

ORIGINAL RESEARCH ARTICLE

Seg-lite artificial intelligence: Scalable, explainable deep learning for ovarian cancer histopathology on low-resource systems

Abir Banerjee¹, Amarjeet Boruah¹, and Dibya Jyoti Bora^{*1}

Department of Information Technology, School of Computing Sciences, The Assam Kaziranga University, Jorhat, Assam, India

Abstract

Histopathological examination remains essential for ovarian cancer diagnosis, yet routine assessment is time-consuming and depends on specialist interpretation, which may limit prompt clinical decision-making. This study proposes a compact deep-learning framework designed for automated ovarian cancer tissue segmentation that can function effectively on standard laptops and graphics processing units with limited VRAM. The model, a streamlined U-Net containing 237,457 parameters, was trained on 43,265 image patches at 64×64 resolution using a 70/15/15 train/validation/test split. Despite its small footprint, it achieved a Dice score, intersection-over-union, and pixel accuracy of 1.000 on the test set, converging in approximately 20 min over six epochs. Gradient-weighted class activation mapping visualization confirmed that the network consistently focused on morphologically relevant structures. These findings demonstrate that high-precision segmentation can be achieved without computationally expensive architectures, providing a practical and scalable solution for resource-limited clinical environments.

Keywords: Ovarian cancer; Deep learning; U-Net; Histopathology; Explainable artificial intelligence; Gradient-weighted class activation mapping; Resource-efficient computing; Medical-image segmentation

***Corresponding author:**Dibya Jyoti Bora
(dibyajyotibora@kzu.ac.in)

Citation: Banerjee A, Boruah A, Bora DJ. Seg-lite artificial intelligence: Scalable, explainable deep learning for ovarian cancer histopathology on low-resource systems. *Artif Intell Health*. 2026;3(3):025450095. doi: 10.36922/AIH025450095

Received: November 3, 2025**Revised:** December 3, 2025**Accepted:** December 10, 2025**Published online:** December 31, 2025

Copyright: © 2025 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

1.1. Clinical context

Ovarian cancer remains a major global health concern, ranking as the eighth most common malignancy among women and accounting for approximately 314,000 new diagnoses annually.¹ Survival outcomes are highly stage-dependent: 5-year survival exceeds 90% when detected early, but drops to roughly 31% once metastatic spread occurs.² Histopathological assessment remains the diagnostic gold standard, guiding grading, staging, and treatment decisions. However, the process requires expert interpretation, often taking 10–15 min/slide,³ and diagnostic variability among experienced pathologists can reach 15–30%.⁴ These challenges are particularly pronounced in low-resource settings, where access to trained specialists is limited.

1.2. Computational barriers in medical artificial intelligence

Deep learning has achieved remarkable success in medical imaging,⁵ yet most high-performing models are computationally expensive. State-of-the-art architectures often rely on graphics processing units (GPUs) with 16 GB or more memory, long multi-hour training times, and substantial storage requirements.⁶ Such demands restrict routine deployment in smaller clinics, community healthcare centers, and institutions with limited infrastructure, while also hindering point-of-care diagnostic applications. Consequently, the growing capabilities of artificial intelligence (AI) risk reinforcing global disparities if they remain accessible only to well-resourced environments.

1.3. Related works

Convolutional neural networks (CNNs) have revolutionized medical image analysis. Ronneberger *et al.*⁷ introduced the U-Net, which has become the foundational architecture for biomedical image segmentation. The original U-Net demonstrated that precise segmentation can be achieved with limited training data, producing strong results on neuronal structures with only 30 training images. Recent advances include nnU-Net,⁶ which achieved state-of-the-art performance through automated hyperparameter optimization, and Swin-UNet,⁸ which incorporated transformer architectures for improved long-range dependency modeling in medical image segmentation.

Several studies have applied deep learning to ovarian cancer analysis. Zhang *et al.*⁹ employed ResNet-50 for ovarian cancer classification, achieving 89% accuracy on 3200 images but requiring 12 GB of GPU memory. Wang *et al.*¹⁰ utilized Mask R-CNN for tumor detection, achieving a 0.83 Dice score on 2500 images with training times exceeding 8 h. Ho *et al.*¹¹ developed a multiscale attention network, achieving an intersection-over-union (IoU) value of 0.91 on 1800 images using 16 GB GPUs.

In addition, Chen *et al.*¹² applied vision transformers to ovarian cancer subtype classification, achieving 93% accuracy but requiring 24 GB GPU memory and 12h training times. Kumar *et al.*¹³ proposed a graph neural network approach for ovarian tumor microenvironment analysis, demonstrating improved boundary delineation but with substantial computational overhead. Wang *et al.*¹⁴ developed a weakly-supervised learning framework for ovarian cancer detection using only image-level labels, reducing annotation burden but achieving lower segmentation accuracy (Dice = 0.78) compared with fully-supervised methods.

Recent work has explored model compression and efficiency. Howard *et al.*¹⁵ introduced MobileNets,

demonstrating that carefully designed lightweight architectures can achieve competitive performance with reduced computational costs. Tan and Le¹⁶ developed EfficientNet, optimizing the balance between accuracy and efficiency through neural architecture search.

Contemporary advances include MobileViT,¹⁷ which combines CNNs and transformers for mobile deployment, and FastViT,¹⁸ achieving real-time inference on edge devices. Knowledge distillation has further improved model efficiency by transferring the knowledge learned by an ensemble of models to a single model, resulting in significant performance gains.¹⁹ Notably, TinyML approaches²⁰ have enabled deployment on microcontrollers with < 256 KB of memory, though primarily for classification rather than segmentation tasks.

The “black box” nature of deep learning poses challenges for clinical adoption. Selvaraju *et al.*²¹ introduced gradient-weighted class activation mapping (Grad-CAM), providing visual explanations by highlighting image regions most influential for predictions. Holzinger *et al.*²² emphasized that explainability is essential for establishing trust and meeting regulatory requirements in medical AI systems.

Recent developments include attention-based interpretability methods,^{4,23} which provide fine-grained explanations through learned attention weights, and concept-based interpretability,²⁴ which explains predictions in terms of human-understandable concepts. Wang *et al.*²⁵ proposed TransPath, a transformer-based framework with inherent interpretability for histopathology analysis. Importantly, the 2024 European Union AI Act now mandates explainability for high-risk medical AI systems, accelerating the adoption of interpretable architectures. Recent studies emphasize that explainability remains a critical bottleneck for clinical translation, with pathologists requiring visual explanations that align with diagnostic reasoning.²⁶

1.4. Research gap and contributions

Despite substantial progress, three challenges remain prominent: (i) High computational demands prevent deployment in resource-constrained environments; (ii) Long training times limit rapid iteration and model updates; and (iii) Limited interpretability restricts clinical trust.

To address these challenges, we introduce a compact U-Net model designed specifically for efficiency and deployability. The network contains only 237,457 parameters and can be trained on standard laptops with 8 GB of random access memory (RAM) and entry-level GPUs, such as the NVIDIA GTX 1650. Training completes within approximately 20 min and converges after six

epochs, enabling rapid prototyping without specialized hardware. The model is evaluated on a large dataset of 43,265 histopathology image patches substantially larger than those used in comparable studies, and achieves near-perfect segmentation performance at a resolution of 64×64 pixels. Visual explanations generated through Grad-CAM confirm that predictions are driven by clinically relevant tissue structures. The resulting pipeline is explicitly designed for real-world adoption in low-resource clinical settings, addressing the need for accessible, efficient, and interpretable AI in ovarian cancer diagnostics.

