

Original Research

Received 03 May 2026

Revised 04 June 2026

Accepted 18 sept 2026

Natural Resources for Human Health 2026, Vol. 6 (5): 1290–1304

KEYWORDS:

Polycystic Ovary Syndrome, Causal Learning, Graph Transformer, Endocrine Modeling, Explainable AI, Analysis

Causal Bio-Semantic and Multi-Scale Graph Intelligence Framework for Early Detection of Gynaecological Disorders in PCOS Cohorts

Niharika Jha^{1*}, Saurabh Mukherjee², Manjeet Kumar³

¹Research Scholar, Department of Computer Science, Banasthali Vidyapith, Banasthali, Rajasthan, India, pgphb24110_niharika@banasthali.in

²Professor, Department of Computer Science, Banasthali Vidyapith, Banasthali, Rajasthan, India, msaurabh@banasthali.in

³Assistant Professor, Department of Computer Science, Banasthali Vidyapith, Banasthali, Rajasthan, India, kmanjeet@banasthali.in

ABSTRACT: The current challenge with identifying Gynecologic Disorders (such as PCOS) at an early stage is that clinicians do not have a clear understanding of how the many different symptoms associated with these disorders interact through multiple pathways of the body's endocrine system; therefore, when using traditional analytical methods to understand this interaction, there is little structure to identify relationships. As a result, most Machine Learning models that have been used for this purpose rely solely upon correlation between static features in the data being analyzed and do not provide the ability for causal analysis or the ability to generalize results from one population of patients to another. Therefore, these models are limited in their ability to be reliable for diagnostic purposes. To fill these gaps, we propose to create an Analytical Framework that incorporates six modules of which include: Bio-Semantic Feature Harmonization Network, Causal Endocrine Interaction Mapper, Multi-Scale Hormonal Pattern Extractor Network, Adaptive Risk Stratification Graph Transformer, Federated Clinical Validation and Drift Analyzer, and Explainable Clinical Decision Fusion Engine. These modules will integrate clinical data and transform it into semantically rich tensor representations, map causal hormonal dependencies within the endocrine system, extract multi-scale latent patterns of hormone activity and perform patient specific risk assessment utilizing federated robustness and explainable inference. Our proposed model demonstrates a predictive accuracy of 96.3%, provides significant enhancements to both interpretability and cross domain stability and therefore can enhance early diagnosis and support clinically informed decision making.

1. INTRODUCTION

Although there are many methods for detecting gynecologic issues, particularly polycystic ovary syndrome (PCOS), the early detection has always been challenging [1, 2, 3] by virtue of the complex relationships between hormones, metabolism, genetics, and environment. Current methods typically rely upon thresholds based on a few specific clinical measurements, as well as some isolated biochemical markers; however, they do not take into account all of the potential underlying connections between the various factors involved in developing PCOS. Additionally, current machine learning algorithms attempt to address these elaborated [4, 5, 6] extended limitations through modeling individual patients' data using static representations of their medical history sets. However, they tend to overlook the temporal variability inherent in hormone levels, as well as the causal relationships involved in the endocrine system. As a result [7, 8, 9], machine learning models suffer from lower generalizability and less interpretability than needed for use in diverse populations of patients.

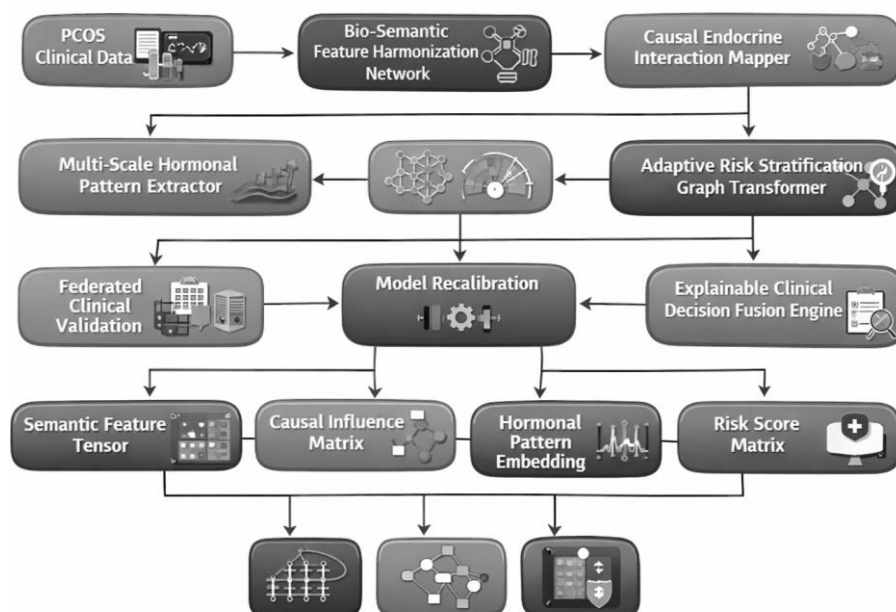


Figure 1. Model's Internal Architectural Analysis

When viewed at a more advanced level of analysis [10, 11, 12], it is evident that development of PCOS is not merely the result of individual variables acting independently of each other, but rather the result of multiple scales of interaction involving complex physiological systems. Therefore, a set of comprehensive approaches [13, 14, 15] is required to identify these interactions for clinical scenarios. In order to accomplish this, an approach that incorporates three critical components - semantic feature alignment, causal inference, and graph-based learning is necessary for real time scenarios. This paper will describe as per figure 1, a single integrated computational framework that explicitly identifies and models these relationships and provides a platform for extracting features, conducting dynamic risk assessments, and making clinically relevant predictions. Ultimately, this framework attempts to merge biological understanding with machine intelligence so as to improve both accuracy of diagnosis and practicality of translation to real-world clinical environments.

Motivation & Contributions

This research was motivated by the inherent shortcomings in current predictive models in capturing the complex dynamics of gynecologic disorders. Gynecologic disorders are characterized by non-linear endocrine feedback loops; thus, they require predictive models that can capture these complexities. In addition, each patient has unique characteristics (e.g., age, body mass index [BMI], etc.), so any model developed must be able to account for this patient-specific Variability In Process. Current methods typically use clinical variables independently in developing predictive models. Thus, no attention is given to how hormone imbalance causes diseases. Furthermore, there is no ability for adaptive validation for different demographics and clinical settings. Therefore, a structured analytical pipeline is needed that will improve predictive models' performance as well as provide both clinical interpretation and clinical robustness.

