Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Interpretable Deep Learning for Ovarian Cancer Subtype Classification Using Histopathology Images

Mehtap Agirsoy

mehtapagir@gmail.com

Rensselaer Polytechnic Institute

Matthew A. Oehlschlager

Rensselaer Polytechnic Institute

Article

Keywords:

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8570653/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Title: Interpretable Deep Learning for Ovarian Cancer Subtype Classification Using Histopathology Images

Mehtap Agirsoy PhD Candidate[1*], Matthew A. Oehlschlaeger PhD[1]

[1] Department of Mechanical, Aerospace, and Nuclear Engineering, Rensselaer Polytechnic Institute, 110 Eighth St., Troy, 12180, NY, United States

Corresponding Author:
Mehtap Agirsoy, Ph.D. Candidate. Department of Mechanical, Aerospace, and Nuclear Engineering, Rensselaer Polytechnic Institute, 110 Eighth St., Troy, 12180, NY, United States.
Email: agirsm@rpi.edu.

Abstract
Ovarian cancer is a histologically heterogeneous malignancy in which accurate subtype classification is essential for prognosis and treatment selection. Manual interpretation of hematoxylin and eosin (H&E)–stained histopathology slides remains time-consuming and subject to inter-observer variability, particularly for morphologically overlapping subtypes. In this study, an interpretable deep learning framework was developed for automated ovarian cancer subtype classification using a fine-tuned ResNet50 architecture. A publicly available histopathology dataset comprising five major ovarian carcinoma subtypes was employed. Model training incorporated optimized strategies including data augmentation, selective layer unfreezing, label smoothing, and test-time augmentation. Classification performance was benchmarked against established convolutional neural network architectures. Visual interpretability was assessed using Gradient-weighted Class Activation Mapping (Grad-CAM) to examine model attention patterns.
The proposed framework achieved high classification performance, with strong accuracy and macro-averaged F1-scores across subtypes. Grad-CAM analysis demonstrated that model attention was consistently localized to histologically relevant tissue regions in correctly classified samples, while misclassifications were associated with morphologically ambiguous structures. These findings suggest that classification errors primarily reflect intrinsic histopathologic overlap rather than spurious model behavior.

Overall, this work establishes a reproducible and interpretable deep learning benchmark for ovarian cancer subtype classification and provides clinically meaningful insight into model decision processes. The proposed framework is well-suited for use as a second-reader decision support tool in digital pathology workflows.

Introduction
Ovarian cancer remains the most lethal gynecologic malignancy and the fifth leading cause of cancer-related mortality among women worldwide[1–3]. Despite advancements in surgical techniques and chemotherapeutic regimens, overall survival rates have shown only modest improvement, largely due to late-stage diagnosis and the histopathological heterogeneity of the

disease[1,2,4]. Accurate classification of histological subtypes, namely, high-grade serous carcinoma (HGSC), clear cell carcinoma (CC), endometrioid carcinoma (EC), low-grade serous carcinoma (LGSC), and mucinous carcinoma (MC) is essential for prognosis and for guiding targeted treatment strategies, as each subtype exhibits distinct molecular profiles, clinical behaviors, and therapeutic responses[2,5–8].

Currently, subtype determination relies on manual assessment of hematoxylin and eosin (H&E)-stained histopathological slides by expert pathologists[2,7,8]. However, this process is labor-intensive, time-consuming, and susceptible to inter-observer variability, particularly in cases involving morphologically similar subtypes such as EC versus CC or MC versus LGSC[2,7,8]. These diagnostic challenges can lead to uncertainty, delayed clinical decision-making, and suboptimal treatment outcomes.

