PROCEEDINGS

ISSN : 2456-1983

International Conference on Multidisciplinary Breakthroughs and NextGen Technologies (ICMBNT - 2025)

26 April 2025

Organized by

SOCIETY FOR CYBER INTELLIGENT SYSTEMS

Reg.No. 78 of 2025

Puducherry - india

www.societycis.org

Melange Publications

www.melangepublications.com

PROCEEDINGS OF

INTERNATIONAL CONFERENCE ON MULTIDISCIPLINARY BREAKTHROUGHS AND NEXTGEN TECHNOLOGIES

(ICMBNT–2025)

26th Apr, 2025

Organized By

SOCIETY FOR CYBER INTELLIGENT SYSTEMS
PUDUCHERRY – INDIA

MESSAGE FROM ORGANIZING COMMITTEE

We feel happy and honored to welcome all the distinguished guests and participants for the International Conference on Multidisciplinary Breakthroughs and NextGen Technologies to be held on 26th April 2025. This conference is organized by Society for Cyber Intelligent Systems, Puducherry – India. The primary mission of our society is to advance cybersecurity and intelligent systems by promoting cutting-edge technologies like AI and machine learning, fostering research in cyber intelligence, and enhancing threat detection and mitigation strategies.

It gives us immense pleasure to present the proceedings of this International Conference (ICMBNT 2025). This conference brings together leading researchers, academicians, industry experts, and students from around the world to share their insights, innovations, and contributions across diverse disciplines that shape our technological future.

The aim of ICMBNT 2025 is to foster collaboration across domains and promote pioneering research that addresses real-world challenges through next-generation technologies. The multidisciplinary nature of this conference provides a unique platform to explore the intersections of engineering, data science, artificial intelligence, biotechnology, sustainable systems, healthcare, social science, education and more.

We express our sincere gratitude to all authors, reviewers, session chairs, and members of the organizing and technical committees who contributed to the success of this conference. We also thank our keynote speakers and panellists for sharing their invaluable perspectives.

ICMBNT 2025 will be a central hub for esteemed research experts worldwide and can anticipate unparalleled opportunities to network, gain invaluable insights, showcase their hidden potential, present significant research findings, receive due credit and recognition for their contributions.

We hope these proceedings serve as a valuable resource and ignite further research and innovation in multidisciplinary and emerging technological fields.

With Best Regards,

Society for Cyber Intelligent Systems
Puducherry – India

Optimizing an Attention-based MIL Architecture for Generalizing on the UBC-OCEAN

Karthik G Shanmugasundaram [1], Prof Dr Ganesh Kumar S[2]
[1]Department of Computer Science, SRMIST Chennai, India
Email: kz1068@srmist.edu.in
[2]Department of Data Science and Business Systems, SRMIST Chennai, India
Email: ganeshk1@srmist.edu.in

Abstract: Early and accurate diagnosis of epithelial ovarian cancer is crucial for timely intervention. AI-driven accurate subtype classification will address subjective variability and is key to treatment management. To improve applicability, rare subtypes need to be flagged as outliers. Whole slide images (WSIs) and Tissue Microarray (TMA) images from the UBC-OCEAN dataset are available for ovarian cancer subtyping. TMA images are more varied in size for the unseen test set. WSI images are large (600MB+), and diseased tissues may be sparsely distributed, requiring careful feature extraction and use. Multiple instance learning approaches, a type of weakly supervised learning model, are key to effectively utilizing WSI for generalizability. Neural approaches have been favoured more than classical MIL because of trainability and interpretability. ABMIL (Attention mechanism-based MIL) has been used to improve the performance. In this work, results from an architecture with a vision transformer backbone and ABMIL is presented and optimized to generalize on the unique challenge of the UBC-OCEAN dataset.

Keywords: Ovarian Cancer, Attention, Vision Transformer, Multiple Instance Learning, UBC-OCEAN, Weakly supervised learning

1. Introduction

Ovarian cancer is one of the deadliest cancers faced by women across the world, according to the World Health Organization (WHO). The burden of the disease is higher in Asia than other countries underscoring the need for timely, accurate diagnostics. The gold standard for diagnosing cancer across the world is microscopic analysis of tissue using Hematoxylin and Eosin (H&E) stained glass slides. [1] [2] With the availability of digital imaging, digitized whole slide images (WSIs) of H&E slides are widely available from hospitals across the world. Typically, smaller cores from multiple WSIs are also assembled as tissue microarrays (TMAs) for comparative analysis.

