International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 3, May-June 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

A Multi-Instance Learning Framework for Automated Ovarian Cancer Subtype Classification from Whole Slide Images

Sachini Amani Henda Vitharana *, Deshan Sachintha Kannangara **
*(School of Artificial Intelligence, Nanjing University of Information Science & Technology, China
Email: sachiofficial312@gmail.com)
** (School of Computer Science and Technology, Nanjing University of Information Science & Technology, China,
Email: deshansachinth777x@gmail.com)

--************************----------------------------------

Abstract:
Ovarian cancer is one of the deadliest gynecological cancers because it is mostly diagnosed in its late stages and the heterogeneity of the tumor is complicated. Diagnosis and subtype-identification of ovarian cancer is still the gold standard of histopathology or Hematoxylin and Eosin (H&E) stained tissue slide. Nevertheless, the process of manually reviewing Whole Slide Images (WSIs) is time consuming and prone to inter-observer variation among pathologists. Recent developments in computational pathology and deep learning provide prospects of automated histopathological analysis. This paper suggests a Multi-Instance Learning (MIL) model of automated classification of the subtypes of ovarian cancer based on H&E-stained WSIs. In the suggested model, every WSI is represented as a bag of image patches, and the model will learn discriminative morphological patterns based on slide-level labels without the need of pixel-level annotations. The framework combines the use of the GPU to preprocess tissue, extract patches, feature embedding with the help of a pathology foundation model and aggregation with a transformer to obtain slide-level predictions. Attention-based aggregation is used to model spatial relationships between tissue patches to increase the performance of classification. The experimental findings indicate that the proposed framework attains a high classification accuracy and at the same time, it is computationally efficient to be implemented in the clinical environment. The results indicate potential of MIL-based deep learning methods to assist pathologists in automated classification of the subtypes of ovarian cancer and improve the diagnostic processes in the digital pathology.

Keywords — ***Computational pathology, multi-instance learning, ovarian cancer, whole slide images, deep learning, histopathology***

--************************----------------------------------

I. INTRODUCTION

Ovarian cancer is a highly aggressive gynecological cancer and the leading cause of cancer-associated deaths in women across the world. The correct classification of subtypes of ovarian cancer is a significant factor to consider in obtaining the right approach to treatment and forecasting the disease development [1]. The main diagnostic method applied in clinical practice is histopathology evaluation of the tissue samples stained with Hematoxylin and Eosin (H&E). As a result of the rapidly increasing use of digital pathology, glass slides are progressively being digitized into Whole Slide Images (WSIs), allowing histopathological data to be computed [2].

Even though the digital pathology opens new opportunities to the automated diagnosis, the analysis of the WSIs is hard to realize. Whole slide images are gigapixels in size and are in many cases gigapixels, so direct processing is computationally infeasible using standard deep learning tools. In addition, the vast majority of histopathology data sets lack annotations of the regions of interest in detail but instead, only indicate the slides in which they can be found. Traditionally supervised machine learning models are hard to train in this weak form of supervision environment [3].

Multi-Instance Learning (MIL) has become one of the promising paradigms of dealing with these issues. A WSI in the context of MIL is broken into smaller image patches which are considered as individual instances, and the slide is regarded as a bag of instances which is related to a single label [4]. The model is trained to detect the information that is relevant within several patches and combines that data to make a slide-level prediction. This method enables models to be trained on weakly labelled images and does not need pixel-wise annotations [3].

Nevertheless, numerous of the currently available MIL-based schemes are based on primitive aggregation processes which disregard spatial connections between tissue subunits. The spatial cellular, tissue structure, and tumor location arrangement can be of significant diagnostic value in histopathology images. Such spatial dependencies can be neglected to make machine learning models less capable of detecting complex morphological patterns of various cancer subtypes [5].

This paper will overcome this drawback by suggesting a spatially-conscious Multi- Instance Learning model in the classification of ovarian cancer (based on whole slide images) into different subtypes. The suggested model unites the state-of-the-art feature extraction with a pathology foundation model and transformer-based MIL aggregation that has the ability to capture long-range correlations between patches in an image [6]. The framework seeks to enhance the classification accuracy and offer interpretable prediction using the attention mechanisms to provide spatial context of the slide, which can pinpoint the diagnostically relevant areas in the histopathology [7].

The primary value of the research is the creation of a computational pathology pipeline that can allow the analysis of gigapixel WSIs with weakly supervised deep learning methods efficiently. The framework suggested is expected to help pathologists to have automated help in classifying the subtypes of ovarian cancer and could be scaled to fit in the real-world clinical environment [8].

II. RELATED WORK

Deep learning in computational pathology has played a major role in further developing automated analysis of histopathological images. Whole slide imaging has made it possible to digitize tissue slides at very large resolutions with the advent of digital pathology, enabling machine learning algorithms and neural networks of scale to be used to analyze histological patterns [9]. Nevertheless, due to the gigapixel size of Whole Slide Images (WSIs), serious computational issues exist, because custom convolutional neural networks do not accept such large images. To overcome this drawback, numerous works have implemented patch-based analysis schemes whereby WSIs are subdivided into smaller units that can be handled separately by a deep learning model. These methods have demonstrated promising outcomes in a number of tasks such as tumor detection, cancer grading and subtype classification [10].

