IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Vision Transformer And Texture-Aware Multi-Instance Learning For Ovarian Cancer Subtype Classification

[1] Pavan Kunchur[1], [2]Vaishnavi Shyam Kulkarni, [3]Soniya Rathod,[4]Vaishnavi Dorikar,[5]Bhavana Kolaki

[1] Professor, [2]Student, [3]Student,[4]Student,[5]Student

[1] Dept of Computer Science and Engineering,

[1] KLS, Gogte Institute of Technology, Belagavi, India

Abstract: Ovarian cancer remains a leading cause of gynaecological cancer mortality, with treatment outcomes heavily dependent on accurate subtype classification. Manual interpretation of gigapixel Whole Slide Images (WSIs) is labour-intensive and challenged by extreme resolution, background noise, and class imbalance, particularly in large datasets such as UBC-OCEAN. This study presents a novel framework for ovarian cancer subtype classification that integrates a Vision Transformer (ViT-B/16) encoder within an attention-based Multi-Instance Learning (MIL) architecture. A hierarchical preprocessing pipeline incorporating texture-aware patch sampling is introduced to prioritize diagnostically informative regions by ranking patches using variance and edge density. The ViT encoder extracts context-rich representations from these selected patches, which are aggregated through single-head attention for slide-level prediction. The proposed system demonstrates strong classification performance across all five subtypes, achieving a balanced accuracy of 0.750, and offers a scalable, objective solution for computational pathology. These results highlight the effectiveness of combining transformer-based feature extraction with targeted sampling strategies to address key challenges in high-resolution medical image analysis and support more accurate personalized diagnostics.

Index Terms - Ovarian Cancer, Subtype Classification, Multi-Instance Learning (MIL), Vision Transformer (ViT), Whole Slide Images (WSI), Computational Pathology, Texture-Aware Sampling

I. INTRODUCTION

Ovarian cancer is one of the leading causes of gynaecological cancer deaths, representing the eighth most frequent disease in women globally [1]. It often carries a dismal prognosis, with approximately 324,000 diagnosed cases equating to 207,000 deaths each year. About 90% of all ovarian cancers are classified into five major carcinoma subtypes: high-grade serous (HGSC, 70%), endometrioid (EC, 11%), clear cell (CCC, 10%), low-grade serous (LGSC, 5%), and mucinous carcinomas (MC, 4%) [2]. These subtypes each display unique morphological patterns, treatment responses, and prognosis outcomes. Personalized treatment planning depends on accurate subtype classification from histopathological images; however, manual pathologist assessment of gigapixel-scale Whole Slide Images (WSIs) and Tissue Microarrays (TMAs) is challenging due to the immense data volume, making it laborious, subjective, and prone to inter-observer variability [1]. While recent developments in deep learning have showed promise in computational pathology, there are still major hurdles in managing the extreme resolution, background noise, and significant class imbalance seen in datasets such as the UBC-OCEAN [3] [4]benchmark from Kaggle [4] [4].Vision Transformers (ViTs), which utilize self-attention mechanisms to assess full images holistically, represent a breakthrough in deep learning for medical imaging. By analyzing images as patch embedding's, ViTs can more accurately capture both local and global feature dependencies than conventional Convolutional Neural

Networks (CNNs). Recent research has demonstrated ViTs' higher performance in lung and breast cancer categorization, underscoring their potential for ovarian cancer detection [5]. Conventional CNNs have difficulty with sparse tissue regions and global context, while simple patch-based approaches frequently fail to prioritize diagnostically important regions amidst large non-informative backdrops. Multi-Instance Learning (MIL) frameworks address these problems by treating slides as "bags" of patches, enabling slide-level predictions without detailed annotations. However, applying MIL still requires complex sampling and aggregation strategies to effectively collect subtype-specific information. This work presents a strong attention-based MIL system for five-class ovarian cancer subtype classification from WSIs and TMAs. To address the data challenges, we propose a hierarchical preprocessing pipeline that incorporates a Vision Transformer (ViT-B/16) encoder, Single-Head Attention Pooling for dynamic instance weighting, and novel texture-aware patch sampling, which ranks patches by variance and edge density to favor cellularly rich regions.