UBC-OCEAN challenge

The University of British Columbia **O**varian **C**ancer sub- typ**E** cl**A**ssification and outlier detectio**N** (UBC-OCEAN) dataset claims to be "the world's most extensive ovarian cancer dataset of histopathology images" collected from more than 20 centers across the world. It was made available to researchers as part of the Kaggle competition which ran from October 2023 to January 2024. A total of 1326 teams from across the world submitted entries to the competition. [3]

The UBC-OCEAN dataset contains 2438 images split be- tween training, unseen public and private test sets. Table I shows the distribution in detail. Training data comprised of 513 WSIs (whole slide images) at 20x magnification and 25 TMAs (tissue microarrays) at 40x magnification. Fig. 1 shows samples of WSI and TMA. Submitted solutions ran on the public test set (437 images, 56% TMAs) and private test set (1488 images, 80% TMAs). [4]

The train dataset is imbalanced on the target class (ovarian cancer subtype labels) as shown by Fig. 2. The main objective of the competition was generalization as measured by balanced accuracy scores on these unseen test sets. The best entries from each team were ranked based on balanced accuracy score on the private test set. The first author was part of a team that submitted an entry with baseline performance that scored 855th during the competition duration.

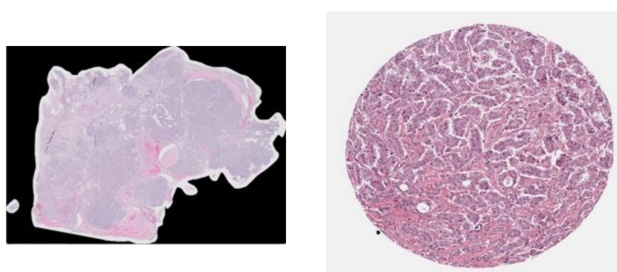
Whole slide images are huge, and disease tissue is typically sparse (Fig. 1a). Any approach to the competition must address this. Effective approaches to the UBC-OCEAN must also additionally address the reduced availability of TMAs in the train data and the need to build a model that can generalize to such a diverse, unseen test set. In particular, the competition organizers had this to say about the test set:

"The test set contains images from different source hospitals than the train set, with the largest area images almost 100,000 x 50,000 pixels. We strongly recommend taking an expansive approach to think- ing about the scenarios your error handling should manage, including differences in image dimensions, quality, slide staining techniques, and more." [3]

These can be pointers to be considered when generalizing.

TABLE 1. WSI, TMA DISTRIBUTION: UBC-OCEAN TRAIN AND TEST

Image type	**Dataset**		
	Train	***Public test***	***Private test***
% of TMAs	5%	56%	80%
WSI	513	194	299
TMA	25	243	1189
Total	538	437	1478



(a) Sample 10077, WSI, EC, 53143x39150, 20x

(b) Sample 13568, TMA, LGSC, 2964x2964, 40x

Figure 1. Sample images from UBC-OCEAN

SCOPE AND APPLICATION OF PRESENT WORK

In this work, the performance of attention mechanism based multiple instance learning (ABMIL) approaches on UBC-OCEAN is studied with a vision transformer architecture whose weights were pretrained on diverse pathology datasets by Lunit Inc., using self-supervised learning (SSL). [5] The latter approach was found effective and avoided heavy training resources neither available in the competition nor available to the authors. The aim was to match a top 5% score on the private test set with the available computational resources as a proof of generalizability. Outlier detection was also another objective of the competition. A standard entropy-based approach was followed to flag outliers but there were no labeled outliers in the training data. No defined measure of success with outliers was provided to the participants along with the competition score. Hence, this was not part of the work's focus.

United Nations Sustainable Development Goals (SDGs) - SDG 3 (Good health and wellbeing) and SDG 10 (reduced inequalities) will be met by this work, as improving performance on such a diverse dataset will enhance availability and diagnostic accuracy in the global south.