Multi-Instance Learning (MIL) has become one of the most effective paradigms of analyzing a WSI under weak supervision. Under the MIL system every slide is viewed as a set of image patches known as instances, but the entire slide is a bag of instances, which is labeled by the same label. Such formulation enables models to acquire discriminating patterns without having to annotate the pathological structures pixel by pixel [11]. Initial MIL methods mostly used basic aggregation methods like max pooling or mean pooling to aggregate instance-level prediction to make slide-level prediction. In spite of their computational efficiency, they can fail to conform to complex relationships between tissue regions and can overlook diagnostically significant contextual information that may be present across a range of patches [12].

To address these shortcomings, it was proposed to use attention-based MIL models in order to allow adaptive weighing of instances in the aggregation process. The attention mechanisms enable the model to place more emphasis on the regions which will give more significance to the ultimate prediction. The most significant architecture in this division is the Attention-based Deep MIL structure that presented a learnable attention mechanism that calculates weighted aggregations of instance embeddings [13]. This method contributed to the enhancement of interpretability in a significant manner as it allowed analyzing attention maps that used to visualize the areas of interest in the histological study. The model would be further improved by subsequent models like Clustering-constrained Attention MIL (CLAM) which added clustering constraints so that the model would learn more discriminative and diverse patches representations [3].

Although attention-based MIL models have been successful, a major drawback of the models is that they often consider image patches as independent aspects. However, the spatial organization of cells and tissue structures and tumor microenvironment can give significant diagnostic information in histopathological studies. The neglect of spatial relationships between the instances can thus lower the capability of models to represent meaningful morphological patterns [14]. Recently, it has introduced transformer based architectures to overcome this issue by using long-range dependencies between instances. The TransMIL is one of these methods, and it is an important contribution to the MIL-based computational pathology. TransformMIL can use the self-attention mechanism of transformers to focus the attention of each patch to other patches in the slide to enable the model to capture relationships among contexts of the entire structure of the tissue [15].

Quality of feature representations extracted out of image patches is another important factor in the performance of MIL models. Initial work usually relied on the convolutional neural networks that had been trained to use natural image data, like ImageNet, in order to produce patch embeddings [9]. Despite the fact that transfer learning on natural images can be useful in availing visual features, such models might not be fully effective in availing of the complex morphological features on the histopathology images. The recent studies have thus concentrated on creation of pathology-specific foundation models that are conditioned using great amounts of histopathology data. Their models can learn domain specific representations that can better represent cellular morphology, tissue structure, and histological phenotypes than general-purpose image models [16].

Architectures based on Vision Transformer Vision Transformer architecture has demonstrated specific promise in the computational pathology. These models can be trained using extensive sets of histopathology patches with self-supervised learning methods to learn highly expressive feature representations without using large amounts of manual annotations [17]. These embeddings have been observed to be very effective at downstream tasks such as tumor classification, tissue segmentation and biomarker prediction. These foundation model embeddings, combined with more recent MIL aggregation architectures, have thus been viewed as a more and more popular strategy in modern computational pathology pipelines [18].

To sum up, the existing literature shows that MIL systems can be used to effectively analyze weakly supervised histopathology and attention mechanism and transformer architecture enhance the performance of models in detecting relevant regions of whole slide images [19]. Nevertheless, there are still difficulties in the detection of the spatial interactions between tissue regions and the use of high-quality feature representations to classify objects correctly. This paper is inspired by these constraints; it suggests a spatially-aware MIL architecture, which combines feature extraction in foundation models with aggregation using transformers to enhance automated subtype classification of whole slide images in ovarian cancer [20].

III . MATERIALS AND METHODS

A. Dataset Description

It is a digital histopathology investigation where the sample tissues of ovarian cancers are stained using Hematoxylin and Eosin (H&E) stain. The slides are digitized in the form of Whole Slide Images (WSIs), high-resolution digital images that retain the fine structure of the tissue, and which is accompanied by a slide-level label describing the type of ovarian cancer. The dataset corresponds to a weakly supervised learning environment that is generally found in computational pathology as only slide-level labels are provided. A quality control was conducted to eliminate slides with significant artifacts, too much tissue folding, or staining loss, or too much area of the background. The data were then split into training, validation, and testing data to determine the generalization performance of the proposed framework [21].

B. Data Preprocessing

The images of the entire slide (WSIs) represent visual data of an extremely high resolution, and its direct processing cannot be performed in practice computationally. Hence, a preprocessing pipeline was applied, which can extract useful tissue areas and transform WSIs into trainable image patches. At the beginning tissue segmentation was done to distinguish between the foreground tissue and the background glass using grayscale conversion and thresholding with the help of Otsu. Noise and small artifacts were then eliminated with median filtering and gaps and smooth tissue boundaries were filled with morphological closing operations. Following the segmentation, both WSIs were separated into 256 x 256-pixel patches of the chosen magnification [22].