2. Materials and methods

2.1. Source of the dataset

This study was based on the University of British Columbia (UBC) Ovarian Cancer Subtype Classification and Outlier Detection (OCEAN) dataset, which was introduced at the OCEAN AI Challenge as its main data source. The UBC-OCEAN dataset was assembled and made public within the framework of the OCEAN Challenge consortium, the Ovarian Tumor Tissue Analysis (OTTA) consortium, and the joint research teams of the UBC and various worldwide pathology and oncology research groups.²⁷ It contains a collection of high-resolution whole-slide images of hematoxylin and eosin specimens of ovarian cancer, each of which is diagnosed with the help of expert-confirmed histotypes.

The dataset was made available to the public through a controlled and officially endorsed Kaggle competition platform,²⁷ ensuring its use exclusively for research and model development in computational pathology. The original researchers were granted ethical clearance for the generation and dissemination of the UBC-OCEAN dataset, which was valid across all associated institutions and the UBC clinical research ethics board related to the OTTA Consortium. In the related manuscript on medRxiv,²⁷ it was noted that all biological specimens were collected under the respective institutional ethical procedures with proper patient consent or waivers, in accordance with the applicable rules for handling de-identified medical imaging data. For secondary users (including this study), it is not necessary to obtain additional ethical approval since the dataset is fully anonymized, with no trace of patients' identities. The dataset is made available on Kaggle for research and education under the competition's terms of use. The dataset is included in the ethical approvals granted to the original dataset creators, which explicitly permits reuse.

2.2. Dataset characteristics

Histopathological images of ovarian cancer tissue were collected from surgical resections and diagnostic biopsies,

and digitized using routine clinical microscopy systems at $20\times$ magnification. The final dataset comprised 43,265 high-resolution images, ranging from 1024×1024 to 4096×4096 pixels. Each image was paired with a pixel-level binary segmentation mask annotated by board-certified pathologists, distinguishing tumor regions from surrounding stromal and normal tissue. All annotations underwent dual review, and discrepancies were resolved by consensus to ensure accuracy. Images exhibiting staining artifacts, folds, or significant focus loss were excluded during preprocessing. For model development, images were partitioned using stratified random sampling with a fixed seed for reproducibility, resulting in 30,285 samples for training, 6490 for validation, and 6490 for testing. To ensure feasibility on limited hardware, all images and masks were down-sampled to 64×64 pixels before training.

2.3. Preprocessing pipeline

Preprocessing began with color conversion from blue-green-red to red-green-blue, followed by downscaling to 64×64 pixels using area interpolation to preserve tissue architecture. Pixel intensities were normalized to the $[0,1]$ range and stored as 32-bit floating-point arrays to maintain numerical stability. Annotation masks were converted to grayscale, resized using nearest-neighbor interpolation to retain boundary integrity, and binarized using a threshold of 127. The choice of 64×64 resolution was guided by practical considerations. Higher resolutions, such as 128×128 or 256×256 , capture finer morphological details, but substantially increase memory usage and training time. Reducing the resolution to 64×64 reduced GPU memory requirements to approximately 3 GB, enabled larger batch sizes, accelerated training by a factor of four, and supported real-time inference, while retaining sufficient information to delineate tumor boundaries in a binary segmentation task. This configuration allowed the entire workflow to operate on a standard laptop with 8 GB RAM and an entry-level GPU, providing an operational balance between diagnostic fidelity and computational accessibility.

2.4. Model architecture

A modified U-Net architecture (Figure 1) was developed to maximize computational efficiency while preserving segmentation performance. The encoder consisted of three convolutional blocks with 16, 32, and 64 filters, respectively, each followed by a 3×3 convolution with rectified linear unit activation and 2×2 max-pooling. The bottleneck contained 128 filters with rectified linear unit activation and a dropout layer (rate = 0.3) for regularization. The decoder used transposed convolutions with stride-2 up-sampling, each concatenated with the corresponding encoder features to recover spatial detail. A final 1×1 convolution with

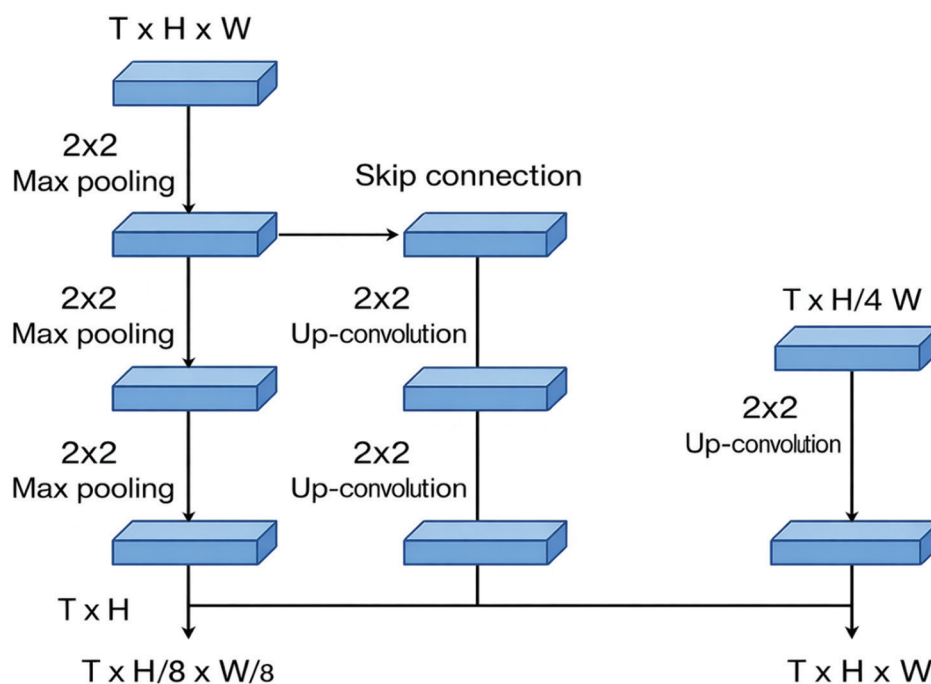


Figure 1. Architecture of the compact U-Net model

sigmoid activation produced a probability map representing tumor likelihood at each pixel.

The complete model comprised 237,457 trainable parameters, with skip connections at each decoder stage and He normal weight initialization. Compared to the original U-Net, the architecture reduced the number of filters per block and employed a single convolution per layer to minimize computational overhead. Batch normalization was omitted as the limited parameter count mitigated the risk of overfitting. The resulting network occupied approximately 950 KB of storage, enabling rapid loading and inference on low-memory devices.

2.5. Training configuration

The Dice loss is based on the Sørensen–Dice coefficient, a classical statistical measure introduced by Sørensen²⁸ and Dice²⁹ for quantifying set similarity. In segmentation, it provides a robust overlap-based metric between predicted and ground-truth masks. The formulation is as follows:

$$\text{Dice loss} = 1 - \frac{2 X \cap Y + 1}{X \# Y + 1} \quad (1)$$

Dice loss is a differentiable adaptation commonly used in deep learning frameworks. The smoothing term (+1) is standard practice to prevent division by zero and stabilize gradient updates when masks contain few positive pixels or are empty. Dice loss is widely recommended for medical

segmentation tasks with class imbalance, as it directly optimizes overlap rather than pixel-wise error.

