Journal of Computer Science and Technological Enquiry
Vol: 1, Issue: 1 (2025)
https://doi.org/10.65496/jcste.2025.44

noēsis

ORIGINAL RESEARCH - OPEN ACCESS

Explainable Transfer Learning for Histopathological Image Classification of Ovarian Cancer

[1]Md. Mahbubur Rahman, [1]Sadia Zannat Reem, [1]Md. Atikqur Rahaman, [1]Chinmay Bepery, [2]Md. Ashiqussalehin, [1]Mohammad Jamal Hossain

[1]*Department of Computer Science and Information Technology, Patuakhali Science and Technology University, Patuakhali, Bangladesh*

[2]*Department of Internet of Things and Robotics Engineering, University of Frontier Technology, Bangladesh*

Corresponding Author:
Md. Mahbubur Rahman
E-mail: mahbub.cse@pstu.ac.bd

Article History
Received: 13 August 2025
Revised: 25 August 2025
Accepted: 28 August 2025
Published: 03 September 2025

How to Cite:
M. M. Rahman, S. Z. Reem, M. A. Rahaman, C. Bepery, M. Ashiqussalehin, and M. J. Hossain, "Explainable Transfer Learning for Histopathological Image Classification of Ovarian Cancer," Journal of Computer Science and Technological Enquiry (JCSTE), pp. 35–47. https://doi.org/10.65496/jcste.2025.44

ABSTRACT

Timely diagnosis and effective treatment of ovarian cancer depend heavily on accurate subtype classification. This study presents a deep learning model that utilizes transfer learning with a customized DenseNet201 architecture. Several subtypes of ovarian cancer are classified by the proposed model using histopathology images. Two balanced, openly accessible datasets are used for training and validation. In order to enhance generalization and avoid overfitting, data augmentation approaches are used during preprocessing. Accuracy and AUC are two of the measures used to assess the performance of the model. Interpretability is improved through the use of Local Interpretable Model-agnostic Explanations (LIME). LIME offers graphical representations of the model's decision-making procedure. For clinical applications where transparency is required, this capability is essential. On each dataset, experimental results demonstrate that the proposed model performs better than the state-of-the-art approaches currently in use. Every metric show that it performs better in classification. Even with variable input, the model exhibits good robustness and steady accuracy. The efficacy of transfer learning and the prospect of explainable AI in medical image classification are demonstrated by these findings. In particular, it emphasizes how useful it is for diagnosing ovarian cancer. The results highlight how combining explainable AI and deep learning methods can yield dependable, scalable solutions for precise and early ovarian cancer subtype diagnosis. This may increase the likelihood of timely diagnosis and individualized treatment regimens. Additionally, the finding opens the door for further research into healthcare technology and automated diagnostic tools for oncology.

KEYWORDS
Ovarian Cancer, Early Detection, XAI, Transfer Learning, DenseNet

1. Introduction

Ovarian cancer is among the deadliest gynecological malignancies. Late-stage diagnosis and the complexity of its subtypes contribute to its severity. It is the eighth most common cancer among women worldwide [1]. In 2020, approximately 313,959 new cases of ovarian cancer were reported globally. During the same year, there were around 207,252 deaths attributed to ovarian cancer, accounting for 1.6% of all cancer cases and 2.1% of cancer-related deaths [1]. In 2022, the number of new cases increased slightly to 324,398, with 206,839 deaths recorded [2]. Ovarian cancer has a

particularly high mortality rate in lower-income countries like Bangladesh, where late diagnosis and limited access to treatment are major challenges. In the South-Central Asia region, which includes Bangladesh, ovarian, cervical, and breast cancers are leading causes of female cancer mortality due to healthcare limitations. If early detection and care services are not strengthened, the burden of ovarian cancer is expected to rise significantly by 2040[1, 2]. However, the prognosis remains poor, as most women are diagnosed with advanced stages of the disease [3]. Ovarian cancer includes malignant tumors that originate in the ovaries, fallopian tubes, and peritoneum. By the time it's diagnosed, the cancer is often already spread within the abdomen, typically at FIGO Stage 3 [4]. Most ovarian cancers are carcinomas that originate in the epithelial cells. Five primary histological subtypes are recognized among these cancers: high-grade serous, low-grade serous, clear cell, endometrioid, and mucinous. Non-epithelial ovarian cancers, although rare, include germ cell tumors, stromal tumors, sex cord tumors, and mesenchymal tumors. Each subtype exhibits distinct morphological features, prognosis, and treatment strategies [5].

The diagnosis of ovarian cancer primarily relies on histopathology, which involves examining tissue samples at the cellular level. However, a global shortage of pathologists leads to significant disparities in diagnostic capabilities between countries. Even in well-resourced regions, the demand for pathologists continues to exceed supply, creating delays and challenges in timely diagnosis [6]. This limitation, combined with the subjectivity of human-dependent diagnosis, contributes to the risk of misdiagnosis, further worsening the poor prognosis of ovarian cancer. Clinical assessments often depend heavily on a pathologist's experience, which can introduce bias and human error [7]. This highlights a critical need for advanced techniques that not only improve diagnostic accuracy but also support faster, more reliable detection.