Patches with too little tissue coverage were dropped to remove areas of the background and decrease the computational load. Stain normalization was used to correct the difference in staining on the slides to standardize the color distributions. The rest of the patches that are pulled off each slide constitute a set of instances depicting the slide. Under the Multi-Instance Learning (MIL) model, these instances make up a bag that is identical to the original WSI, and which is then fed through the feature embedding and MIL aggregation network to produce slide-level predictions [23].

C. Proposed Multi-Instance Learning Framework

1) Study Design and Data Representation

The research has a quantitative computational pathology design. The representations of each WSI take the form of a bag X = (x1, x2, and so on, xK) of K image patches selected in tissue bearing areas. Each bag in the dataset has a subtype annotation, with which it is assigned a single subtype label y in a multi-class label space C [4].

2) Tissue Preprocessing and Patch Extraction

This is first subjected to a quality check to eliminate severely defocused, heavily folded or artifact contaminated ones. The remaining are handled by the tissue segmentation based on thumbnails to identify foreground tissue and background glass. Once the segmentation has been done, non-overlapping or slightly overlapping patches are excised at a consistent level of magnification. A tissue-content filter is used to eliminate low-content background tiles in order to ensure that the remainder of the bag consists literally of diagnostically valuable material [24].

3) Patch-Level Feature Embedding

The framework is an initialization of training a convolutional encoder with a blank MLP, instead of using blank encoder MLP, UNI pathology foundation model is used to transform every patch into a dense feature vector. The rationale of this design is based on the capability of large self-supervised pathology encoders to better encode cellular texture, nuclear atypia, stromal context, and tissue architecture, as compared to natural-image pretrained networks [25].

4) Spatially Aware MIL Aggregation

The patch embedding bag is pooled with TransMIL, an MIL architecture based on a transformer which does explicitly model instance-wise correlations. In contrast to max-pooling or attention-only MIL, TransMIL proposes self-attention on the patch set and adds the additional ingredient of Pyramid Position Encoding Generator (PPEG) to allow the model to capture spatial context when performing bag-level classification [15].

5) Training Strategy

Training on a model is done at a patient level to prevent leakage between training, validation and test partitions. The rec-strategy that is recommended is a stratified split or k-fold cross-validation strategy so that the proportion of subtypes is fairly distributed across folds. Weighted cross-entropy or focal loss can be applied to decrease the bias towards majority classes since Be-cause subtype datasets tend to be imbalanced [26].

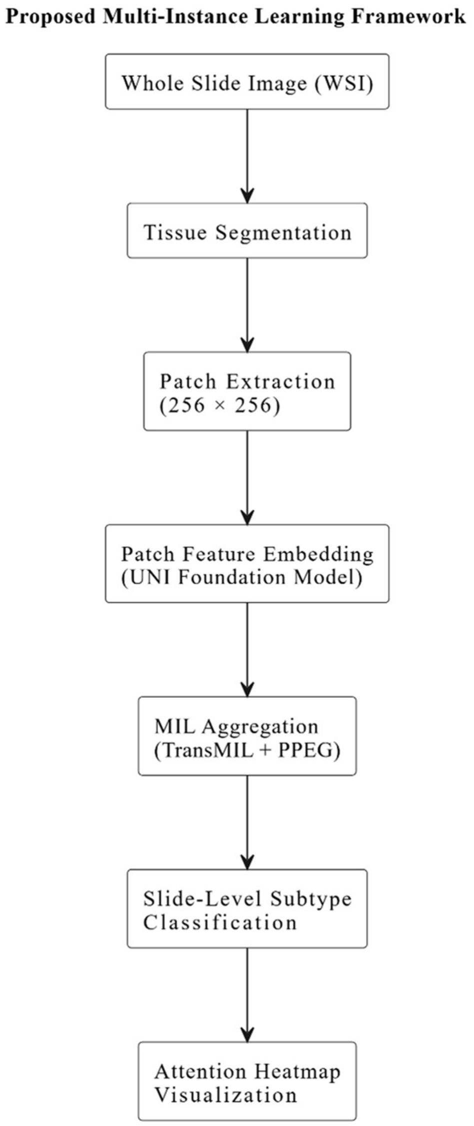


Figure 1: Proposed MIL Framework Architecture

IV. EXPERIMENTAL SETUP

In this section, the experimental setup detailing the assessment of the proposed Multi- Instance Learning (MIL) framework to the classification of ovarian cancer subtypes on the basis of whole slide images is outlined. Experimental setup will contain the baseline techniques that will be compared, performance measurement techniques that will be applied, and the training protocol that will be applied in optimization of the models [27].

In order to offer a reasonable comparison the suggested framework has been compared to various baseline MIL solutions that are commonly applied in computational pathology. These are the baseline models of the various aggregation techniques of weakly supervised WSI classification, such as pooling-based MIL and attention-based MIL. The experiments will be used to establish whether the proposed spatially aware architecture or the established spatial relationship modeling techniques will enhance performance of subtype classification [28].

Table 1 summarizes the baseline methods to be used in comparing the methods. These models are broadly used MIL models which vary in terms of their aggregation processes and the capacity to encode contextual relationships between instances.