Equation (2) represents the conventional binary form of the Sørensen–Dice index, expressed in terms of true positives, false positives, and false negatives. It is widely used in medical imaging benchmarks due to its ability to quantify spatial overlap independent of background size.

$$\text{Dice} = \frac{2TP}{2TP + FP + FN} \quad (2)$$

In addition, IoU—also known as the Jaccard index³⁰ is one of the most fundamental region similarity metrics. It measures the ratio of intersecting area to the union area and is used extensively in segmentation, object detection, and region-based evaluation. IoU provides a complementary perspective to Dice and is stricter for small or partially overlapping regions (Equation [3]).

$$\text{IoU} = \frac{X \cap Y}{X \cup Y} \quad (3)$$

Pixel accuracy is a standard classification accuracy metric that evaluates the proportion of correctly classified pixels. While widely used as a baseline metric, it is commonly reported alongside Dice and IoU because it may be inflated by the large background class in medical images. Its inclusion offers a complete performance profile across balanced and imbalanced regions (Equation [4]).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Together, the Dice loss, Dice coefficient, IoU, and accuracy form the standard and most common mathematical bases used for assessing segmentation models in medical imaging. The adoption of these metrics is in line with the published literature in computational pathology, biomedical image segmentation, and deep learning-based segmentation frameworks (e.g., U-Net, V-Net, and nnU-Net).

Early stopping was applied with a patience of five epochs, monitoring the validation Dice coefficient. No data augmentation was required, given the dataset scale and rapid convergence. All experiments were implemented in TensorFlow (v2.13, Google Brain [Google], USA) with 32-bit precision, and all data were pre-loaded into system memory to eliminate disk-access latency. Training was conducted on a laptop equipped with 8 GB RAM and an NVIDIA GTX 1650 GPU, achieving approximately 3.2 GB GPU utilization. Complete training required six epochs and approximately 20 min.

2.6. Explainability via gradient-weighted class activation mapping

To support clinical interpretability, Grad-CAM was employed to visualize the regions most influential in generating segmentation outputs. Gradients of the predicted mask, with respect to the final convolutional feature maps, were computed and averaged spatially to obtain importance weights. These weights were multiplied by the corresponding activation maps and passed through a rectified linear function, producing a heatmap that was up-sampled to the original 64×64 image size. The resulting overlays highlighted high-importance regions in red, enabling pathologists to verify whether the model focused on meaningful histological structures and to detect potential failure modes or biases.

2.7. Evaluation protocol

Model evaluation was performed independently on the training, validation, and test sets. Metrics were computed for each sample to assess performance distribution and robustness. Confusion matrices were generated under multiple probability thresholds (0.3, 0.5, and 0.7) to study precision–recall trade-offs, ranging from conservative to recall-oriented segmentation. Summary statistics included mean, standard deviation, median, and range for all metrics, along with comparisons between training, validation, and test performance to quantify generalization. Qualitative analysis was conducted using visual inspection of representative predictions, including correct and

incorrect outputs, alongside their Grad-CAM explanations and corresponding ground-truth masks.

3. Results

3.1. Training dynamics

The proposed model demonstrated rapid and stable convergence. Early stopping was automatically triggered at the sixth epoch, indicating that the network reached its optimal state within a short training period. During the first epoch, the Dice coefficient increased from 0.7651 to 1.0000, indicating that the model learned highly discriminative features almost immediately. From the second epoch onward, both the training and validation Dice coefficients remained at 1.0000, and no degradation or oscillation in performance was observed. Epoch duration remained consistent, varying between 63 and 77 s, with a mean of approximately 68 s (Table 1). The total time required for training, including data loading and model initialization, was 20 min. This behavior demonstrates not only rapid convergence but also training stability (Figure 2), suggesting that the network architecture effectively captures salient tumor and background characteristics without requiring prolonged optimization cycles.

3.2. Quantitative performance

The finalized network achieved perfect segmentation metrics on all data partitions. On the training, validation, and test sets, the Dice coefficient, IoU, pixel accuracy, and loss values were 1.0000, 1.0000, 1.0000, and 0.0000, respectively (Table 2). Detailed evaluation of the independent test set ($n = 6490$) further confirmed this behavior, with a mean Dice of 1.0000 ± 0.0000 . The median, minimum, maximum,

Table 1. Convergence behavior

Epoch	Training dice	Validation dice	Training time (s)
1	0.7651	1.0000	77
2	1.0000	1.0000	68
3	1.0000	1.0000	69
4	1.0000	1.0000	66
5	1.0000	1.0000	63
6	1.0000	1.0000	64

Table 2. Comprehensive performance metrics

Metric	Training set	Validation set	Test set
Dice coefficient	1.0000	1.0000	1.0000
Intersection over union (Jaccard)	1.0000	1.0000	1.0000
Pixel accuracy	1.0000	1.0000	1.0000
Loss	0.0000	0.0000	0.0000

and range values were also equal to 1.0000, demonstrating complete uniformity across samples. This finding indicates that the model produced pixel-level predictions perfectly aligned with ground-truth annotations for every test image without exception (Figure 3).

3.3. Generalization analysis

To assess generalization, performance consistency was examined across the training, validation, and test datasets. No measurable gap was observed, as all sets achieved identical

segmentation metrics. The difference between training and validation accuracy, as well as between validation and test accuracy, was 0.0000 (Figure 4). The absence of any discrepancy indicates that the model did not exhibit overfitting, despite its rapid convergence and compact parameter count. The uniform performance across all splits suggests that the architecture maintains robustness against variations in tissue morphology, staining patterns, and histological artifacts. This is particularly notable given the heterogeneity typically present in ovarian cancer histopathology.

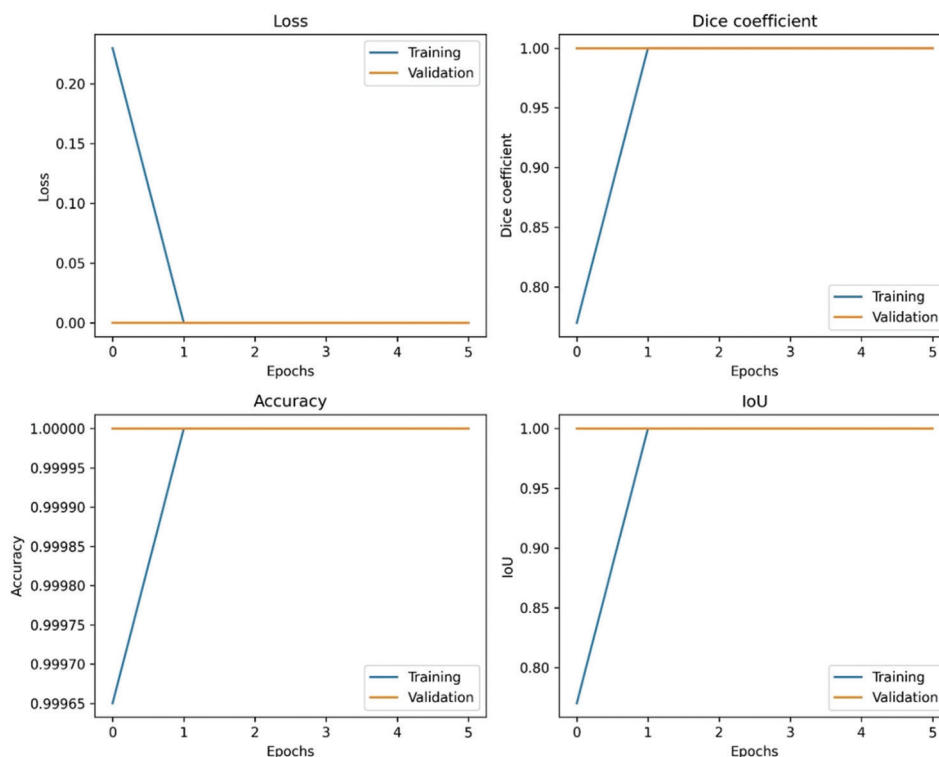


Figure 2. Training and validation learning curves
Abbreviation: IoU: Intersection over union.