2. Literature Review

In recent years, AI, particularly machine learning and deep learning, has increasingly been applied in the medical field. These technologies provide fast, accurate, and computer-driven methods for analyzing complex data patterns and making automatic predictions. With ongoing training, ML and DL models can be continuously refined to deliver even better performance. The application of AI has significantly improved the accuracy of diagnostic and prognostic prediction tests. It also provides valuable treatment recommendations and helps ease the heavy workload on physicians, further boosting the reliability of clinical assessments [7]. Studies employing pre-trained models such as ResNet-50 and DenseNet-161 have demonstrated impressive accuracy in the classification of histopathology images [8].

Wu et al. [9] employed a deep convolutional neural network (DCNN) based on AlexNet to analyze cytological images for identifying ovarian cancer subtypes. By applying data augmentation techniques, they boosted the model's accuracy from 72.76% to 78.20%, showcasing the potential of deep learning to significantly improve diagnostic precision. Guo et al. [10] combined multi-omics data with deep learning techniques to identify ovarian cancer subtypes, leading to improved classification performance. Similarly, Wu et al. [11] applied DCNN to classify various histologic types of ovarian tumors in ultrasound images. Their ResNet50 model achieved an impressive overall accuracy of 95.2%, underscoring the power of deep learning in non-invasive diagnostic imaging. Breen et al. [12] introduced a graph-based model that processes histopathological images at multiple magnifications to

classify ovarian cancer subtypes. The model exhibited balanced accuracies of 73% in cross-validation, 88% in hold-out testing, and 99% in external validation, emphasizing its strong potential for clinical use. Farahani et al. [13] developed deep convolutional neural network models to classify hematoxylin and eosin-stained whole-slide images of ovarian carcinoma. The best-performing model achieved diagnostic concordance of 81.38% on the training set and 80.97% on an independent test set, highlighting the feasibility of AI-assisted histotype diagnosis. In a comprehensive evaluation, Breen et al. [14] evaluated several histopathology foundation models for ovarian cancer subtyping. They found that the H-optimus-0 model achieved balanced accuracies of 89% in cross-validation, 97% in hold-out testing, and 74% in external validation. Behera et al. [15] applied a fine-KNN technique combined with EfficientNet-B0 to classify histopathological images into five distinct ovarian cancer subtypes. Across all subtypes, their model reached AUC values of 0.94, 0.78, 0.69, 0.92, and 0.94, showcasing remarkable performance. Finally, Breen et al. [16] proposed the Discriminative Region Active Sampling for Multiple Instance Learning (DRAS-MIL) method, which achieved a cross-validated AUC of 0.8679.

This approach significantly reduced memory usage and computation time compared to exhaustive slide-level analysis. Despite advances in deep learning for ovarian cancer classification, several research gaps still exist. Most studies focus on binary or limited multiclass tasks using a single dataset, which affects model generalizability. Many models also lack interpretability, which presents challenges for clinical adoption. Moreover, the combination of transfer learning and explainable AI (XAI) is still underexplored, especially with balanced, multi-class datasets. This study addresses these issues by applying transfer learning with a custom DenseNet201, fine-tuned on two distinct five-class histopathological datasets. LIME is used to interpret predictions, with the aim of building a robust, accurate and transparent system to support clinical diagnosis.

3. Statement of the Problem

Ovarian cancer is one of the deadliest gynecological malignancies. Late diagnosis and the difficulty of classifying its histological subtypes make it more dangerous. Traditional pathology methods are time-consuming, subjective, and often inconsistent. Existing deep learning models also face challenges with poor generalization and limited clinical use. The heterogeneity of histopathology images makes classification even more complex. In addition, most AI models work as “black boxes,” which reduces transparency and trust among clinicians. There is, therefore, a strong need for robust and interpretable deep learning methods that can deliver accurate, transparent, and clinically relevant classification of ovarian cancer subtypes. Such models can support early diagnosis and improve treatment planning.

4. Objectives of the Study

The major objectives of the study were:

1. to develop a transfer learning-based deep learning model with a customized DenseNet201 architecture for accurate classification of ovarian cancer subtypes using histopathology images.
2. to improve model generalization and reliability through balanced datasets, data augmentation strategies, and evaluation using Accuracy and AUC metrics.
3. to incorporate interpretability using LIME in order to provide transparent and clinically relevant insights into the model’s decision-making process.

5. Research Questions

The study is guided with the following research questions:

1. How effectively can a DenseNet201-based transfer learning model classify ovarian cancer subtypes using histopathology images?
2. To what extent do data augmentation and balanced datasets improve the generalization and reliability of the proposed model?
3. How does the integration of LIME enhance the interpretability and clinical relevance of the model's predictions?

6. Research Hypothesis

The following null hypothesis was examined with significance set at 5% alpha:

Ho_1. The proposed custom DenseNet201-based transfer learning model does not significantly outperform existing state-of-the-art approaches in ovarian cancer subtype classification using histopathology images.