BACKGROUND

Until October 2024, there was a publication embargo on the UBC-OCEAN challenge which is now lifted. The present literature review studies the methods adopted for cancer subtyping of whole slide images, especially MIL approaches.

Ovarian cancer is one of the deadliest cancers faced by women all over the world. Asian countries face a significant burden of this disease and early diagnosis and intervention is crucial. Demand for qualified histopathologists is also high and artificial intelligence could be a valuable assistant in this regard helping expand availability and to address the problem of inter-observer variability. Ovarian cancer typically affects the skin of the ovaries and epithelial ovarian cancer presents itself as a heterogeneous disease with five common types: high-grade serous carcinoma (HGSC), clear cell ovarian carcinoma (CCOC), endometrioid carcinoma (ENOC), low- grade serous carcinoma (LGSC) and mucinous carcinoma (MUC). HGSC is the most common accounting for 70% of the cases. [1] [4] Accurate classification is important as treatment management is dependent on the histotype with each type exhibiting distinctive characteristics.

A major benchmarking study that addresses the challenge of subtyping, tumour localization and grading using whole slide images "implemented and systematically compared six different methods" that include the use of variety of feature extractors (ResNets, EfficientNet, Vision Transformers) as well as varieties of weakly supervised (WSL) and multiple instance learning (MIL) methods such as classical MIL, attention based MIL and CLAM (Clustering-constrained attention MIL). [6] They also evaluated these methods on six different WSI datasets covering different cancerous tissues (renal, colorectal, gastric, bladder).

Another study [2] focused on the development of deep learning-based MIL systems that used only the labels and showed high performance on an extensive variety of cancer datasets without requiring pixel-wise annotations. This showed the potential for applying deep learning to drive decision- making in clinical settings. The effectiveness of MIL techniques for histopathological image classification on breast

Figure 2. Imbalance in ovarian cancer subtypes
(UBC-OCEAN training)

cancer images has been demonstrated by [7]. In [8], different transfer learning approaches for classifying epithelial ovarian cancer WSI using a VGG19 backbone were studied.

Attention based MIL which produces a scalar attention score for each tile with two fully connected layers was introduced by [9].

CLAM, introduced by [10] has been empirically shown to outperform classical MIL by them. Among the five datasets considered by [6], their results show that ABMIL outperformed CLAM on five datasets as measured by the area under the receiver operating curve (AUROC). Both ABMIL and CLAM were shown to outperform classical MIL.

Transfer learning has now been accepted as a crucial part of tasks on medical images and [11] have studied how CNNs and ViTs can be used effectively for domain transfer of weights.

Choice of pre-trained domain specific weights over standard IMAGENET weights has been shown to be critical to achieve the best performance.

2. Methodology and Experiments

Methodological considerations

Generalization performance is the main consideration for this dataset. As already noted, the training dataset has 5% TMAs while the private test set has up to 80% TMAs. A variety of approaches were adopted by the top entries in the competition. One approach was to use 60% of the TMAs for validation as adopted by [12]. The choice of appropriate feature extractors and the use of transfer learning plays an important role in generalization performance as well as in reducing training time which is a relevant constraint for this work. [13]

Feature extractors*:* Choice of feature extractor decides the details of the architecture. Some of the common choices available that are state of the art in many use cases are: ResNets, Inception networks, ViTs, EfficientNets.

Evaluation*:* Evaluation of the multi-class classification was done using balanced accuracy score as required by the competition. [3] Balanced accuracy is computed as the average recall across all classes and is necessary to accurately assess the performance of imbalanced classification.

Baseline

For baseline, an InceptionV3 network was chosen. The idea was to choose something different from ViTs used in our target architecture for the sake of comparison.

Our initial choice for baseline and target architecture was efficient nets as these are designed to perform well in resource constrained environments. The difficulty with these experiments was that the way they scaled up did not fit well with the GPU/memory constraints especially when our ABMIL algorithm which requires heavy parallelization was applied.

Eventually, InceptionV3 was used as it could also fit a minimum number of tiles required to demonstrate ABMIL. These experimental choices were necessary to study the performance improvements of key modules of the target architecture.