II. Literature Survey

The 2020 WHO Classification defines five primary histotypes for ovarian cancer subtypes based on histopathological characteristics: high-grade serous carcinoma (HGSC, ~70%), endometrioid carcinoma (EC, ~11%), clear cell carcinoma (CCC, ~10%), low-grade serous carcinoma (LGSC, ~5%), and mucinous carcinoma (MC, ~4%). Since different histotypes have different cell origins, molecular changes, and therapeutic responses, accurate subclassification influences prognosis and treatment. [6] [2]

2.1 Histotype Evolution

Following WHO changes in 2014, histotype definitions stabilized, classifying serous carcinomas as either LGSC (MAPK-mutated, indolent) or HGSC (TP53-mutated, aggressive). Reclassifications improved repeatability to over 90% by reducing "mixed" or "undifferentiated" diagnoses made using immunohistochemistry (IHC). Rare forms, such as mesonephric-like adenocarcinoma, have surfaced; these are frequently aggressive and KRAS-mutated. [6]

2.1 Diagnostic Approaches

Principal histotypes can be distinguished with around 90% accuracy using a four-marker IHC panel (WT1/p53/napsin A/PR); for example, WT1+/p53 abnormal validates HGSC over EC. On datasets like TCGA-OV, deep learning (DL) models such as CNNs and hybrids (ResNet, VGG-16) categorize subtypes from histology with 94-100% accuracy. Current research uses morpho-genomics (CNN-ViT) to predict subtypes and mutations from H&E slides. [6] [7] [8] [9]

2.3 Diagnostic Approaches

Principal histotypes can be distinguished with around 90% accuracy using a four-marker IHC panel (WT1/p53/napsin A/PR); for example, WT1+/p53 abnormal validates HGSC over EC. On datasets like TCGA-OV, deep learning (DL) models such as CNNs and hybrids (ResNet, VGG-16) categorize subtypes from histology with 94-100% accuracy. Current research uses morpho-genomics (CNN-ViT) to predict subtypes and mutations from H&E slides. [6] [7] [8] [9]

2.4 DL Advances

CNNs (such as DenseNet-121 and YOLO) that achieve AUCs of 0.91-0.99 for subtype recognition across ultrasound, CT, MRI, and histology are highlighted in DL surveys. In multi-class problems (HGSC, EC, CCC, LGSC, MC), hybrid models (CNN-LSTM, fine-KNN) outperform conventional techniques. Foundation models emphasize the need for data augmentation while evaluating histopathology for subtypes. [6] [2] [10] [7]

III. Methodology

The methodology for this study focused on developing a robust computer-aided diagnosis system for ovarian cancer subtype classification using Whole Slide Images (WSIs) and Tissue Microarrays (TMAs). The approach integrated specialized image preprocessing, a patch-based feature extraction pipeline, and an attention-based deep learning architecture to handle the high resolution and heterogeneity of histopathological data. The dataset consists of histopathological images from the UBC-OCEAN [4] obtained from Kaggle [3], comprising both Whole Slide Images (WSI) and Tissue Microarrays (TMA).

3.1 Hierarchical Data Acquisition and Preprocessing

A multi-stage, hierarchical patch extraction strategy was employed to efficiently manage the high-resolution histopathological data and prepare consistent inputs for the model. The extraction pipeline prioritized patches from available low-resolution thumbnails. If thumbnails were absent, the system implemented a regional cropping strategy, extracting 1024 ×1024 pixel regions from the full WSI. Crucially, deterministic cropping was used for the validation set to ensure evaluation reproducibility, while stochastic cropping was used during training to enhance generalization. To isolate informative cellular regions from the background, a binary tissue mask was generated by converting images to the HSV colour space and applying a threshold (> 20)to the saturation channel. Patches extracted from WSIs were standardized to 224×224 pixels. TMA cores, often captured at higher magnifications, were initially extracted at 448×448 pixels and subsequently down sampled to $224 \times 224$to maintain input consistency. Finally, a density filtering step was applied, retaining patches only if their tissue coverage density, as determined by the binary mask, exceeded 30%.