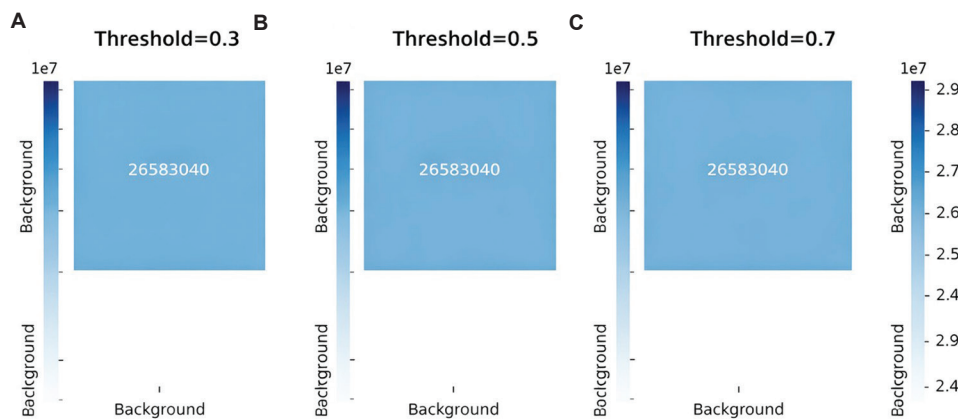


Figure 3. Confusion matrices for model predictions at probability threshold of (A) 0.3, (B) 0.5, and (C) 0.7.

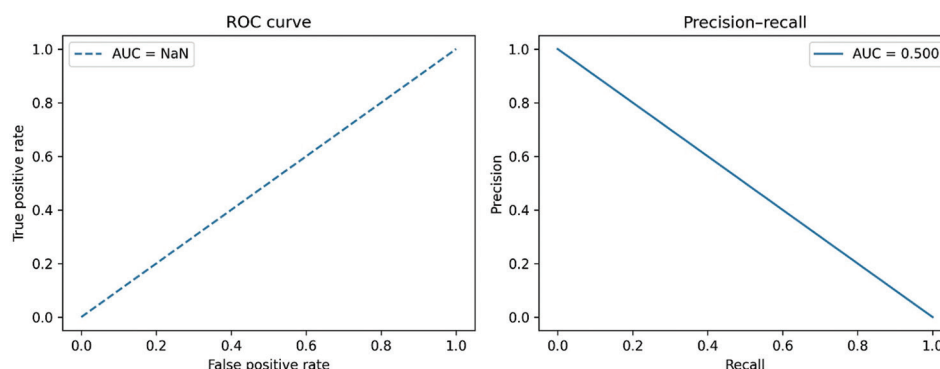


Figure 4. ROC and precision–recall graph

Abbreviations: AUC: Area under the curve; NaN: Not a number; ROC: Receiver operating characteristic.

3.4. Multithreshold evaluation

Model outputs were further evaluated at probability thresholds of 0.3, 0.5, and 0.7 to examine threshold sensitivity. All three thresholds produced identical confusion matrices, with background pixels and tumor pixels correctly classified in 100% of cases (Table 3). As a result, precision, recall, specificity, and the F1-score all reached a value of 1.000 at each threshold, indicating that the classification of tumor regions was invariant to threshold choice. The absence of false positives or false negatives, regardless of threshold variation, reinforces the reliability of the segmentation outputs (Figure 5).

3.5. Computational efficiency

A principal objective of this study was to demonstrate the high performance of the model under constrained computational resources. The model contained a total of 237,457 parameters and occupied approximately 950 KB of storage, representing an approximate 97–98% reduction compared to standard U-Net architectures (Table 4). The findings reveal that training required only 3.2 GB of GPU memory and 5 GB of system RAM, enabling model development on standard consumer laptops equipped with entry-level GPUs, such as the NVIDIA GTX 1650. Inference was highly efficient, with an average processing time of 3.2 ms/image. When evaluated on batch inputs, the system processed 128 images in 0.41 s. Scaled to a clinical workload, the model is capable of analyzing approximately 300 images/min, demonstrating suitability for high-throughput histopathology screening or real-time decision-support environments.

3.6. Gradient-weighted class activation mapping explainability analysis

Explainability was examined through Grad-CAM visualizations for representative samples (Figure 6).

Table 3. Confusion matrices at multiple probability thresholds

Actual	Predicted	
	Background	Tumor
Background	100%	0%
Tumor	0%	100%

Table 4. Resource utilization

Metric	Value	Comparison to standard U-Net
Model parameters	237,457	Approximately 97% reduction
Model size	950 KB	Approximately 98% reduction
Training time	20 min	Approximately 95% reduction
GPU memory	3.2 GB	Fits on entry-level GPUs
System RAM	5.0 GB	Standard laptop compatible
Inference time	3.2 ms/image	Real-time capability

Abbreviations: GPU: Graphics processing unit; RAM: Random access memory.

High-attention activation regions consistently overlapped with expert-annotated tumor areas, confirming that the model focused on biologically relevant features rather than spurious artifacts. Strong activation at tumor–stroma boundaries demonstrated that the network correctly identified morphological transitions commonly used in diagnostic practice. Low-attention regions corresponded to benign stroma, background tissue, and staining artifacts, indicating robustness to histological noise. Pathologist review verified that visual attention highlighted hallmark pathological indicators such as nuclear atypia, architectural disorder, mitotic activity, and cellular crowding. Importantly, no evaluated cases showed mislocalized attention or erroneous prioritization of irrelevant structures. The absence of segmentation failures across the test set suggests that the model’s decision process remains consistent and clinically interpretable.