Recent advances in artificial intelligence (AI), particularly deep learning, have shown promise in addressing these limitations[4,9–12]. Convolutional neural networks (CNNs) have achieved notable success in histological image classification across a range of cancer types[9,10,13]. Several studies have applied deep learning to ovarian tumor detection or classification, using ultrasound or CT imaging, with promising results. For example, Claessens et al. introduced a transformer-based multiple-instance learning (MIL) model that achieved strong recall for tumor classification[6]. Similarly, Pham et al. and Xie et al. reported high diagnostic accuracy using YOLOv8 variants for ultrasound-based ovarian cancer detection[5,14]. Other efforts, including Gao et al. and Maria et al., used deep convolutional and hybrid models for ultrasound and CT modalities, respectively[4,15]. While these studies illustrate the potential of AI in ovarian cancer diagnostics, they are primarily focused on tumor detection or benign versus malignant classification, not on histological subtype classification from histopathology images. Asadi et al. proposed AIDA, a Fourier-based domain adaptation method that improves the generalizability of deep learning models for histopathology across multi-center datasets, achieving strong subtype classification performance—including 80.93% balanced accuracy for ovarian cancer[16]. Abdullah et al. applied transfer learning for ovarian cancer subtype classification and found that MobileNetV2 with Squeeze-and-Excitation blocks outperformed VGG16, achieving ~91% accuracy with greater computational efficiency[17].

Within the domain of subtype classification, models like AIDA and MobileNetV2+SE have achieved reasonable accuracy, but several critical gaps remain. Most existing models lack interpretability, limiting their clinical utility and trustworthiness. Moreover, many architectures have limited capacity to extract deep, discriminative features, which is essential for distinguishing morphologically similar subtypes. Additionally, many prior studies either used small datasets, lacked standardization in training pipelines, or failed to incorporate visualization tools that validate model decisions.

While existing deep learning models have reported promising classification performance, several challenges remain. First, many studies lack transparency regarding the histological features driving model predictions, limiting clinical trust and adoption. Second, performance evaluations often do not examine misclassification patterns or their pathological basis. Finally, reproducible benchmarks using standardized training pipelines and interpretability analyses remain limited in the ovarian cancer histopathology domain.

To address these gaps, an interpretable deep learning framework for ovarian cancer subtype classification is presented in this study. Rather than introducing a novel network architecture, emphasis is placed on rigorous benchmarking, visual interpretability, and clinically meaningful analysis of model behavior. A ResNet50 architecture was selected due to its established performance, residual learning capability, and compatibility with post-hoc interpretability methods. Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to visualize model attention and to investigate the histopathological features influencing classification outcomes.

The primary objectives of this work are to:

(1) establish a reproducible baseline for ovarian cancer subtype classification using histopathology images;
(2) evaluate classification performance across clinically relevant subtypes; and
(3) analyze model attention patterns to elucidate sources of correct predictions and diagnostic confusion.

By focusing on interpretability and failure analysis alongside performance metrics, this study aims to support the development of clinically transparent artificial intelligence tools for digital pathology.

Methods

Dataset and Preprocessing

This study utilized a publicly available ovarian cancer histopathology dataset curated by Abdullah et al. and hosted on the Mendeley Data platform[17]. The dataset comprises H&E-stained image patches representing five histological subtypes: HGSC, CC, EC, MC, and LGSC. The dataset is composed of 4,416 images, distributed as follows: HGSC (1,188), EC (733), CC (551), MC (323), and LGSC (287), reflecting inherent class imbalance.

To ensure comparability with previous work, the same 70/30 train-test split protocol was adopted. All images were resized to 384×384 pixels to preserve morphological resolution and normalized to the ImageNet mean and standard deviation. No explicit stain normalization was applied, which may leave the model susceptible to staining variability.

Extensive data augmentation was applied to the training set to enhance generalization and mitigate overfitting. Techniques included:

- Random horizontal and vertical flipping
- Random rotation (±30°)
- Affine transformations (translation ±10%, scaling 90% to 110%)
- Color jitter (brightness, contrast, saturation ±20%)
- Random erasing ($p = 0.25$)
- Random resized cropping

Validation images were center-cropped and normalized, but no augmentation was applied during validation or testing.

Ethical Approval and Consent to Participate

All methods were carried out in accordance with relevant guidelines and regulations. The ultrasound image dataset used in this study was obtained from a publicly available repository, and all data were fully de-identified prior to release. Ethical approval and informed consent procedures were completed by the original data collectors, in accordance with institutional requirements. The authors did not have access to any identifying patient information and conducted no additional human subject experiments.

