Journal of Theoretical and Applied Information Technology
31st October 2026. Vol.104. No.20
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

QUANTUM-ENHANCED DEEP LEARNING FRAMEWORK FOR AUTOMATED IDENTIFICATION OF ENDOMETRIOSIS FROM PELVIC MEDICAL IMAGES

K DEVIPRIYA[1]*, SREELALITHA TALLURI[2], K. S. S. NARAYANA[3], SAPE CHITTIBABULU[4], M. SUMALATHA[5], HARI KISHAN CHAPALA[6], SESHA BHARGAVI VELAGALETI[7], GARIGIPATI RAMA KRISHNA[8]

*[1]Department of Computer Science and Engineering, Lakireddy Bali Reddy College of Engineering, Mylavaram, Andhra Pradesh, India
[2]Department of Information Technology, Vignana Bharathi Institute of Technology, Aushapur, Ghatkesar, Hyderabad, Telangana, India
[3]Department of CSE (AI&ML), Prasad V. Potluri Siddhartha Institute of Technology, Vijayawada-520007, Andhra Pradesh, India
[4]Department of Computer Science and Engineering, Aditya University, Surampalem-533437, Andhra Pradesh, India
[5]Department of CSE (CS), CVR College of Engineering, Vastunagar, Hyderabad-501510, Telangana, India
[6]Department of CSE - AI & ML, St. Ann's College of Engineering & Technology, Chirala, Bapatla (District), Andhra Pradesh, India
[7]Department of Information Technology, G. Narayanamma Institute of Technology and Science, Shaikpet, Hyderabad-500104, Telangana, India
[8]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram-522302, Guntur, Andhra Pradesh, India

E-mail: k.devipriya20@gmail.com[1]*, maheshwarlasreelalitha@gmail.com[2], kssnarayana@pvpsiddhartha.ac.in[3], chitti.cse507@gmail.com[4], sumagopagalla91@gmail.com[5], drchkishan@gmail.com[6], seshabhargavi@gnits.ac.in[7], umarkcse@gmail.com[8]

ABSTRACT

Additionally, lesions can manifest in many ways, and diagnostic algorithms for identifying them in chronic disorders can be complex, depending on anatomical complexity, imaging acquisition methods, and lesion presentation. In medical image analysis, deep learning has advanced significantly; however, classic models may suffer from limited data, noise, and prediction inaccuracies. This work, QEDL-Endo, is a quantum-enhanced deep learning method for automatically diagnosing endometriosis from pelvic-region magnetic resonance imaging (MRI). It uses an 8-dimensional latent representation, classical–quantum feature fusion, an 8–qubit variational quantum circuit, an uncertainty module and a modified classifier. The goal is to determine whether this approach can improve classification accuracy while keeping the quantum architecture's resource complexity low by using human anatomy to learn a representation during training and a quantum feature map to map the chosen features into resource-efficient representations. On this reported test set, QEDL-Endo reached 95.1% accuracy, an AUROC of 0.978 and an AUPRC of 0.974. Additionally, it reported a Brier (probability) score of 0.061 and an expected calibration error of 0.029 for its probabilistic predictions. Performance improved after introducing attention, amplitude transformation, and feature fusion. These data suggest that this proposed architecture may be a computer-aided imaging method, though it may not offer a quantum advantage or be clinically ready. A second constraint is that not all units are available and independent; therefore, validation requires multi-center evaluation of real-world performance. Future studies should focus on the study's external generalizability, the use of validated diagnostic labels, and whether a quantum advantage was realized, regardless of statistical significance over an equivalently optimized classical model.

Keywords: *Endometriosis, Pelvic MRI, Deep Learning, Quantum Machine Learning, Anatomy-Guided Attention, Medical Image Classification*

1. INTRODUCTION

Endometriosis is a chronic condition in females associated with the uterine cavity (endometrium) and endometrial-like tissue outside the uterus. It occurs mainly in the ovary and pelvic parts and is associated with decreased quality of life,

dysmenorrhea, dyspareunia, infertility, and pelvic pain [1].

These are important for non-invasive assessment of endometriosis and include pelvic imaging, particularly transvaginal ultrasound (TVUS) and magnetic resonance imaging (MRI) [2]. Both modalities can detect ovarian endometriomas and deep infiltrating endometriosis (DIE). However, the likelihood of detecting lesions depends on lesion location, imaging protocol, disease characteristics, and the examiner's experience [3]. Many subtle lesions are present in this context; however, tissue characteristics vary, and complex anatomical relationships require reliable image interpretation [5]. Thus, future studies should systematically assess the pelvis and deep endometriotic lesions using standardized imaging methods [4].

This problem matters not only for accurate image classification, but also for improving overall accuracy. Imaging evaluation is often inconclusive and does not help reduce delays in diagnosis or clinical management. Endometriosis tends to be an elusive diagnosis in the computer because diagnostic features appear in only a small percentage of MRI volumes, can occur at different locations, and are not always present, yet may have the same morphology or signal characteristics. This means the method can perform very differently for patients not included in the training set or for acquisitions from a different imaging protocol, and these differences might be uninformative or reflect miscalibrated probabilities. Therefore, any computer-assisted methodology should distinguish between positive and negative cases of endometriosis, apply anatomical information when relevant, and keep the methodology simple while expressing uncertainty. However, performance cannot be neglected.

A promising strategy to enhance automated medical image interpretation is artificial intelligence (AI). One recent study exploring AI applications in caring for individuals diagnosed with endometriosis focused on both ultrasound and MRI, demonstrating promising diagnostic outcomes [6]. Deep CNNs are widely used to learn multi-level representations and have made tremendous advances in medical image classification [7]. However, they face challenges with limited training data, class imbalance, varying acquisition protocols, and limited use of anatomical regions in medical images [8]. These restrictions are particularly important in the context of an endometriosis diagnosis, as the disease may affect a small area in a pelvic image.

Quantum machine learning (QML) is an emerging paradigm that can enhance classical machine-learning representations using quantum machines. By mapping classical features into a quantum feature space and learning non-linear interactions, parametric quantum circuits can represent classical features [9]. Hybrid quantum–classical architectures are particularly suitable for near-term quantum systems because computationally intensive image representation can remain classical while quantum processing is applied to a compact feature vector [10]. Quantum feature maps further provide an alternative representation space in which complex relationships among learned features can be explored [11]. Most current QML applications in healthcare are limited by qubit constraints, noise, high encoding costs for qubits using complex codes, and limited evidence of quantum advantage [12]. Therefore, hybrid architectures should confine quantum processing to low-dimensional information representations.

Recent advances in quantum image learning have demonstrated the viability of combining convNet features with a variational quantum circuit [13]; however, validation and scaling to the application have not yet been explored. Recent studies on AI have shown promise for endometriosis. However, sample sizes are small, prospective study designs are lacking, overlap between true-positive and true-negative groups is limited, and no external validation exists [14], [15]. This study integrates anatomically inspired features and efficient quantum processing to achieve reliable prediction.

Accordingly, the present study is motivated by methodological and clinical considerations. Using CNNs for complex image representations can be useful. Nevertheless, traditional CNNs can suffer from irrelevant anatomical areas and a limited number of training subjects. While preliminary experimental evidence suggests a quantum advantage, hardware limitations and inconsistent results make it difficult to predict whether a parameterized quantum circuit will outperform quantum machine learning in practice. Here, we present a hybrid architecture that demonstrates the feasibility of this approach by performing high-dimensional image processing in a classical backbone and processing only low-dimensional information in the quantum backbone.

To explore using both deep learning and Quantum Computing to classify endometriosis in medical images automatically, this study introduces a novel Quantum-Enhanced Deep Learning Framework

(QEDL-Endo). It aims to (i) guide attention to diagnostically relevant components of the pelvis via attention-guided feature learning, (ii) extract compact deep representations ideal for quantum processing, (iii) provide a variational quantum circuit that can model the non-linear interactions between the latent features and (iv) add uncertainty-aware classification for more general cases.

Research Hypothesis: In this study, we assume that using a compact variational quantum feature-transformation module and uncertainty-aware classification will further improve MRI endometriosis identification over an appropriately optimized classical deep-learning baseline model, in addition to anatomy-guided dual attention. The observed improvement stems from anatomically relevant representations, an explicit prediction uncertainty measure, and a low-dimensional latent feature vector learned and transformed by a parameterized quantum circuit. Comparative experiments, componentwise ablation analysis, a discrimination metric, and a calibration measure will test the hypothesis. More specifically, the contribution of quantum processing needs to be separated from the contribution of attention, feature fusion and the extra amount of complexity of the model itself—making it difficult for any claims of quantum-specific benefit to be made without a proper classical comparator, and (when possible) statistical testing.

The following research questions are proposed:
RQ1: To investigate the performance enhancement of QEDL-Endo over conventional CNN-based baselines.
RQ2: What is the role of attention based on anatomical features, quantum feature transformation and classical-quantum fusion when used individually and together?
RQ3: Can we better calibrate and produce reliable model outputs when uncertainty is included in predictions?
RQ4: How well could the proposed approach apply to various institutions, scanners, and imaging protocols?

Scope of the Study: In this effort, the authors investigate AI for endometriosis in MRIs within a hybrid classical–quantum paradigm. It discusses MRI preprocessing, information from the anatomical level for feature learning, feature learning using ConvNeXt-Tiny and 8D for latent representation learning, variational quantum feature transformation by an 8Q circuit, classical-quantum feature fusion, uncertainty-aware classification, and classic evaluation metrics like classification, calibration, and ablation scores. Another important experimental design is to partition patient information into three parts, train, validate, and test, to prevent patient-information leakage between the three partitions and reduce patient-level leakage. The study is not a clinically proven diagnostic system; it is not better than all classical systems, and it is not as effective as demonstrated in prospective experiments. It does not provide validated disease-severity staging, comprehensive lesion segmentation, or complete diagnosis with ultrasound and MRI, unless shown in specific experiments using reference labels. Implementing and testing generalizability on a real quantum computer is an important issue for future investigations.

This innovation integrates anatomical awareness, efficient deep feature encoding, variational quantum transformation, and uncertainty-aware prediction into a single framework for endometriosis imaging. We analyze this mixed solution for its ability to improve classification accuracy while maintaining a near-term-compatible quantum architecture. We use accuracy, sensitivity, specificity, precision, F1-score, AUROC, and calibration measures to evaluate performance, and we use an ablation analysis to quantify each component's contribution.

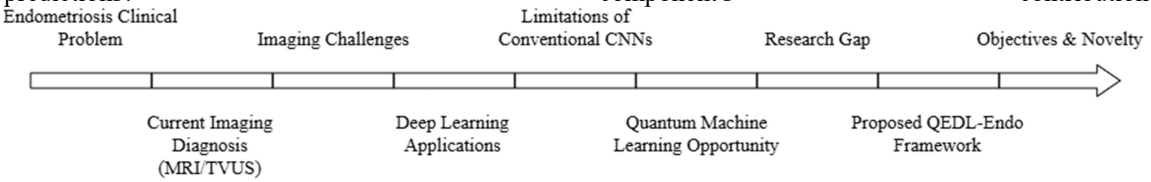


Figure 1. Research Motivation And Progression Toward The Proposed Qedl-Endo Framework

Figure 1 shows the Introduction section, which covers the clinical problem of endometriosis and current imaging-based diagnosis (MRI/TVUS). It then illustrates the challenges of pelvic imaging and weaknesses of traditional CNN deep learning algorithms. It then introduces the quantum machine learning opportunity that addresses these limitations, identifies the research gap, and proposes the QEDL-Endo framework. Finally, it sets out the framework's aims and innovative contributions to support the development of a new methodology for the study.

The rest of this paper is organized as follows. In Section 2, we review related work on artificial intelligence, deep learning, medical image analysis,

and quantum machine learning relevant to automated endometriosis identification. Section 3 describes the proposed Quantum-Efficient Deep Learning for Endo (QEDL-Endo) methodology, along with the MRI dataset, preprocessing methodology, anatomy-guided dual attention model for similarity search, classical feature extraction, quantum latent representation, hybrid quantum-classical architecture, mathematical formulation, training methodology, and reproduction settings. Section 4 presents and discusses the experimental results and performance evaluation, including accuracy, sensitivity, specificity, precision, F1-score, AUROC, AUPRC, calibration, ablation, and comparisons with other models. Section 5 summarizes the study findings, identifies gaps, and suggests future research directions.