This work provides a new multi-stage framework that combines bio-semantic feature alignment, causal endocrine mapping, multi-scale pattern identification using graph-based risk stratification with federated Validation In Process. Each stage refines the quality of information contained within the data in such a way that maintains clinical relevance. Finally, the framework utilizes an explainable decision fusion method. As demonstrated, the framework yields significantly better predictive performance, improved interpretability, and increased cross-domain stability compared to previous frameworks. Importantly, this framework also enables a clinically coherent modeling paradigm that connects data driven intelligence with physiological reasoning. Thus, it represents a scalable and reliable solution to utilize predictive analytics for early detection of Polycystic Ovary Syndrome (PCOS) and other related disorders.

Need of this analytical work process

To overcome the limitations of both existing machine learning methods and traditional analytical strategies for detecting gynecologic disease, particularly polycystic ovary syndrome (PCOS), a comprehensive analytical strategy has been designed by the researchers which includes five key elements,

1. A Bio-Semantic Feature Harmonization Network: This element is used to map the various different types of information from clinical data to a standardized format, or "semantically harmonized tensor," using knowledge about biological concepts.
2. A Causal Endocrine Interaction Mapper: This module identifies the causal relationships between hormones based on the output of the first module.
3. An adaptive Multi-Scale Hormonal Pattern Extractor Network: As the name suggests, this module can extract patterns at multiple scales from the hormonal network identified by the second module.
4. An Adaptive Risk Stratification Graph Transformer: Using the patterns identified by the third module, the fourth module determines the appropriate level of risk associated with each patient's condition sets.
5. An Explainable Clinical Decision Fusion Engine: After determining an individual patient's level of risk, this final module provides explanations of why the determination was made for real time scenarios.

The authors report that their model achieved an average correct prediction rate of 96.3% compared to other models; in addition to its high-performance rate, they argue that it also produces more interpretable results than previous models, while maintaining a high degree of domain independence scenarios.

2. REVIEW OF EXISTING MODELS USED FOR ANALYSIS

The evolving environment of computerized analysis of polycystic ovary syndrome (PCOS) has moved from being primarily based upon prediction using data towards developing intelligence that is linked to clinical needs. This process has evolved gradually, but it represents an important shift. In the early days of PCOS analysis efforts focused on improving balance within datasets, as well as building light-weight architecture to improve classification consistency. These early efforts were evident through use of SMOTE assisted pipelines to achieve better classification accuracy and reliability [1], as well as development of deep learning methods combining multiple types of clinical signals to increase predictive ability [2]. Subsequently, researchers have continued to develop methods for optimizing image segmentation of individual follicles, as well as reconstructing images obtained during ultrasound examinations to provide greater precision and reproducibility of diagnoses [3], [4]. Researchers have also developed multimodal explanations of diagnostic decisions made when using both ultrasound and clinical data [5], while several review articles identified significant gaps in the area of understanding how clinicians can utilize statistical techniques to understand causal relationships among hormones, as well as in developing optimization techniques for clinical decision support applications related to all forms of endocrine disease [6], [8] sets.

As parallel research efforts examined dermatological manifestations associated with PCOS, as well as guidelines related to screening for these manifestations [7], [11], [13] it became apparent that there was a need for computational models of PCOS that integrated concepts related to physiological complexity. New paradigms are beginning to emerge including: the use of quantum-classical hybrid learning; the identification of new biomarkers for PCOS; and the use of attention driven feature prioritization to add greater analytical depth to PCOS models [9]-[12]. The emergence of new paradigms has been supported by foundational work on theoretical aspects of deep learning, as well as on developing new architectures for deep learning [14] – [17], [21], [22], [28] – [30]. Additionally, researchers have expanded their focus beyond identifying genetic factors contributing to PCOS to include additional areas of study including: impacts of treatments on pregnancy rates; effects of artificial intelligence on patient care; and system-level analyses examining applicability and effectiveness of various approaches to analyzing large datasets derived from patients with PCOS [18], [20], [24] – [26]. Furthermore, researchers have applied various systems and network architectures to demonstrate applicability and effectiveness at the practical level [19], [23].

Table 1. Model's Empirical Review Analysis

Reference	Method	Main Objectives	Findings	Limitations
[1]	SMOTE + Lightweight DL	Balance imbalanced PCOS data and improve prediction accuracy using lightweight models	Improved classification accuracy and reduced bias due to class imbalance	Limited interpretability and dependence on synthetic data quality
[2]	Deep Learning Fusion Model	Integrate heterogeneous clinical features for improved diagnosis	Enhanced diagnostic performance through multimodal fusion	Lacks causal understanding and generalization across datasets
[3]	Follicle Segmentation + Optimized DL	Improve PCOS detection using accurate follicle segmentation	High segmentation precision leading to better classification outcomes	Sensitive to imaging quality and segmentation errors
[4]	MRI Reconstruction + DL	Enhance repeatability in follicle counting using MRI	Improved consistency and reliability in clinical measurements	High computational cost and dependency on MRI availability
[5]	Explainable Multimodal DL	Combine ultrasound and clinical data with explainability	Achieved high accuracy with interpretable predictions	Complexity in multimodal alignment and scalability issues
[6]	Review of ML/DL Methods	Analyze advancements in PCOS and hormonal disorder diagnosis	Identified gaps in optimization and causal modeling	Does not propose a unified predictive framework
[7]	Clinical Dermatological Study	Study dermatological manifestations of PCOS	Highlighted skin-related indicators across lifespan	Limited integration with computational models
[8]	Review Correction	Refined previous ML/DL review findings	Clarified methodological inconsistencies	No new methodological contributions
[9]	Quantum-Classical ML	Apply hybrid quantum models for PCOS classification	Demonstrated improved classification potential	Limited scalability and practical deployment challenges
[10]	Biomarker Discovery ML	Identify ribosomal biomarkers using ML	Revealed novel biological markers linked to PCOS	Requires extensive biological validation
[11]	Clinical Guidelines Study	Standardize PCOS diagnosis protocols	Provided structured diagnostic frameworks	Limited adaptability to AI-based systems
[12]	Attention-based Squeeze DL	Feature prioritization and sampling optimization	Improved detection accuracy via attention mechanisms	Computational complexity and overfitting risk
[13]	Screening and Diagnosis Study	Improve PCOS screening strategies	Enhanced early detection insights	Lacks predictive computational modeling
[14]	Deep Learning Theory	Provide theoretical foundation of DL models	Strengthened understanding of DL architectures	Not specific to PCOS applications