Architecture

Fig. 3 shows the overall training architecture.

Preprocessing:

A significant proportion of the WSIs in the training set had repetitions of the same image and these were cropped out. Both WSIs and TMAs were also cropped to fit to improve training efficiency.

Splitting:

60% of the TMAs were reserved for validation in line with the need to generalize on TMAs expected in the public and private test sets. WSIs were split at 70/30 which is a more conventional split.

Feature extraction:

Pre-trained vision transformer weights from Lunit DINO with image sizes of 224x224 and using two patch sizes (8 and 16) were used to extract features from each image and were concatenated to produce the input to the ABMIL classifier. [5] The DINO training methodology was introduced by [14] and is noted to be a natural fit with vision transformers.

Transformers were developed for sequential data such as language where tokens of data are readily understood. Vision Transformers use patches to emulate this by converting a 2D image into sequence of "patches" of different sizes (here 8x8 and 16x16). [15]

ABMIL solution:

The key idea behind the ABMIL solution is to break the large images into tiles and to use an attention network to learn the mapping between the individual tiles and the target.

Attention Based Multiple Instance Learning

Multiple instance learning is a form of weakly supervised learning in which an image is divided into tiles and each tile inherits the label of the whole "bag" of instances. Typically, MIL works through aggregation of the tiles using some form of pooling (average, max etc) which is not trainable. With WSI being huge and disease tissue may be sparse and variable, neural networks such as attention mechanism have been used to handle this challenge. [6] [9]

In this design illustrated in Fig. 4, each tile is first passed through a feature extractor — in this case, a double Vision Transformer (ViT) backbone to obtain tile-level embeddings. Instead of averaging or pooling these embeddings, the model uses an attention mechanism to assign different weights to each tile.

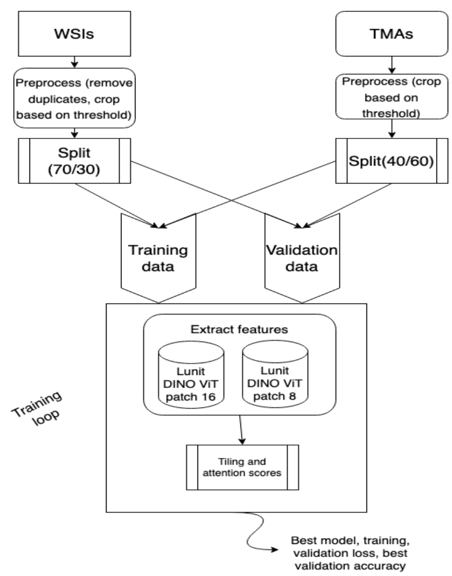


Figure 3. Training: Lunit DINO ViTs (two extractors 224x224, with patches 8 and 16) with ABMIL

By learning to focus on the most informative tiles in the training data, the attention mechanism can highlight critical regions while downplaying background patches that do not contribute to the label. Specifically, the model concatenates features from two different ViT patch sizes (patch8 and patch16), projects them into a lower-dimensional embedding via a linear layer and then computes tile-level attention scores using Attention A and Attention B layers. These scores are converted into weights via a softmax operation over all tiles. Finally, the weighted sum of tile embeddings is fed into a classification head, yielding the predicted label for type of ovarian cancer.