3.2 Texture-Based Informative Patch Sampling

To ensure the model focused on diagnostically relevant regions and to manage the computational cost of WSIs, a custom, texture-aware ranking sampler was implemented to select a representative "bag" of instances for each slide. Patches were ranked based on a linear combination of pixel variance (σ^2)and edge density (∇) where edge density was calculated as the mean absolute gradient in the x and y directions:

$$Score = 0.7\sigma^2 + 0.3\nabla \quad (1)$$

Patches with a mean pixel intensity ≥ 0.90 were excluded as near-white background or blank space. For each slide, the top 32 patches with the highest texture scores were selected to form the Multi-Instance Learning (MIL) bag, concentrating the model's effort on areas exhibiting high cellularity and architectural complexity.

3.3 Architecture: Attention-Based Multi-Instance Learning

The classification was structured as a Multi-Instance Learning (MIL) [11] task, where each slide (the bag) was represented by the set of 32 selected patches (the instances). A Vision Transformer (ViT-B/16) [12], pre-trained on ImageNet-1K [13], was used as the feature encoder. The standard classification head was replaced by a linear projection layer that mapped the CLS token feature into a 768-dimensional embedding space. The resulting patch embeddings were then aggregated using a Single-Head Attention Pooling mechanism [11]. This learned module dynamically computes a scalar weight for each patch, reflecting its diagnostic importance, and calculates a weighted sum to produce the final, aggregated slide-level feature vector. This aggregated feature was passed through Layer Normalization and a final linear layer to predict one of the five ovarian cancer subtypes.

3.4 Training Protocol and Optimization

The model was trained for 20 epochs using a highly optimized protocol designed to handle severe class imbalance and maximize computational efficiency. Class Balancing was achieved using inverse-frequency class weights, which were integrated into two mechanisms: a Weighted Random Sampler in the DataLoader to ensure equal class representation across epochs, and direct application within the loss calculation. The effective training optimization was governed by a Focal Loss [14] ($\gamma = 2.0$) applied within the step_batch function, utilizing the inverse-frequency class weights(α).Although a standard $CrossEntropyLoss$ with label smoothing ($\epsilon = 0.05$)was defined, the model was primarily optimized by the class-balanced Focal Loss. The AdamW optimizer [15] ($LR = 3 \times 10^{-5}$, Weight Decay = 0.01 was utilized, paired with a Cosine Annealing Learning Rate Scheduler. Finally, training leveraged Automatic Mixed Precision (AMP) for accelerated performance, with gradient updates performed using gradient accumulation over 4 steps, effectively increasing the batch context for each optimization step. The implementation utilized the PyTorch [16] and PyTorch Image Models (timm) [17] libraries.

IV. Results and Analysis

4.1 Training Dynamics and Model Selection

The training process for the attention-based Multi-Instance Learning (MIL) model spanned 20 epochs, utilizing a class-weighted Focal Loss to address the severe class imbalance in the training set. The Training vs Validation Loss curve (Figure 1) shows a sharp decrease in training loss (approaching 0.0000), while the validation loss stabilizes around 0.006–0.008. This divergence reflects the intense optimization of the weighted training loss. The Training vs Validation Accuracy curves (Figure 2) illustrate rapid convergence on the training data, but generalization performance was governed by the Balanced Accuracy metric. The model achieved its peak generalization performance at Epoch 7, yielding a Balanced Accuracy of 0.750 (Val Accuracy 0.741) (Figure 3). This epoch was selected as the final checkpoint for all subsequent analysis and evaluation, as further training led to a decline in balanced accuracy, indicative of increasing overfitting.