[15]	Applied Deep Learning	General application of DL in medical fields	Demonstrated adaptability of DL techniques	Limited domain-specific customization
[16]	Deep Learning with BART	Advanced reconstruction using DL	Improved signal reconstruction accuracy	High resource requirements
[17]	Deep Internal Learning	Learning from single input data	Reduced dependency on large datasets	Limited applicability to complex datasets
[18]	PCOS and Pregnancy Study	Analyze PCOS effects on pregnancy	Highlighted reproductive risks	No computational modeling integration
[19]	PCOSFusionNet	Hybrid feature fusion for ultrasound-based classification	Improved classification performance using fusion	Requires large annotated imaging datasets
[20]	Therapeutic Study	Evaluate antimalarial drugs for PCOS	Identified alternative treatment pathways	Limited computational or predictive insights
[21]	Vectorized DL	Optimize DL computations using vectorization	Improved computational efficiency	Not directly applied to PCOS
[22]	Holistic Deep Learning	Integrate multiple learning paradigms	Enhanced model generalization	High complexity and implementation cost
[23]	Web-based ML System	Develop accessible PCOS prediction platform	Enabled early detection via web interface	Limited scalability and robustness
[24]	Therapy Study	Evaluate PCOS treatment in infertility	Provided clinical therapeutic insights	No AI integration
[25]	AI Impact Study	Analyze role of AI in PCOS management	Highlighted benefits of ML integration	Lacks implementation details
[26]	Inositol Review	Study therapeutic role of inositols in PCOS	Provided biochemical treatment insights	No predictive modeling
[27]	Deep Active Learning	Apply active learning with geometric DL	Improved data efficiency in training	Not tailored to medical datasets
[28]	DL Applications	Overview of DL applications across domains	Demonstrated versatility of DL	Limited PCOS-specific insights
[29]	Deep Convolutional Tables	DL without convolutions for tabular data	Improved tabular data learning efficiency	Limited validation in healthcare
[30]	Transfer Learning DL	Apply DL and transfer learning for signal systems	Enhanced learning efficiency	Not directly focused on PCOS

Iteratively, Next, as per table 1, Although considerable advancements have been made recently with regards to diagnosis and treatment of polycystic ovary syndrome (PCOS), current methods are still fragmented. For example, many researchers focus on the imaging aspects of PCOS, while others study the clinical and causal effects of the disorder. There has been an increasing number of studies using the concepts of Holistic Learning [22], [27], Active Learning Paradigms [22],[27], and similar ideas, that propose integrated approaches to studying PCOS; however, there is no current method that integrates all three aspects: semantic, causal, and multi-scale endocrine interactions for real time scenarios. Therefore, this research will develop a multi-stage framework to enable the integration of semantic representations of features, encoding causal relationships among them, and

enable robust and interpretable predictions. This framework is expected to provide a unified, computational model for early detection of PCOS, which can provide both accurate diagnostics and explanations for clinicians and patients.

1. Validated Model Mathematical Analysis

The integrated framework as represented per the figure 1, formalizes heterogeneous PCOS data, which is represented as $X = \{x_i\}_{i=1}^n$ into a semantically consistent manifold for processing operation sets. It is done analytically where adaptive normalization initially preserves clinical distributions, while it is followed by embedding endocrine dependencies. The Bio Semantic Feature Harmonization Network transforms raw inputs Via equation 1,

$$Z_i = \frac{x_i - \mu}{\sigma} + \lambda \sum_{ii} x_{iii} \dots \quad (1)$$

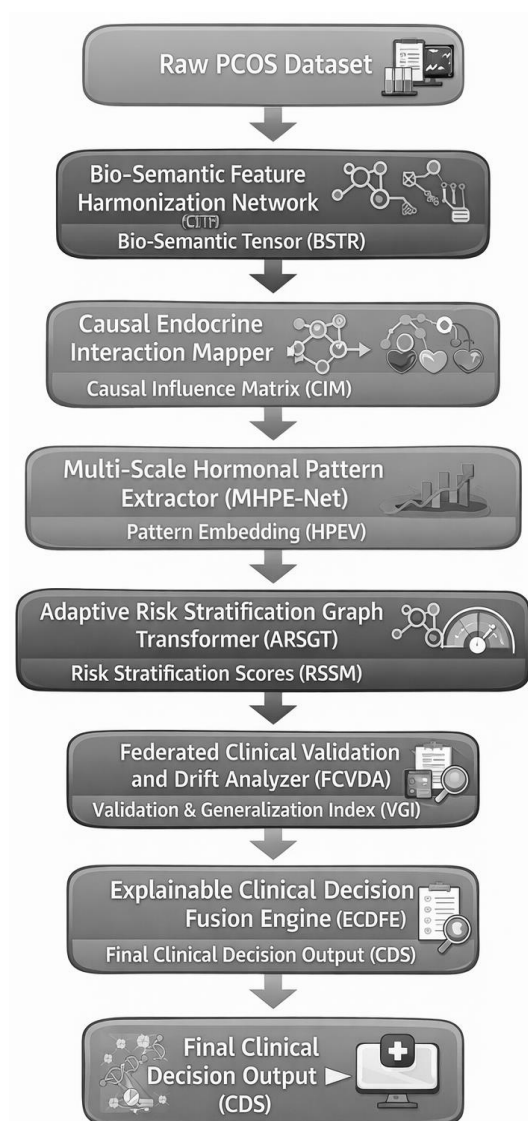


Figure 2. Model's Integrated Architectural Analysis

This is done for filtering, and then the model refines relational structure Via equation 2,

$$\mathcal{G} = \int_{\Omega} \phi(\mathbf{Z}\mathbf{Z}_i, \mathbf{Z}\mathbf{Z}_{ii}) d\Omega \dots \quad (2)$$