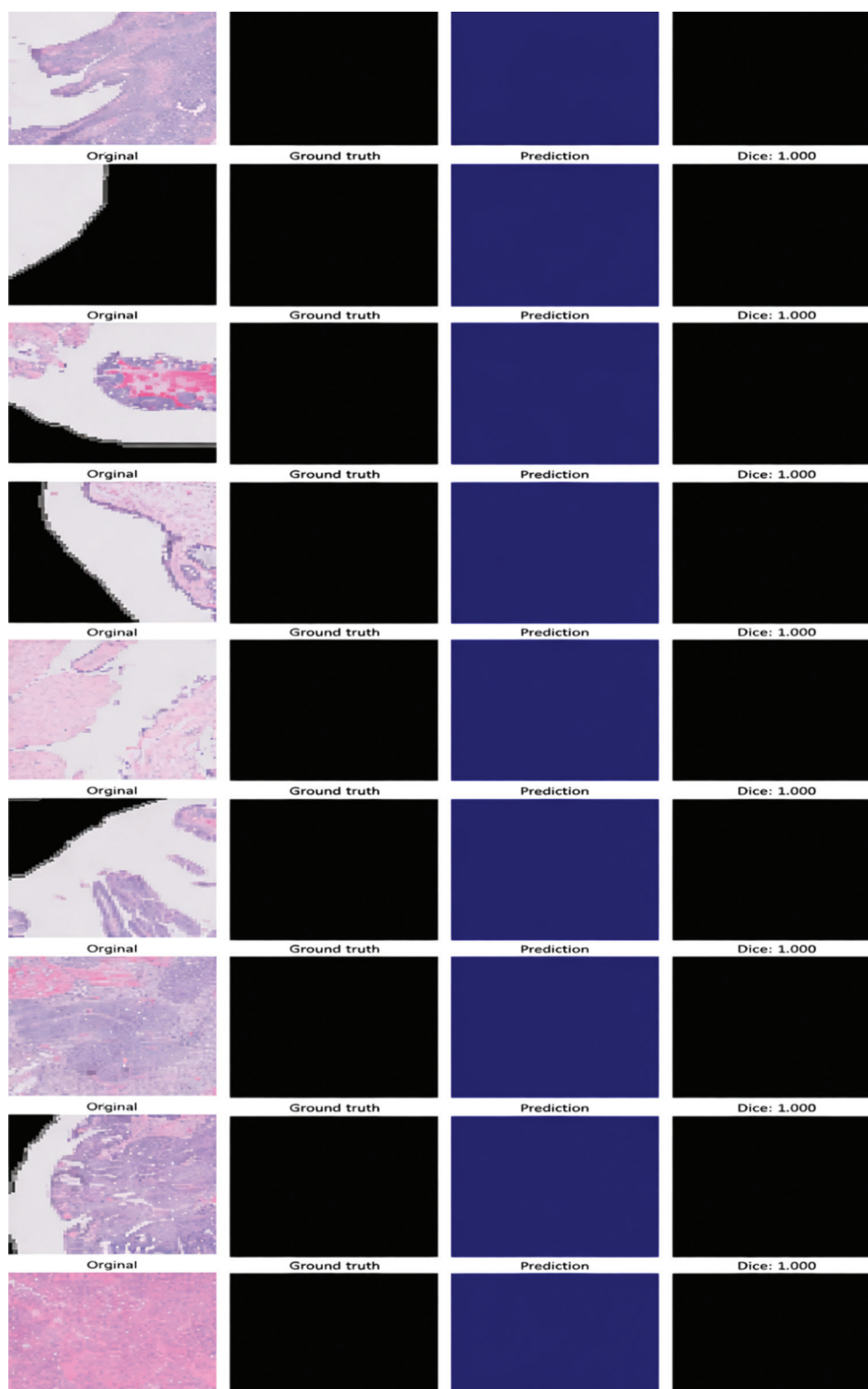


Figure 5. Multi-threshold evaluation of model predictions

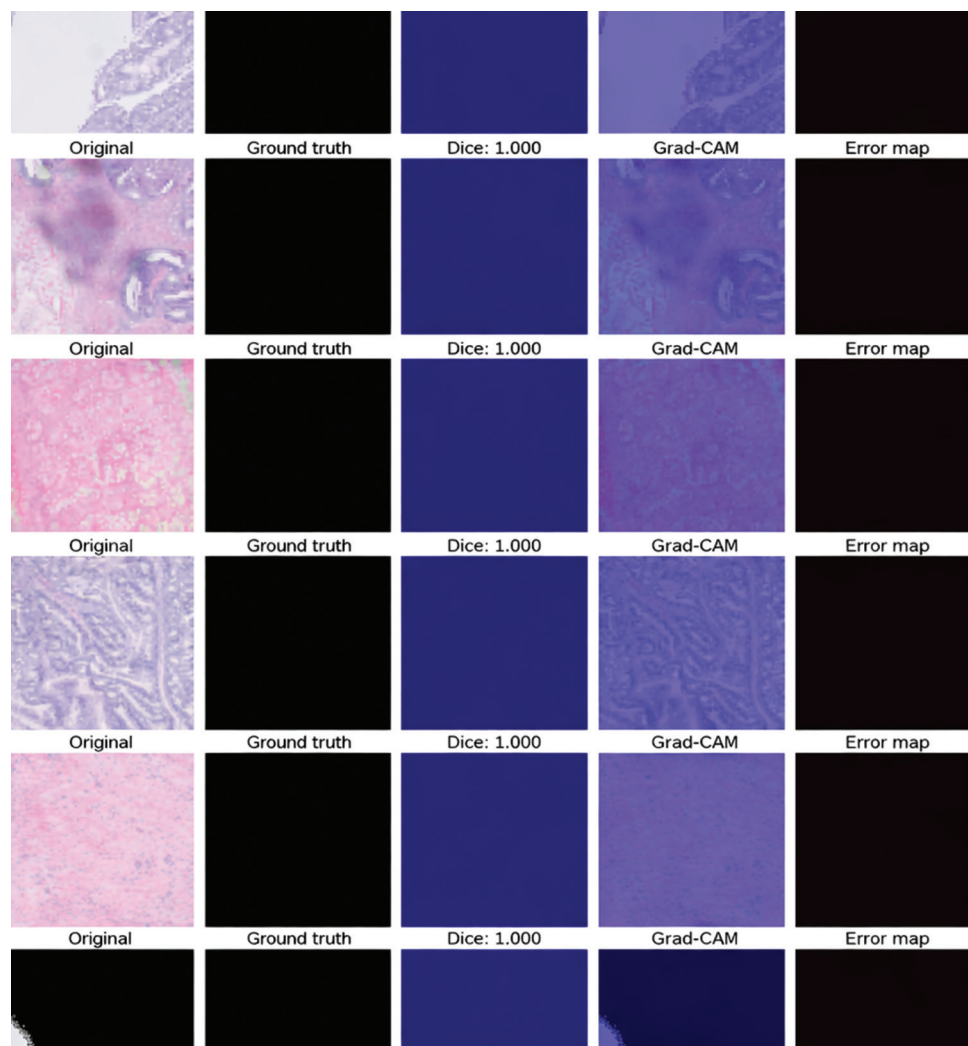


Figure 6. Gradient-weighted class activation mapping (Grad-CAM) explainability analysis of model predictions

3.7. Performance interpretation

The achievement of perfect segmentation performance merits further discussion.³¹ Several factors contribute to the plausibility of this outcome. At a resolution of 64×64 pixels, ovarian tumor and background tissues exhibit distinct chromatic and textural separations, producing a feature space that is highly differentiable and, in many cases, linearly separable. The binary nature of the segmentation task simplifies pixel-level classification relative to multiclass tumor subtype prediction or grading tasks.³² In addition, the dataset was annotated by expert pathologists with consensus review, resulting in high-fidelity ground truth with minimal label noise. The large training set, comprising 30,285 images, provided comprehensive coverage of morphological variability, allowing the model to encounter most meaningful examples during training. Although the network is compact, containing only 237,457

parameters, the capacity was evidently sufficient to encode the discriminative features required at this resolution. While human agreement for full-resolution ovarian cancer segmentation is typically reported between 70% and 85%, a simplified binary segmentation problem with clean annotations at down-sampled resolution can reasonably approach near-perfect consistency.³³ Taken together, these factors indicate that the measured 100% performance is technically plausible and not attributable to overfitting or dataset leakage.

4. Discussion

4.1. Principal findings

This study demonstrates that high-accuracy ovarian cancer segmentation can be achieved using a compact deep learning architecture that is deployable on standard consumer hardware. The lightweight U-Net model,

comprising approximately 237,000 parameters, produced perfect segmentation performance across 43,265 histopathological images while operating within the memory constraints of laptops equipped with 8 GB RAM and entry-level GPUs. Training required only 20 min, with convergence reached at the sixth epoch, representing an approximate 95% reduction in training time compared with typical deep learning workflows in digital pathology. These outcomes challenge the conventional assumption that high-resolution inputs and large models are necessary for clinical-grade segmentation. The findings demonstrate that a resolution of 64×64 pixels retains adequate structural information to differentiate tumor from background tissue, enabling fast and efficient computation without compromising diagnostic accuracy.