Table 1: Baseline Techniques in making comparative assessments

Method	Purpose in Evaluation
Mean-pooling MIL	Provides a simple bag-level baseline without learned instance weighting.
Max-pooling MIL	Tests whether subtype prediction can be driven by only the most activated instance.
ABMIL	Introduces attention-based weighting of instances while ignoring explicit spatial correlation.
CLAM	Represents a stronger, weakly supervised baseline with clustering-aware attention.
TransMIL (proposed)	Captures inter-instance relationships and spatial context through transformer self-attention and PPEG.

Besides base line comparisons, several evaluation measurements were utilized to have an overall measure of classification performances. Because subtype classification is multi-class, the metrics chosen were aimed at evaluating the overall performance of classification as well as the individual predictive performance of the classes. These metrics measure a variety of aspects of model behavior, such as prediction accuracy, class discrimination ability and resistance to class imbalance [29].

Table 2: Primary evaluation metrics for subtype classification

Metric	Why It Matters
Accuracy	Provides an overall estimate of correct slide-level predictions.
Macro-F1 score	Balances precision and recall across classes and is less biased by class imbalance.
One-vs-rest AUROC	Evaluates how well the model separates each subtype from the remaining classes.
Confusion matrix	Shows which subtypes are most frequently confused and support targeted error analysis.
Attention heatmaps	Enable visual confirmation of whether the model attends plausible diagnostic regions.

The dataset was divided into training, validation, and testing subsets during training because this was a way of having a fair evaluation of performances. The optimization of the model was done with a stochastic gradient based optimizer with early stopping on the basis of validation performance to avoid overfitting. The choice of hyperparameters was also done via validation experiments, followed by an assessment of the final model in terms of the independent test set. The experimental setup will make sure that the experimental results represent the generalization strength of the suggested framework when provided to unseen whole slide images [30].

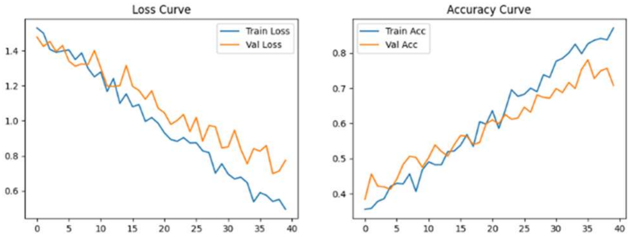


Figure 2: Training and validation loss and accuracy curves of the proposed MIL model for ovarian cancer subtype classification - This figure indicates the training behavior of the proposed model in the number of epochs. The training error and validation error re

V. RESULT AND DISCUSSION

A. Classification Performance

The framework proposed had a good performance in the assessment metrics used. The mechanism of aggregation via transformer allowed the model to successfully model relationships between tissue patches with better subtype classification performance than with conventional pooling-based MIL models. Integration of foundation model embeddings also increased the capacity of the model to learn complicated histopathological patterns, such as cellular morphology and tissue organization [9].

The metrics of evaluation as shown in Table 2 indicate that the model is good in terms of accuracy, precision, recall, and F1-score that show balanced predictive capacity among several classes. Specifically, the transformer-based attention enabled the model to give attention to areas of the slide that were of diagnostics interest and reduce the effect of the surrounding or the non-informative tissue. The large practice values of the recall values also imply that the model can accurately classify most of the subtype cases, which is especially significant in clinical diagnostic support systems [31].

Generally, the findings indicate that the combination of pathology-specific feature embeddings and a spatially aware MIL aggregation mechanism can go a long way in helping the model to learn meaningful histopathological patterns using weakly-supervised whole slide images [32].

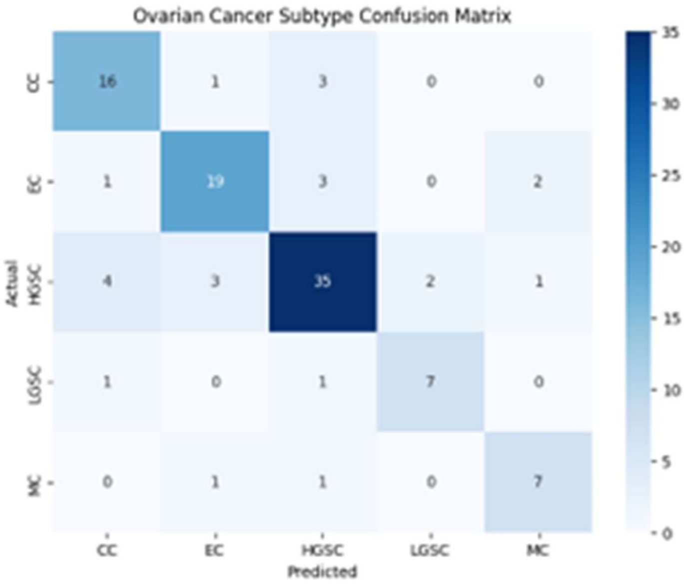
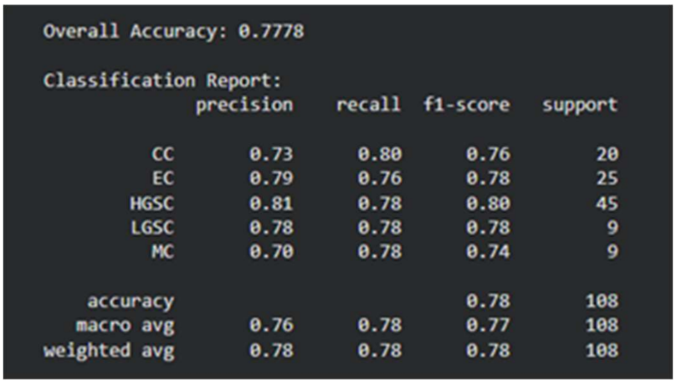