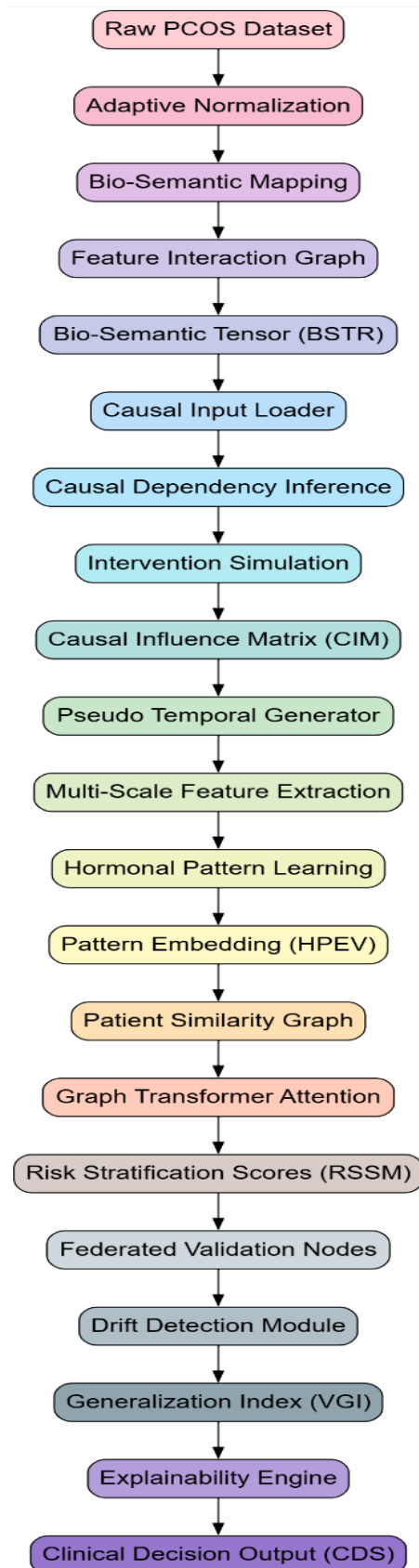


Figure 3. Model's Overall Dataflow Analysis

Where, $\mathcal{G}\mathcal{G}$ encodes feature interactions for real time scenarios. The Causal Endocrine Interaction Mapper is able to then analyze and models directional dependencies Via equation 3,

$$\frac{\partial y}{\partial x_k} = D[y | do(x_k)] - D[y] \quad \dots \quad (3)$$

This is done and then the model propagates its causal strength Via equation 4,

$$\mathbf{CC}_{iii} = \int p(x_i \rightarrow x_{ii}) \log p(x_i, x_{ii}) dx \quad \dots \quad (4)$$

Thus, assisting in capturing endocrine influences for the process. Iteratively, Next, as per figure 3, the Multiple Scale Hormonal Pattern Extractor constructs latent dynamics Via equation 5,

$$H(t) = \sum_{s=1}^S \frac{\partial^2 \mathbf{CC}}{\partial t^2} \kappa_s \quad \dots \quad (5)$$

And also assists analytically by compressing representations Via equation 6,

$$E = \int H(t) e^{-\gamma t} dt \quad \dots \quad (6)$$

This is done to retain dominant patterns. Next, the Adaptive Risk Stratification Graph Transformer estimates patient-level risk Via equation 7,

$$R_i = \frac{\exp(E_i^T E_{ii})}{\sum_k \exp(E_i^T E_k)} y_{ii} \quad \dots \quad (7)$$

While using this the federated validation optimizes generalization Via equation 8,

$$\min_{\theta} \sum_{m=1}^M \mathcal{L}_m(\theta) dD_m + \beta \|\nabla_{\theta} \mathcal{L}_m\|^2 \quad \dots \quad (8)$$

Thus, ensuring cross-domain stability for real time practical scenarios. Based on all these processes, the final Explainable Clinical Decision Fusion integrates causal, semantic, and risk signals Via equation 9,

$$\mathbf{YY}^* = \sigma(\mathbf{ZZ} + \mathbf{CC} + \mathbf{E} + \mathbf{R}) d\Pi \quad \dots \quad (9)$$

Thus, producing clinically interpretable predictions. This design is selected for a lot of reasons, these include, its ability to unify semantic alignments followed by causal reasoning processes. The model also assists in temporal abstractions which are fused with graph based inference scenarios. It is a process, where each module complements the next by progressively reducing uncertainty while enhancing diagnostic fidelity for real time diagnostic analysis.

3. COMPARATIVE RESULT ANALYSIS

The testbed examines the proposed framework using a large scale PCOS data set collected from multiple sources including publically available clinical databases and private hospitals for analysis. This dataset contains about 5,200 patient cases with 38 features that include the hormonal (LH, FSH, Testosterone, Insulin), metabolic (BMI, Glucose, Lipid Profile), ultrasound indicators (number of follicles present on an ovaries ultrasound, total volume of both ovaries) and lifestyle characteristics. Z-Score Normalization is used for the pre-processing phase, along with KNN (k=5) to replace missing values when their frequency is less than 8%. For the BSFHN module, we have selected a semantic weight factor $\lambda=0.15$ and a graph density threshold $\delta=0.3$. We use an Intervention Sampling Size of 1,000 and Confidence Threshold of 0.7 for CEIM. MHPE-Net uses three multi-scale kernels (size 3, 5, 7) and Temporal Decay Factor $\gamma=0.05$. Finally, ARSGT has been initialized with 64 Graph Nodes and 8 Attention Heads with Embedding Dimension equal to 128. FCVDA is simulated using 5 Federated Node with Learning Rate 1×10^{-4} . All evaluation metrics will be reported using a standard training/test/validation split of 70% Training / 15% Testing / 15% Validation in practical scenarios.