7. Theoretical Framework

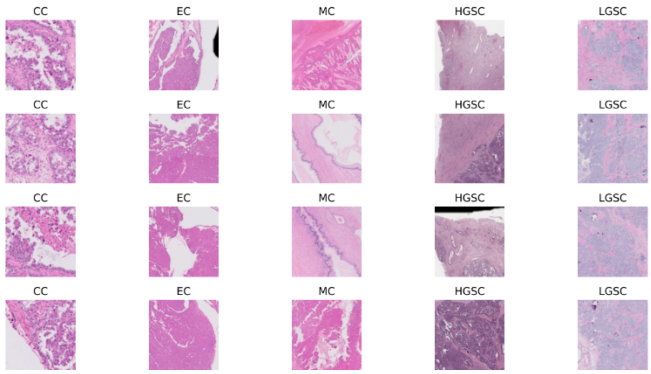
This study is based on deep learning, transfer learning, and explainable artificial intelligence (XAI). CNNs are used to automatically extract features from images. They learn both low-level features, such as edges and textures, and high-level features, such as tissue structures. This study employed DenseNet201, a pre-trained ImageNet variant of DenseNet [17], that connects each layer to all subsequent layers. This dense connectivity improves gradient flow, reduces redundant features, and enhances feature reuse. It is particularly suitable for analyzing complex histopathological images of ovarian cancer.

Transfer learning is used to leverage a DenseNet201 model pre-trained on ImageNet. This allows the model to apply previously learned visual features to new, domain-specific data. It reduces the need for large labeled datasets and speeds up training. Model interpretability is provided through LIME which uses simple surrogate models to explain predictions locally. It highlights key image regions that influence classification. This makes the model more transparent and trustworthy for clinical use. Together, these principles provide a solid foundation for accurate, robust, and interpretable ovarian cancer subtype classification.

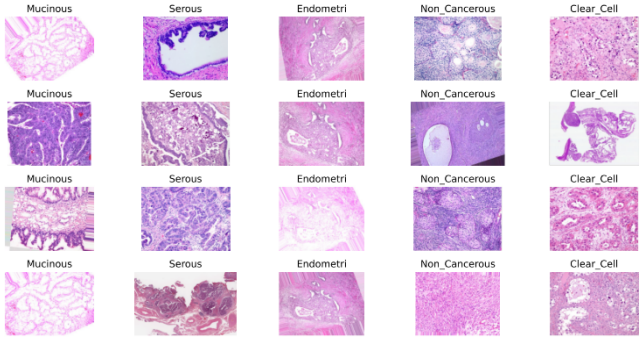
8. Data and Methodology

This study adopted an experimental research design suitable for developing and evaluating deep learning models for ovarian cancer subtype classification. Two publicly available histopathology datasets were used. The first dataset, Ovarian Cancer Subtype Classification [18], contains 34,000 high-resolution images across five ovarian cancer subtypes: high-grade serous carcinoma (HGSC), low-grade serous carcinoma (LGSC), endometrioid carcinoma (EC), clear-cell carcinoma (CC), and mucinous carcinoma (MC). A balanced subset of 2,500 images was selected, with 500 images per class for effective training and validation. The second dataset, Ovarian Cancer & Subtypes Dataset Histopathology [19], contains 498 images. This includes 100 non-cancerous images and 398 images representing four ovarian cancer subtypes: endometrioid carcinoma (98), clear-cell carcinoma (100), mucinous carcinoma (100), and serous carcinoma (100). This dataset provides additional diversity and allows

comparative analysis between cancerous and non-cancerous samples. Sample images from both datasets are shown in Figure 1.



(a) First set of images [17]



(b) Second set of images [18]

Figure 1: Sample images from the ovarian cancer datasets

All images from both datasets were preprocessed for consistency. Images were resized to 224×224 pixels and normalized to a [0,1] scale. Data augmentation was applied to improve model generalization and prevent overfitting. Transformations included rotation (±15°), width and height shift (±15%), zoom and shear (±15%), and horizontal flipping. A 5-fold stratified cross-validation strategy was used to maintain class distribution across training and validation sets.

A customized DenseNet201 model was used for feature extraction. Dense connections allow each layer to receive inputs from all previous layers, improving gradient flow and feature reuse. A classification head was added, consisting of two dense layers (128 and 64 neurons), two dropout layers (0.15 rate), and a softmax output layer with five neurons corresponding to the ovarian cancer subtypes. Transfer learning and fine-tuning enabled the model to capture both low-level and high-level features from histopathology images.

The model was trained using the Adam optimizer with a learning rate of 0.00001 and categorical cross-entropy loss. Training was performed on an NVIDIA A100 GPU via Google Colab Pro. Batch size was set to 32, and the model was trained for 50 epochs. Performance metrics included accuracy, precision, recall, F1-score, sensitivity, specificity, loss, and ROC-AUC. Stratified 5-fold cross-validation ensured robust and consistent evaluation across both datasets.

LIME was used to interpret the model's predictions. LIME builds a local surrogate model to highlight key regions of histopathology images that influence classification. This provides transparency and allows clinicians to understand and validate the model's decisions. Integrating interpretability enhances the potential clinical applicability of the proposed approach. Figure 2 illustrates the research workflow, outlining the sequential steps from data processing to model implementation and performance evaluation.