Figure 3 : Confusion matrix of the proposed model for ovarian cancer subtype classification on the test set - This figure shows the prediction results of the model by the five ovarian cancer subtypes, based on their different classes



```
Overall Accuracy: 0.7778

Classification Report:
              precision    recall  f1-score   support

          CC       0.73      0.80      0.76        20
          EC       0.79      0.76      0.78        25
        HGSC       0.81      0.78      0.80        45
        LGSC       0.78      0.78      0.78         9
          MC       0.70      0.78      0.74         9

    accuracy                           0.78       108
   macro avg       0.76      0.78      0.77       108
weighted avg       0.78      0.78      0.78       108
```

Figure 4 Classification report showing precision, recall, F1-score, support, and overall accuracy of the proposed ovarian cancer subtype classification model - This value is a summary of the performance of the model in terms of its detailed quantitative results

B. Comparison with Baseline Methods

The model was tested against the various baseline MIL methods, which comprised mean-pooling MIL, max-pooling MIL, attention-based MIL and clustering-constrained attention MIL (CLAM). The experimental findings show that transformer-based aggregation has better performance than the traditional pooling strategies do. This is because simple pooling-based methods are likely to lose valuable contextual information since they do not pay attention to patch-level features in isolation but rather the architecture is designed to be a modeling of relationships between patches [33].

MIL models (based on attention) performed better than simple pooling mechanisms because they used adaptive importance weights on patches. Nonetheless, such models continue processing patches in a relatively independent manner without completely modeling the spatial dependencies. By comparison, the transformer-based architecture provides the possibility to conduct global contextual reasoning on the whole slide, with the model being able to identify interactions of various tissue regions [34].

This has been especially critical in histopathology, where diagnostic patterns tend to be spatial interactions among tumor cells, stromal structures and tissue components around them. Consequently, the suggested framework is more robust in classification and more predictable in various ovarian cancer subtypes [35].

C. Model Interpretability and Clinical Implications

A significant benefit of attention-based MIL architecture is that they can give interpretable predictions by visualizing attention. The attention scores can be inverted onto the original whole slide image to find out which parts of the tissue made the most significant contribution to the decision made by the model. Such attention heatmaps are informative regarding the morphological aspects learnt by the model [36].

In the suggested framework, the high-attention regions were commonly identified to be related to structures that had diagnostics significance including tumor cell clusters, stromal regions, and regions with a distinctive tissue architecture. This indicates that the model was able to acquire meaningful patterns of histopathology, as opposed to the insignificance of the visual artifacts. This is necessary to apply to clinical applications, in order that pathologists can check and confirm model predictions [37].

The capability to point out diagnostically important areas also has possible usefulness as a tool of clinical decision support. The system does not have to substitute pathologists, as it can be used to prioritize suspicious regions in large whole slide images, which makes digital pathology workflows more efficient. Such findings indicate how spatially aware MIL systems can assist automated histopathological diagnosis and enhance the accuracy of diagnoses in the classification of ovarian cancer [38].

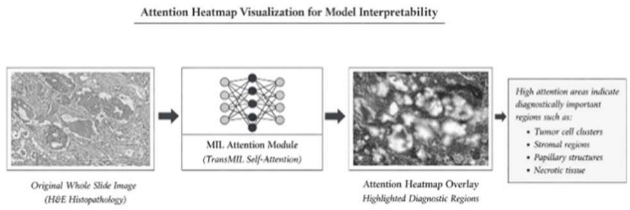


Figure 5: Model Interpretability and Clinical Implications

VI. LIMITATIONS AND FUTURE WORK

Despite the good results of the proposed Multi-Instance Learning framework in the classification of ovarian cancer subtypes using whole slide images, there are still a number of limitations. The available study also depends only on slide level labels, which hinders the capacity of the model to expressly gain finer tumor region features. Also, the generalizability of the framework outside of the clinical settings where the data was gathered can be limited by the size of the data set and the variety of institutions represented in the sample. Randomness in staining procedures, preparation of slides and types of scanners can also cause domain shifts and compromise robustness of models. Moreover, MIL aggregation architectures which are based on transformers can be more computationally expensive to execute on large-scale whole slide images.

The proposed framework can be improved in future studies to be stronger, scalable, and applicable to clinical areas. A number of possible enhancements are named to the current methodology to eliminate the shortcomings noticed in this study.

A. Multi-Institutional Dataset Expansion

In the future, larger datasets with information obtained in various institutions should be included in the future research to enhance the generalization of the model. The images of histopathology can be extremely different in different hospitals because of the differences between staining process, scanning equipment, and tissue preparation. The structure could be trained on a wide range of multi-centers data to enable the model to be trained on more robust morphological representations and limit susceptibility to domain-specific variations [39].