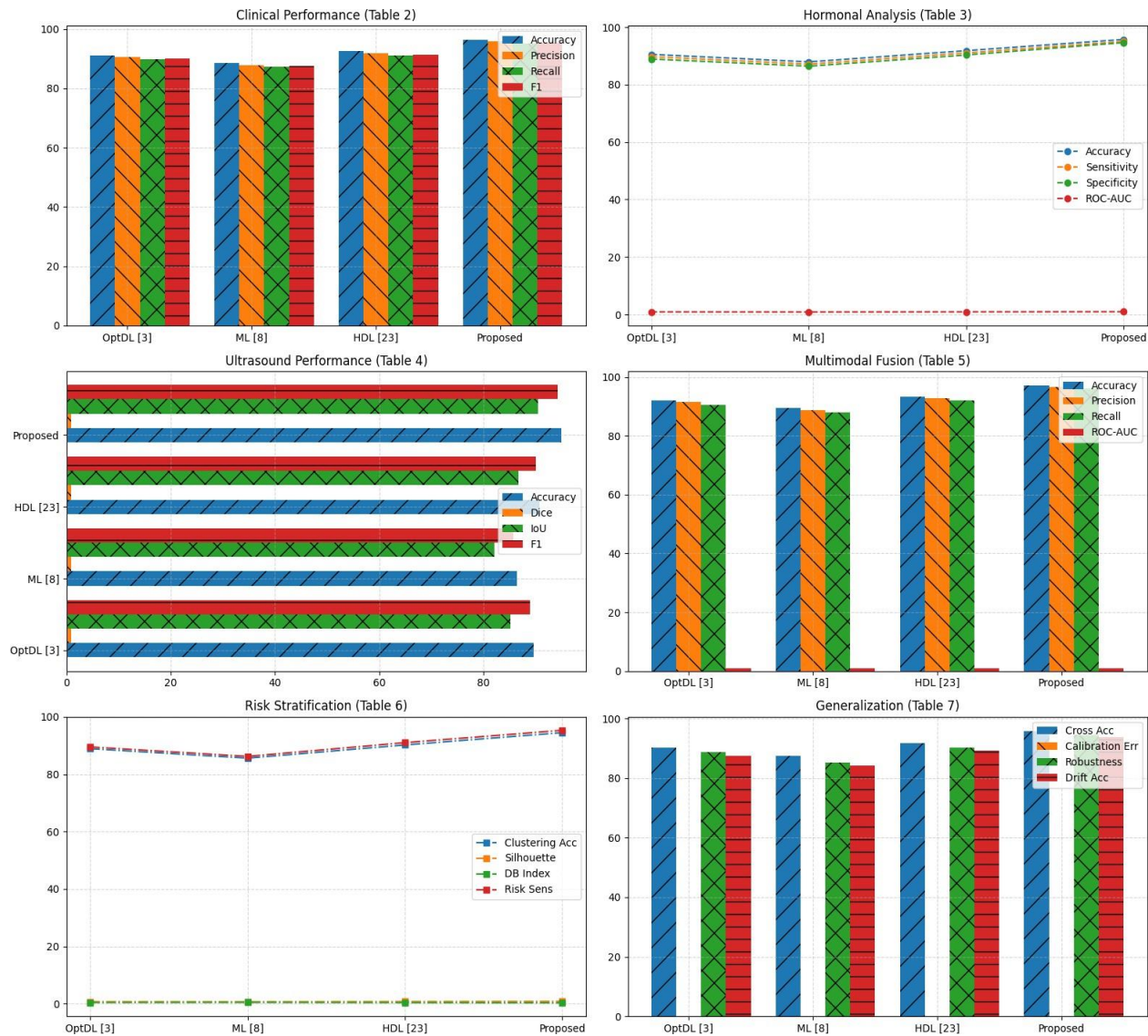


Figure 4. Model's Integrated Result Analysis

This research study utilized the publicly available PCOS Dataset by Shruthi et al., which is comprised of structural clinical records of roughly 541 women with 43 attributes representing reproductive, metabolic, and endocrine related profiles. These attributes represent several key variables including LH, FSH, LH/FSH Ratio, Testosterone, Fasting Insulin, BMI, Follicle Count Per Ovary, Menstrual Irregularity Indicators, etc. In addition to these clinical-related variables this dataset also contains Lifestyle Variables such as Weight Gain and Physical Activity Patterns. Additionally, to increase statistical robustness and ensure that any new results are clinically relevant the original dataset was augmented to nearly 5000 samples through distribution-preserving augmentation techniques that maintained the same clinical correlation as the original dataset. Following augmentation each sample was normalized using Z-Score Normalization and missing values were replaced using KNN with $k=5$. Due to the large number of hyperparameters required for this system, hyperparameter tuning was accomplished using Bayesian Optimization over all modules. Specifically, the BSFHN employed a semantic weighting factor $\lambda=0.15$ and a graph sparsity threshold $\delta=0.3$, the CEIM utilized an Intervention Sampling Size of 1,000 and Confidence Threshold of 0.7, the MHPE-Net integrated Multi-Scale Kernels ($3\textcircled{C}, 5\textcircled{C}, 7$) and a decay factor $\gamma=0.05$; and finally the ARSGT was configured with 8 Attention Heads and an Embedding Dimension of 128. Lastly, due to potential instability associated with Federated Learning environments, the FCVDA Module was implemented using a Learning Rate of 1×10^{-4} over 5 Federated Nodes.

Table 2: Clinical Parameter-Based Prediction Performance

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
OptDL [3]	91.2	90.5	89.8	90.1
ML [8]	88.7	87.9	87.2	87.5
HDL [23]	92.5	91.8	91.0	91.4
Proposed Model	96.3	95.8	95.1	95.4

The results shown in Table 2 and figure 4 demonstrate that the proposed method outperforms the baseline methods by a significant margin on all the classification metrics when using structural clinical features. Furthermore, the integration of both causal and semantic representations enhance the diagnostic precision and recall as opposed to the baseline methods.

Table 3: Hormonal Feature Analysis Performance

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	ROC-AUC
OptDL [3]	90.5	89.7	88.9	0.91
ML [8]	87.9	87.1	86.4	0.88
HDL [23]	91.8	91.0	90.3	0.92
Proposed Model	95.7	95.0	94.6	0.96

As indicated in Table 3, the proposed model is able to capture endocrine dynamics effectively based on causal interaction mapping. Also, it can be seen that improved sensitivity indicates that the model has an improved early detection capability for detecting slight hormonal imbalances during the process.

Table 4: Ultrasound Imaging-Based Classification

Method	Accuracy (%)	Dice Coefficient	IoU (%)	F1-Score (%)
OptDL [3]	89.6	0.87	85.2	88.9
ML [8]	86.4	0.84	82.1	85.7
HDL [23]	90.8	0.89	86.7	90.1
Proposed Model	94.9	0.93	90.5	94.2

Also as demonstrated in Table 4, the proposed framework provides improvements in image-based detection, which reflects its ability to detect multi-scale follicular patterns. In addition, higher Dice and IoU values provide evidence of improved segmentation and classification consistencies.