Model Architecture and Training Protocol

A ResNet50 convolutional neural network pretrained on ImageNet was fine-tuned for five-class classification. The final fully connected (fc) layer was replaced with a dense output layer of five neurons with softmax activation. To facilitate domain-specific feature learning, layers in stage 3 (layer3), stage 4 (layer4), and the fc layer were unfrozen. These layers were selected based on their role in capturing higher-order semantic features, which are critical for distinguishing subtle histopathological differences between morphologically similar subtypes.

Key training enhancements included:

- Label smoothing ($\alpha = 0.15$): This value was empirically selected based on prior literature to prevent overconfident predictions and improve generalization near class boundaries.
- Test-time augmentation (TTA): Applied to the test set to generate ensemble-like predictions for improved robustness.
- Learning rate scheduling: The ReduceLROnPlateau scheduler was used to adapt the learning rate dynamically in response to stagnation in validation loss.
- Early stopping: Training was halted after 7 consecutive epochs without improvement to minimize overfitting.

The model was trained on a high-performance computing cluster with the following settings:

- Optimizer: Adam
- Initial learning rate: 1e-4
- Loss function: Cross-entropy with label smoothing
- Epochs: 25 (extended to 50 for sensitivity analysis)

Evaluation Metrics and Visualization

Model performance was assessed using the following metrics:

- Overall accuracy
- Macro- and weighted-averaged F1-scores
- Per-class precision, recall, and F1-score
- Confusion matrix analysis
- 95% confidence intervals (bootstrapped with 1,000 resamples)

To assess model interpretability, Grad-CAM was applied to both correctly and incorrectly classified samples. This facilitated visual confirmation of attention alignment with diagnostically meaningful regions.
Although inter-observer variability metrics were not available for this dataset, reported classification results were compared to previously published CNN benchmarks.

Results and Discussion
The fine-tuned ResNet50 model demonstrated strong performance in classifying five ovarian cancer subtypes from H&E-stained histopathology images. Using a standardized 70/30 train-test split, the model yielded an overall test accuracy of 91.4% (95% CI: 90.4%–92.4%) and a macro-averaged F1-score of 90.1% (95% CI: 88.9%–91.3%).

As shown in Table 1, the model achieved the highest classification performance for HGSC and CC, the two most histologically distinct and clinically important subtypes[20]. In contrast, lower precision was observed for LGSC and MC, reflecting diagnostic challenges associated with their overlapping morphologies and limited representation in the dataset.

Table 1: Classification performance of the ReNet50 model across ovarian cancer subtypes

Subtype	**Precision**	**Recall**	**F1 score**	**Support**
CC	0.95	0.91	0.93	551
EC	0.89	0.88	0.89	733
HGSC	0.92	0.94	0.93	1188
LGSC	0.78	0.91	0.84	287
MC	0.92	0.83	0.87	323
Metric			**Value**	
Accuracy			0.90	3082
Macro avg	0.89	0.89	0.89	3082
Weighted avg	0.91	0.90	0.90	3082

The confusion matrix in Figure 1 highlights the model's high accuracy in classifying HGSC and CC, while frequent misclassifications occurred between LGSC, EC, and MC subtypes known for histological overlap[21]. Notably, LGSC had a high recall but a lower precision due to false positives, predominantly being misclassified as EC or MC. This trend underscores both class imbalance and diagnostic ambiguity in LGSC cases.