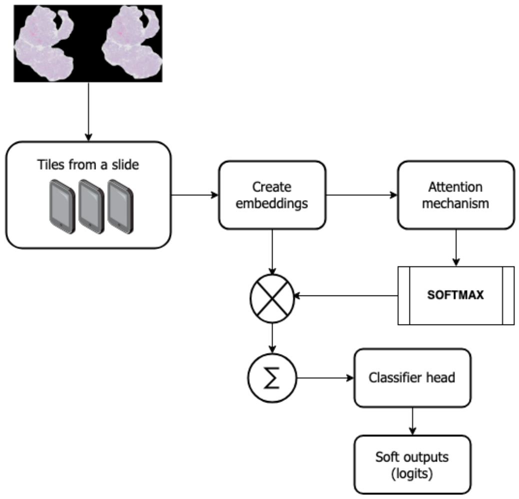


Figure 4. ABMIL classifier

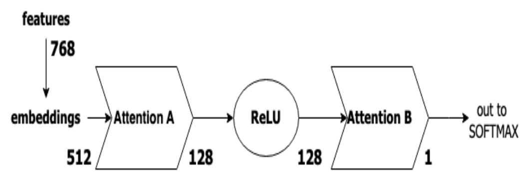


Figure 5. Tuned attention mechani

Attention mechanism implemented: Fig. 5 shows a schematic of the attention mechanism implemented and tuned for our problem. Two attention layers (A & B) were used following [16]. First, input features for each tile are embedded into a lower dimensional neural network (512 input, 128 output). A ReLu layer is applied to this and passed to a second attention layer B which produces a single attention logit for each tile, resulting in BxN outputs. As the network is trained on the data using backpropagation, the attention scores can be expected to reflect which tiles are relevant to the classification task. The parameters of this attention mechanism were manually tuned to improve generalization performance of the final task. The overall network was simpler than used in [16].

Algorithm 1 Forward Pass – Lunit DINO with ABMIL

Require: A batch of tile images $X \in \mathbb{R}^{B\times N\times C\times H\times W}$, where B is the batch size and N is the number of tiles per image
Ensure: Logits $\ddot{Y} \in \mathbb{R}^{B\times \text{num\_classes}}$

for $b \leftarrow 1$ to B **do**
    Initialize an empty list: featureList
    for $n \leftarrow 1$ to N do
        imgTile ← ***X[b, n, :, :, :]***
        Feature Extraction:
            tileFeats ← DoubleViTExtractor(imgTile)
            featureList.append(tileFeats)
    end for
    F[b] ← concatenate(featureList) $\in \mathbb{R}^{N\times 1536}$

```
end for
F ← concatenate all F[b] ∈ ℝ^(B×N×1536)
E ← LinearEmbed(F) (E ∈ ℝ^(B×N×d_embed))
attIntermediate ← ReLU(Linear(E, 128))
attLogits ← Linear(attIntermediate, 1)
A ← Softmax(attLogits, axis=1)
Z ← Σ_{n=1}^{N} A[:, n] · E[:, n] (Z ∈ ℝ^(B×d_embed))
Ÿ ← Classifier(Z)
return Ÿ
```

Compute and software

Kaggle provides 30 hours of training time per week on a Nvidia GPU P100. One training notebook can be run for 12 hours at a time after which it ends automatically. For inference, Kaggle provides 100 submissions per day that uses 2x NVIDIA GPU T4. Each inference submission can be run for a maximum of 9 hours. These are the constraints of the competition although participants were free to use external hardware for training. Ultimately, the constraint is mainly on inference time. The present work used the available training GPUs on Kaggle for training purposes as well.

Kaggle's notebook environment makes available most of the software required. The neural code was implemented in *PyTorch* and the package *timm* was used to load pre-trained models. For loading and preprocessing images, OpenCV, PIL were used. To improve preprocessing performance, another package called *pyvips* was used. No external data was used for modeling.

3. Result

Table II shows the results obtained during training and inference, comparing the baseline (InceptionV3) and the ABMIL with the target architecture.

It is to noted that increasing the number of tiles was not possible with InceptionV3 on the available hardware. ABMIL was first demonstrated using 20 tiles with an image feature size of 384x384 and improved performance over the baseline (which used 512x512 InceptionV3).

After this, 100 tiles were randomly sampled from each image and the target architecture with double Lunit DINO as detailed above. This improved validation accuracy and public test score, but the improvement in the private test score wasn't significant enough. More significant generalization improvements were achieved by tiling the image thumbnails completely using 400 tiles - 0.48 in public test score, 0.42 in private test score with the validation accuracy remaining roughly the same.