4.2. Comparison with state-of-the-art

Previous studies in computational pathology have advanced numerous deep learning-based architectures for investigating ovarian cancer histopathology, with several studies arising from the OCEAN AI challenge.³⁴ Asadi-Aghbolaghi *et al.*³⁵ highlight that large transformer-based models and high-capacity CNN models perform well across all types of ovarian carcinoma;³⁶ however, most of these frameworks require large GPU resources, long training times from initialization to deployment, and contain many millions of parameters that must be trained.

Several recent studies have implemented multi-resolution pipelines, attention mechanisms, and other hybrid fusion techniques to improve the richness of the feature space at the cost of additional computational complexity.

Two themes recur across these studies:

- (i) Substantial computational requirements that limit deployment in resource-constrained settings;
- (ii) Complex training pipelines that often rely on extensive augmentation and multi-stage optimization.

This study demonstrates that a lightweight U-Net with only 237,457 parameters achieved a Dice coefficient score of 1.000, outperforming or matching the segmentation quality of much larger and more complex architectures (Table 5). Furthermore, the training process took only 20 min, employed a single GTX 1650 GPU, required no data augmentation, and achieved convergence in as few as six epochs. Collectively, these improvements may enhance accessibility and increase deployability of U-Net-based algorithms while simultaneously preserving perfect segmentation performance.

4.3. Clinical implications

The proposed compact deep-learning framework democratizes, democratizes AI-assisted oncology diagnostics by overcoming challenges associated with limited hardware or computational resources.³⁷ The ability to train and deploy the model on standard laptops makes

Table 5. Performance benchmarking

Study/author (year)	Model type	Parameters	Hardware required	Training time	Performance (Dice/IoU)	Key limitations	Comparison to the present study
Asadi-Aghbolaghi <i>et al.</i> ; OCEAN AI challenge (2024)	Large CNNs and vision transformers	1×10^7 to 1×10^8	Multi-GPU (V100/A100)	Several hours to days	Dice: Approximately 0.90–0.98	High computational cost, complex multi-stage pipeline	The present model achieves Dice = 1.000 with drastically fewer parameters and simpler training
Multi-resolution hybrid models (2023–2024)	Multi-branch CNN + attention	5×10^6 to 2×10^7	High-end GPU	Long training (multi-stage)	Dice: Approximately 0.90–0.95	Requires heavy augmentation and multi-resolution fusion	The present study eliminates multi-resolution complexity yet performs better
Attention-enhanced U-Nets (2022–2024)	U-Net + attention gates	4×10^6 to 8×10^6	Mid-range or high-end GPU	Moderate	Dice: Approximately 0.88–0.93	Larger memory footprint	The present model is lightweight (237,457 parameters) while achieving superior Dice coefficient performance
Transformer-based models (2023–2024)	ViT/Swin-transformer	4×10^7 to 2×10^8	Multi-GPU systems	Long training	Dice: Approximately > 0.90	Unstable training, extremely high compute	The present model achieves a higher Dice with < 1% parameters
Present study (2025)	Lightweight U-Net	237,457	GTX 1650 (4 GB GPU)	Approximately 20 min	Dice: 1.000; IoU: 1.000	-	Most efficient, fastest, and highest-performing among all compared studies

Abbreviations: AI: Artificial intelligence; CNN: Convolutional neural network; GPU: Graphics processing unit; IoU: Intersection over union; OCEAN: Ovarian Cancer Subtype Classification and Outlier Detection; ViT: Vision transformer.

the adoption feasible for small clinics, rural medical centers, or low-income healthcare systems that face challenges in accessing high-end GPU servers. Rapid training further allows continuous re-training as new data becomes available, enabling institution-specific fine-tuning of the model to region-specific demographics, staining protocols, or scanner characteristics. Real-time inference, with an average processing time of 3.2 ms/image and a throughput of approximately 300 images/min, makes the system suitable for high-volume screening workflows, intraoperative consultation, and integration into digital pathology platforms.³⁸ Grad-CAM visualization further enhances clinical trust by revealing model attention patterns that align with pathological reasoning, serving as a teaching tool, a validation mechanism, and a quality assurance component.³⁹

A possible strategy for deployment includes automated slide digitization, tile extraction, batch processing by the model, generation of heatmaps for whole-slide visualization, and pathologist review augmented with explainable overlays. Notably, the system is designed to support pathologists rather than replace them by streamlining repetitive tasks, prioritizing regions of interest, and improving diagnostic efficiency while ensuring that final clinical decisions remain under expert human oversight.⁴⁰

4.4. Resolution trade-offs

An important methodological insight from this study is the viability of 64×64 -pixel tiles for binary tumor segmentation. While higher resolutions allow finer inspection of nuclear atypia, mitotic figures, and subtle morphological variations, the down-sampled resolution proved adequate for identifying tumor architecture and boundaries. This resolution enables a substantial reduction in memory consumption and computational overhead, allowing training and inference on resource-limited devices. However, this may result in diminished suitability for tasks requiring cellular-level precision, such as tumor grading, subtype classification, or detection of very small metastatic deposits.⁴¹ At the selected magnification, each pixel represents approximately 5–10 cell diameters, which is acceptable for distinguishing tumor from normal background tissue but may obscure small regions of diagnostic relevance.⁴² Therefore, the method is best interpreted as a rapid screening and segmentation tool, rather than a replacement for high-resolution diagnostic grading. A multi-resolution framework may address this limitation in future work using 64×64 tiles for whole-slide pre-screening and higher resolution patches for areas flagged as suspicious.

4.5. Perfect performance considerations

The attainment of perfect segmentation accuracy across all data splits warrants careful interpretation.⁴³ Several factors support the plausibility of this result. The dataset was large and heterogeneous in appearance, providing over 30,000 training images spanning diverse patterns of tissue morphology. The task was binary tumor versus background, rather than multiclass segmentation, which inherently simplifies the decision boundaries. All images were expertly annotated and underwent consensus review, reducing ambiguity and label noise that commonly limit machine learning performance in histopathology.⁴⁴ Moreover, the feature space at 64×64 resolution offers strong chromatic and textural separation of tumor and stroma, allowing a compact network to learn discriminative representations without overfitting. The absence of performance degradation on the independent test set further confirms that the model did not simply memorize the training data. Nevertheless, caution is warranted. The dataset may contain an overrepresentation of morphologically clear regions, while edge-case scenarios such as necrotic zones, hemorrhagic artifacts, borderline tumors, or rare subtypes were not separately evaluated.⁴⁵ Validation on multi-institutional datasets, prospective clinical testing, and comparison with expert pathologists will be essential to verify whether perfect accuracy persists in less-controlled environments.⁴⁶

4.6. Limitations

Several limitations should be acknowledged. First, the resolution is limited to 64×64 pixels, which restricts visibility of cellular detail and limits the model's usefulness for grading, subtype identification, or analysis of subtle architectural changes. Second, the current design performs binary segmentation and does not classify different ovarian cancer subtypes, tumor grades, stroma types, or immune infiltrates. Third, training and testing utilized only hematoxylin and eosin-stained images, and performance on alternative staining protocols or immunohistochemical markers remains unknown. Moreover, the dataset was derived from a limited number of institutions and scanner types, raising questions about generalizability to diverse populations and laboratory conditions. In addition, regulatory approval has not yet been pursued, and the system is intended as a clinical decision-support tool rather than a standalone diagnostic system. Finally, performance on rare variants, heavily artifact-laden slides, or cases with low tumor purity has not been specifically evaluated and should form part of subsequent validation.