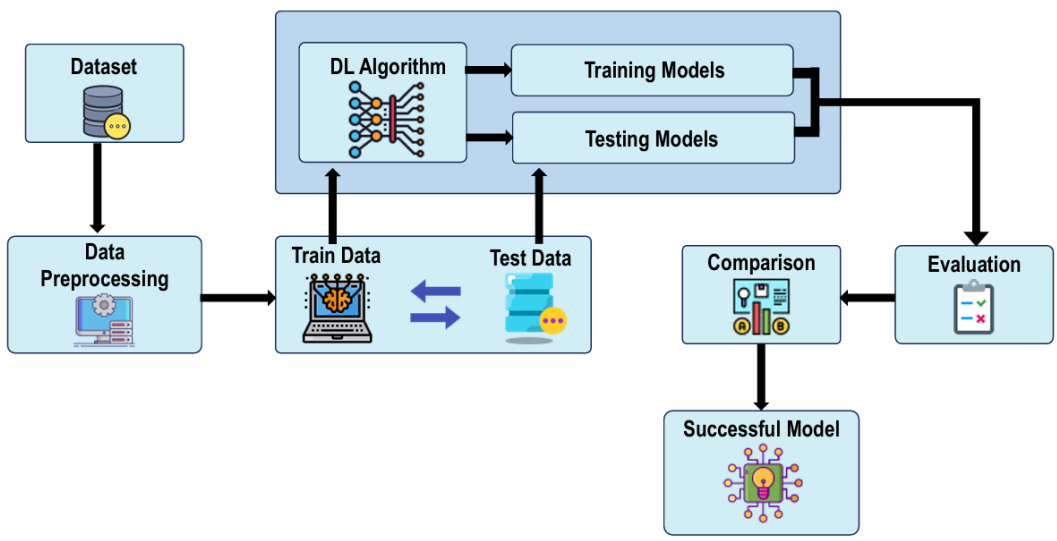


Figure 2: Flow diagram of the study

9. Results

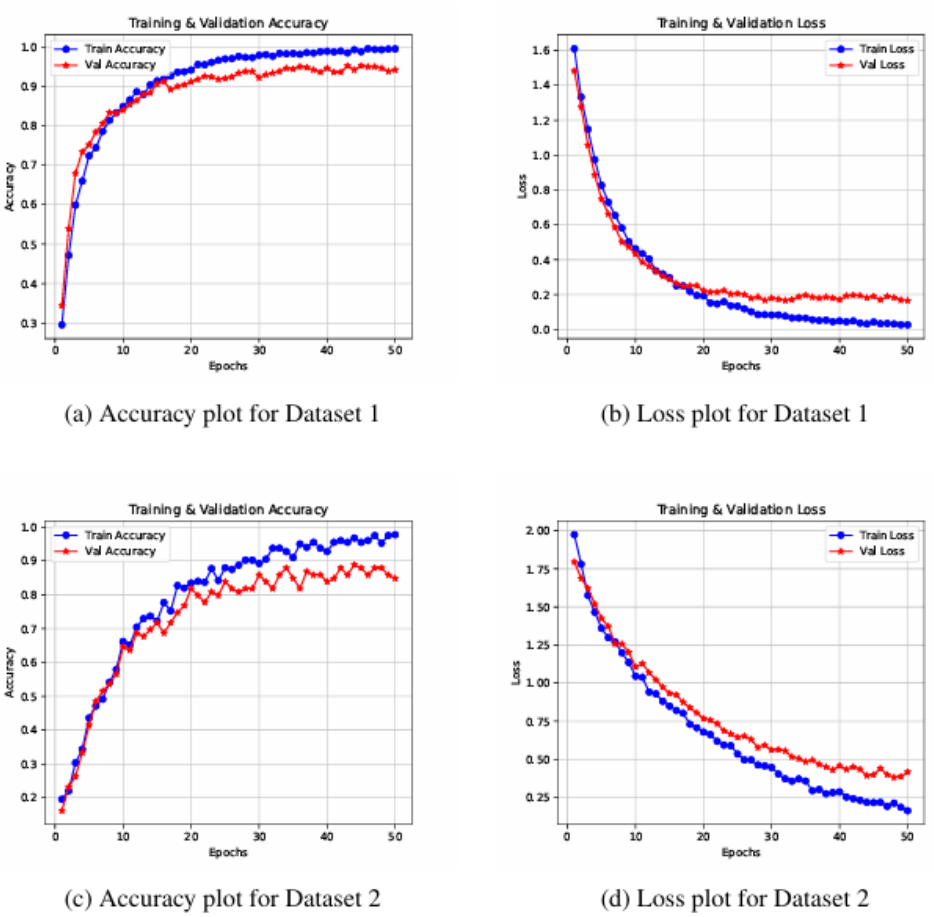
Research Question 1: How effectively can a DenseNet201-based transfer learning model classify ovarian cancer subtypes using histopathology images?

To investigate the first research question, the performance of the proposed DenseNet201-based model in classifying ovarian cancer subtypes was evaluated. The model was assessed using key metrics such as accuracy, AUC, precision, recall, f1-score, and loss [20], and its predictions were compared with baseline models. Detailed results are visualized through confusion matrices and performance plots, while the accompanying table summarizes the quantitative outcomes. These analyses provide a comprehensive view of the model's classification capability and demonstrate its effectiveness in distinguishing between ovarian cancer subtypes.