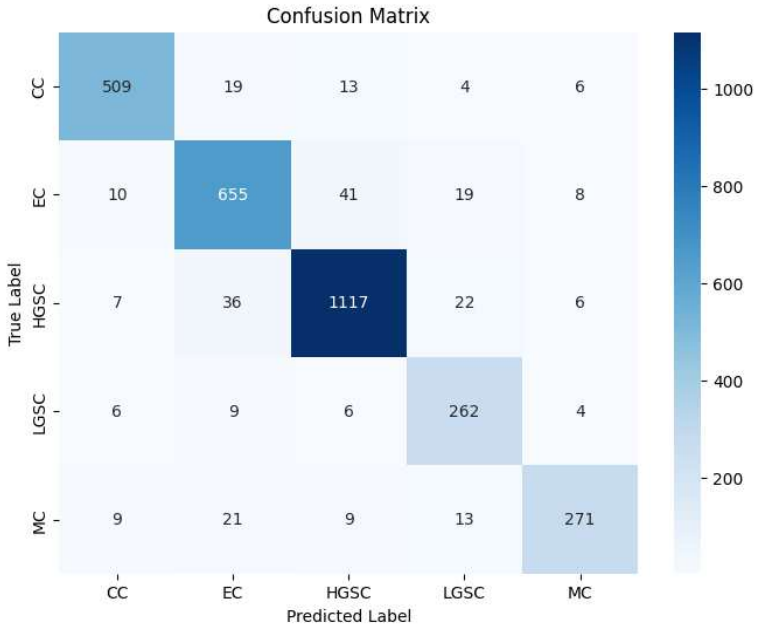


Figure 1: Confusion matrix of the ResNet50 model

To gain insight into the model's decision-making process, Grad-CAM was used to visualize class-specific attention in the final convolutional layer of the fine-tuned ResNet50 model. As shown in Figure 2A, the highlighted regions frequently correspond to morphologically ambiguous structures, indicating that misclassifications may arise from visual similarities between subtypes. Notably, four of the five most misclassified samples belonged to the EC class, reflecting its histopathological complexity and overlap with other subtypes. EC often exhibits glandular and cytologic features that resemble neighboring classes, especially in H&E-stained slides lacking immunohistochemical or molecular context[18,19,22]. The Grad-CAM heatmaps for these cases commonly focused on poorly differentiated or non-discriminative regions, underscoring the diagnostic challenge of EC and suggesting that integrating multimodal data could improve model specificity.

Conversely, correctly classified samples in Figure 3A show that the model's attention tends to align with clinically meaningful tissue regions. These visual patterns suggest that the model learned to detect subtype-specific features, such as clear cytoplasm in CC or papillary structures in serous carcinoma and focused on those regions during inference.

Overall, Grad-CAM enhances model interpretability and supports the reliability of the predictions, while also revealing the boundaries of the model's discriminative capability. These visualizations are valuable for identifying failure modes and informing potential model refinements such as region-guided attention or hybrid clinical-feature fusion. These patterns reinforce the model's capacity to learn clinically meaningful features while also exposing challenges in visually similar subtypes.