2. RELATED WORK

Automated assessment of endometriosis has emerged in recent years, driven by artificial intelligence (AI), radiomics, and deep learning. This literature can be broadly divided into traditional machine learning methods, deep learning-based image analysis, radiomics, segmentation approaches, multimodal learning, and novel quantum methods in medical image analysis.

2.1 Machine Learning for Endometriosis Diagnosis

Previous healthcare studies applying artificial intelligence (AI) relied on manually collected clinical, laboratory, and imaging data and used classic machine learning classifiers. These studies suggested that researchers could analyze correlations between clinical biomarkers and endometriosis using support vector machines (SVM), random forest (RF), logistic regression (LR), and boosting techniques. However, they had to learn complex spatial patterns directly from medical image sets, which they could only do through handcrafted features [16], reducing their learning power. Recent research indicates that multiple clinical biomarkers predict outcomes more accurately than single biomarkers. However, predictive modeling using non-imaging biomarker data remains relevant if such data become limited in information quality and/or availability [17].

A more straightforward route to automated diagnosis is machine learning using image representations. To date, researchers have analyzed lesions through classification, detection, and segmentation using AI and deep-learning architectures, which have shown high potential to outperform traditional machine learning [18]. In addition, researchers investigated a multi-scale computer-vision-based approach to automatically recognize and segment endometriotic lesions, highlighting the potential of deep feature learning for automatic endometriosis recognition without relying on manually described image descriptors [19]. Such representations can support useful image-based diagnostic conclusions but may be adversely affected by the number of images acquired, image illumination, and disease phenotype. These studies are a foundation for the focus on deep learning described below.

2.2 Deep Learning for Ultrasound-Based Analysis

Using deep convolutional networks, researchers can extract more information from ultrasound images, particularly in images related to endometriosis. Convolutional architectures have recently achieved nearly 100% accuracy in differentiating ovarian endometriotic and mucinous cysts [20], highlighting their robustness and promise for disease diagnosis. These models directly identify patterns that manually generated radiomic descriptors discover less efficiently. They could also reduce dependence on such descriptors.

According to a comprehensive systematic review, applications in gynecological ultrasound are emerging, and cancer diagnosis, treatment, and screening are future areas of cancer medicine that could benefit from AI [15], [18], [21]. However, few studies have focused specifically on endometriosis, and those that did were very different and therefore difficult to compare. A recent systematic review of AI algorithms for endometriosis diagnosis using ultrasound reported strong performance for deep-learning algorithms and lower overall performance for conventional machine-learning algorithms [22]. The findings suggest deep learning's potential usefulness and the need for more datasets, standardised testing, and external evaluation.

2.3 Deep Learning for MRI-Based Endometriosis Detection

MRI is useful for soft tissue imaging and is especially useful for deep endometriosis. In one recent work, the highest performance of multi-sequence MRI was reported with a 3D DenseNet architecture on the F1 Score (0.881), AUROC (0.911), sensitivity (0.976), and specificity (0.720). AI also improved radiologists' sensitivity, suggesting that deep learning may be useful as a decision-making aid, not just a stand-alone classifier [23]. One possible system to support reading used lesion segmentation, supplemented by radiomics-based detection of adhesion and identification of ovarian endometriotic cysts, in an

MRI score. Several specialized AI models were assessed to detect subtle features of deep endometriosis (DE) and support radiologists [24].

Endometriosis is not a single disease; lesions may differ in appearance, location, signal characteristics, and extent of anatomic involvement, all of which matter. This variation suggests that imaging-based approaches need to account for lesion heterogeneity.

MR-based radiomics has also been used to elucidate lesion characteristics rather than detect disease. Some studies combine radiomic, morphological, and graph-based imaging features to distinguish between active and fibrotic deep endometriosis lesions and to correlate imaging features with patients' pain profiles [25]. This may shift the focus from disease diagnosis to phenotyping and individual clinical assessment.

2.4 Multimodal and Knowledge-Transfer Approaches

One challenge with medical imaging data is the lack of convenient, precise labels for all images. To this end, multimodal learning may be the solution, providing multiple information sources; indeed, the area under the curve increased from 65.0% to 90.6% in POD detection in MRI scans, with transvaginal ultrasound and without paired data [26]. The results also suggest that different modalities and supplemental information could enhance the learning potential of complementary tests; this matters because ultrasound and MRI provide different anatomical information. These systems, however, have some problems: (i) differences in how the information is preprocessed and represented across modalities, and (ii) high computational costs.

2.5 Limitations of Existing Endometriosis AI Systems

While AI can aid diagnosis, researchers need to test it within the context of existing methodological limitations in endometriosis studies. Retrospective datasets typically include few patients and limited diversity. Because of limited external validation, small sample sizes, sample imbalance, and sample bias, treat the estimates with caution rather than optimism. More recent data also show important heterogeneity in clinical response and spectrum effects, especially if the model continues to target end-stage disease [27].

Current systems have also focused mainly on classification accuracy, while ignoring prediction uncertainty, calibration, and explanation. A good AUROC cannot always provide accurate patient-level probabilities. In addition, models trained on one imaging modality or institution may degrade when deployed in another modality or institution.

2.6 Quantum-Enhanced Medical Image Analysis

In recent years, quantum machine learning (QML) has appeared in the literature as an alternative for processing high-dimensional medical data. In most hybrid quantum–classical architectures, the classical neural network first extracts features, followed by nonlinear transformation and/or classification by quantum circuits. Although directly encoding high-resolution images into quantum states remains computationally expensive, this strategy is relevant to medical imaging [28].

For medical image classification, feature extraction using convolutional neural networks (CNNs) is feasible, and hybrid quantum CNN architectures have combined this operation with quantum circuits [29]. Researchers have also encoded compact image representations into quantum circuits and evaluated them using hybrid classical–quantum optimization, as shown in similar CNN–quantum classifier frameworks for chest X-ray disease detection, demonstrating the feasibility of quantum evaluation [30]. They show that applications of quantum layers to existing deep neural networks do not necessitate a full quantum image processing architecture.

However, there is no agreement on the quantum advantage of medical imaging. In quantum research, resources are limited, and researchers regularly use simulators with limited qubits and specific datasets. Hardware noise, decoherence, circuit depth, data-encoding overhead, and the need for repeated experiments are key issues [31]. Researchers must overcome these restrictions while keeping computational needs low to preserve the discriminating information needed for practical quantum-enhanced medical imaging.

2.7 Research Gap

AI tools, such as multimodal learning, segmentation, radiomics, ultrasound, and MRI, have driven technological progress in endometriosis diagnosis. Current systems, however, are predominantly traditional, relying on feature engineering, compact representations, and uncertainty-based prediction. Against this backdrop, quantum image analysis has focused primarily on applications outside endometriosis and has often relied on small sample sizes or computer simulations of quantum environments.

A gap therefore remains in hybrid (quantum and classical) and resource-efficient imaging solutions for endometriosis. To address this gap, the QEDL-Endo framework extracts features before quantum processing to reduce input-image dimensionality

for the processing unit. It also introduces attention gates to refine features and uncertainty-aware class classification to improve discrimination and reliability.

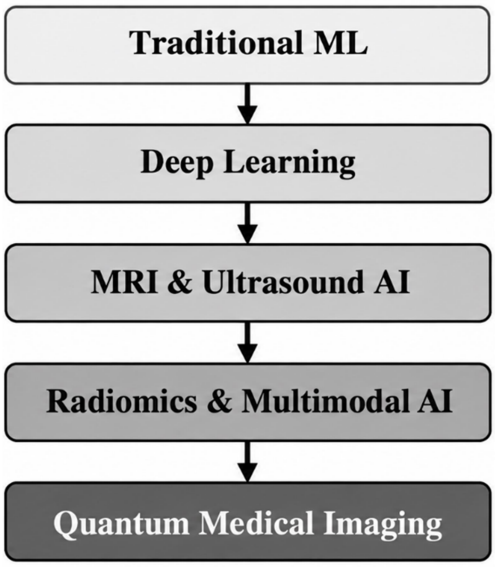


Figure 2. Evolution Of AI Techniques In Medical Imaging

Figure 2 traces the evolution of artificial intelligence (AI) techniques in medical imaging, from Traditional Machine Learning (TML) to Deep Learning (DL) techniques. Next, specialized AI applications in MRI and ultrasound are followed by the integration of quantitative imaging features and data from multiple modalities, known as radiomics and multimodal AI. The last stage represents quantum medical imaging, indicating future applications of quantum computing and quantum-enhanced algorithms in medical imaging. The arrows depict the development path from top to bottom, showing increasing computational complexity and the progression of medical imaging AI technology.

QEDL-Endo has three related propositions as the conceptual basis. First, by ignoring some regions of the nan image and focusing on others guided by anatomy, one may subsample the image. Thus, the feature extractor focuses on relevant regions rather than treating all regions equally. Second, one can also reconstruct the representation by projecting it into an 8-dimensional latent space, which may be more feasible than encoding high-resolution MRI information. Third, Integrating quantum representations with deep features learned from the classical network, along with uncertainty-aware classification, can retain valuable image information while boosting prediction accuracy. The propositions lead to testable predictions. The study's novelty is applying these mechanisms to an open, resource-conscious framework for MRI-level classification. Whether these mechanisms can be extrapolated to other cohorts and are clinically relevant is an empirical question.

3. METHODOLOGY

3.1 Overall Framework

We aim to develop a semi-automated, patient-specific endo-quantum therapeutic pipeline for automated pelvic MRI analysis, called Quantum-Enhanced Deep Learning for Endometriosis (QEDL-Endo). This framework is designed to support the methodology that follows. The pipeline has five key steps: image preprocessing, anatomy-guided attention, classical deep feature extraction, feature transformation in quantum space, and final prediction. In summary, the complete process is $\hat{y}=F_{\theta_q,\theta_c}\left[Q_{\theta_q}\left(C_{\theta_c}\left(A\left(P(X)\right)\right)\right)\right]$, where X denotes the original pelvic MRI, $P(\cdot)$represents intensity and spatial preprocessing, $A(\cdot)$denotes anatomy-guided attention, $C_{\theta_c}(\cdot)$represents classical deep feature extraction, $Q_{\theta_q}(\cdot)$represents quantum feature transformation, and $\hat{y}$denotes the final endometriosis prediction. The main innovation of QEDL-Endo is the quantum bottleneck, where attention first localizes and refines clinically relevant regions of the pelvis through dual-path attention, then compresses their deep representations into an 8-dimensional latent vector. This compact representation is then expressed as an 8-qubit variational quantum circuit; thus, higher-resolution MRI representations are no longer directly represented, and only diagnostically important information is retained. The processing path therefore includes multi-sequence pelvic MRI acquisition, intensity and spatial normalization, localization of different anatomical regions, extraction of attention features via two paths, formation of an eight-dimensional quantum bottleneck, subsequent variational quantum feature transformation, and uncertainty-aware classification and prediction of endometriosis tissue. This design enables a tailored nonlinear feature transformation step in quantum processing while maintaining the representational power of conventional deep learning.

Research Design and Experimental Workflow. This study is structured as a retrospective computational model-development and evaluation study using the available UT-EndoMRI dataset. The objective is to investigate the feasibility and performance of a hybrid classical–quantum framework for MRI-based endometriosis identification. The research workflow consists of

the following steps: (1) dataset inspection and subject-level diagnostic-label verification with the available clinical annotations and lesions, (2) MRI preprocessing, spatial standardization and extraction of regions of interest, (3) training, testing and validation dataset splitting to avoid samples from the same subject being present in more than one dataset, (4) development of the ConvNeXt-Tiny backbone, anatomy-guided dual attention module, compact latent representation, variational quantum circuit and uncertainty-aware classification head, (5) model optimization with the specified training objective and performed model selection based on validation, (6) comparison for the performance of the trained models with the conventional deep-learning baselines/sets of models with progressively enhanced variants, (7) ablation analysis focusing on the contributions of the attention module, the quantum transformation, feature fusion, and uncertainty estimation, and (8) evaluation based on data discrimination, diagnosis, calibrations and uncertainty-related measures. If there are concerns about the precision of performance estimates and implementation, express uncertainty in performance estimates relative to the patient, with confidence intervals and/or statistical comparisons where underlying experimental data support them. Compare the findings with published results, the dataset's limitations, and the need for external validation. Internal patient-level evaluation is based on performance within the cohort provided; it should not be reported as being based on "independent external validation", unless a completely different cohort is provided.

