026-02312-4.
- [27] S. Jeong et al., "Enhanced ovarian cancer diagnosis using deep learning on pelvic ultrasound with integrated clinical data: retrospective multicenter study," *Journal of Gynecologic Oncology*, vol. 37, no. 4, art. e84, 2026, doi: 10.3802/jgo.2026.37.e84.
- [28] W. Li, Y. Xia, Y. Li, X. Wu, and L. Shi, "Diagnostic performance of deep learning models on ultrasound images for distinguishing benign from malignant ovarian cysts," *Journal of Ovarian Research*, vol. 19, art. 190, 2026, doi: 10.1186/s13048-026-02090-1.
- [29] S. L. Lim, H. Dong, A. J. Darling, J. Moyett, A. Swartz, M. A. Mazurowski, and A. Song, "Development of an image-based deep learning algorithm to predict ovarian torsion," *European Journal of Obstetrics & Gynecology and Reproductive Biology*, vol. 326, art. 115383, 2026, doi: 10.1016/j.ejogrb.2026.115383.
- [30] P.-C. Chiu, C.-Y. Lee, H.-C. Hsu, Y.-J. Tai, Y.-C. Chiang, and T.-P. Lu, "Development of a novel multimodal deep learning approach to improve diagnostic precision in ovarian cancer," *Briefings in Bioinformatics*, vol. 27, no. 3, art. bbag224, 2026, doi: 10.1093/bib/bbag224.
- [31] M. H. Sadeghi et al., "Longitudinal deep learning models for tracking disease progression in ovarian cancer using PET/CT imaging and clinical reports," *Physical and Engineering Sciences in Medicine*, vol. 49, no. 1, pp. 333–344, 2026, doi: 10.1007/s13246-025-01669-0.
- [32] G. Funston, N. Park, M. Thompson, M. Yetisgen, B. A. Goff, L. Kessler, and F. M. Walter, "Using large language models to identify prediagnostic clinical features of ovarian cancer from healthcare records: a population-based case-control study," *British Journal of General Practice*, vol. 76, no. 768, pp. e544–e551, 2026, doi: 10.3399/BJGP.2025.0366.
- [33] E. L. Barber, R. Garg, C. Persenaire, and M. Simon, "Natural language processing with machine learning to predict outcomes after ovarian cancer surgery," *Gynecologic Oncology*, vol. 160, no. 1, pp. 182–186, 2021, doi: 10.1016/j.ygyno.2020.10.004.
- [34] G. Litjens et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017, doi: 10.1016/j.media.2017.07.005.
- [35] R. F. Andreotti et al., "O-RADS US risk stratification and management system: a consensus guideline from the ACR Ovarian-Adnexal Reporting and Data System Committee," *Radiology*, vol. 294, no. 1, pp. 168–185, 2020, doi: 10.1148/radiol.2019191150.
- [36] R. R. Selvaraju et al., "Grad-CAM: visual explanations from deep networks via gradient-based localization," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.
- [37] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 4765–4774.
- [38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- [39] M. Tan and Q. V. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," in *Proc. International Conference on Machine Learning (ICML)*, 2019, pp. 6105–6114.
- [40] A. Dosovitskiy et al., "An image is worth 16x16 words: transformers for image recognition at scale," in *Proc. International Conference on Learning Representations (ICLR)*, 2021.
- [41] A. Vaswani et al., "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [42] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186, doi: 10.18653/v1/N19-1423.
- [43] E. Alsentzer et al., "Publicly available clinical BERT embeddings," in *Proc. 2nd Clinical Natural Language Processing Workshop*, 2019, pp. 72–78, doi: 10.18653/v1/W19-1909.
- [44] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.
- [45] A. Kirillov et al., "Segment anything," in *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 4015–4026.