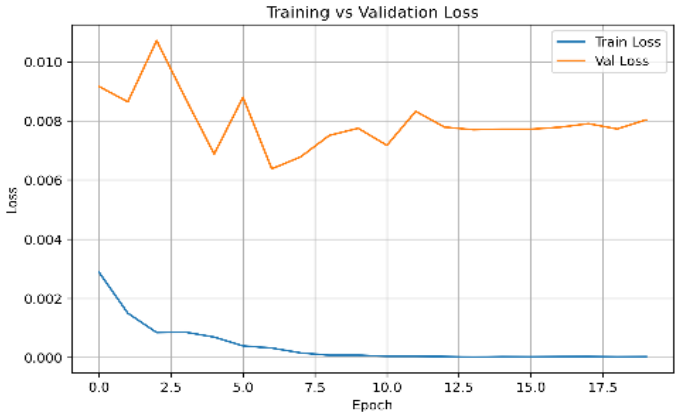

fig 4.1.1 training and validation loss

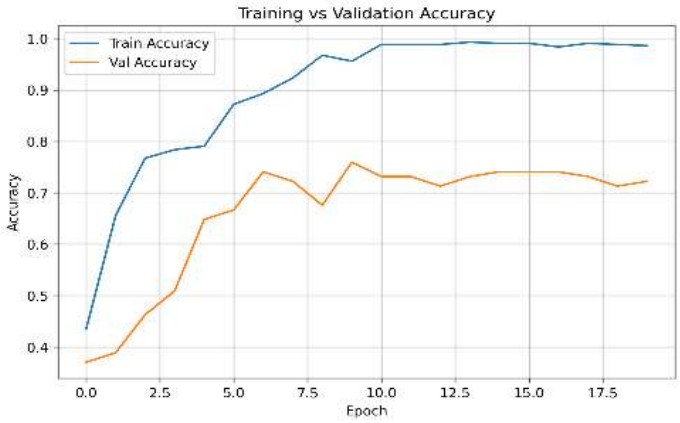

fig 4.1.2. training and validation accuracy

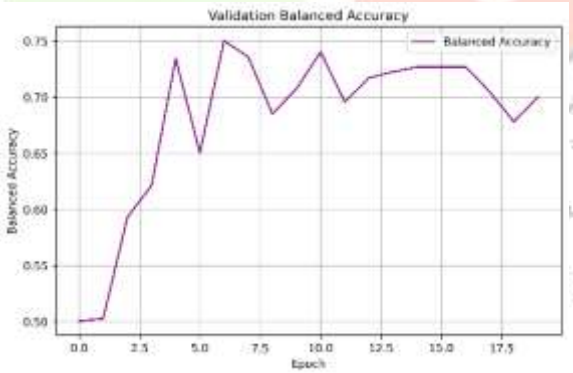

fig 4.1.3. validation balanced accuracy

4.2 Subtype-Specific Classification Performance

The selected model demonstrated a robust capacity for distinguishing between the five ovarian cancer subtypes, achieving a Macro F1-Score of 0.74. The detailed performance metrics are summarized in the Classification Report (Table 1), and the sample-level performance is visually represented in the Confusion Matrix

table 4.2.1. classification report metrics

Class	Precision	Recall	F1-Score	Support
CC	0.80	0.80	0.80	20
EC	0.64	0.84	0.72	25
HGSC	0.81	0.67	0.73	45
LGSC	0.67	0.89	0.76	9
MC	0.83	0.56	0.67	9

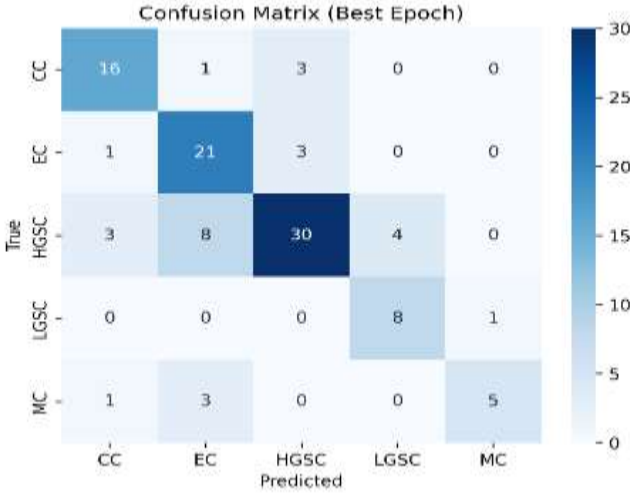

fig 4.2.1. confusion matrix (best epoch 7)

4.2.1 High-Grade Serous Carcinoma (HGSC)

As the most prevalent class (Support: 45), the model demonstrated strong predictive quality for HGSC, achieving the highest precision at 0.81 among all subtypes. However, the recall was lower at 0.67, suggesting that the model had difficulty distinguishing approximately one-third of the HGSC samples, primarily misclassifying them as Endometrioid Carcinoma (EC) in the confusion matrix. This common misclassification highlights the architectural and cytological overlap between the two subtypes.