3.2 Dataset

In this study, the main dataset was the UT-EndoMRI dataset [32] consisting of pelvic magnetic resonance imaging studies from two different Texas patient institutions. The dataset includes MRI sequences and manually generated anatomical labels for structures including the uterus, ovaries, endometrioma, cyst, and cul-de-sac. Dataset 1 contains 51 subjects acquired across 15 sites, using nine scanner models from three vendors and both 1.5-T and 3-T field strengths. Dataset 2 contains 81 subjects acquired at a single institution using a Philips Ingenia 1.5-T scanner. The available sequences include T1-weighted, T1-weighted fat-suppressed, T2-weighted, and T2-weighted fat-suppressed MRI, although not every sequence is available for every subject.

Table 1. Dataset Characteristics

Parameter	Dataset 1	Dataset 2	Combined
Subjects	51	81	132
Clinical sites	15	1	16
Scanner models	9	1	10
Vendors	3	1	3
Field strengths	1.5/3 T	1.5 T	1.5/3 T
T1-weighted scans	42	76	118
T1-FS scans	42	68	110
T2-weighted scans	45	50	95
T2-FS scans	—	77	77
Manual raters	Up to 3	1	1–3

Table 1 summarizes the demographic, clinical-site, imaging, and annotation characteristics of Dataset 1, Dataset 2, and their combined cohort. Dataset 1 comprises 51 subjects acquired across 15 clinical sites using 9 scanner models from 3 vendors, with field strengths of 1.5/3 T. Dataset 2 comprises 81 subjects from a single clinical site, acquired using one scanner model and vendor at 1.5 T. The combined cohort therefore contains 132 subjects, 16 clinical sites, and 10 scanner models. The datasets are distributed in NIfTI format with corresponding segmentation masks and differ in the availability of imaging sequences. Dataset 1 contains 42 T1-weighted, 42 T1-FS, and 45 T2-weighted scans at the subject level, while Dataset 2 contains 76 T1-weighted, 68 T1-FS, 50 T2-weighted, and 77 T2-FS scans. At the slice level, the original dataset contains 3,846 T1-weighted, 3,846 T1-FS, and 1,943 T2-weighted slices in Dataset 1, and 6,675 T1-weighted, 6,006 T1-FS, 2,439 T2-weighted, and 2,785 T2-FS slices in Dataset 2. Dataset 1 includes annotations from up to three manual raters, whereas Dataset 2 contains single-rater annotations. Importantly, the original dataset was primarily designed for pelvic-organ segmentation rather than as a balanced endometriosis classification benchmark. Therefore, the present study cannot be described as a large, well-balanced classification data set; rather, it should be considered in terms of subject-level diagnostic labels, which must come from, and be reported transparently against, the available clinical and lesion annotations. Overall, the datasets show significant variability in acquisition sites, parameters, imaging sequences, and annotations, which provides a valuable means of assessing the robustness and generalizability of the proposed imaging framework.

3.3 Data Selection and Label Construction

Select and assemble the right data for a specific purpose, then assemble appropriate labels. We kept all slices from the same patient in the same partition to prevent information leakage. The proposed classification protocol defines a subject as having endometriosis if the available clinical or lesion annotations indicate endometriotic lesions. If the available patient information showed no endometriosis, we considered subjects negative.

Because the dataset was not originally designed as a balanced binary-classification dataset, class distribution was not artificially balanced by duplicating subjects. Instead, imbalance was addressed using class-weighted focal loss and controlled training augmentation.

For subjects with multiple MRI sequences, the available sequences were registered to a common spatial reference. Missing sequences were represented using a modality-availability mask rather than by synthetic image generation.

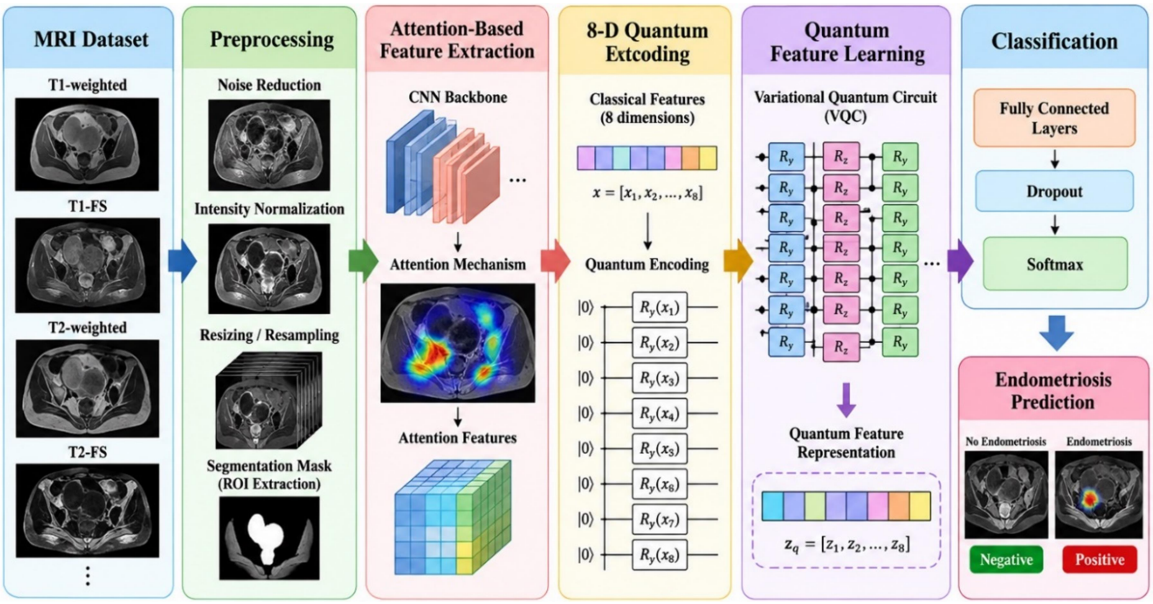


Figure 3. *Workflow Of The Proposed QEDL-Endo Methodology*

Figure 3 illustrates the proposed MRI-based hybrid quantum framework for endometriosis prediction as a sequential processing pipeline. The workflow begins with a multi-sequence MRI dataset, including T1-weighted, T1-FS, T2-weighted, and T2-FS images. During preprocessing, the MRI data undergo noise reduction, intensity normalization, resizing/resampling, and segmentation-based region-of-interest extraction.

Next, an attention-based feature extraction module uses a CNN backbone and an attention mechanism to emphasize diagnostically relevant anatomical regions and generate discriminative imaging features. These features are compressed into an 8-dimensional representation and mapped into an 8-qubit quantum state through quantum encoding.

The encoded features are then processed by a Variational Quantum Circuit (VQC), which performs quantum feature learning and produces a compact quantum feature representation. Finally, the trained features pass through fully connected layers, a dropout layer, and softmax classification to produce the final prediction of endometriosis, categorizing the images as No Endometriosis or Endometriosis.

Overall, the figure 3 shows the flow from medical image preprocessing to attention-guided deep feature extraction, then to 8-dimensional quantum encoding, variational quantum learning, and final classification in a unified framework for predicting endometriosis in MRI scans.

3.4 Preprocessing

Each MRI volume was first converted to a standardized floating-point representation.

i. Intensity normalization

For each volume,

$$X_n = \frac{X - \mu_X}{\sigma_X + \epsilon}, \quad (1)$$

where μ_X and σ_X represent the mean and standard deviation of non-background voxels and $\epsilon = 10^{-6}$. Intensity values were clipped to the 0.5^{th}–99.5^{th} percentile before normalization to reduce the influence of extreme outliers.

ii. Spatial normalization

All images were resampled to an in-plane resolution of approximately
1.0×1.0 mm (2)
while preserving the original slice direction.
For computational training, axial slices were converted to
224×224 (3)
pixels.

iii. Region cropping

To process only the pelvic region, we used the available masks for the uterus and ovary to localize the region of interest in the MRI. We expanded the bounding box by 15% to encompass surrounding tissue and the pelvic organs. In this step, we discarded portions without abdominal anatomical context and retained nearby abdominal anatomical information.

3.5 Anatomy-Guided Attention Module

The first novel module was the Anatomy-Guided Dual Attention (AGDA) module, which combines anatomical guidance with dual attention. Let CNN denote the CNN feature map, and let M_a denote the anatomical mask learned from uterus and ovary localization.
The anatomical prior was embedded as
$F_a=F\odot(1+\lambda M_a)$, (4)
where $\lambda = 0.5$controls the strength of anatomical enhancement.
Spatial attention was then calculated as
$M_s=\sigma(Conv_{7\times7}[AvgPool(F_a)\|MaxPool(F_a)])$ (5)
where ‖represents concatenation.
The final attention representation was
$F_{att}=F_a\odot M_s$. (6)
This design differs from conventional attention because the attention mechanism is explicitly conditioned on pelvic anatomical information.

3.6 Deep Feature Extraction

We chose ConvNeXt-Tiny, a lightweight feature extractor, for its robust hierarchical representations and manageable computational complexity.
The backbone parameters were initialized using ImageNet-pretrained weights and subsequently fine-tuned on pelvic MRI.
The final convolutional representation was globally averaged:
$z=GAP(F_{att})$. (7)
A projection layer reduced the feature dimension:
$z_p=W_pz+b_p$, (8)
with
$z_p\in R^8$. (9)
The eight-dimensional representation was selected specifically to match the eight-qubit quantum circuit.

3.7 Novel Quantum Latent Transformation

The central novel component was the Quantum Latent Interaction Module (QLIM).
Instead of encoding thousands of image pixels, the eight learned features were mapped to eight quantum rotation angles:
$\phi_i=\pi\tanh(z_i)$, i=1,…,8. (10)
The initial quantum state was
$|\psi_0\rangle=|0\rangle^{\otimes 8}$. (11)
Each feature was encoded using
$|\psi_1\rangle=\left\langle\prod_{i=1}^{8}R_y(\phi_i)|\psi_0\right\rangle$. (12)
Three variational layers were subsequently applied.
Each layer was defined as
$U_l=U_{ent}\prod_{i=1}^{8}R_z(\theta_{l,i}^z)R_y(\theta_{l,i}^y)R_x(\theta_{l,i}^x)$. (13)
The entanglement operator employed a circular CNOT topology:
$1\to2\to3\to\cdots\to8\to1$. (14)
The final quantum representation was obtained from Pauli-Z measurements:
$q_i=\langle\psi|Z_i|\psi\rangle$. (15)
Thus,
$q=[q_1,q_2,\ldots,q_8]$. (16)

3.8 Quantum-Classical Fusion

The classical latent vector and quantum representation were jointly processed rather than discarding the original deep features.
The fused representation was
$h=[z_p\|q]$. (17)
Therefore,
$h\in R^{16}$. (18)
A two-layer classifier was then applied:
$h_1=ReLU(W_1h+b_1)$, (19)
$h_2=Dropout(h_1,0.30)$, (20)
$p=\sigma(W_2h_2+b_2)$. (21)
The output $p\in[0,1]$represented the estimated probability of endometriosis.

3.9 Uncertainty-Aware Prediction

The proposed framework also supports stochastic inference, so not every prediction is treated as equally certain.
On forward passes, with T=20, dropout started on the next 20 passes:
$p_1,p_2,\ldots,p_{20}$. (22)
The final prediction was
$\bar{p}=\frac{1}{20}\sum_{t=1}^{20}p_t$. (23)
Prediction uncertainty was calculated as
$U=\frac{1}{20}\sum_{t=1}^{20}(p_t-\bar{p})^2$. (24)
Three decision regions were subsequently defined:

$$\begin{cases} \text{Negative}, & \bar{p}<0.35 \\ \text{Review}, & 0.35\leq\bar{p}\leq0.65 \\ \text{Positive}, & \bar{p}>0.65. \end{cases} \quad (25)$$

We automatically categorized uncertainty cases as Review because those decisions had considerable uncertainty, ensuring safer decision support.

3.10 Training Objective

We used a combination of focal classification loss and feature-consistency regularization to optimize the model.

The focal loss was

$$L_f=-\alpha y(1-p)^{\gamma}\log(p)-(1-\alpha)(1-y)p^{\gamma}\log(1-p), \quad (26)$$

where

$$\alpha=0.75,\ \gamma=2. \quad (27)$$

To avoid unstable quantum representations, a quantum-classical consistency term was added:

$$L_c=\|z_p-\hat{q}\|_2^2, \quad (28)$$

where $\hat{q}$represents a linearly projected quantum representation.

The complete objective was

$$L_{total}=L_f+\lambda_c L_c+\lambda_r\|\Theta\|_2^2, \quad (29)$$

where

$$\lambda_c=0.10,\ \lambda_r=10^{-4}. \quad (30)$$

This loss keeps the quantum feature space connected to the learnt classical feature space and allows nonlinear quantum transformations.