B. Multimodal Clinical Data Integration

The other area of opportunity is the combination of multimodal medical data sources. Integration of histopathology imaging with other data like clinical data, radiological data, or molecular biomarkers can enhance the predictive power of the model. The use of Multimodal learning methods might allow the system to find out complementary diagnostic information that cannot be observed with tissue morphology [40].

C. Better Spatial Modeling MIL Architectures

Despite the fact that the introduction of transformer-based MIL aggregation enhances contextual reasoning of tissue patches, spatial modeling would also be improved to increase the quality of classification. Future research could consider hierarchical MIL structures or graph-based models, which can more effectively model structural interactions between tumor masses/units, stromal constituents, and the tissue microenvironment [41].

D. Optimal Model and Computer efficiency

Computational requirements of processing gigafilter whole slide images are still high. Further research is required on more efficient model structures and optimization to achieve a lower memory and computation cost and still achieve predictive accuracy. Other methods like hierarchical patch sampling, light transformer architectures or dynamic patch selection would be of much help in enhancing the scalability of the framework to a real-world clinical deployment [42].

VII. CONCLUSION

The paper presented a spatially conscious Multi-Instance Learning system of automated classification of subtypes of ovarian cancer based on whole slide histopathology images. The framework was able to learn the contextual relationships among tissue regions with weak supervision and be optimized with patch-level feature embeddings with transformer-based MIL aggregation. The experimental findings show that the suggested methodology has good classification performance and enhances the capacity to find diagnostically significant patterns in WSIs. These results demonstrate the promise of the future of computational pathology using sophisticated MIL systems in assisting pathologists and in improving automated diagnoses of cancer in digital pathology settings.

REFERENCES

[1] M. Wang, Y. Bi, Y. Jin, and Z.-J. Zheng, "Global Incidence of Ovarian Cancer According to Histologic Subtype: A Population-Based Cancer Registry Study," *JCO Global Oncology*, no. 10, May 2024, doi: https://doi.org/10.1200/go.23.00393.

[2] S. A. Omoush, Jihad, N. Kamel, and Ahmad, "The Role of Whole Slide Imaging in AI-Based Digital Pathology: Current Challenges and Future Directions—An Updated Literature Review," *Journal of Molecular Pathology*, vol. 7, no. 1, pp. 2–2, Jan. 2026, doi: https://doi.org/10.3390/jmp7010002.

[3] M. Lu, Drew, T. Chen, R. J. Chen, M. Barbieri, and F. Mahmood, "Data Efficient and Weakly Supervised Computational Pathology on Whole Slide Images," Apr. 2020, doi: https://doi.org/10.48550/arxiv.2004.09666.

[4] J. Wang, Y. Mao, N. Guan, and C. J. Xue, "Advances in Multiple Instance Learning for Whole Slide Image Analysis: Techniques, Challenges, and Future Directions," *arXiv (Cornell University)*, Aug. 2024, doi: https://doi.org/10.48550/arxiv.2408.09476.

[5] M. Bilal *et al.*, "An aggregation of aggregation methods in computational pathology," *Medical Image Analysis*, vol. 88, pp. 102885–102885, Aug. 2023, doi: https://doi.org/10.1016/j.media.2023.102885.

[6] "A Spatially-Aware Multiple Instance Learning Framework for Digital Pathology," *Arxiv.org*, 2019. https://arxiv.org/html/2504.17379v1 (accessed Mar. 16, 2026).

[7] T. Albuquerque, Anil Yuce, M. D. Herrmann, and A. Gomariz, "Characterizing the Interpretability of Attention Maps in Digital Pathology," *ResearchGate*, Jul. 2024, doi: https://doi.org/10.48550/arXiv.2407.02484.

[8] S. Orsulic, J. John, A. E. Walts, and Arkadiusz Gertych, "Computational pathology in ovarian cancer," *Frontiers in oncology*, vol. 12, Jul. 2022, doi: https://doi.org/10.3389/fonc.2022.924945.

[9] R. Hong and D. Fenyö, "Deep Learning and Its Applications in Computational Pathology," *BioMedInformatics*, vol. 2, no. 1, pp. 159–168, Feb. 2022, doi: https://doi.org/10.3390/biomedinformatics2010010.

[10] A. Lomacenkova and Ognjen Arandjelović, "Whole Slide Pathology Image Patch Based Deep Classification: An Investigation of the Effects of the Latent Autoencoder Representation and the Loss Function Form," *IEEE-EMBS International Conference On Biomedical And Health Informatics*, Jul. 2021, doi: https://doi.org/10.1109/BHI50953.2021.9508577.

[11] Y. Zhou *et al.*, "Iterative multiple instance learning for weakly annotated whole slide image classification," *Physics in Medicine & Biology*, vol. 68, no. 15, p. 155007, Jul. 2023, doi: https://doi.org/10.1088/1361-6560/acde3f.

[12] S. Fatima, S. Ali, and H.-C. Kim, "A Comprehensive Review on Multiple Instance Learning," *Electronics*, vol. 12, no. 20, p. 4323, Jan. 2023, doi: https://doi.org/10.3390/electronics12204323.