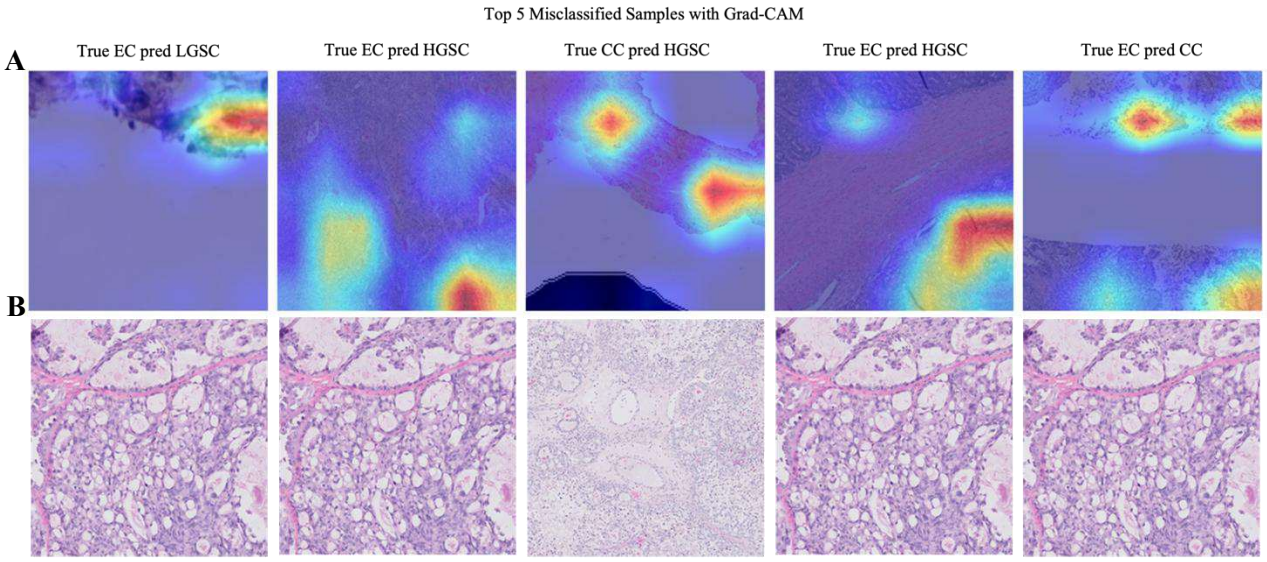


Figure 2: A) Grad-CAM overlays of the top five misclassified samples, illustrating model attention on ambiguous histologic regions likely contributing to subtype confusion. **B)** Corresponding ground truth H&E-stained slides without overlays, providing reference morphology.

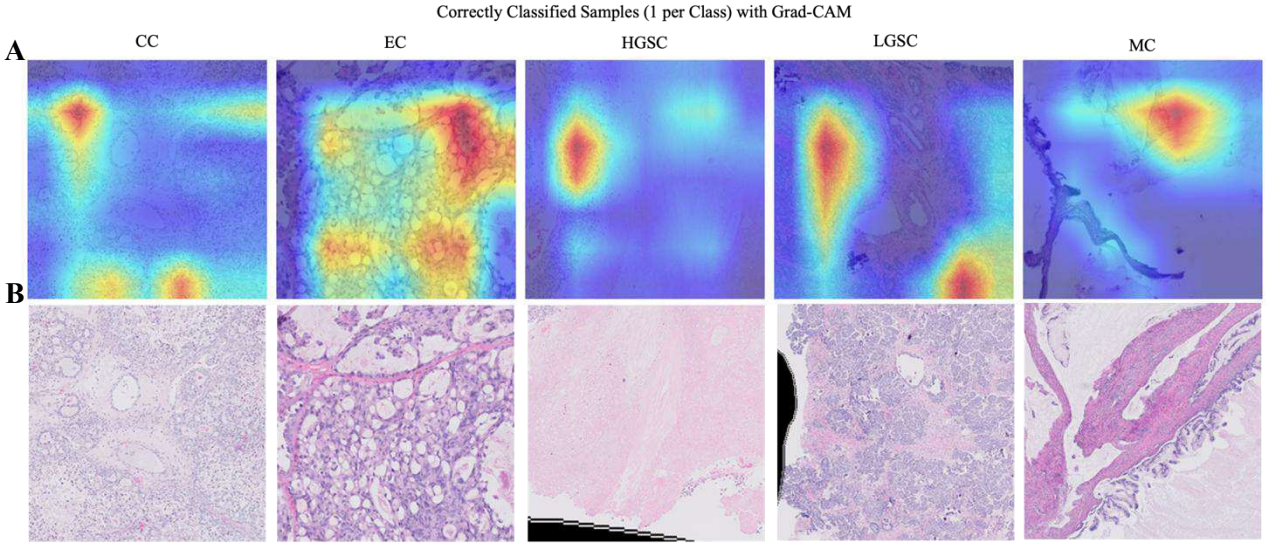


Figure 3. A) Grad-CAM heatmaps of correctly classified samples, showing attention localized to subtype-defining structures. **B)** Ground truth images for verification of visual alignment.

Table 2 compares ResNet50 against prior architectures. ResNet50 achieved an accuracy of 91% and a macro-averaged F1-score of 90%, matching the highest performance reported by MobileNetV2+SE, while offering additional interpretability and architectural flexibility.

Table 2: Comparison of performance metrics and architectural characteristics among VGG16, MobileNetV2 with SE blocks, and ResNet50 models for ovarian cancer subtype classification

Model	Depth	Accuracy (%)	Macro F1 (%)	95%CI (Accuracy)	Strengths
VGG16	16	~88	~86	Not reported	Simplicity
MobileNetV2+SE	~53	91	~89	Not reported	Efficient, attention-enabled
ResNet50	50	**91.4**	**90.1**	**90.4–92.4**	Best balance of accuracy and F1

VGG16, though historically significant, underperformed due to its high parameter count and absence of residual connections, limiting its ability to learn deep hierarchical features. MobileNetV2+SE demonstrated strong accuracy with lower computational cost, making it well-suited for deployment in resource-constrained environments. However, its relatively shallow architecture may limit its capacity to capture fine-grained histological distinctions among morphologically similar subtypes.