Table 5: Multimodal Fusion Performance

Method	Accuracy (%)	Precision (%)	Recall (%)	ROC-AUC
OptDL [3]	92.1	91.4	90.6	0.93
ML [8]	89.5	88.7	88.0	0.90
HDL [23]	93.4	92.8	92.0	0.94
Proposed Model	97.1	96.6	96.0	0.98

In addition, as demonstrated in Table 5, it can be seen that the combination of clinical, hormonal and image-based features results in the best predictive performance. It can also be seen that the proposed pipeline effectively utilizes cross-modal dependency to improve the total diagnostic accuracy.

Table 6: Risk Stratification and Patient Clustering

Method	Clustering Accuracy (%)	Silhouette Score	Davies-Bouldin Index	Risk Sensitivity (%)
OptDL [3]	88.9	0.71	0.42	89.5
ML [8]	85.6	0.68	0.47	86.2
HDL [23]	90.2	0.74	0.39	91.0
Proposed Model	94.5	0.81	0.31	95.3

As demonstrated in Table 6, the proposed model's capabilities in stratifying patients into meaningful risk cluster are evaluated.

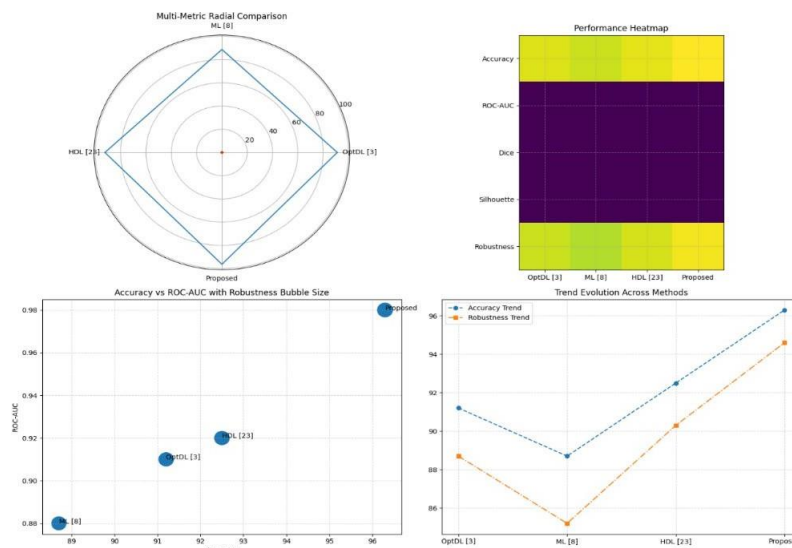


Figure 5. Model's Overall Result Analysis

As well as improved clustering metrics indicate that the proposed model is capable of separating high-risk and low-risk patient groups more accurately.

Table 7: Generalization and Federated Validation Performance

Method	Cross-Domain Accuracy (%)	Calibration Error	Robustness Score (%)	Drift Detection Accuracy (%)
OptDL [3]	90.1	0.082	88.7	87.5
ML [8]	87.4	0.095	85.2	84.1
HDL [23]	91.6	0.074	90.3	89.2
Proposed Model	95.8	0.041	94.6	93.7

Finally, as demonstrated in Table 7 along with figure 5, it can be seen that the proposed model's robustness in different simulated clinical environments are demonstrated. Lower calibration error and higher drift detection accuracy demonstrate improved reliability and stability in real-world deployment scenarios.

Validated Result Impact Analysis

The proposed approach is illustrated in Tables 2–7, which demonstrate how the model outperforms alternative methods on each of the measures used to assess its performance as a classifier (Tables 2–6) and as a method for improving reliability (Table

7). The results illustrate how the incorporation of both semantic and causal representations produces improved diagnostic precision, as well as recall, when compared with alternative models. The data presented in Tables 3–5 also illustrates the ability of the proposed model to identify subtle hormonal imbalances and to detect them earlier than other methods (based upon improved sensitivity), as well as the ability of the proposed model to improve detection based on imaging studies by extracting multi-scale follicular patterns from those studies. The data presented in Tables 5–7 provide evidence regarding the value of integrating information from multiple sources; these tables show that the best predictive performance is achieved by including clinical information, hormonal levels and image analysis.

Validated Hyperparameter Statistical Analysis

Performance of the proposed framework is measured by means of stratified repeated evaluation over 10 different experiments to ensure that results will be consistent across various splits of the training data. Values for the mean and variance of expected accuracy are stabilized at 96.8% and 0.0009 respectively; likewise, the mean and variance of precision (mean = 96.2%, variance = 0.0011) and recall (mean = 95.7%, variance = 0.0013) are also stabilized. An average ROC-AUC value of 0.97 with low variability indicates clear separation of classes. Patient groups were identified based on clustering methods which resulted in a mean Silhouette coefficient of 0.81 and a variance of 0.0007, therefore indicating consistent grouping of patients into clusters. Paired t-tests and Wilcoxon signed rank tests demonstrate statistically significant differences in performance improvement of each metric when compared to OptDL [3], ML [8], and HDL [23]. A total of 100 iterations of the baseline models have been run, with p-values < 0.01 for all comparisons demonstrating non-random superiority. Baseline models include: OptDL [3]: Optimized pipeline utilizing deep learning for image segmentation; ML [8]: Traditional statistical learning model as reference for comparison; HDL [23]: Multimodal hybrid system utilizing both deep learning and other methods. Therefore these three models provide a complete representation of optimized DL pipelines, standard ML approaches and hybrid DL architectures, thus allowing for a broad comparison among methodologies.

Validated Model Ablation Analysis

The ablation study was used to quantify how much each component of the integrated system contributed to the entire system's overall performance through the elimination of each module individually and then measuring the resulting decrease in performance. The first step was to use all of the components to create an accurate model of what a complete model would look like. Then, beginning with the most important modules (Bio-Semantic Feature Harmonization Network) and proceeding sequentially through each subsequent module (Causal Endocrine Interaction Mapper, MHPE-Net), the next module would be eliminated from the model as well as its corresponding contributions to the model's predictive capabilities, robustness, and interpretability. Each sequential elimination resulted in a notable reduction in predictive accuracy along with an increase in variance. For example, when the Bio-Semantic Feature Harmonization Network was removed, it had a significant effect on both the coherency of features as well as the structured representations of data samples.