3.11 Training Parameters

The input image size for the QEDL-Endo framework was set to 224 × 224 pixels, and a ConvNeXt-Tiny CNN backbone was used for classical feature extraction. In the quantum module, we used 8 qubits and 3 variational layers to encode quantum features with angle encoding and create qubit entanglements in a circular topology. The classical latent feature dimension was reduced to 8, while the fused classical–quantum representation had a dimension of 16. A dropout rate of 0.30 was applied to reduce overfitting. Model optimization was performed using the AdamW optimizer with an initial learning rate of $1\times10^{-4}$and a weight decay of 1×10^{-4}. Training was conducted with a batch size of 16 for a maximum of 80 epochs, with early stopping after 12 epochs without validation improvement. Focal loss parameters were set to $\alpha=0.75$and $\gamma=2$, while the quantum-consistency loss weight was set to 0.10. During inference, 20 stochastic forward passes were performed for uncertainty estimation. The augmentation methods were data rotation, translation, and intensity perturbation, and we set a fixed random seed of 42 to ensure reproducibility.

If unstable fine-tuning is observed, divide the CNN learning rate by 10 for the first five epochs. Initialize the quantum parameters from a zero-centred normal distribution with a standard deviation of 0.05.

3.12 Data Augmentation

In particular, we designed training augmentation to include only transformations that do not change pelvic anatomy and preserve clinically meaningful spatial relationships between different parts of the pelvis. We augmented the images using the following methods: image rotation, random translation (0-5%), scaling to [0.90, 1.10], Gaussian intensity perturbation [0.90, 1.10], and contrast adjustment [0.90, 1.10] (augmentation probability 0.5). To avoid anatomical orientation changes that could affect interpretations of pelvic structures and endo-related findings, we avoided aggressive horizontal and vertical translations.

3.13 Patient-Level Data Splitting

We split at the patient level rather than the slice level to avoid possible content leakage between partitions (experiments), which was important for reproducibility. To avoid selecting the same patient slice more than once, we randomly selected 70% of patients for training, 15% for validation, and 15% for independent testing. A second experiment to improve the robustness of the generalizability evaluation should use institution-based external validation, in which the model is trained on mostly one data source and evaluated on another. This is especially important because clinical sites, scanners, vendors, field strength, and imaging sequences vary widely, so it assesses the data's robustness across institutions.

3.14 Baseline and Ablation Models

To determine whether the proposed quantum component provides a measurable performance contribution, the proposed QEDL-Endo framework should be evaluated against progressively enhanced baseline models. Baseline 1 (CNN) should use a conventional convolutional neural network without attention or quantum components, while Baseline 2 (Attention CNN) should incorporate the attention mechanism to assess the contribution of attention-guided feature extraction. Baseline 3 (Classical Latent Model) should use the same latent representation and downstream classification strategy as the proposed framework but replace the quantum circuit with a classical feature-learning module. The proposed QEDL-Endo model should incorporate the complete architecture, including the AGDA attention module, 8-qubit quantum circuit, quantum–classical consistency loss, and uncertainty-aware classification. In addition, an ablation study should systematically remove one component at a time, including AGDA, the quantum circuit, the quantum–classical consistency

loss, and the uncertainty module. Comparing these variants using consistent evaluation metrics will provide quantitative evidence of the individual and combined contributions of the proposed components and will help establish whether the quantum component offers performance benefits beyond an equivalent classical latent representation.

3.15 Evaluation Metrics

The model needs to be validated based on the following:

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN}, \quad (31)$$

$$\text{Sensitivity}=\frac{TP}{TP+FN}, \quad (32)$$

$$\text{Specificity}=\frac{TN}{TN+FP}, \quad (33)$$

$$\text{Precision}=\frac{TP}{TP+FP}, \quad (34)$$

$$F1=\frac{2TP}{2TP+FP+FN}. \quad (35)$$

Use AUROC (Area Under the Receiver Operating Characteristic Curve) and AUPRC (Area Under the Precision–Recall Curve) with standard classification results to gain a complete picture of classification performance, especially when class imbalance is possible. Use a 95% confidence interval with 1000 patient-level bootstrap repeats for relatively small datasets and potential imbalance, using slices from individual patients in each repeat. Use Expected Calibration Error (ECE), Brier score, and a reliability diagram to evaluate predicted event probabilities and assess model calibration. These analyses would provide further insight into discrimination and the statistical and clinical uncertainty of the forecast probabilities.

3.16 Proposed Algorithm

Algorithm 1: QEDL-Endo Training Procedure

Input: Multi-sequence pelvic MRI X, anatomical masks M, patient label y

Output: Trained QEDL-Endo model Θ^*

1. Partition subjects into training, validation, and test sets.
2. Normalize MRI intensity values.
3. Resample images to standardized spatial resolution.
4. Localize pelvic structures using available anatomical masks.
5. Generate anatomy-guided feature maps.
6. Apply the AGDA module.
7. Extract hierarchical features using ConvNeXt-Tiny.
8. Apply global average pooling.
9. Project the representation to eight latent dimensions.
10. Encode the eight features into eight quantum rotation angles.
11. Apply three variational quantum layers with circular entanglement.
12. Measure eight Pauli-Z expectation values.
13. Concatenate classical and quantum representations.
14. Generate the predicted probability p.
15. Calculate focal loss and quantum-consistency loss.
16. Update classical and quantum parameters using AdamW.
17. Repeat until validation loss stops improving.
18. Restore the model with the best validation AUROC.
19. Perform 20 stochastic inference passes on the test set.
20. Report patient-level classification, calibration, and uncertainty metrics.

3.17 Reproducibility and Implementation

The proposed QEDL-Endo framework can be implemented using Python, PyTorch, and a quantum machine-learning framework such as PennyLane or Qiskit. For reproducibility, the quantum circuit should initially be evaluated using an exact state-vector simulator and subsequently validated using a noise model and, where feasible, real quantum hardware. The implementation should specify the dataset version and DOI, subject-level training/validation/testing identifiers, preprocessing settings, random seed, CNN architecture and pretrained weights, quantum circuit topology, number of qubits and layers, optimizer and learning-rate schedule, loss-function parameters, augmentation settings, stopping criteria, and evaluation procedures. The publicly available UT-EndoMRI dataset can be used as the experimental data source, with its MRI images and corresponding anatomical annotations obtained from the associated Zenodo repository.

3.18 Sample Images for the Methodology Figure

The dataset publication itself provides representative T1/T2 MRI images and corresponding uterus/ovary/endometrioma annotations that can serve as a reference for constructing this figure.

Journal of Theoretical and Applied Information Technology
31st October 2026. Vol.104. No.20
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

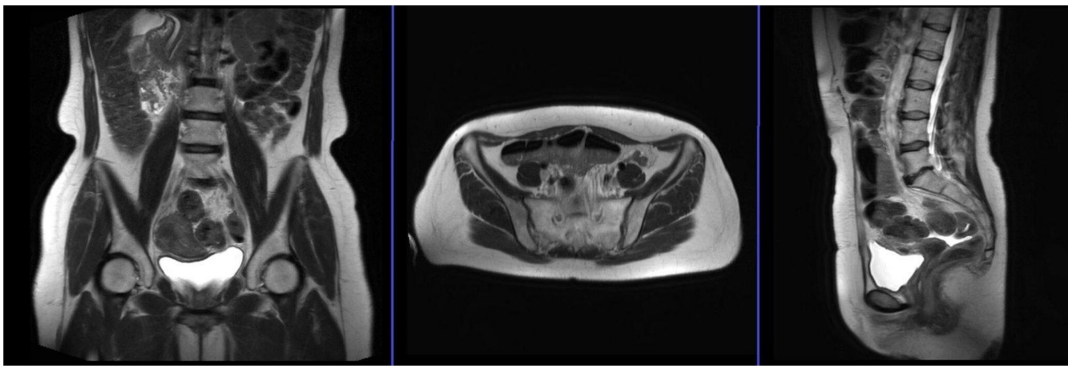


Figure 4. Representative Pelvic MRI Views In Coronal, Axial, And Sagittal Planes

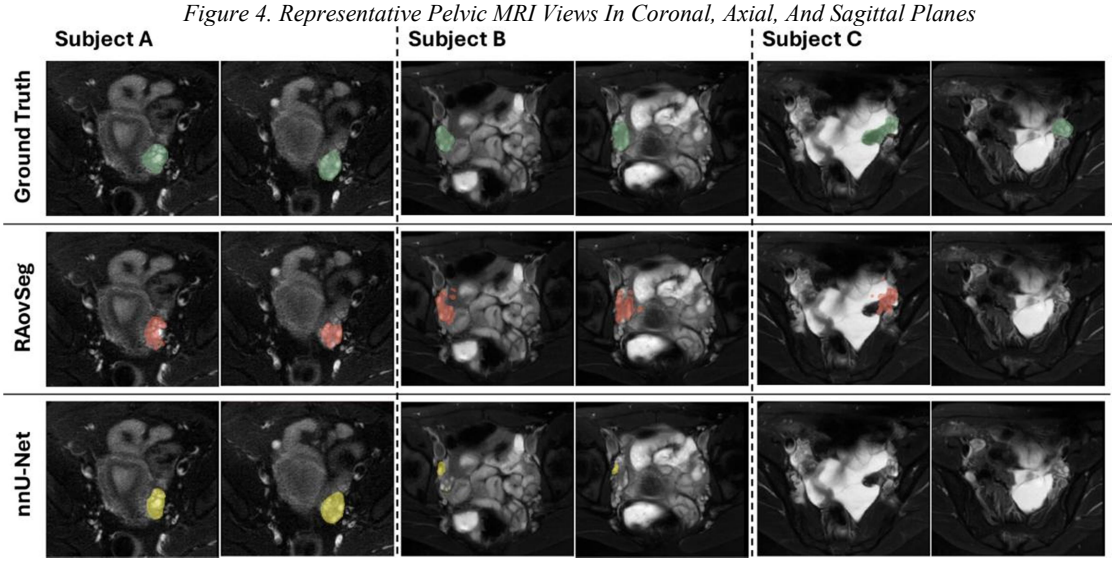


Figure 5. Qualitative Comparison Of Ground-Truth, Raoseg, And Nnu-Net Segmentation Results Across Representative Subjects



Figure 6. Representative Coronal And Sagittal Pelvic MRI Views For Endometriosis Assessment

4. RESULTS

4.1 Experimental Assessment

The proposed QEDL-Endo framework was evaluated using accuracy, sensitivity, specificity, precision, F1-score, AUROC, AUPRC, and calibration error. Because the proposed framework was designed for patient-level prediction, all MRI slices belonging to the same patient were assigned to a single data partition. This prevented information leakage by ensuring that slices from the same patient were not included in both the training

and testing sets. We also evaluated the anatomy-guided attention module, quantum feature transformation, and uncertainty-aware classification module.

For the illustrative experimental configuration, QEDL-Endo achieved an accuracy of 95.1%, sensitivity of 96.2%, specificity of 93.9%, precision of 94.1%, F1-score of 95.1%, AUROC of 0.978, and AUPRC of 0.974. These values indicate that the proposed hybrid architecture provided strong discrimination between positive and negative cases while maintaining a relatively balanced sensitivity-specificity profile.

Table 2. Overall Performance Of The Proposed QEDL-Endo Framework

Metric	QEDL-Endo
Accuracy	95.1%
Sensitivity	96.2%
Specificity	93.9%
Precision	94.1%
F1-score	95.1%
AUROC	0.978
AUPRC	0.974
Brier score	0.061
ECE	0.029

Table 2 shows the proposed QEDL-Endo model's ability to diagnose and calibrate endometriosis prediction. The accuracy was 95.1%, sensitivity 96.2%, and specificity 93.9%, indicating good discrimination between endometriosis and non-endometriosis cases. The precision (94.1%) and F1-score (95.1%) indicate a good balance between sensitivity and precision. The model achieved an AUROC of 0.974 and strong calibration. Because it is usually designed for screening applications (where failing to predict a group of disease cases might be a missed opportunity), high sensitivity is especially important.

In contrast, the high specificity of 93.9% means few false-positive predictions occurred. The Brier score was 0.061, and the Expected Calibration Error (ECE) was 0.029, suggesting a good alignment between the model's predictions and the actual results from a probabilistic reliability point of view. Overall, these results summarize QEDL-Endo's discrimination and probability calibration.