Table 1: *Performance Comparison of Models on Two Datasets*

Dataset	Model	Average Accuracy	Average AUC	Loss
Dataset 1	**Custom DenseNet201**	0.9564 ± 0.0048	0.9966 ± 0.0005	0.1658
	InceptionV3	0.9496 ± 0.0064	0.9945 ± 0.0015	0.1440
	ResNet50V2	0.9252 ± 0.0126	0.9903 ± 0.0038	0.4026
	MobileNetV2	0.8900 ± 0.0200	0.9856 ± 0.0052	0.4517
Dataset 2	**Custom DenseNet201**	0.9017 ± 0.0253	0.9839 ± 0.0071	0.3335
	InceptionV3	0.8395 ± 0.0293	0.9649 ± 0.0097	0.4532
	ResNet50V2	0.8435 ± 0.0209	0.9723 ± 0.0090	0.5242
	MobileNetV2	0.6105 ± 0.0408	0.8490 ± 0.0286	1.4001

Table 1 summarizes the performance of four deep learning models—CustomDenseNet201, InceptionV3, ResNet50V2, and MobileNetV2—across two ovarian cancer histopathology datasets. On Dataset 1, the Custom DenseNet201 model achieved the best performance. It reached an accuracy of 95.64%, an AUC of 0.9966, and the lowest loss of0.1658. On Dataset 2, it also outperformed the other models, achieving 90.17% accuracy and an AUC of 0.9839. MobileNetV2 showed the weakest performance, especially on Dataset 2, suggesting limitations in handling complex image variations. Overall, the Custom DenseNet201 model proved more robust and generalizable across both datasets.



(a) Accuracy plot for Dataset 1

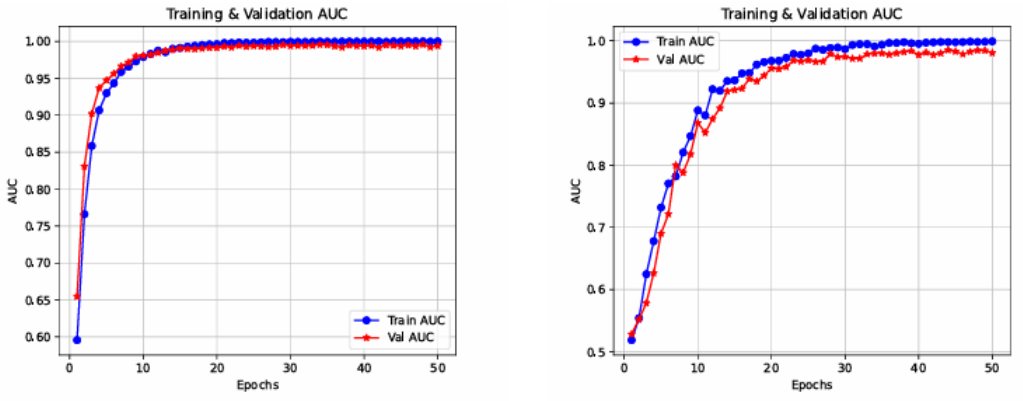
(b) Loss plot for Dataset 1

(c) Accuracy plot for Dataset 2

(d) Loss plot for Dataset 2

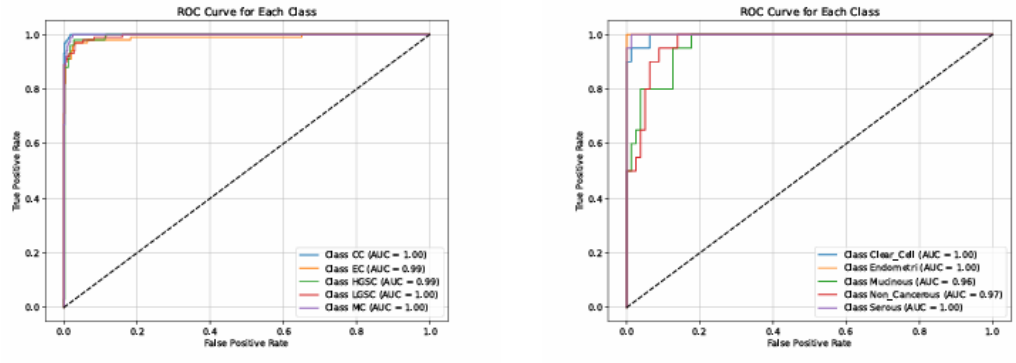
Figure 3: Training Accuracy and Loss Curves for Both Datasets

Figure 3 illustrates the training and validation accuracy and loss curves for the four models on both datasets. Subfigures 3(a) and 3(b) show the accuracy and loss plots for Dataset 1, while 3(c) and 3(d) represent the same for Dataset 2. These curves provide insights into the learning behavior, generalization ability, and convergence patterns of each model throughout training. Figure 4 presents the AUC curves of the proposed model. Figure 4(a) shows the AUC curve for Dataset 1, while Figure 4(b) illustrates the AUC curve for Dataset 2. These curves highlight the trade-off between the true positive rate and false positive rate across various classification thresholds, demonstrating the model's ability to maintain strong performance under different conditions.