4.2.2 Endometrioid Carcinoma (EC)

The EC subtype (Support: 25) achieved a high recall of 0.84, meaning the majority of true EC cases were correctly identified. However, its precision was the lowest among all classes at 0.64. This indicates that while the model rarely missed an EC case, a substantial number of its EC predictions were incorrect, often being confused with HGSC, thus contributing significantly to the overall false positive rate for this class.

4.2.3 Clear Cell Carcinoma (CC)

The performance for Clear Cell Carcinoma (CC, Support: 20) was highly balanced, with both a precision and recall of 0.80. This consistent performance resulted in the highest F1-Score among the three most common classes (CC, EC, HGSC) and suggests that the texture-based patch sampling strategy successfully captured the distinct pathological features of clear cell morphology.

4.2.4 Low-Grade Serous Carcinoma (LGSC)

Despite being a minority class (Support: 9), the LGSC subtype demonstrated the highest classification sensitivity, achieving a remarkable recall of 0.89 (8 out of 9 samples correctly identified). This exceptional performance is a direct result of the integrated class-balancing mechanisms (Weighted Sampler and Focal Loss) used during training, which successfully prioritized the correct identification of rare, diagnostically critical subtypes [14]. The precision stood at 0.67, contributing to a strong F1-Score of 0.76.

4.2.5 Mucinous Carcinoma (MC)

Mucinous Carcinoma (MC, Support: 9) was also a challenging minority class. It achieved the highest overall precision across all classes at 0.83, indicating that when the model classified a slide as MC, the prediction was highly reliable. However, it exhibited the lowest recall at 0.56, suggesting that almost half of the true MC cases were missed, which points to a potential limitation in the feature encoding or sampling process for this particular subtype's architectural patterns.

4.3 Discriminative Power and Robustness

The overall discriminative capability of the model was further validated through the analysis of the Area Under the Curve (AUC) from the Receiver Operating Characteristic (ROC) curves and the Precision-Recall (PR) curves. The ROC analysis (Figure 5) showed that the model maintains a strong separation margin across all classes, with AUC values consistently above 0.88. Notably, MC achieved the highest AUC of 0.952, followed by CC at 0.920. This indicates that despite the low recall for MC, the model's confidence scores are highly reliable in distinguishing positive from negative cases for this subtype. The Precision-Recall curves (Figure 6) also confirmed the model's robustness against class imbalance, as the curves for all classes, especially the minority ones (LGSC, MC), were significantly elevated above the baseline, reinforcing the reliability of the classification system.

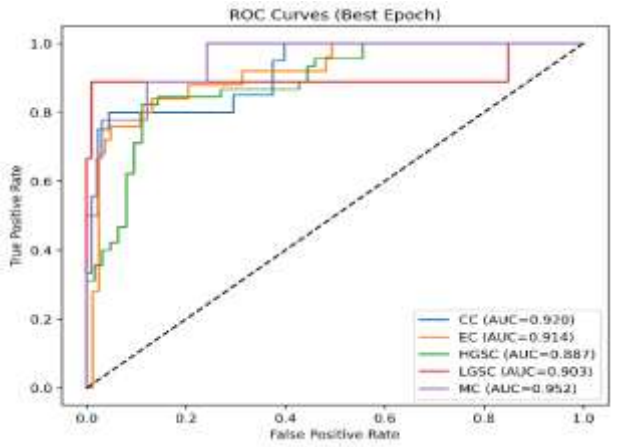

fig 4.3.1. roc curves (AUC) for all subtypes

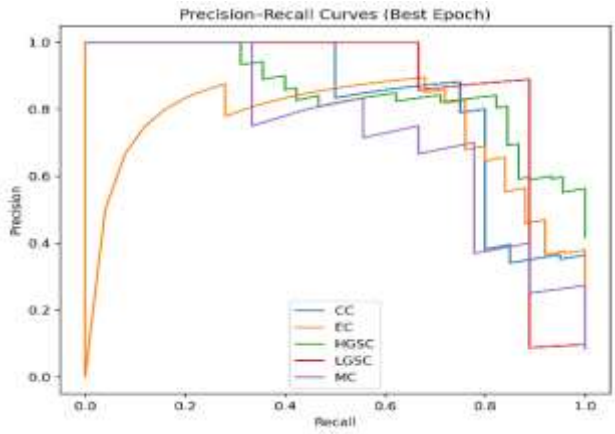

fig 4.3.2. precision-recall (PR) curves

V. Conclusion

This study developed an attention-based Multiple Instance Learning (MIL) framework for five-class ovarian cancer subtype classification from Whole Slide Images. By combining texture-based informative patch sampling, a ViT [12] encoder, and Single-Head Attention Pooling [11], the model effectively handled gigapixel-scale variability.