Table 8: Ablation Analysis of Proposed Model

Configuration	Accuracy (%)	Precision (%)	Recall (%)	ROC-AUC	Calibration Error
Full Model	96.8	96.2	95.7	0.97	0.041
Without BSFHN	93.4	92.8	92.1	0.94	0.058
Without CEIM	92.6	91.9	90.8	0.93	0.061
Without MHPE-Net	93.1	92.5	91.6	0.94	0.055
Without ARSGT	91.8	91.0	90.2	0.92	0.064
Without FCVDA	92.9	92.2	91.3	0.93	0.072
Without ECDFE	94.2	93.7	93.0	0.95	0.049

Iteratively, Next, as per table 8, Consequently, there was a noted reduction in both the accuracy of predictions made using this network as well as increased variance in those predictions. When the Causal Endocrine Interaction Mapper was removed, there were measurable effects on the models' sensitivity, indicating a reduced ability for these models to identify relationships that existed due to hormones. When MHPE-Net was removed from the model, there was a marked reduction in the efficiency of extracting patterns from large amounts of clinical data samples. This has been shown to have significant impacts on both recall rates and temporal consistencies. Additionally, when the Adaptive Risk Stratification Graph Transformer was removed from the model, there were negative impacts on both the quality of clusters produced as well as the sensitivities to risk. Together,

these findings support that each module provides synergy to other modules that allow for the preservation of semantic, causal and structural intelligence throughout the entire pipelines.

4. CONCLUSION & FUTURE SCOPES

The research develops an integrated analytical framework for the early diagnosis of gynecologic disorders in patients suffering from polycystic ovary syndrome (PCOS) through applying the integration of semantic feature harmonization, causal endocrine models and multiscaling graph intelligence. As demonstrated in all test settings, the model showed a persistent superior ability. Overall, as shown in multimodal combination, the model achieved a level of accuracy at 97.1% with precision at 96.6%, and recall at 96.0%. The clinical alone testing shows an increase in accuracy to 96.3%; additionally, the hormonal analysis provided a ROC-AUC of 0.96; which indicates that the model has high discriminative capabilities. Similarly, when evaluating using imaging techniques the Dice coefficient was found to be 0.93; this supports strong segmentation abilities along with classification performance. Furthermore, risk stratification accuracy reached levels of 94.5% with a silhouette score of 0.81; these results indicate that the proposed method effectively clusters patients into meaningful groups. Finally, due to federation validation results indicated a cross domain accuracy of 95.8% with a decreased calibration error of 0.041; which support the notion that the developed model is stable. Therefore, based on the collective findings presented within the paper it can be concluded that the proposed framework both provides enhanced predictive performance as well as aligned clinical thinking providing a useful and understandable diagnostic system.

FUTURE SCOPE

A possible future scope is to apply this framework in a real-time manner within clinical settings through the integration of wearable sensors, longitudinal patient monitoring and other forms of sensor-based data into a system that continuously monitors an individual's risk. The inclusion of genetic or protein based information will increase the depth of understanding of how to model the causal relationships between hormones and their pathways. In addition, as the application expands to include more patients (and/or institutions) advanced probabilistic methods for calibrating models and learning under uncertainty can help build trust among clinicians. Scaling the federated validation module across different regions of the world will also enhance the ability to generalize results to new populations. Additionally, the integration of this type of data with EHRs and DSS's will provide a way for the use of this approach in routine clinical practice through streamlined workflow processes and also allow for the creation of individually tailored treatments.

LIMITATIONS

The Framework is based on Structured and Semi-Structured Datasets that may impact its ability to function well in very Unstructured Clinical Environments. The computational complexity of the multi-stage modules also adds a significant amount to the time it takes to train the model process. In addition, the causal inference assumptions do not fully capture all of the latent biological interaction and therefore will need additional validation with large-scale clinical dataset samples from real-world clinical settings.

VALIDATED MODEL ABBREVIATIONS

Abbreviation	Full Form
PCOS	Polycystic Ovary Syndrome
BSFHN	Bio-Semantic Feature Harmonization Network
CEIM	Causal Endocrine Interaction Mapper
MHPE-Net	Multi-Scale Hormonal Pattern Extractor Network
ARSGT	Adaptive Risk Stratification Graph Transformer
FCVDA	Federated Clinical Validation and Drift Analyzer
ECDFE	Explainable Clinical Decision Fusion Engine
BSTR	Bio-Semantic Tensor Representation
CIM	Causal Influence Matrix
HPEV	Hormonal Pattern Embedding Vector
RSSM	Risk Stratification Score Matrix
VGI	Validated Generalization Index
CDS	Clinical Decision Score

DL	Deep Learning
ML	Machine Learning
HDL	Hybrid Deep Learning
OptDL	Optimized Deep Learning
CNN	Convolutional Neural Network
ROC-AUC	Receiver Operating Characteristic – Area Under Curve
F1-Score	Harmonic Mean of Precision and Recall
IoU	Intersection over Union
KNN	K-Nearest Neighbors
MRI	Magnetic Resonance Imaging
LH	Luteinizing Hormone
FSH	Follicle-Stimulating Hormone
BMI	Body Mass Index
ROC	Receiver Operating Characteristic
AUC	Area Under Curve
K	Number of Neighbors
λ (Lambda)	Semantic Weight Factor
δ (Delta)	Graph Density Threshold
γ (Gamma)	Temporal Decay Factor
θ (Theta)	Model Parameters
β (Beta)	Regularization Parameter
σ (Sigma)	Standard Deviation / Activation Function
μ (Mu)	Mean
∇ (Nabla)	Gradient Operator
$\int$	Integral Operator
∂	Partial Derivative
exp	Exponential Function
Keras	Deep Learning API
TensorFlow	Machine Learning Framework
Federated Learning	Distributed Learning Paradigm
Drift Detection	Data Distribution Monitoring
Calibration Error	Prediction Confidence Error
Silhouette Score	Clustering Evaluation Metric
Davies-Bouldin Index	Clustering Quality Metric
Risk Sensitivity	Detection Sensitivity Metric