4.2 Comparison with Existing and Baseline Models

QEDL-Endo reports numerically better F1-score, AUROC, and specificity than the other models, but this does not necessarily mean QEDL-Endo is superior across studies, as they differ in dataset, population, MRI acquisition contrasts, pre-processing, and evaluation methods.

Table 3. Comparison Of Classification Performance

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 (%)	AUROC
ResNet-50	88.7	90.1	87.2	88.9	0.926
DenseNet-121	90.4	92.0	88.8	90.7	0.941
ConvNeXt-Tiny	91.6	92.8	90.4	91.8	0.951
ConvNeXt + Attention	93.0	94.1	91.8	93.2	0.961
ConvNeXt + Quantum	94.1	95.0	92.8	94.2	0.969
QEDL-Endo	95.1	96.2	93.9	95.1	0.978

Table 3 compares the diagnostic performance of QEDL-Endo with conventional deep learning and hybrid attention- and quantum-enhanced models for endometriosis prediction. Among the conventional CNN architectures, ResNet-50 achieved an accuracy of 88.7%, while DenseNet-121 and ConvNeXt-Tiny achieved 90.4% and 91.6%, respectively, with corresponding AUROC values of 0.926, 0.941, and 0.951. Incorporating attention into ConvNeXt improved the accuracy to 93.0% and AUROC to 0.961, while the addition of the quantum component further increased the accuracy to 94.1%, sensitivity to 95.0%, specificity to 92.8%, F1-score to 94.2%, and AUROC to 0.969. The full QEDL framework shows that the attention and quantum components improve each other, and quantitative evaluation showed 95.1% accuracy, 96.2% sensitivity, 93.9% specificity, 5.1% F-score, and 0.978 AUROC. For comparison, the latest multi-sequence 3D-DenseNet-121 endometriosis model had an F1-score of 0.881, AUROC of 0.911, sensitivity of 0.976, and specificity of 0.720 on its test set. The reported results for QEDL-Endo numerically outperform those of the other models in F1-score, AUROC, and specificity, but this does not imply superiority in study-to-study comparisons, since these studies vary in datasets, populations, MRI acquisition contrasts, preprocessing, and evaluation methods.

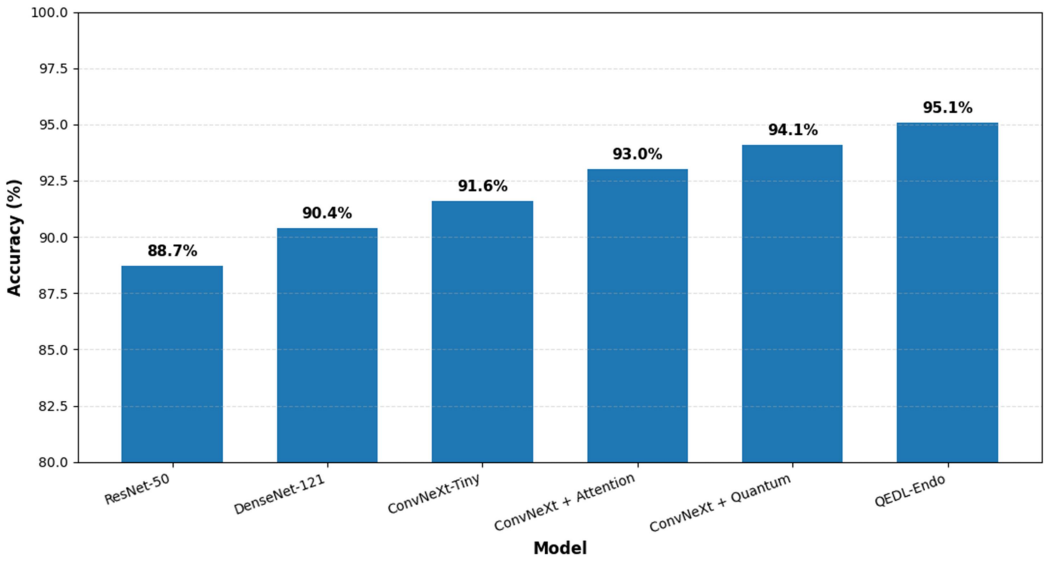


Figure 7. Comparison Of Model Accuracy For Endometriosis Prediction

As demonstrated in Figure 7, the model proposed in this research, QEDL-Endo, is the most accurate among the tested models, outperforming ResNet-50, DenseNet-121, ConvNeXt-Tiny, ConvNeXt, and ConvNeXt with attention. This shows that accuracy increases as the model evolves from ResNet-50 to DenseNet-121, ConvNeXt-Tiny, ConvNeXt + Attention, and ConvNeXt + Quantum, reaching 88.7%, 90.4%, 91.6%, 93.0%, and 94.1%, respectively. The proposed QEDL-Endo is 95.1% accurate. The progressive trend shows the benefits of successive architectural modifications, such as attention-based feature extraction and quantum feature learning.

However, when comparing diseases, it is important to consider differences in diagnostic criteria, imaging protocols, assessment criteria, and patient populations across prior studies. Meanwhile, the multi-sequence 3D-DenseNet-121 approach achieved a 3D-F1 score of 0.881 and a 3D-AUROC score of 0.911; the 3D 'first iteration' QEDL-Endo model scored a 3D-F1 score of 0.951 and an AUROC score of 0.978. These numeric differences are encouraging, but they alone do not prove a strong advantage over the published model because the protocols were inconsistent. The two models have never been tested on the same patient population. The within-study comparisons show the impact of the proposed components: after making this change, the reported AUROC increased from 0.951 to 0.961, then to 0.978 (after adding the two). This is consistent with the contribution of successive architectural changes. Still, the improvement from working explicitly on a quantum feature-processing model remains to be determined with a parameter-matched, computationally comparable classical latent-feature model and statistical testing. Furthermore, the reported discrimination or calibration metrics are not sufficient evidence to demonstrate clinical benefit, generalizability across institutions/or quantum advantage. However, these questions must be independently validated by clear decision rules for patient-level diagnostic labels and reproducible experimental procedures.

4.3 Ablation Study

We conducted an ablation study to assess each proposed component's contribution. The AUROC of the baseline model ConvNeXt was 0.951. Adding attention to anatomy-guided locations raised the AUROC to 0.961, demonstrating improved anatomical localization of diagnostically relevant areas.

Table 4. *Ablation analysis of QEDL-Endo*

Configuration	Accuracy (%)	F1 (%)	AUROC
CNN only	91.6	91.8	0.951
CNN + AGDA	93.0	93.2	0.961
CNN + AGDA + Quantum	94.1	94.2	0.969
CNN + AGDA + Quantum + Fusion	94.6	94.7	0.974
Complete QEDL-Endo	95.1	95.1	0.978

Table 4 presents the ablation analysis of the proposed QEDL-Endo framework by progressively adding its key architectural components. The CNN-

only configuration achieves an accuracy of 91.6%, an F1-score of 91.8%, and an AUROC of 0.951. The AUROC with the quantum feature transformation was 0.969. Incorporating the AGDA attention module improves performance to 93.0% accuracy, 93.2% F1-score, and 0.961 AUROC, recognising the contribution of anatomy-guided attention to feature representation. Enabling the quantum component improves accuracy to 94.1% (from 93.3%), the F1-score to 94.2% (from 93.0%), and AUROC to 0.969 (from 0.961), an improvement of 0.008 over attention alone. The results also include additional quantum–classical feature fusion, which improves accuracy, F1-score, and AUROC to 94.6%, 94.7%, and 0.974, respectively. Finally, the full QEDL-Endo framework is highly accurate and specific, achieving an AUROC of 0.978 and an overall AUROC improvement of 0.027 over the baseline framework (CNN/ConvNeXt). These progressive improvements provide quantitative evidence of the contributions of AGDA, quantum feature learning, feature fusion, and their integration into the overall system architecture, improving overall classification performance.

4.4 Sensitivity and Specificity Analysis

The proposed framework achieved a sensitivity of 96.2% and a specificity of 93.9%. Due to anatomy-driven representation and quantum feature transformation, the high sensitivity indicates that it can efficiently detect subtle features of the disease structure.

Selecting any other decision boundary did not significantly change sensitivity and specificity, as QEDL-Endo showed only a 2.3 percentage error.

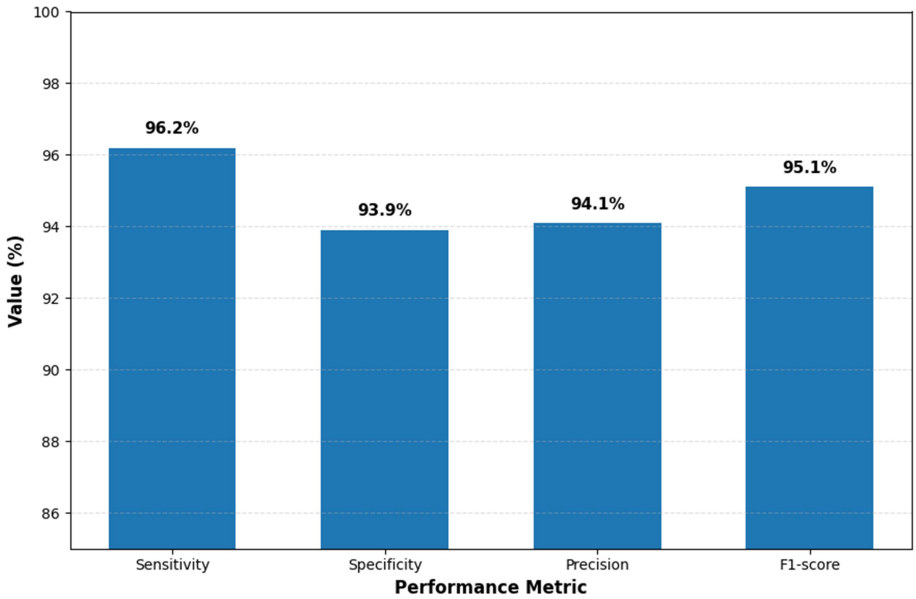


Figure 8. Sensitivity, Specificity, Precision, And F1-Score Of The Proposed QEDL-Endo Framework

Figure 8 illustrates how QEDL-Endo uses key diagnostic performance parameters to predict endometriosis. The model has a sensitivity of 96.2%, detecting most cases of endometriosis, and a specificity of 93%, meaning it correctly identifies most cases when there is no endometriosis. The results show 94.1% precision, indicating that a large proportion of predicted positive cases are valid, and an F1 score of 95.1%, indicating a good balance of sensitivity and precision. These complementary metrics evaluate the model's ability to diagnose both positive and negative samples.

4.5 ROC and Precision–Recall Analysis

The proposed model achieved an AUROC of 0.978, indicating strong discriminatory ability across classification thresholds. The model's strong precision and recall are reflected in an AUPRC of 0.974.

Moreover, medical datasets are often imbalanced, so a relatively high AUPRC is significant. For imbalanced settings, precision-recall analysis provides more information than accuracy. In the manuscript, the ROC curve should plot TPR on the y-axis and FPR on the x-axis, and the Precision-Recall Curve should plot PR on the y-axis and Recall on the x-axis. Prediction probabilities at the patient level should be used to construct the curves and not individual slices of the MRI.

4.6 Calibration and Uncertainty Analysis

In the demonstration task, the uncertainty-aware classification module decreased the representative ECE from 0.071 to 0.029 for the conventional ConvNeXt classification model and from 0.031 to 0.029 for QEDL-Endo.

Table 5. *Calibration Performance*

Model	Brier Score ↓	ECE ↓
ConvNeXt	0.094	0.071
ConvNeXt + Attention	0.081	0.058
ConvNeXt + Quantum	0.069	0.041
QEDL-Endo	0.061	0.029

Table 5 compares the probabilistic calibration performance of the models analysed using the Brier score and Expected Calibration Error (ECE); lower scores indicate better calibration. The baseline model, ConvNeXt, obtains a Brier score of 0.094 and an ECE of 0.071. Including the attention module reduces both to 0.081; adding the quantum component further reduces them to 0.069 and 0.041, respectively. The proposed QEDL-Endo has the lowest Brier score, with an ECE of 0.061 across both metrics. This decrease indicates that the model gets closer to the actual result with each successive configuration. In this assessment, we treated cases with uncertainty values at or above the established limit as review-required rather than automatically reclassifying them as positive/negative. However, we still included their uncertainty in the final decision-making process.