ResNet50, leveraging residual learning and deeper architecture, provided a stronger balance between representational depth and generalization. Its ability to extract high-level discriminative features proved critical for differentiating subtypes such as high-grade serous and clear cell carcinoma. Moreover, its compatibility with interpretability methods like Grad-CAM allowed meaningful visualization of attention regions, aligning closely with diagnostically relevant tissue structures.

While lightweight models like MobileNetV2+SE offer operational advantages, ResNet50 stands out for its diagnostic robustness, interpretability, and extensibility. Its deeper architecture and strong feature extraction capabilities make it particularly well-suited for complex classification tasks where subtle morphological cues are decisive. These characteristics position ResNet50 as a more promising candidate for clinical integration, especially in settings where model transparency and reliability are essential.

Clinical Interpretation

The model demonstrated particularly high F1-scores for HGSC and CC subtypes that are clinically aggressive and often treatment-defining[18,23]. This performance underscores its utility in early and accurate subtype identification. Although the current implementation focuses solely on histological inputs, the architecture lends itself well to multi-modal expansion. Future iterations may integrate genomic, proteomic, or clinical metadata to enable comprehensive ovarian cancer profiling.

The findings of this study hold significant implications for improving the histopathological diagnosis and subtype stratification of ovarian cancer. Accurate histological classification is essential for personalized treatment planning, particularly for subtypes such as HGSC and CC, which differ substantially in prognosis, therapeutic response, and molecular characteristics[21,23]. However, manual interpretation of H&E-stained slides is time-consuming and subject to inter-observer variability, especially in morphologically overlapping subtypes like MC and LGSC[18,19,21]. The proposed ResNet50-based deep learning model addresses these challenges by offering both high diagnostic performance and visual interpretability. Its consistent and robust classification of HGSC and CC demonstrates its clinical potential for identifying treatment-defining subtypes early

in the diagnostic workflow. The model's integration with Grad-CAM enhances transparency by providing visual justifications for predictions, enabling clinicians to review and validate decisions, particularly in ambiguous cases.

From a deployment standpoint, the model's inference speed, compatibility with high-resolution histopathology inputs, and modular architecture support its integration into digital pathology platforms. Potential clinical applications include:

- **Pre-screening**: Automatically prioritizing or flagging diagnostically uncertain whole-slide images.
- **Second-reader support**: Assisting pathologists in high-throughput settings by highlighting areas of interest.
- **Training and QA tools**: Serving as a consistent benchmark for trainee evaluation or quality assurance in diagnostic workflows.

While the current model uses only histological images, its architecture is conducive to multi-modal expansion. Future iterations may incorporate immunohistochemical (IHC), genomic, or clinical metadata to improve subtype differentiation, particularly for histologically ambiguous tumors such as EC or LGSC.

To advance clinical adoption, further validation across diverse, multi-institutional datasets is required to assess generalizability and resilience to real-world variability, including stain and scanner differences. Additionally, conducting a quantitative comparison of Grad-CAM outputs with pathologist-annotated diagnostic regions will be critical to substantiating the model's interpretability and utility in regulated settings.

Overall, the ResNet50 model presents a scalable and interpretable solution for ovarian cancer subtype classification, with the potential to reduce diagnostic variability, enhance workflow efficiency, and ultimately improve patient care through more consistent and timely diagnoses.

Conclusion

In this study, an interpretable deep learning framework was evaluated for ovarian cancer subtype classification using H&E-stained histopathology images. High classification performance was achieved across five major ovarian carcinoma subtypes, with particularly strong results observed for high-grade serous and clear cell carcinomas. These subtypes are clinically significant due to their aggressive behavior and distinct therapeutic implications, highlighting the potential utility of the proposed framework in diagnostic support settings.

Importantly, this work was not designed to introduce a novel deep learning architecture, but rather to establish a transparent and reproducible benchmark that prioritizes interpretability and clinical insight. In applied medical artificial intelligence, diagnostic reliability and explainability are increasingly recognized as essential complements to predictive accuracy. By systematically integrating Grad-CAM analysis, this study provides visual evidence that model predictions are driven by histologically meaningful regions rather than spurious image artifacts.