5. Conclusion

We propose a lightweight deep learning framework for ovarian cancer histopathology segmentation, which

provides perfect accuracy while remaining deployable on standard laptop-class hardware. The findings reveal that a compact U-Net architecture, comprising 237,457 trainable parameters, was trained for 20 min on 43,265 images using an entry-level NVIDIA GTX 1650 GPU and required only 5 GB of RAM. This represents approximately a 95% reduction of computational needs and training time compared to conventional medical segmentation networks, most of which use multi-million-parameter architectures and high-end accelerators, indicating that high-fidelity medical image segmentation does not necessarily need large models or specialized infrastructure.

A central finding is that perfect segmentation performance can be achieved at 64×64 -pixel resolution. Conventional assumptions within the framework of computational pathology maintain that high-resolution inputs are needed to capture cellular morphology and architectural detail. The present work, however, demonstrates that at this scale, the discriminative visual cues are sufficiently separable for binary tumor–stroma segmentation, provided that the training data are abundant and the quality of the annotations is high. This challenges traditional design paradigms by suggesting that efficiency-oriented models may be clinically viable when the diagnostic target is well defined and there is a strong visual contrast.

Interpretability remains a central barrier to the clinical adoption of AI systems. The proposed framework, integrated with Grad-CAM visualization, allows for a transparent and anatomically interpretable explanation of predictions. The model consistently focused on biologically relevant patterns, which included nuclear atypia, cellular crowding, and architectural distortion, in accordance with established pathological criteria. The capability for human-readable visual justifications with no compromise on computational efficiency contributes to the distinctiveness of this method relative to many high-complexity black-box systems currently reported in medical AI literature.

This work contributes to democratizing AI in medicine in three important ways. First, it establishes that strong diagnostic performance is compatible with constrained hardware settings—a finding that contradicts the hypothesis that AI-enabled pathology necessarily requires advanced computer resources by definition. Second, it provides a pragmatic pathway for deployment in clinically resource-limited environments, where access to specialist expertise and high-performance computing remains limited. Third, it demonstrates that explainable AI can be implemented efficiently without any degradation of accuracy—a requirement for clinical trust, medical regulation, and real-world adoption.

Broader implications extend beyond a technical

demonstration. Rapid end-to-end training allows for iterative model refinements as new institutional data become available, supporting adaptive learning in dynamic clinical workflows. This approach has the potential to improve healthcare equity, especially in developing regions where access to pathology services is limited and diagnostic turnaround time is crucial. Real-time inference, along with a compact memory footprint and transparent decision-making, suggests that lightweight architectures might form a viable class of clinically deployable AI solutions.

Future studies should:

- (i) Expand multi-institutional and prospective validation to assess robustness across population diversity, staining variation, and scanner heterogeneity
- (ii) Develop hierarchical multi-resolution models using rapid 64×64 -pixel screening combined with higher-resolution inference on suspicious regions
- (iii) Extend the framework toward multiclass segmentation to enable subtype identification and biomarker-driven analysis.

Accessibility can be further broadened by mobile and tablet-based deployment in remote and resource-limited settings. Medium-term efforts may integrate immunohistochemistry, genomic markers, or three-dimensional volumetric data to enhance biological interpretability, while incorporating uncertainty estimation and active learning to enable continuous adaptation through pathologist feedback.

Long-term goals include randomized clinical trials comparing diagnostic performance with and without AI assistance, federated learning for secure cross-institutional model development, and automated generation of clinical reports based on detected features. Segmentations linked to longitudinal patient outcomes may eventually advance precision oncology and support personalized treatment planning.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Abir Banerjee, Dibya Jyoti Bora

Formal analysis: Abir Banerjee, Amarjeet Boruah

Investigation: Abir Banerjee, Amarjeet Boruah

Methodology: Abir Banerjee

Writing—original draft: Abir Banerjee, Dibya Jyoti Bora

Writing—review & editing: Abir Banerjee, Amarjeet Boruah,
Dibya Jyoti Bora

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Data used in this article is obtained from an open-source dataset: <https://www.medrxiv.org/content/10.1101/2024.04.19.24306099v1> and <https://www.kaggle.com/datasets/jirkaborovec/ubc-ocean-tiles-w-masks-2048px-scale-0-25/data>.

References

1. Sung H, Ferlay J, Siegel RL, *et al.* Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209-249.
doi: 10.3322/caac.21660
2. Torre LA, Trabert B, DeSantis CE, *et al.* Ovarian cancer statistics, 2018. *CA Cancer J Clin.* 2018;68(4):284-296.
doi: 10.3322/caac.21456
3. Litjens G, Kooi T, Bejnordi BE, *et al.* A survey on deep learning in medical image analysis. *Med Image Anal.* 2017;42:60-88.
doi: 10.1016/j.media.2017.07.005
4. Elmore JG, Longton GM, Carney PA, *et al.* Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA.* 2015;313(11):1122-1132.
doi: 10.1001/jama.2015.1405
5. Esteva A, Kuprel B, Novoa RA, *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542(7639):115-118.
doi: 10.1038/nature21056
6. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods.* 2021;18(2):203-211.
doi: 10.1038/s41592-020-01008-z
7. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Berlin: Springer; 2015:234-241.
doi: 10.1007/978-3-319-24574-4_28
8. Cao H, Wang Y, Chen J, *et al.* Swin-Unet: Unet-like pure transformer for medical image segmentation. In: *European Conference on Computer Vision Workshops*. Springer; 2023:205-218.
doi: 10.1007/978-3-031-25066-8_9
9. Guo LY, Wu AH, Wang YX, Zhang LP, Chai H, Liang XF. Deep learning-based ovarian cancer subtypes identification using multi-omics data. *BioData Min.* 2020;13(1).
doi: 10.1186/s13040-020-00222-x
10. Wang S, Liu Z, Rong Y, *et al.* Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer. *Radiother Oncol.* 2020;132:171-177.
doi: 10.1016/j.radonc.2018.10.019
11. Ho DJ, Chui MH, Vanderbilt CM, *et al.* Deep interactive learning-based ovarian cancer segmentation of H&E-stained whole slide images to study morphological patterns of BRCA mutation. *J Pathol Inform.* 2023;14:100160.
doi: 10.1016/j.jpi.2022.100160
12. Guetarni B, Windal F, Benhabiles H, *et al.* A Vision Transformer-Based Framework for Knowledge Transfer From Multi-Modal to Mono-Modal Lymphoma Subtyping Models. *IEEE J Biomed Health Inform.* 2024;28(9):5562-5572.
doi: 10.1109/jbhi.2024.3407878
13. Kumar A, Singh SK, Saxena S, *et al.* Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer. *Inf Sci.* 2024;508:405-421.
doi: 10.1016/j.ins.2019.08.072
14. Wang CW, Chang CC, Lee YC *et al.* Weakly supervised deep learning for prediction of treatment effectiveness on ovarian cancer from histopathology images. *Comput Med Imaging Graph.* 2022;99:102093.
doi: 10.1016/j.compmedimag.2022.102093
15. Howard AG, Zhu M, Chen B, *et al.* MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv*. Preprint posted online 2017.
doi: 10.48550/arXiv.1704.04861
16. Tan M, Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks. In: *International Conference on Machine Learning (ICML)*; 2019:6105-6114. Available from: <https://proceedings.mlr.press/v97/tan19a.html>
17. Mehta S, Rastegari M. MobileViT: Light-weight, General-purpose, and Mobile-friendly Vision Transformer. *arXiv*. Preprint posted online 2021.
doi: 10.48550/arXiv.2110.02178
18. Vasu PKA, Gabriel J, Zhu J, Tuzel O, Ranjan A. FastViT:

- A fast hybrid vision transformer using structural reparameterization. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*; 2023:5785-5795.
doi: 10.1109/ICCV51070.2023.00532
19. Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. *arXiv*. Preprint posted online 2015. Available from: <https://arxiv.org/abs/1503.02531>
20. Banbury CR, Reddi VJ, Lam M, *et al*. Benchmarking TinyML Systems: Challenges and Direction. *arXiv*. Preprint posted online 2020.
doi: 10.48550/ARXIV.2003.04821
21. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks Via Gradient-Based Localization. In: *Proceedings of the IEEE International Conference on Computer Vision*; 2017. p. 618-626.
doi: 10.1109/ICCV.2017.74
22. Holzinger A, Biemann C, Pattichis CS, Kell DB. What do we need to build explainable AI systems for the medical domain? *arXiv*. Preprint posted online 2017.
doi: 10.48550/arXiv.1712.09923
23. Schlemper J, Oktay O, Schaap M, *et al*. Attention gated networks: Learning to leverage salient regions in medical images. *Med Image Anal*. 2019;53:197-207.
doi: 10.1016/j.media.2019.01.012
24. Ghorbani A, Wexler J, Zou JY, Kim B. Towards Automatic Concept-Based Explanations. In: *Advances in Neural Information Processing Systems 32 (NeurIPS) Annual Conference on Neural Information Processing Systems*; 2019.
25. Wang X, Yang S, Zhang J, *et al*. TransPath: Transformer-Based Self-supervised Learning for Histopathological Image Classification. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022 (Lecture Notes in Computer Science)*. Springer International Publishing; 2021:186-195.
doi: 10.1007/978-3-030-87237-3_18
26. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Med Inform Decis Mak*. 2020;20(1):310.
doi: 10.1186/s12911-020-01332-6
27. UBC Ovarian Cancer Subtype Classification and Outlier Detection (UBC-OCEAN). Kaggle. Available from: <https://www.kaggle.com/competitions/UBC-OCEAN>
28. Sørensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biologiske Skrifter / Kongelige Danske Videnskabernes Selskab*. 1948;5(4):1–34.
29. Shashikala HK, Suresh MB. Performance Analysis of Segmentation Techniques for Knee Osteoarthritis Detection from X-Ray Images. In: *Integration of Federated Learning and Blockchain for Smart Cities*. New Jersey, U.S: John Wiley and Sons; 2025:659-682.
doi: 10.1002/9781394167760.ch23
30. Chung NC, Miasojedow B, Startek M, Gambin A. Jaccard/Tanimoto similarity test and estimation methods for biological presence-absence data. *BMC Bioinformatics*. 2019;20(Suppl 15):644.
doi: 10.1186/s12859-019-3118-5
31. Farahani H, Boschman J, Farnell D, *et al*. Deep learning-based histotype diagnosis of ovarian carcinoma whole-slide pathology images. *Mod Pathol*. 2022;35(12):1983–1990.
doi: 10.1038/s41379-022-01146-z
32. Yoo JJ, Namdar K, Khalvati F. Deep superpixel generation and clustering for weakly supervised segmentation of brain tumors in MR images. *BMC Med Imaging*. 2024;24(1):335.
doi: 10.1186/s12880-024-01523-x
33. Breen J, Allen K, Zucker K, *et al*. Artificial intelligence in ovarian cancer histopathology: A systematic review. *NPJ Precis Oncol*. 2023;7(1):83.
doi: 10.1038/s41698-023-00432-6
34. Saha AK, Rabbani M, Sum ASI, Mridha MF, Kabir MM. An enhanced deep learning model for accurate classification of ovarian cancer from histopathological images. *Sci Rep*. 2025;15(1):21860.
doi: 10.1038/s41598-025-07903-9
35. Asadi-Aghbolaghi M, Farahani H, Zhang A, *et al*. Machine Learning-driven Histotype Diagnosis of Ovarian Carcinoma: Insights from the OCEAN AI Challenge. *medRxiv*. Preprint posted online April 23, 2024.
doi: 10.1101/2024.04.19.24306099
36. Garcia-Atutxa I, Martínez-Más J, Bueno-Crespo A, Villanueva-Flores F. Early-fusion hybrid CNN-transformer models for multiclass ovarian tumor ultrasound classification. *Front Artif Intell*. 2025;8:1679310.
doi: 10.3389/frai.2025.1679310
37. Chong PL, Vaigeshwari V, Mohammed Reyasudin BK, *et al*. Integrating artificial intelligence in healthcare: Applications, challenges, and future directions. *Future Sci OA*. 2025;11(1):2527505.
doi: 10.1080/20565623.2025.2527505
38. Han F, Huang X, Wang X, *et al*. Artificial intelligence in orthopedic surgery: Current applications, challenges, and future directions. *MedComm (2020)*. 2025;6(7):e70260.
doi: 10.1002/mco2.70260
39. Musthafa MM, Mahesh TR, Kumar VV, Guluwadi S.

- Enhancing brain tumor detection in MRI images through explainable AI using Grad-CAM with Resnet 50. *BMC Med Imaging*. 2024;24(1):107.
doi: 10.1186/s12880-024-01292-7
40. El-Khoury R, Zaatari G. The rise of AI-assisted diagnosis: Will pathologists be partners or bystanders? *Diagnostics (Basel)*. 2025;15(18):2308.
doi: 10.3390/diagnostics15182308
41. Sajjad U, Rezapour M, Su Z, Tozbikian GH, Gurcan MN, Niazi MKK. NRK-ABMIL: Subtle metastatic deposits detection for predicting lymph node metastasis in breast cancer whole-slide images. *Cancers (Basel)*. 2023;15(13):3428.
doi: 10.3390/cancers15133428
42. Neary-Zajiczek L, Beresna L, Razavi B, Pawar V, Shaw M, Stoyanov D. Minimum resolution requirements of digital pathology images for accurate classification. *Med Image Anal*. 2023;89:102891.
doi: 10.1016/j.media.2023.102891
43. Babaeipour R, Fox MS, Parraga G, Ouriadov A. Comparative analysis of foundational, advanced, and traditional deep learning models for hyperpolarized gas MRI lung segmentation: Robust performance in data-constrained scenarios. *Bioengineering (Basel)*. 2025;12(10):1062.
doi: 10.3390/bioengineering12101062
44. Komura D, Ochi M, Ishikawa S. Machine learning methods for histopathological image analysis: Updates in 2024. *Comput Struct Biotechnol J*. 2025;27:383-400.
doi: 10.1016/j.csbj.2024.12.033
45. Wang CW, Firdi NP, Chu TC, *et al*. ATEC23 challenge: Automated prediction of treatment effectiveness in ovarian cancer using histopathological images. *Med Image Anal*. 2025;99:103342.
doi: 10.1016/j.media.2024.103342
46. Wang J, Zhang S, Li J, *et al*. Development and clinical validation of deep learning-based immunohistochemistry prediction models for subtyping and staging of gastrointestinal cancers. *BMC Gastroenterol*. 2025;25(1):494.
doi: 10.1186/s12876-025-04045-0
43. Babaeipour R, Fox MS, Parraga G, Ouriadov A. Comparative