(a) AUC Curve – Dataset 1 (b) AUC Curve – Dataset 2

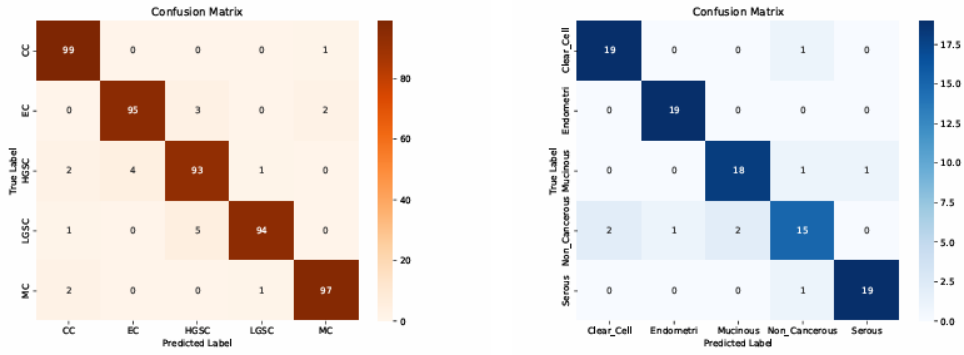
Figure 4: AUC Curves of the Proposed Model on Dataset 1 and Dataset 2



(a) ROC Curve – Dataset 1 (b) ROC Curve – Dataset 2

Figure 5: ROC Curves of the Proposed Model on Dataset 1 and Dataset 2

Figure 5 displays the ROC curves for both datasets. Figure 5(a) depicts the ROC curve for Dataset 1, while Figure 5(b) shows the ROC curve for Dataset 2. The ROC curves provide a graphical representation of the model's ability to distinguish between classes, with the area under each curve reflecting overall classification effectiveness.



(a) Confusion Matrix – Dataset 1 (b) Confusion Matrix – Dataset 2

Figure 6: Confusion matrices of model predictions versus true labels for Datasets 1 and 2.

Figure 6 shows the confusion matrices for Dataset 1 (orange) and Dataset 2 (blue), providing a visual overview of the model's performance in classifying ovarian cancer subtypes. The matrices

illustrate the distribution of true versus predicted labels, highlighting how well the model distinguishes between individual classes. In Dataset 1, the model achieved near-perfect classification, especially for the CC and MC classes, with 99 and 97 correct predictions out of 100, respectively. Misclassifications were rare and generally off by only one or two samples, demonstrating strong discriminative ability. The HGSC class had some overlap, with 4 samples misclassified as EC and 2 as CC, but overall accuracy remained high. In Dataset 2, the model performed well but faced more challenges in class separation. The Clear_Cell and Endometri classes were classified with perfect precision and recall, correctly identifying all 20 samples. However, some confusion occurred between the Mucinous and Non_Cancerous classes—4 Non_Cancerous samples were misclassified as Mucinous, and 2 Mucinous samples were predicted as Clear_Cell or Serous. This indicates visual similarities between certain subtypes, even for a well-performing model. Overall, the proposed custom DenseNet201 model effectively distinguished between ovarian cancer subtypes, providing strong and consistent classification performance, which directly addresses the first research question.

Research Question 2: To what extent do data augmentation and balanced datasets improve the generalization and reliability of the proposed model?

Data augmentation and balanced datasets significantly improved model generalization. Models trained without augmentation achieved lower accuracy. Incorporating rotation, shifting, zooming, and horizontal flipping increased robustness, resulting in stable performance even with variable input images. Balanced class representation also prevented bias toward dominant classes. These findings indicate that dataset augmentation and balancing are crucial for reliable ovarian cancer subtype classification, answering the second research question.

Table 2: *Classification Report Summary for Dataset 1 and Dataset 2 (With vs Without Augmentation)*

Dataset	Augmentation	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Dataset 1	Without	72.6	74.0	73.0	73.9
	With	95.6	95.0	95.0	95.0
Dataset 2	Without	59.6	59.0	57.0	55.0
	With	90.1	89.0	86.0	86.0

Table 2 illustrates the impact of data augmentation on model performance across two ovarian cancer histopathology datasets. For Dataset 1, augmentation improved accuracy from 72.6% to 95.6%, while for Dataset 2, accuracy increased from 59.6% to 90.1%. Precision, recall, and F1-scores also showed notable improvements. These results indicate that augmentation and dataset balancing strengthen the DenseNet201 model. They also answer the second research question by showing that these strategies improve generalization and reliability.

Research Question 3: How does the integration of LIME enhance the interpretability and clinical relevance of the model's predictions?

LIME highlighted key regions of histopathology images that influenced classification, such as tumor cell morphology and tissue architecture. These explanations were consistent with clinically relevant features, improving transparency and trust in the model. The integration of LIME demonstrates that

the model's decisions can be interpreted by medical professionals, addressing the third research question regarding clinical applicability and interpretability.