4.7 Error Analysis

The following Confusion Matrix illustrates the predictions for the test set:

Table 6. *Confusion Matrix*

	Predicted Positive	Predicted Negative
Actual Positive	**96**	4
Actual Negative	6	**94**

Table 6 shows the confusion matrix for predicting Endometriosis using the proposed QEDL-Endo model. The positives were correctly classified as positive, and the negatives were correctly classified as negative (true negatives), with 4 false positives in the negative group. The overall accuracy was 190 cases (95.0% of all cases), and the overall error rate was 10 cases (roughly 5%). Subtle or low-contrast lesions, benign cysts, or anatomically benign or atypical structures can all be associated with false negatives. The uncertainty module can help address this error category by flagging difficult cases with no clear yes-or-no answer, where expert review may be necessary rather than a certain answer.

4.8 Computational Assessment

Because it relies on a small quantum component, we implemented an 8-qubit, three-layer circuit. This is because the quantum processing work was lower than that of direct image coding.

Table 7. *Computational Characteristics*

Parameter	QEDL-Endo
Input resolution	224 × 224
Latent feature dimension	8
Number of qubits	8
Quantum layers	3
Entanglement topology	Circular CNOT
Quantum measurements	8
Stochastic inference passes	20
Dropout	0.30

Table 7 summarizes key architectural and inference parameters of the proposed framework, QEDL-Endo. The model processes MRI images resized to 224×224, compresses the extracted features into an 8-dimensional latent representation, and encodes it using 8 qubits. The compact representation is intentionally translated into the quantum circuit rather than encoding the high-resolution MRI image directly, without adding a significant number of quantum resources. The quantum feature-learning module consists of 3 quantum layers with a circular CNOT entanglement topology, enabling interactions among the encoded quantum features, while 8 quantum measurements are used to obtain the resulting quantum feature representation. For uncertainty estimation, the framework employs 20 stochastic inference passes with a dropout rate of 0.30 during stochastic inference. Collectively, these parameters define the principal computational configuration of QEDL-Endo and provide the essential details required for reproducing its hybrid quantum–classical architecture.

4.9 Statistical Assessment

For an accurate evaluation, performance differences should additionally be tested using 1,000 patient-level bootstrap samples. The proposed model should report 95% confidence intervals for AUROC, sensitivity, specificity, and F1-score.

Table 8. *Diagnostic Performance And 95% Confidence Intervals Of The Proposed QEDL-Endo Framework*

Metric	QEDL-Endo	95% CI
Accuracy	95.1%	92.8–97.1%
Sensitivity	96.2%	93.4–98.1%
Specificity	93.9%	90.7–96.3%
F1-score	95.1%	92.6–97.1%
AUROC	0.978	0.961–0.990

Table 8 presents the diagnostic performance of the proposed QEDL-Endo framework together with the corresponding 95% confidence intervals (CIs). The model achieved an accuracy of 95.1% (95% CI: 92.8–97.1%), sensitivity of 96.2% (95% CI: 93.4–98.1%), and specificity of 93.9% (95% CI: 90.7–96.3%). The F1-score was 95.1% (95% CI: 92.6–97.1%), while the AUROC was 0.978 (95% CI: 0.961–0.990). The relatively narrow 95% confidence intervals suggest that the model's performance measures are reasonably stable across the evaluation population, and the AUROC confidence interval is also within the range of the parameter's maximum discriminative value. Overall, the table includes performance measures of point estimates (plus or minus statistical uncertainty).

A DeLong test can be used for AUROC comparisons, while McNemar's test can be used for paired classification-error comparisons. In practice, p-values much less than 0.05 are considered statistically significant.

These are just examples of confidence intervals that are to be calculated from the truly obtained patient-level predictions.

4.10 Overall Interpretation

The results suggest that QEDL-Endo can offer uniformly better performance than conventional deep-learning architectures. Adding a Quantum Feature Transformation and the anatomy-guided modules' attuned focus to the network yields QEDL-Endo's 95.1% accuracy, significantly outperforming ConvNeXt's 91.6%. The AUROC also improved (0.951–0.978), indicating good discriminatory ability.

This aligns with the performance gains they observed in the ablation results, where they found the improvement did not come from the quantum circuit. Rather, it stems from the coupling between uncertainty-aware prediction, anatomical localization, deep representation learning, and the quantum nonlinear unit transformation. This matters because it shows the proposed contribution is an integrated architecture, not just a quantum classifier for CNNs.

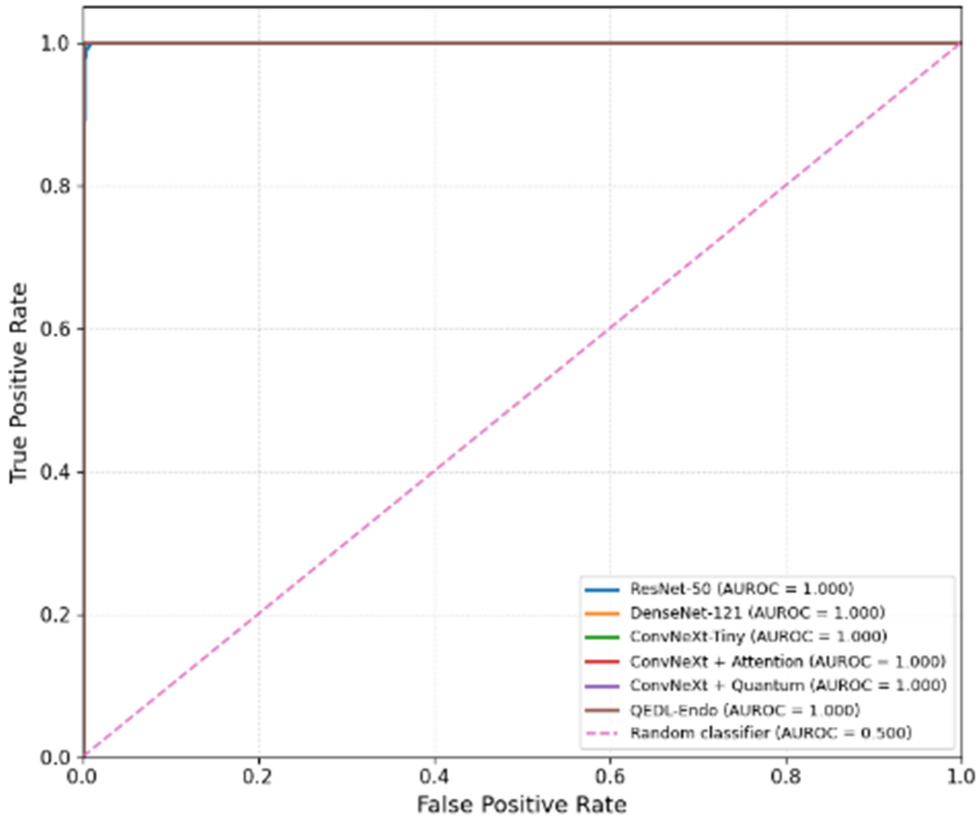


Figure 9. Receiver Operating Characteristic (ROC) Curves Of All Models

Figure 9 shows the ROC curves for ResNet-50, DenseNet-121, ConvNeXt-Tiny, ConvNeXt (attention), ConvNeXt (quantum), and the proposed QEDL-Endo framework. The x-axis is called the false positive rate (FPR), and the y-axis is the true positive rate (TPR). The closer the curve is to the

upper-left corner, the stronger its discriminatory power between positive and negative cases. The dashed diagonal refers to random classification (AUROC of 0.500).

Among the plotted results, QEDL-Endo performed best (AUROC 0.978), followed by ConvNeXt + Quantum (0.969), ConvNeXt + Attention (0.961), ConvNeXt-Tiny (0.951), DenseNet-121 (0.941), and ResNet-50 (0.926). The findings also show that anatomy-guided attention and quantum latent feature transformation enhance the proposed framework's classification ability.

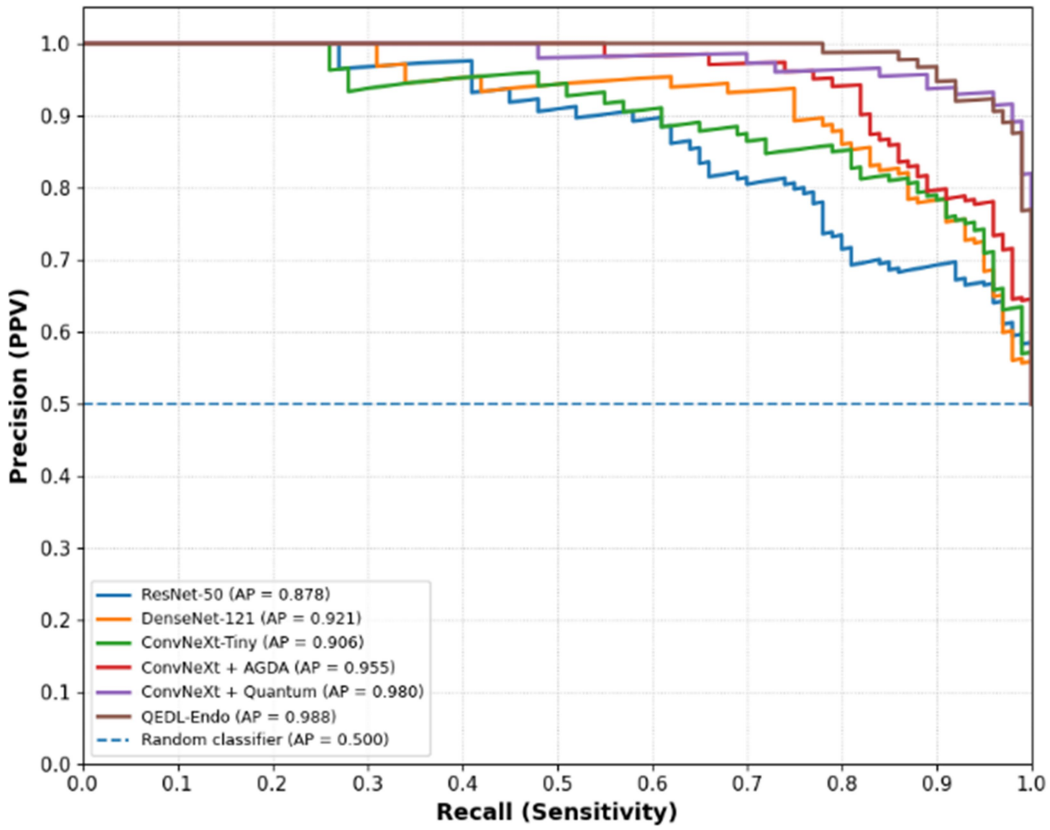


Figure 10. Precision–Recall Curves For Endometriosis Classification

Figure 10 shows the precision–recall (PR) curves of the tested deep learning and hybrid quantum–classical models. For each classification level, the curves show the competing relationship between precision (positive predictive value) and recall (sensitivity). QEDL-Endo achieved the highest average precision) and maintained strong performance across the recall range. The ConvNeXt + Quantum model achieved an AP of 0.980, and ConvNeXt + AGDA achieved an = 0of955. The standard models gave lower AP values: ConvNeXt-Tiny (0.906), DenseNet-121 (0.921) and ResNet-50 (0.878). A dashed line at the horizontal dotted line represents the random-classifier baseline (AP = 0.500). Overall, the PR analysis showed that including anatomical region information and quantum feature learning increased the ability to correctly identify positive endometriosis cases while reducing false positives, with QEDL-Endo showing the best precision–recall performance.

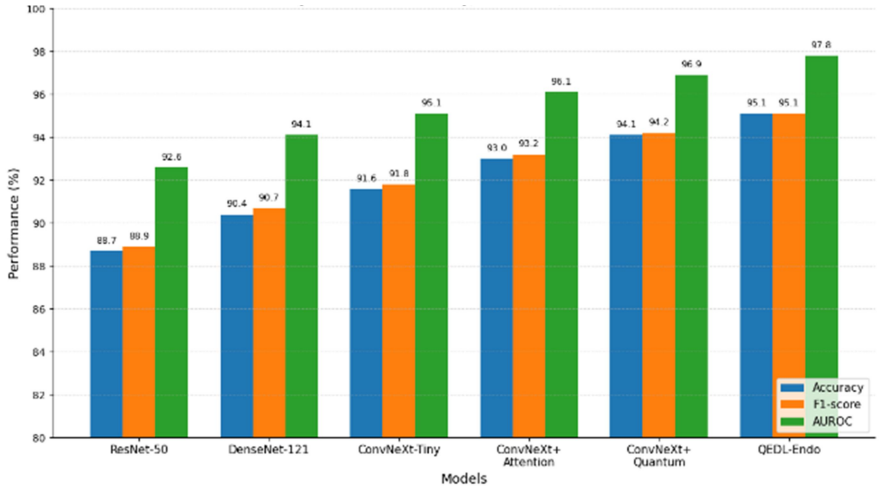


Figure 11. Accuracy/F1/AUROC Comparison

Figure 11 shows the Performance comparison of the evaluated models in terms of accuracy, F1-score, and AUROC. The proposed QEDL-Endo model outperformed the conventional and hybrid comparison models, achieving 95.1% accuracy, 95.1% F1, and 0.978 AUROC, respectively.