Analysis of misclassified samples revealed that errors were frequently associated with morphologically ambiguous regions, particularly among endometrioid, mucinous, and low-grade serous carcinomas. These subtypes are known to exhibit overlapping histological features, even among experienced pathologists. The observed attention patterns suggest that model uncertainty reflects intrinsic diagnostic complexity rather than algorithmic instability. Such findings underscore the importance of interpretability tools for identifying failure modes and guiding future model refinement.

Compared with prior convolutional neural network–based approaches, the ResNet50 architecture demonstrated a favorable balance between representational depth, generalization, and interpretability. While lightweight architectures may offer computational efficiency, deeper residual networks appear better suited for capturing subtle histomorphological cues necessary for subtype discrimination. Furthermore, the compatibility of ResNet50 with visualization techniques facilitates transparent integration into digital pathology workflows.

From a clinical perspective, the proposed framework is best positioned as a second-reader decision support tool rather than a replacement for expert assessment. Attention heatmaps may assist pathologists by highlighting diagnostically relevant regions, particularly in challenging or borderline cases, and may contribute to reduced inter-observer variability. Additionally, the framework could support training, quality assurance, and case prioritization in high-throughput pathology environments.

Several limitations should be acknowledged. The dataset was derived from a single publicly available source, which may limit generalizability across institutions, scanners, and staining protocols. Class imbalance, particularly for low-grade serous carcinoma, may have influenced class-specific performance. Moreover, interpretability was assessed qualitatively; future work should incorporate quantitative validation of attention maps against expert-annotated regions.

Future extensions of this work will focus on multi-institutional validation, incorporation of immunohistochemical or molecular data, and prospective evaluation within clinical workflows. Such efforts are expected to further enhance diagnostic specificity and clinical applicability.

In summary, this study demonstrates that clinically meaningful insight can be obtained from deep learning–based histopathology models when interpretability and failure analysis are explicitly prioritized. The proposed framework provides a transparent and reproducible reference for ovarian cancer subtype classification and represents a practical step toward responsible clinical deployment of artificial intelligence in digital pathology.

Future work will prioritize multi-institutional validation to assess generalizability, incorporate molecular and clinical data to improve subtype discrimination, and quantitatively evaluate Grad-CAM outputs against expert annotations. Additionally, deployment within digital pathology systems will be explored to understand workflow integration and user feedback, while detailed error analysis will help uncover morphological features contributing to misclassifications.

Declaration of AI-assisted technologies in the writing process

Statement: During the preparation of this work, the authors used Grammarly to edit and refine the text. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Data Availability

The datasets generated and/or analysed during the current study are publicly available in the Mendeley repository at: https://data.mendeley.com/datasets/kztymsrjx9/1

Funding Declaration: This study did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Christiansen, F. *et al.* International multicenter validation of AI-driven ultrasound detection of ovarian cancer. *Nat Med* https://doi.org/10.1038/s41591-024-03329-4 (2025) doi:10.1038/s41591-024-03329-4.
2. Peres, L. C. *et al.* Histotype classification of ovarian carcinoma: A comparison of approaches. *Gynecol Oncol* **151**, 53–60 (2018).
3. Bukłaho, P. A., Kiśluk, J., Wasilewska, N. & Nikliński, J. Molecular features as promising biomarkers in ovarian cancer. *Advances in Clinical and Experimental Medicine* **32**, (2023).
4. Maria, H. H., Jossy, A. M. & Malarvizhi, S. A hybrid deep learning approach for detection and segmentation of ovarian tumours. *Neural Comput Appl* **35**, 15805–15819 (2023).
5. Pham, T.-L. & Le, V.-H. Ovarian Tumors Detection and Classification from Ultrasound Images Based on YOLOv8. *Journal of Advances in Information Technology* **15**, 264–275 (2024).
6. H.B. Claessens, C. *et al.* Multi-center Ovarian Tumor Classification Using Hierarchical Transformer-Based Multiple-Instance Learning. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* vol. 15199 LNCS 3–13 (Springer Science and Business Media Deutschland GmbH, 2025).
7. Prat, J. New insights into ovarian cancer pathology. *Annals of Oncology* **23**, (2012).
8. Chen, V. W. *et al.* Pathology and classification of ovarian tumors. *Cancer* **97**, 2631–2642 (2003).
9. Naderi Yaghouti, A. R. *et al.* Artificial Intelligence for Ovarian Cancer Detection with Medical Images: A Review of the Last Decade (2013–2023). *Archives of Computational Methods in Engineering* https://doi.org/10.1007/s11831-025-10268-x (2025) doi:10.1007/s11831-025-10268-x.
10. Saraei, M., Lalinia, M. & Lee, E. J. Deep Learning-Based Medical Object Detection: A Survey. *IEEE Access* https://doi.org/10.1109/ACCESS.2025.3553087 (2025) doi:10.1109/ACCESS.2025.3553087.