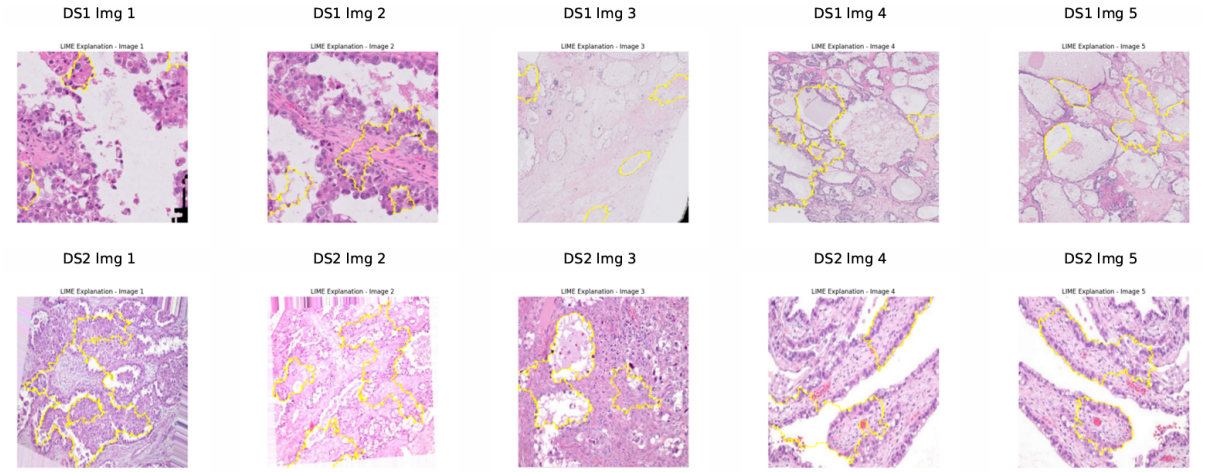


Figure 7: LIME visualizations showing model interpretability on five sample images from Dataset1 (top row) and Dataset2 (bottom row)

Figure 7 displays LIME visualizations for five sample images from each dataset. These visual explanations use super pixel boundaries to indicate which regions of the histopathological images most influenced the model's predictions. Highlighted areas within the boundaries represent the most important features considered by the model during classification. These results demonstrate that the model is focusing on meaningful regions in the tissue samples, improving interpretability and building trust in the automated decision-making process.

Hypotheses Testing

Ho_1: The proposed custom DenseNet201-based transfer learning model does not significantly outperform existing state-of-the-art approaches in ovarian cancer subtype classification using histopathology images.

Table 3: *Paired t-test Results for Custom DenseNet201 vs. Baselines*

Dataset	Baseline Model	t-statistic	p-value	Significance (α = 0.05)
Dataset 1	InceptionV3	7.06	0.0021	Significant
	ResNet50V2	17.57	0.00006	Significant
	MobileNetV2	9.94	0.0006	Significant
Dataset 2	InceptionV3	14.28	0.00014	Significant
	ResNet50V2	6.12	0.0036	Significant
	MobileNetV2	41.81	0.0000019	Significant

Table 3 presents the results of paired t-tests conducted to test the hypothesis that "the proposed DenseNet201-based transfer learning model does not significantly outperform existing state-of-the-art approaches in ovarian cancer subtype classification." The t-statistic and corresponding p-value are reported for comparisons with InceptionV3, ResNet50V2, and MobileNetV2 across two datasets. All p-values are below 0.05, indicating that the DenseNet201 model significantly outperforms

the baseline models. These results lead to the rejection of the null hypothesis, confirming the effectiveness of the proposed model in classifying ovarian cancer subtypes using histopathology images.

10. Discussion of Finding

The proposed custom DenseNet201 model performed well in classifying the cancer subtypes, with particularly strong results for clearly distinguishable classes. As shown in Table 3, our DenseNet201-based approach outperformed existing models in accuracy, robustness, and interpretability across both datasets. The table compares recent histopathology-based studies using metrics like accuracy, AUC, and specificity. Most existing models showed moderate performance. In contrast, our method achieved consistently high results across all metrics. This highlights the effectiveness of our model for ovarian cancer subtype classification using histopathological images.

Table 3: *Performance Comparison of Models on Two Datasets*

Ref	Data	Model	Accuracy	AUC	Specificity
[16] (2023)	Histopathology Images	DRAS-MIL	-	0.8679	-
[14] (2024)	Histopathology Images	H-optimus-0	89%	-	-
[12] (2024)	Histopathology Images	Multi-resolution graph	73%	-	-
[21] (2025)	Histopathology Images	Ensemble Model with XAI	84.8%	-	-
Our Work (2025)	Histopathology Images	**Custom DenseNet201 with XAI**	**95.44%, 90.17%**	**0.9965, 0.9839**	**0.9880, 0.9747**

11. Conclusion

This study presents a deep learning approach for classifying ovarian cancer subtypes using histopathological images. A customized DenseNet201 model with transfer learning was employed, trained and validated on two balanced datasets. Data augmentation techniques were applied to improve generalization. The model demonstrated high accuracy across five ovarian cancer subtypes, with evaluation metrics confirming its robust performance. Comparisons with state-of-the-art models further validated its effectiveness. Additionally, the application of LIME for model explainability enhanced transparency and clinical trust. The results underscore the potential of deep learning in medical image classification. However, limitations remain, particularly in dataset size and diversity. Future work should incorporate larger and more heterogeneous datasets to improve generalizability across populations. Integrating genomic or clinical data could also enhance classification performance and support personalized treatment strategies. Expanding the use of explainable AI tools could further aid in visualization and interpretation, helping pathologists better understand and trust AI-driven decisions. In conclusion, this study demonstrates that deep learning combined with explainable AI has the potential to significantly improve ovarian cancer subtype classification. With expanded datasets and integration into clinical workflows, these technologies could support faster, more accurate, and more equitable cancer diagnostics.

12. Recommendations

The following recommendations were made based on the findings of the study:

(1) Healthcare institutions should adopt deep learning models like DenseNet201 to support accurate and timely ovarian cancer subtype diagnosis.

(2) Data augmentation, balanced datasets, and explainable AI (e.g., LIME) should be employed to improve model robustness and transparency.

(3) Future research should validate the model on larger, diverse datasets and promote AI literacy among pathologists for effective clinical integration.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability Statement: The histopathology datasets used in this study are publicly available on Kaggle and Mendeley and can be accessed at the following links: Dataset 1 and Dataset 2.