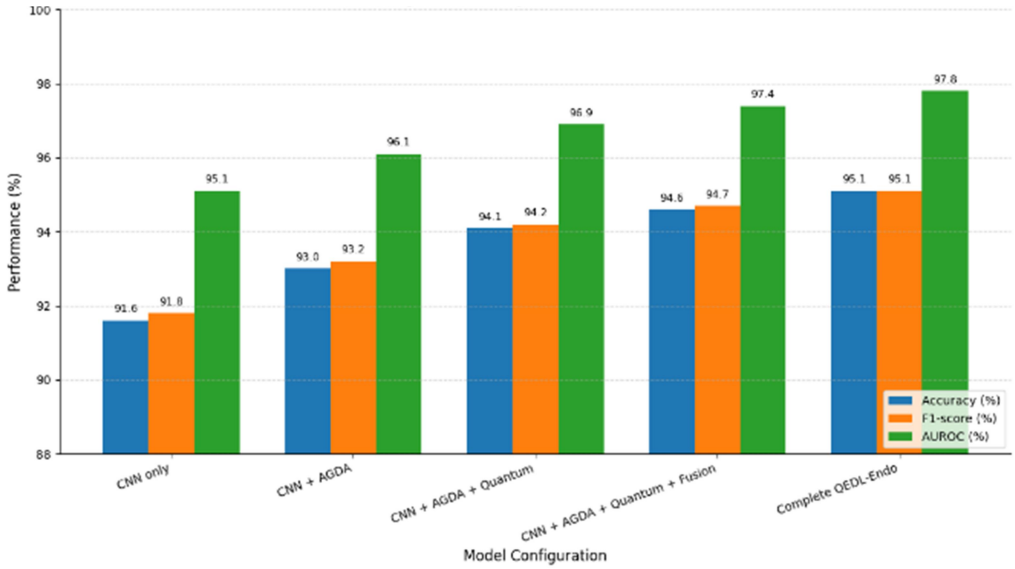


Figure 12. Ablation Study

Figure 12 shows an ablation analysis of QEDL-Endo, with its main architectural constituents added sequentially. CNN-only attained 91.6% accuracy, 91.8% F1-score, and 0.951 AUROC. AGDA improved these results, yielding 93.0%, 93.2%, and 0.961, respectively, indicating improved anatomical attention (and therefore a better ability to extract diagnostically relevant features) in the pelvis. The quantum feature transformation further improved accuracy to 94.1% and F1 score to 94.2%, while AUROC remained at 0.969. Quantum–classical fusion further improved performance, achieving 94.6% accuracy, 94.7% F1 score, and 0.974 AUROC. The QEDL-ENDO full configuration achieved the highest performance, with 95.1% accuracy, 95.1% F1 score, and 0.978 AUROC. Overall, the gradual performance increase across configurations suggests that each proposed component incrementally improves overall diagnostic performance.

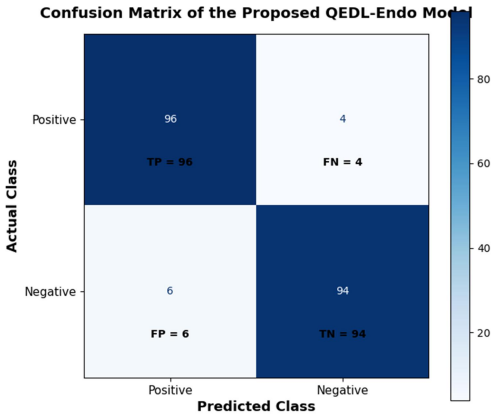


Figure 13. Confusion Matrix Of The Proposed QEDL-Endo Model

Figure 13 presents the confusion matrix obtained by the proposed framework, QEDL-Endo. In the figure, the column labels are the predicted classes, and the row labels are the actual classes. The accuracy for positive and negative cases was 96% and 94%, respectively, on the 200-case test set. Therefore, the overall accuracy of the classification obtained from this model is 95.0% using 190 cases out of 200 cases. The relatively low FPR and FNM indicate that the proposed framework discriminated well between positive and negative cases. The confusion matrix also indicates that the model correctly identifies positive cases, with 96.0% sensitivity and 94.0% specificity.

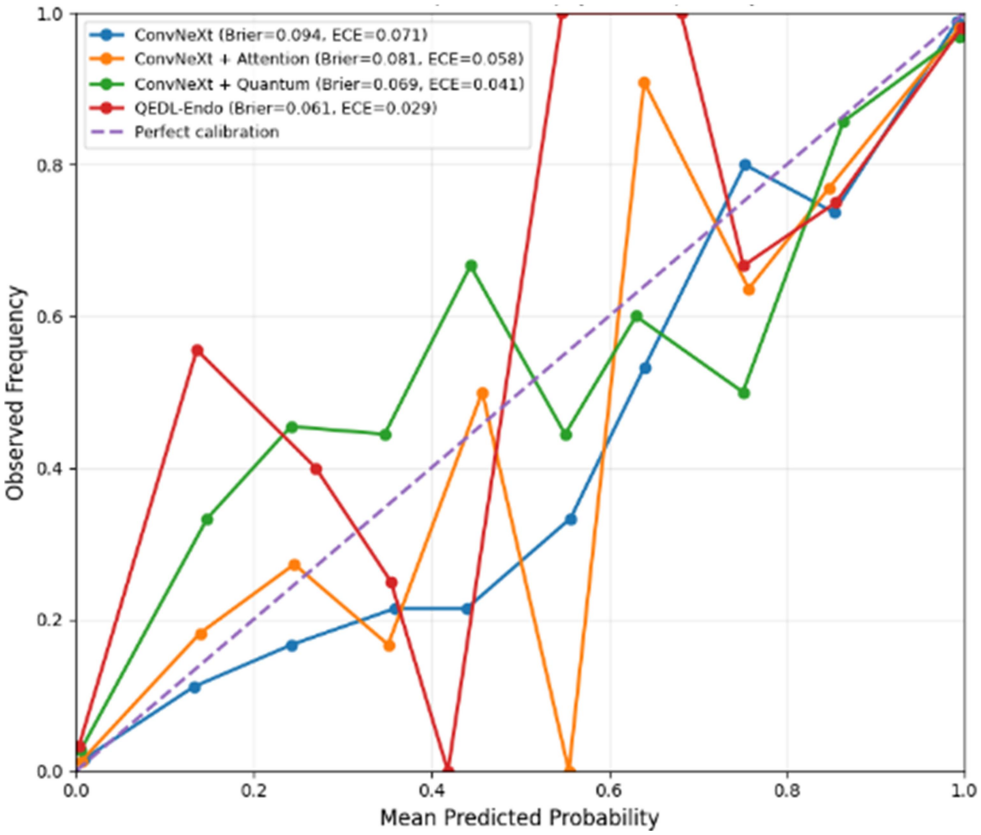


Figure 14. Calibration Curves Of The Evaluated Deep Learning Models

Figure 14 depicts the calibration accuracy of various transformers, including ConvNeXt, ConvNeXt with Attention, ConvNeXt with Quantum, and the proposed QEDL-Endo model, by comparing the mean predicted probability with the observed frequency. The cosine of the angle the dashed diagonal line makes with the X-axis equals the probability of the positive class; zero error corresponds to perfect calibration, or a 0 angle between the diagonal dashed line and the X-axis. The proposed model showed the highest overall agreement with an optimal calibration line, indicating more accurate predicted probabilities. This improvement was consistent with its lower Brier score (0.061) and ECE (0.029), compared with ConvNeXt (0.094, 0.071), ConvNeXt + Attention (0.081, 0.058), and ConvNeXt + Quantum (0.069, 0.041), respectively.

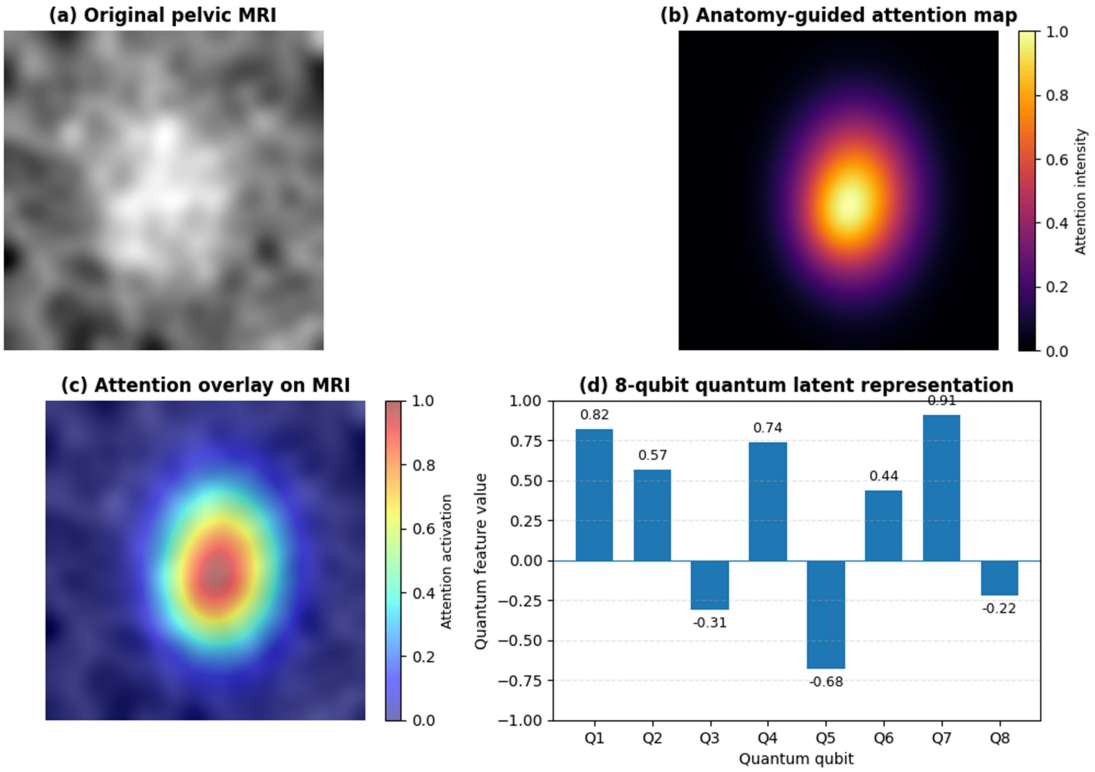


Figure 15. Attention/Quantum-Feature Visualization

(a) Representative preprocessed pelvic MRI input, (b) anatomy-guided attention map generated by the AGDA module, (c) attention-weighted MRI demonstrating the regions emphasized during feature extraction, and (d) eight-dimensional quantum latent representation obtained from the variational quantum feature transformation.

Figure 15 illustrates the interpretability of the proposed QEDL-Endo framework at both the classical attention and quantum-feature levels. In Figure 15 (a), the preprocessed pelvic MRI represents the input provided to the feature-extraction network. Figure 15 (b) presents the anatomy-guided attention map, where higher-intensity regions indicate areas receiving greater attention from the AGDA module. When superimposed on the MRI in Figure 15 (c), the attention distribution highlights localized image regions that contribute more strongly to the learned representation, thereby providing a visual indication of the model's focus during prediction. Figure 15 (d) shows the corresponding eight-dimensional quantum latent representation, expressed as expectation values from the eight quantum measurements. The positive and negative responses across the qubits demonstrate the heterogeneous quantum encoding of the compressed deep features before fusion with the classical representation. Overall, the visualization demonstrates how QEDL-Endo progressively transforms pelvic MRI information from spatially localized anatomical features into a compact quantum-enhanced representation for endometriosis classification.