The best performance was achieved at Epoch 7, with a Balanced Accuracy of 0.750 and a Macro F1-Score of 0.74. Class-balancing strategies enabled high sensitivity for rare subtypes, including a recall of 0.89 for LGSC, while all subtypes achieved AUC values above 0.88. Overall, this work demonstrates a reliable and clinically meaningful computational pathology system capable of accurately distinguishing major ovarian cancer subtypes

References

[1] S. Lv and G. Zhang, "A Robust Deep Learning Framework for Ovarian Cancer Subtype Classification," in *20th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 2024.

[2] J. Breen et al., "A comprehensive evaluation of histopathology foundation models for ovarian cancer subtype classification," *npj Precision Oncology*, vol. 9, p. 33, 2025.

[3] M. Asadi-Aghbolaghi and e. al., "Machine Learning-Driven Histotype Diagnosis of Ovarian Carcinoma: Insights from theOCEAN AI Challenge," *medRxiv*, 2024.

[4] M. Asadi-Aghbolaghi and e. al., "UBC Ovarian Cancer Subtype Classification and Outlier Detection (UBC-OCEAN)," Kaggle, 2023. . [Online]. Available: https://kaggle.com/competitions/UBC-OCEAN.

[5] A. B. Apeksha and S. B. Gowda, "A Vision Transformer-Based Approach For Ovarian Cancer Detection And Classification," *Int. J. Creative Res. Thoughts (IJCRT)*, vol. 13, no. 5, pp. C377-C385, 2025.

[6] M. Köbel and E. Y. Kang, "The evolution of ovarian carcinoma subclassification," *Cancers (Basel)*, vol. 14, no. 2, p. 416, 2022.

[7] S. M. Abdullah, A. A. Masum, N. U. Prince and L. A. Mim, "Deep learning-based ovarian cancer subtype classification using VGG16 and MobileNetV2 with Squeeze-and-Excitation Blocks," *J. Angiotherapy*, vol. 8, no. 8, pp. 1-11, 2024.

[8] 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," *J. Cancer Res. Clin. Oncol.*, vol. 150, no. 7, p. 361, 2024.

[9] A. K. Saha et al., "An enhanced deep learning model for accurate classification of ovarian cancer from histopathological images," *Sci. Rep.*, vol. 15, no. 1, p. 21860, 2025.

[10] M. H. Sadeghi, S. Sina, H. Omidi, A. H. Farshchitabrizi and M. Alavi, "Deep learning in ovarian cancer diagnosis: a comprehensive review of various imaging modalities," *Pol. J. Radiol.*, vol. 89, p. e30–e48, 2024.

[11] M. Ilse, J. M. Tomczak and P. O. Tiňo, "Attention-based Deep Multiple Instance Learning," in *International Conference on Machine Learning (ICML)*, 2018.

[12] A. Dosovitskiy, L. Beyer, N. Kolkin and et al, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *International Conference on Learning Representations (ICLR)*, 2021.

[13] J. Deng, W. Dong, R. Socher, L. Li, K. Li and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[14] T.-Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal Loss for Dense Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, p. 318–327, 2020.

[15] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," in *International Conference on Learning Representations (ICLR)*, 2019.

[16] A. Paszke, S. Gross, F. Massa and et al, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[17] R. Wightman, "PyTorch Image Models," 2019. [Online]. Available: https://github.com/rwightman/pytorch-image-models.

[18] N. Youssef, K. Gabriel, L. El Saadawi, F. Simaika and H. Issa, "Ovarian cancer subtype classification using convolutional neural networks: an evaluation of deep learning techniques for histopathological image analysis," *Neural Comput. & Applic.*, vol. 37, no. 33, p. 28107–28123, 2025.