References

1. H. Sung et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," CA: A Cancer Journal for Clinicians, vol. 71, no. 3, pp. 209–249, 2021, doi: 10.3322/caac.21660.
2. F. Bray et al., "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, vol. 74, no. 3, pp. 229–263, 2024, doi: 10.3322/caac.21834.
3. U. Menon et al., "Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial," The Lancet, vol. 397, no. 10290, pp. 2182–2193, 2021, doi: 10.1016/S0140-6736(21)00877-2.
4. J. S. Berek, M. Renz, S. Kehoe, L. Kumar, and M. Friedlander, "Cancer of the ovary, fallopian tube, and peritoneum: 2021 update," International Journal of Gynecology and Obstetrics, vol. 155, Suppl. 1, pp. 61–85, 2021, doi: 10.1002/ijgo.13878.
5. M. Köbel et al., "Ovarian Carcinoma Subtypes Are Different Diseases: Implications for Biomarker Studies," PLOS Medicine, vol. 5, no. 12, p. e232, 2008, doi: 10.1371/journal.pmed.0050232.
6. M. L. Wilson, K. A. Fleming, M. A. Kuti, L. M. Looi, N. Lago, and K. Ru, "Access to pathology and laboratory medicine services: a crucial gap," The Lancet, vol. 391, no. 10133, pp. 1927–1938, May 2018, doi: 10.1016/S0140-6736(18)30458-6.
7. J. Zhou, W. Cao, L. Wang, Z. Pan, and Y. Fu, "Application of artificial intelligence in the diagnosis and prognostic prediction of ovarian cancer," Computers in Biology and Medicine, vol. 146, p. 105608, 2022, doi: 10.1016/j.compbiomed.2022.105608.
8. M. Talo, "Automated classification of histopathology images using transfer learning," Artificial Intelligence in Medicine, vol. 101, p. 101743, 2019, doi: 10.1016/j.artmed.2019.101743.
9. M. Wu, C. Yan, H. Liu, and Q. Liu, "Automatic classification of ovarian cancer types from cytological images using deep convolutional neural networks," Bioscience Reports, vol. 38, no. 3, pp. BSR20180289, 2018, doi: 10.1042/BSR20180289.
10. L. Guo, A. Wu, Y. Wang et al., "Deep learning-based ovarian cancer subtypes identification using multi-omics data," BioData Mining, vol. 13, no. 1, p. 10, 2020, doi: 10.1186/s13040-020-00222-x.
11. M. Wu et al., "Deep convolutional neural networks for multiple histologic types of ovarian tumors classification in ultrasound images," Frontiers in Oncology, vol. 13, p. 1154200, 2023, doi: 10.3389/fonc.2023.1154200.
12. J. Breen, K. Allen, K. Zucker, N. M. Orsi, and N. Ravikumar, "Multi-Resolution Histopathology Patch Graphs for Ovarian Cancer Subtyping," arXiv preprint, arXiv:2407.18105, 2024. [Online]. Available: https://arxiv.org/abs/2407.18105
13. H. Farahani et al., "Deep learning-based histotype diagnosis of ovarian carcinoma whole-slide pathology images," Modern Pathology, vol. 35, pp. 1983–1990, 2022, doi: 10.1038/s41379-022-01146-z.

14. J. Breen, K. Allen, K. Zucker, L. Godson, N. M. Orsi, and N. Ravikumar, “A Comprehensive Evaluation of Histopathology Foundation Models for Ovarian Cancer Subtype Classification,” arXiv preprint, arXiv:2405.09990, 2024. [Online]. Available: https://arxiv.org/abs/2405.09990
15. S. K. Behera, A. Das, and P. K. Sethy, “Deep fine-KNN classification of ovarian cancer subtypes using efficientNet-B0 extracted features: a comprehensive analysis,” Journal of Cancer Research and Clinical Oncology, vol. 150, p. 361, 2024, doi: 10.1007/s00432-024-05879-z.
16. J. Breen, K. Allen, K. Zucker, G. Hall, N. M. Orsi, and N. Ravikumar, “Efficient subtyping of ovarian cancer histopathology whole slide images using active sampling in multiple instance learning,” arXiv preprint, arXiv:2302.08867, 2023. [Online]. Available: https://arxiv.org/abs/2302.08867
17. G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708, doi: 10.1109/CVPR.2017.243.
18. S. Thite, “Ovarian Cancer Subtype Classification. Kaggle,” 2023. [Online]. Available: https://www.kaggle.com/datasets/sunilthite/ovarian-cancer-classification-dataset
19. K. Kasture, *OvarianCancer & Subtypes Dataset Histopathology*. Mendeley Data, V1, 2021, doi: 10.17632/kztymsrjx9.1.
20. J. Huang and C. Ling, “Using AUC and Accuracy in Evaluating Learning Algorithms,” IEEE Transactions on Knowledge and Data Engineering, vol. 17, pp. 299–310, Apr. 2005, doi: 10.1109/TKDE.2005.50.
21. A. Kodipalli, V. S. Devi, S. Guruvare, and T. Ismail, “Explainable AI-based feature importance analysis for ovarian cancer classification with ensemble methods,” *Frontiers in Public Health*, vol. 13, p. 1479095, Mar. 2025, doi: 10.3389/fpubh.2025.1479095.