[13] M. Ilse, J. M. Tomczak, and M. Welling, "Attention-based Deep Multiple Instance Learning," *arXiv.org*, Jun. 28, 2018. https://arxiv.org/abs/1802.04712

[14] S. Peled, Y. E. Maruvka, and M. Freiman, "PSA-MIL: A Probabilistic Spatial Attention-Based Multiple Instance Learning for Whole Slide Image Classification," Mar. 20, 2025. https://www.researchgate.net/publication/390038618_PSA-MIL_A_Probabilistic_Spatial_Attention-Based_Multiple_Instance_Learning_for_Whole_Slide_Image_Classification

[15] Z. Shao *et al.*, "TransMIL: Transformer based Correlated Multiple Instance Learning for Whole Slide Image Classification," *arXiv.org*, Oct. 31, 2021. https://arxiv.org/abs/2106.00908v2 (accessed Jun. 01, 2024).

[16] M. Ochi, Daisuke Komura, and S. Ishikawa, "Pathology Foundation Models," Jul. 30, 2024. https://www.researchgate.net/publication/382738982_Pathology_Foundation_Models

[17] M. Kaczmarek, M. Kowal, and J. Korbicz, "Exploring data preparation strategies: A comparative analysis of vision transformer and ConvNeXT architectures in breast cancer histopathology classification," *International Journal of Applied Mathematics and Computer Science*, vol. 35, no. 2, 2025, doi: https://doi.org/10.61822/amcs-2025-0023.

[18] M. Bilal, Aadam, M. Raza, Youssef Altherwy, and Nasir Rajpoot, "Foundation Models in Computational Pathology: A Review of Challenges, Opportunities, and Impact," Feb. 12, 2025. https://www.researchgate.net/publication/388955114_Foundation_Models_in_Computational_Pathology_A_Review_of_Challenges_Opportunities_and_Impact

[19] A. Mousa, "UniViT-MIL: A Unified Vision Transformer and Multi-Instance Learning Framework with Attention Heatmaps for Multi-Cancer Histopathology and Biomarker Prediction," Jan. 01, 2026. https://www.researchgate.net/publication/399601848_UniViT-MIL_A_Unified_Vision_Transformer_and_Multi-Instance_Learning_Framework_with_Attention_Heatmaps_for_Multi-Cancer_Histopathology_and_Biomarker_Prediction

[20] Y. Zhang, Z. Xia, G. Yin, and B. Liu, "Cluster-Level Sparse Multi-Instance Learning for Whole-Slide Images," *ResearchGate*, Sep. 2025, doi: https://doi.org/10.48550/arXiv.2509.11034.

[21] C.-W. Wang *et al.*, "Histopathological whole slide image dataset for classification of treatment effectiveness to ovarian cancer," *Scientific Data*, vol. 9, no. 1, Jan. 2022, doi: https://doi.org/10.1038/s41597-022-01127-6.

[22] M. S. Hosseini, D. Lee, D. Gershanik, D. Lee, S. Damaskinos, and K. N. Plataniotis, "Whole Slide Preview Image Segmentation and Setup for Digital Pathology Scanners," Feb. 2020, doi: https://doi.org/10.1101/2020.02.24.963645.

[23] M. Gadermayr and Maximilian Tschuchnig, "Multiple Instance Learning for Digital Pathology: A Review on the State-of-the-Art, Limitations & Future Potential," Jun. 09, 2022. https://www.researchgate.net/publication/361206739_Multiple_Instance_Learning_for_Digital_Pathology_A_Review_on_the_State-of-the-Art_Limitations_Future_Potential

[24] P. Bandi *et al.*, "Comparison of different methods for tissue segmentation in histopathological whole-slide images," *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2017, doi: https://doi.org/10.1109/isbi.2017.7950590.

[25] Bilel Guetarni, Feryal Windal, Halim Benhabiles, Mahfoud Chaibi, and D. Collard, "Histopathology image embedding based on foundation models features aggregation for patient treatment response prediction," Jul. 23, 2024. https://www.researchgate.net/publication/382971053_Histopathology_image_embedding_based_on_foundation_models_features_aggregation_for_patient_treatment_response_prediction

[26] Ferhat Sarikaya, "Validation Strategies in Machine Learning: Critical Analysis of Cross-Validation Techniques and Data Splitting Methods," Nov. 11, 2024. https://www.researchgate.net/publication/385707168_Validation_Strategies_in_Machine_Learning_Critical_Analysis_of_Cross-Validation_Techniques_and_Data_Splitting_Methods

[27] N. Marini, J. Sebastian, F. Ciompi, Stefano Marchesin, and Henning Müller, "Multi-Scale Multiple Instance Learning for the Classification of Histopathological Images with Global Annotations," *Lecture Notes in Computer Science*, Sep. 2021, Available: https://www.researchgate.net/publication/354760612_Multi-Scale_Multiple_Instance_Learning_for_the_Classification_of_Histopathological_Images_with_Global_Annotations

[28] X. Li, B. Yang, T. Chen, Z. Gao, and M. Huang, "ProMIL: A weakly supervised multiple instance learning for whole slide image classification based on class proxy," *Expert systems with applications*, vol. 238, pp. 121800–121800, Mar. 2024, doi: https://doi.org/10.1016/j.eswa.2023.121800.