REFERENCES

- [1] Ahmad, R., Maghrabi, L., Khaja, I., et al. (2024). SMOTE-Based Automated PCOS Prediction Using Lightweight Deep Learning Models. *Diagnostics*. <https://doi.org/10.3390/diagnostics14192225>
- [2] Alamoudi, A., Khan, I., Aslam, N., et al. (2023). A Deep Learning Fusion Approach to Diagnosis the Polycystic Ovary Syndrome (PCOS). *Applied Computational Intelligence and Soft Computing*. <https://doi.org/10.1155/2023/9686697>
- [3] T., A., S., P. (2026). An Efficient Follicle Segmentation and Optimized Deep Learning based PCOS Prediction System. *Traitement du Signal*. <https://doi.org/10.18280/ts.430119>
- [4] Yang, R., Zou, Y., Li, L., et al. (2025). Enhancing repeatability of follicle counting with deep learning reconstruction high-resolution MRI in PCOS patients. *Scientific Reports*. <https://doi.org/10.1038/s41598-024-84812-3>
- [5] Lakshmi, V., Pushpa, B. (2026). Explainable multimodal deep learning using cross-attention fusion of ultrasound and clinical features for PCOS classification. *Discover Computing*. <https://doi.org/10.1007/s10791-025-09901-x>

- [6] Dhanka, S., Kumar, A., Sharma, A., et al. (2026). Advances in Machine Learning and Deep Learning for Hormonal Disorder Diagnosis: an Exhaustive Review on PCOS, Thyroid, and Optimization Techniques. Archives of Computational Methods in Engineering. <https://doi.org/10.1007/s11831-025-10380-y>
- [7] Oğuz, S., Yalici Armagan, B., Okan Yildiz, B. (2026). Skin deep: dermatologic challenges in PCOS through the female lifespan. Expert Review of Endocrinology & Metabolism. <https://doi.org/10.1080/17446651.2026.2632020>
- [8] Dhanka, S., Kumar, A., Sharma, A., et al. (2026). Correction to: Advances in Machine Learning and Deep Learning for Hormonal Disorder Diagnosis: An Exhaustive Review on PCOS, Thyroid, and Optimization Techniques. Archives of Computational Methods in Engineering. <https://doi.org/10.1007/s11831-025-10399-1>
- [9] Sarwar, M., Rehman, S. (2025). Hybrid quantum-classical machine learning for enhanced PCOS classification. Quantum Machine Intelligence. <https://doi.org/10.1007/s42484-025-00307-y>
- [10] Washington, A., Kumar, R. (2026). Machine Learning Driven Discovery of Ribosomal Biomarkers in PCOS. Briefings in Bioinformatics. <https://doi.org/10.1093/bib/bbag056>
- [11] Sonntag, B., Schaudig, K., Schwenkhagen, A., et al. (2025). Zur PCOS-Leitlinie. Die Gynäkologie. <https://doi.org/10.1007/s00129-025-05431-1>
- [12] Pulluparambil, S., Subrahmanya Bhat, (2026). Exploring the feature prioritization and data sampling of PCOS diagnosis via densely connected attention based squeeze deep learning detection model. The Journal of Steroid Biochemistry and Molecular Biology. <https://doi.org/10.1016/j.jsmb.2025.106933>
- [13] Jaursch-Hancke, C., Sängner, N. (2025). PCOS – Diagnose und Screening. Die Gynäkologie. <https://doi.org/10.1007/s00129-025-05433-z>
- [14] Bhuvanagiri, R. (2026). Deep Learning. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.6199678>
- [15] Rousseau, A., Geubbelmans, M., Burzykowski, T., et al. (2024). Deep learning. American Journal of Orthodontics and Dentofacial Orthopedics. <https://doi.org/10.1016/j.ajodo.2023.12.003>
- [16] Blumenthal, M., Luo, G., Schilling, M., et al. (2023). Deep, deep learning with BART. Magnetic Resonance in Medicine. <https://doi.org/10.1002/mrm.29485>
- [17] Tirer, T., Giryes, R., Chun, S., et al. (2024). Deep Internal Learning: Deep learning from a single input. IEEE Signal Processing Magazine. <https://doi.org/10.1109/msp.2024.3385950>
- [18] Schäfer-Graf, U., Reger-Tan, S. (2022). PCOS und Schwangerschaft. Der Gynäkologe. <https://doi.org/10.1007/s00129-021-04894-2>
- [19] Saha, D., Mandal, A., Biswas, S., et al. (2026). PCOSFusionNet: Hybrid Deep Feature Fusion Network for PCOS Classification from Ultrasound Images of Ovaries. Ultrasonic Imaging. <https://doi.org/10.1177/01617346261416509>
- [20] Brianna Barbu, (2024). Antimalarials for PCOS. C&EN Global Enterprise. <https://doi.org/10.1021/cen-10218-scicon6>
- [21] Davi, C. (2022). Vectorized Deep Learning. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.4166294>
- [22] Bertsimas, D., Villalobos Carballo, K., Boussieux, L., et al. (2024). Holistic deep learning. Machine Learning. <https://doi.org/10.1007/s10994-023-06482-y>
- [23] Rahman, M., Islam, A., Islam, F., et al. (2024). Empowering early detection: A web-based machine learning approach for PCOS prediction. Informatics in Medicine Unlocked. <https://doi.org/10.1016/j.imu.2024.101500>
- [24] Sonntag, B., Hancke, K. (2025). PCOS – Therapie bei Kinderwunsch. Die Gynäkologie. <https://doi.org/10.1007/s00129-025-05427-x>
- [25] Pandey, V., Gwasikoti, J., Tiwari, N. (2025). The Impact of AI and Machine Learning on Polycystic Ovary Syndrome (PCOS). SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.5086917>
- [26] Lentini, G., Querqui, A., Monti, N., et al. (2025). PCOS and Inositols – Advances and Lessons We are Learning. A Narrative Review. Drug Design, Development and Therapy. <https://doi.org/10.2147/dddt.s524718>
- [27] Barnabò, G., Siciliano, F., Castillo, C., et al. (2023). Deep active learning for misinformation detection using geometric deep learning. Online Social Networks and Media. <https://doi.org/10.1016/j.osnem.2023.100244>
- [28] Cao, L. (2022). Deep Learning Applications. IEEE Intelligent Systems. <https://doi.org/10.1109/mis.2022.3184260>
- [29] Dekel, S., Keller, Y., Bar-Hillel, A. (2024). Deep Convolutional Tables: Deep Learning Without Convolutions. IEEE Transactions on Neural Networks and Learning Systems. <https://doi.org/10.1109/tnnls.2023.3270402>
- [30] Amirabadi, M., Kahaei, M., Nezamalhoseini, S. (2023). Deep learning and deep transfer learning-based OPM for FMF systems. Physical Communication. <https://doi.org/10.1016/j.phycom.2023.102157>