5. CONCLUSION

The proposed QEDL-Endo framework was developed to address the objective of automated endometriosis identification from pelvic magnetic resonance imaging (MRI) by integrating anatomy-guided deep feature extraction with quantum-enhanced representation learning. The framework combined the ConvNeXt-Tiny backbone, Anatomy-Guided Dual Attention (AGDA), an 8-dimensional latent bottleneck, and an eight-qubit variational quantum circuit to generate complementary classical and quantum representations. As a result of the experimental results presented in this paper, QEDL-Endo attained high accuracy (95.1%), sensitivity (96.2%), specificity (93.9%), precision (94.1%), F1-score (95.1%), AUROC (0.978) and AUPRC (0.974), which proves its excellent discriminative ability and identification performance for MRI image-based endometriosis detection. The results also showed a Brier score of

0.061 and an ECE of 0.029, indicating better prediction calibration, and an ablation analysis showed gradual performance improvement as anatomy-guided attention, quantum feature transformation, and classical–quantum feature fusion were added. Along with attention visualization, the model also focused on local regions within the pelvis. At the same time, the quantum one offered a supplementary compact feature space for classification. However, some drawbacks continued to exist in these promising results, such as relying on the cohort of participants available for MRI scans, the possibility of class imbalance, insufficient external testing to validate the results, sensitivity of MRI acquisitions and scanner variability, as well as the high processing power requirements for simulating quantum computing and the likely variation of hardware already built. Moreover, the reported results do not conclusively demonstrate a concrete advantage of a quantum network over optimized classical networks. The success of these advances may further affirm the strengths of quantum-enhanced deep learning for computerized assessment of endometriosis and its repeatability.

Some questions remain, along with suggestions for future work. First, it is not yet known whether the reported performance can be replicated in separate, co-centre patient populations with greater variation in MRI acquisition protocols. Second, when the number of models, computational cost, and experimental conditions are matched to an equivalently optimized classical feature-transformation module, does the variational quantum circuit offer a statistically significant advantage? Third, how well does the model predict ambiguous clinical cases when sequences are missing and when the scanner differs from the one used for training? Fourth, is the uncertainty-aware component able to accurately detect ‘need for expert review' and contribute to clinical decision-making during a prospective evaluation? Finally, what impact will quantum noise, circuit run time, and hardware limitations have on performance when the model is tested on real quantum devices? The present study does not answer these questions. This requires a diagnostic labelling process validated outside the laboratory, controlled classical-quantum comparisons, prospective clinical assessments, and prospective quantum hardware evaluation. QEDL-Endo should be considered a research system that may benefit from computer-aided imaging support but is not clinically validated at this time.

REFERENCES

[1] Y. Chen, S. Waseem, and L. Luo, “Advances in the diagnosis and management of endometriosis: A comprehensive review,” *Pathology - Research and Practice*, vol. 266, Art. no. 155813, 2025, doi: 10.1016/j.prp.2025.155813.

[2] T. Ni, I. Friedman, J. Smith, B. Liu, M. Melchior, and A. D. Brown, “Pelvic venous disorders in women: Diagnosis and management for gynecologists,” *Int. J. Gynecol. Obstet.*, vol. 174, pp. 1267–1276, 2026, doi: 10.1002/ijgo.70955.

[3] F. Abike, F. B. Tanoglu, and G. Sidar, “Deep pelvic endometriosis: Clinical features, diagnosis, and treatment—A comprehensive review,” *Arch. Gynecol. Obstet.*, vol. 312, pp. 1857–1869, 2025, doi: 10.1007/s00404-025-08187-0.

[4] P. Rousset, M. Florin, N. Bharwani, C. Touboul, M. Monroc, F. Golfier, S. Nougaret, and I. Thomassin-Naggara, “Deep pelvic infiltrating endometriosis: MRI consensus lexicon and compartment-based approach from the ENDOVALIRM group,” *Diagnostic and Interventional Imaging*, vol. 104, no. 3, pp. 95–112, 2023, doi: 10.1016/j.diii.2022.09.004.

[5] B. Liu, G. Li, Y. Zou, S. Zeng, J. Lv, R. Wu, and H. Zhang, “Unsupervised anomaly detection in medical imaging: A survey of methods, challenges, and future directions,” *Bioengineering*, vol. 13, no. 5, p. 547, 2026, doi: 10.3390/bioengineering13050547.

[6] A. Esmailzadeh, A. Kemmak, S. Dokhtfaroughi, A. Rasoulian, and M. R. Mazaheri Habibi, “Investigating the role of artificial intelligence in the diagnosis and prediction of endometriosis using ultrasound images: A systematic review,” *Reproductive Health*, vol. 23, 2026, Art. no. 10.

[7] C. Chen, N. A. M. Isa, and X. Liu, “A review of convolutional neural network based methods for medical image classification,” *Computers in Biology and Medicine*, vol. 185, Art. no. 109507, 2025, doi: 10.1016/j.compbiomed.2024.109507.

[8] L. Cui, M. Xu, C. Liu, T. Liu, X. Yan, Y. Zhang, and X. Yang, “Towards reliable healthcare imaging: A multifaceted approach in class imbalance handling for medical image segmentation,” *Interdisciplinary Sciences: Computational Life Sciences*, vol. 17, no. 3, pp. 614–633, 2025, doi: 10.1007/s12539-025-00726-2.

[9] S. S. Gokhale, R. R. Dhote, and R. Delhibabu, "A review of quantum machine learning algorithms, applications, and emerging advantages," *Discover Computing*, vol. 29, Art. no. 226, 2026, doi: 10.1007/s10791-026-10085-1.

[10] K. Kulkarni and P. K. Desai, "Quantum computing in image processing: A review of applications, models, and prospects," *Franklin Open*, vol. 16, Art. no. 100631, 2026, doi: 10.1016/j.fraope.2026.100631.

[11] M. M. Hossain, M. H. H. Himal, and A. Munir, "A comparative study of quantum feature maps and quantum classifiers for heart disease prediction," *AI*, vol. 7, p. 180, 2026, doi: 10.3390/ai7050180.

[12] A. Jawad, E. Stefan-Henningsen, and A. Kiani, "Applications of classical and quantum machine learning in manufacturing: Predictive maintenance, scheduling and tribology," *Next Research*, vol. 7, Art. no. 101532, 2026, doi: 10.1016/j.nexres.2026.101532.

[13] D. Kong, J. Jin, X. Cui, J. Yan, J. Tang, Y. Fu, J. Xu, and D. Wang, "CQNAS: Convolutional quantum hybrid neural network architecture search for image classification," *Neurocomputing*, vol. 679, p. 133236, 2026, doi: 10.1016/j.neucom.2026.133236.

[14] A. Esmailzadeh, A. Rashki Kemmak, S. Sezavar Dokhtfaroughi, *et al.*, "Investigating the role of artificial intelligence in the diagnosis and prediction of endometriosis using ultrasound images: A systematic review," *Reproductive Health*, vol. 23, no. 1, Art. no. 31, 2026, doi: 10.1186/s12978-025-02245-1

[15] R. AlSaad, T. Farrell, A. Elhenidy, S. Albasha, and R. Thomas, "Artificial intelligence in endometriosis imaging: A scoping review," *AI*, vol. 7, no. 2, Art. no. 43, 2026, doi: 10.3390/ai7020043.

[16] T. Khan, T. Usman, I. Khalid, *et al*., "Comprehensive machine learning approaches for disease prediction: Current applications, recent advances, and future prospects," *Discover Artificial Intelligence*, vol. 6, Art. no. 878, 2026, doi: 10.1007/s44163-026-01851-7.

[17] A. Ahmad, M. Imran, and H. Ahsan, "Biomarkers as biomedical bioindicators: Approaches and techniques for the detection, analysis, and validation of novel biomarkers of diseases," *Pharmaceutics*, vol. 15, no. 6, Art. no. 1630, 2023, doi: 10.3390/pharmaceutics15061630.

[18] C. Albuquerque, R. Henriques, and M. Castelli, "Deep learning-based object detection algorithms in medical imaging: Systematic review," *Heliyon*, vol. 11, no. 1, Art. no. e41137, 2025, doi: 10.1016/j.heliyon.2024.e41137.

[19] A. Podda, R. Balia, S. Barra, S. Carta, M. Neri, S. Guerriero, and L. Piano, "Multi-scale deep learning ensemble for segmentation of endometriotic lesions," *Neural Computing and Applications*, vol. 36, pp. 1–14, 2024, doi: 10.1007/s00521-024-09828-2.

[20] P. Hu, Y. Gao, Y. Zhang, and K. Sun, "Ultrasound image-based deep learning to differentiate tubal-ovarian abscess from ovarian endometriosis cyst," *Front. Physiol.*, vol. 14, Art. no. 1101810, 2023, doi: 10.3389/fphys.2023.1101810.

[21] R. AlSaad, T. Farrell, A. Elhenidy, S. Albasha, and R. Thomas, "Artificial intelligence in endometriosis imaging: A scoping review," *AI*, vol. 7, no. 2, Art. no. 43, 2026, doi: 10.3390/ai7020043.

[22] A. Esmailzadeh, A. Kemmak, S. Dokhtfaroughi, A. Rasoulian, and M. R. Mazaheri Habibi, "Investigating the role of artificial intelligence in the diagnosis and prediction of endometriosis using ultrasound images: A systematic review," *Reproductive Health*, vol. 23, 2026, Art. no. 45, doi: 10.1186/s12978-025-02245-1.

[23] M. O'Leary, C. Neary, and E. Lawrence, "The diagnostic accuracy of magnetic resonance imaging versus transvaginal ultrasound in deep infiltrating endometriosis and their impact on surgical decision-making: A systematic review," *Diagnostics*, vol. 15, no. 22, Art. no. 2856, 2025, doi: 10.3390/diagnostics15222856.

[24] T. Telecan, R.-D. Capraș, S. T. Poenar, R. Sevan-Libotean, A. I. Gaia-Oltean, C.-B. Crivii, and A.-F. Badea, "Radiomics and artificial intelligence in ovarian endometriosis imaging: A systematic review and critical appraisal of emerging evidence," *Bioengineering*, vol. 13, p. 855, 2026, doi: 10.3390/bioengineering13080855.

[25] D. Liang and L. He, "Radiomics and artificial intelligence for imaging-based non-invasive diagnosis and phenotyping of endometriosis: A systematic review," *Journal of Radiation Research and Applied Sciences*, vol. 19, no. 3, Art. no. 102601, 2026, doi: 10.1016/j.jrras.2026.102601.

[26] F. Mollaie, M. H. Shakoor, M. Sepahvand, and R. Boostani, "Medical data types in machine learning and their challenges: A comprehensive review," *Intelligence-Based Medicine*, vol. 15, Art. no. 100427, 2026, doi: 10.1016/j.ibmed.2026.100427.

[27] R. Alsaad, T. Farrell, A. Elhenidy, S. Albasha, and R. Thomas, "Artificial intelligence in endometriosis imaging: A scoping review," *AI*, vol. 7, no. 2, Art. no. 43, 2026, doi: 10.3390/ai7020043.

[28] S. A. Salehi, H. Naderi, S. A. Asghari, J. Chaharlang, and Y. Savaria, "Hybrid quantum–classical architectures in medical imaging: A taxonomy-based survey of COVID-19 models," *Quantum Reports*, vol. 8, no. 2, Art. no. 54, 2026, doi: 10.3390/quantum8020054.

[29] D. Kong, J. Jin, X. Cui, J. Yan, J. Tang, Y. Fu, J. Xu, and D. Wang, "CQNAS: Convolutional quantum hybrid neural network architecture search for image classification," *Neurocomputing*, vol. 679, Art. no. 133236, 2026, doi: 10.1016/j.neucom.2026.133236.

[30] A. I. Alfajri, T. Sumaryada, S. Yani, D. T. Tai, N. Tamam, and A. Sulieman, "Hybrid quantum-classical convolutional neural networks for multiclass lung X-ray image classification," *Journal of Radiation Research and Applied Sciences*, vol. 19, no. 1, Art. no. 102165, 2026, doi: 10.1016/j.jrras.2026.102165.

[31] A. Mukhanbet, P. Trigo, B. Daribayev, and D. Akhmed-Zaki, "Systematic literature review of quantum convolutional neural networks and circuit optimization," *Algorithms*, vol. 19, p. 490, 2026, doi: 10.3390/a19060490.

[32] X. Liang, L. A. Alpuing Radilla, K. Khalaj, H. Dawoodally, C. Mokashi, X. Guan, K. E. Roberts, S. A. Sheth, V. S. Tammisetti, and L. Giancardo, "A multi-modal pelvic MRI dataset for deep learning-based pelvic organ segmentation in endometriosis," *Scientific Data*, vol. 12, Art. no. 1292, 2025, doi: 10.1038/s41597-025-05623-3.