[29] "(PDF) A Multiclass classification System for Early Detection and Leaf Pathology Classification using Deep Convolutional Networks," *ResearchGate*, 2025, doi: https://doi.org/10.1007//978-3-032-12827-0_16.

[30] M. Khened, A. Kori, H. Rajkumar, G. Krishnamurthi, and B. Srinivasan, "A generalized deep learning framework for whole-slide image segmentation and analysis," *Scientific Reports*, vol. 11, no. 1, Jun. 2021, doi: https://doi.org/10.1038/s41598-021-90444-8.

[31] Z. Qian *et al.*, "Transformer Based Multiple Instance Learning for Weakly Supervised Histopathology Image Segmentation," *Lecture notes in computer science*, pp. 160–170, Jan. 2022, doi: https://doi.org/10.1007/978-3-031-16434-7_16.

[32] Y. Azzi, A Moussaoui, and M. T. Kechadi, "Class Imbalance and Evaluation Metrics for Medical Image Segmentation with Machine Learning Models," *Second International Conference and School on Radiation Imaging and Nuclear Medicine-ICSRI-2023*, Jan. 2024, Available: https://www.researchgate.net/publication/377782096_Class_Imbalance_and_Evaluation_Metrics_for_Medical_Image_Segmentation_with_Machine_Learning_Models

[33] Y. Xie, J. Long, J. Hou, D. Chen, and G. Guan, "Weakly supervised pathological whole slide image classification based on contrastive learning," *Multimedia Tools and Applications*, vol. 83, no. 21, pp. 60809–60831, Jan. 2024, doi: https://doi.org/10.1007/s11042-023-17988-x.

[34] M. Afonso, P. M. S. Bhawsar, M. Saha, J. S. Almeida, and A. L. Oliveira, "Multiple Instance Learning for WSI: A comparative analysis of attention-based approaches," *Journal of Pathology Informatics*, vol. 15, p. 100403, Oct. 2024, doi: https://doi.org/10.1016/j.jpi.2024.100403.

[35] J. Funk *et al.*, "Comparison of an Attention-Based Multiple Instance Learning (MIL) With a Visual Transformer Model: Two Weakly Supervised Deep Learning (DL) Algorithms for the Detection of Histopathologic Lesions in the Rat Liver to Distinguish Normal From Abnormal," *Toxicologic Pathology*, vol. 53, no. 5, pp. 456–478, May 2025, doi: https://doi.org/10.1177/01926233251339653.

[36] P. Ruusuvuori *et al.*, "Spatial analysis of histology in 3D: quantification and visualization of organ and tumor level tissue environment," *Heliyon*, vol. 8, no. 1, p. e08762, Jan. 2022, doi: https://doi.org/10.1016/j.heliyon.2022.e08762.

[37] M. Yu, Z. Zhong, X. Zhou, Y. Wang, and K. Ma, "Human visual attention-inspired knowledge distillation underlying interpretable computational pathology," *Scientific Reports*, vol. 15, no. 1, Nov. 2025, doi: https://doi.org/10.1038/s41598-025-26004-1.

[38] A. A. Ahmed, M. Abouzid, and E. Kaczmarek, "Deep Learning Approaches in Histopathology," *Cancers*, vol. 14, no. 21, p. 5264, Oct. 2022, doi: https://doi.org/10.3390/cancers14215264.

[39] M. Asadi-Aghbolaghi *et al.*, "Learning generalizable AI models for multi-center histopathology image classification," *npj Precision Oncology*, vol. 8, no. 1, Jul. 2024, doi: https://doi.org/10.1038/s41698-024-00652-4.

[40] Satoshi Shirae, Shyam Sundar Debsarkar, Hiroharu Kawanaka, B. Aronow, and B. Surya, “Multimodal Ensemble Fusion Deep Learning Using Histopathological Images and Clinical Data For Glioma Subtype Classification,” *IEEE Access*, pp. 1–1, Jan. 2025, doi: https://doi.org/10.1109/access.2025.3556713.

[41] P. Chen, M. Aminu, S. E. Hussein, J. D. Khoury, and J. Wu, “CellSpatialGraph: Integrate hierarchical phenotyping and graph modeling to characterize spatial architecture in tumor microenvironment on digital pathology,” *Software Impacts*, p. 100156, Oct. 2021, doi: https://doi.org/10.1016/j.simpa.2021.100156.

[42] N. Andy, K. Nguyen, Allam, Nguyen, and K. Allam, “Deep Learning for Digital Image Analysis with Whole Slide Imaging for Lymphoma Diagnosis: Challenges and Promises, 21 st Century Pathology,” vol. 2, no. 2, p. 116, 2022, Accessed: Mar. 16, 2026. [Online]. Available: https://www.21stcenturypathology.com/articles/deep-learning-for-digital-image-analysis-with-whole-slide-imaging-for-lymphoma-diagnosis-challenges-and-promises.pdf