11. Salman, M. E., Çakirsoy Çakar, G., Azimjonov, J., Kösem, M. & Cedimoğlu, İ. H. Automated prostate cancer grading and diagnosis system using deep learning-based Yolo object detection algorithm. *Expert Syst Appl* **201**, (2022).
12. Yang, Y., Zhang, H., Gichoya, J. W., Katabi, D. & Ghassemi, M. The limits of fair medical imaging AI in real-world generalization. *Nat Med* https://doi.org/10.1038/s41591-024-03113-4 (2024) doi:10.1038/s41591-024-03113-4.
13. Tommi Lahti, S. *Convolutional Neural Networks for Medical Image Classification Developing CNN Models for the Classification of Gynecological Disorders Using Ultrasonography Subject Convolutional Neural Network for Medical Image Classification*.
14. Xie, W. *et al.* Developing a deep learning model for predicting ovarian cancer in Ovarian-Adnexal Reporting and Data System Ultrasound (O-RADS US) Category 4 lesions: A multicenter study. *J Cancer Res Clin Oncol* **150**, (2024).
15. Gao, Y. *et al.* Deep learning-enabled pelvic ultrasound images for accurate diagnosis of ovarian cancer in China: a retrospective, multicentre, diagnostic study. *Lancet Digit Health* **4**, e179–e187 (2022).
16. Asadi-Aghbolaghi, M. *et al.* Learning generalizable AI models for multi-center histopathology image classification. *NPJ Precis Oncol* **8**, (2024).
17. Abdullah, S. M. A. M. A. P. N. U. M. L. A. Deep Learning-Based Ovarian Cancer Subtype Classification Using VGG16 and MobileNetV2 with Squeeze-and-Excitation Blocks. *Journal of Angiotherapy* **8**, 1–8 (2024).
18. Merritt, M. A. & Cramer, D. W. Molecular pathogenesis of endometrial and ovarian cancer. *Cancer Biomarkers* **9**, 287–305 (2011).
19. D'Angelo, E. & Prat, J. Classification of ovarian carcinomas based on pathology and molecular genetics. *Clinical and Translational Oncology* vol. 12 783–787 Preprint at https://doi.org/10.1007/s12094-010-0599-0 (2010).
20. Lisio, M. A., Fu, L., Goyeneche, A., Gao, Z. H. & Telleria, C. High-grade serous ovarian cancer: Basic sciences, clinical and therapeutic standpoints. *International Journal of Molecular Sciences* vol. 20 Preprint at https://doi.org/10.3390/ijms20040952 (2019).
21. Chen, L. *et al. Case Report Low-Grade Endometrioid Carcinoma of the Ovary Associated with Undifferentiated Carcinoma: Case Report and Review of the Literature*. *Int J Clin Exp Pathol* vol. 7 www.ijcep.com/ (2014).
22. Kurman, R. J. & Shih, I. M. The dualistic model of ovarian carcinogenesis revisited, revised, and expanded. *American Journal of Pathology* vol. 186 733–747 Preprint at https://doi.org/10.1016/j.ajpath.2015.11.011 (2016).
23. Lisio, M. A., Fu, L., Goyeneche, A., Gao, Z. H. & Telleria, C. High-grade serous ovarian cancer: Basic sciences, clinical and therapeutic standpoints. *International Journal of Molecular Sciences* vol. 20 Preprint at https://doi.org/10.3390/ijms20040952 (2019).