TABLE 2. EXPERIMENTAL RESULTS (BASELINE, ABMIL)

Solution	**Public Test Score**	**Private Test Score**	**Best Validation Balanced Accuracy**
Baseline (InceptionV3), 512x512	0.27	0.24	0.54
ABMIL (InceptionV3), 384x384, 20 tiles	0.36	0.33	0.72

Solution	Public Test Score	Private Test Score	Best Validation Balanced Accuracy
ABMIL (Double ViT, Lunit DINO), 224x224, 100 tiles random	0.42	0.35	.778
ABMIL (Double ViT, Lunit DINO), 224x224, 400 tiles, optimized, structured and complete	0.48	0.42	0.77

4. Conclusion and future works

As of the writing of this paper, a top 10% performance score on the competition (private test score) was matched. Model training is constrained by the size of the image, especially WSIs being huge. Current modelling uses thumbnails only because of memory constraints. To address this, WSI images are proposed to be loaded directly at 25% scaling. Heavy parallelization is required, and this is proposed to be achieved by rewriting the code to be efficient using advanced loading features of packages such as pyvips and reworking the training loop.

In addition to improving generalization performance to match the top scores, future work should address real world application which involves foundation models tailored to this domain, as studied by [17].

5. References

[1] H. Farahani et al., 'Deep learning-based histotype diagnosis of ovarian carcinoma whole-slide pathology images', *Modern Pathology*, vol. 35, no. 12, pp. 1983–1990, 2022.

[2] G. Campanella et al., 'Clinical-grade computational pathology using weakly supervised deep learning on whole slide images', *Nature Medicine*, vol. 25, no. 8, pp. 1301–1309, 2019.

[3] UBC (Kaggle Profile). UBC-OCEAN research competition homepage. Retrieved Mar 12, 2025 - https://www.kaggle.com/competitions/UBC-OCEAN

[4] M. Asadi-Aghbolaghi et al., 'Machine Learning-driven Histotype Diag- nosis of Ovarian Carcinoma: Insights from the OCEAN AI Challenge', *medRxiv*, 2024.

[5] Kang, M., et al, 'Benchmarking self-supervised learning on diverse pathology datasets', in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 3344–3354.

[6] N. G. Laleh et al., 'Benchmarking weakly-supervised deep learning pipelines for whole slide classification in computational pathology', *Medical Image Analysis*, vol. 79, p. 102474, 2022.

[7] P. J. Sudharshan, C. Petitjean, F. Spanhol, L. E. Oliveira, L. Heutte, and P. Honeine, 'Multiple instance learning for histopathological breast cancer image classification', *Expert Systems with Applications*, vol. 117, pp. 103–111, 2019.

[8] Y. Wang et al., 'Classification of epithelial ovarian carcinoma whole- slide pathology images using deep transfer learning', *arXiv preprint* arXiv:2005.10957, 2020.

[9] M. Ilse, J. Tomczak, and M. Welling, 'Attention-based deep multiple instance learning', in *International Conference on Machine Learning*, 2018, pp. 2127–2136.

[10] M. Y. Lu, D. F. K. Williamson, T. Y. Chen, R. J. Chen, M. Barbieri, and F. Mahmood, 'Data-efficient and weakly supervised computational pathology on whole-slide images'. *Nature Biomedical Engineering*, vol. 5, pp. 555–570, 2021.

[11] Christos Matsoukas, Johan Fredin Haslum, Moein Sorkhei, Magnus Söderberg, and Kevin Smith. 'What makes transfer learning work for medical images: Feature reuse & other factors', in *CVPR*, pp. 9225–9234, 2022.

[12] Owkin. *Kaggle 1st place solution report*. Retrieved Jan 31, 2025 from https://www.kaggle.com/competitions/UBC-OCEAN/discussion/466455

[13] Filiot et al., 'Scaling Self-Supervised Learning for Histopathology with Masked Image Modeling', *medRxiv*, 2023.

[14] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. 'Emerging properties in self-supervised vision transformers.' in *ICCV*, pp. 9650–9660, 2021.

[15] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 'An image is worth 16x16 words: Transformers for image recognition at scale.' in *ICLR*, 2020.

[16] Zznznb (Kaggle Profile). *Kaggle 2nd place solution report*. Retrieved Mar 6, 2025 from https://www.kaggle.com/competitions/UBC-OCEAN/discussion/465410

[17] J. Breen, K. Allen, K. Zucker, L. Godson, N. M. Orsi, and N. Ravikumar, 'Histopathology Foundation Models Enable Accurate Ovarian Cancer Subtype Classification', *arXiv preprint* arXiv:2405.09990, 2024.