Journal Pre-proofs

Research Paper

Whole slide image-level classification of malignant effusion cytology using clustering-constrained attention multiple instance learning

Dongwoo Kim, Jongwon Lee, Minsoo Jung, Kwangil Yim, Gisu Hwang, Hongjun Yoon, Daeky Jeong, Won June Cho, Mohammad Rizwan Alam, Gyungyub Gong, Nam Hoon Cho, Chong Woo Yoo, Yosep Chong, Kyung Jin Seo

PII: S0169-5002(25)00444-1
DOI: https://doi.org/10.1016/j.lungcan.2025.108552
Reference: LUNG 108552

To appear in: *Lung Cancer*

Received Date: 16 December 2024
Revised Date: 16 April 2025
Accepted Date: 20 April 2025

Please cite this article as: D. Kim, J. Lee, M. Jung, K. Yim, G. Hwang, H. Yoon, D. Jeong, W.J. Cho, M.R. Alam, G. Gong, N.H. Cho, C.W. Yoo, Y. Chong, K.J. Seo, Whole slide image-level classification of malignant effusion cytology using clustering-constrained attention multiple instance learning, *Lung Cancer* (2025), doi: https://doi.org/10.1016/j.lungcan.2025.108552

TITLE

Whole slide image-level classification of malignant effusion cytology using clustering-constrained attention multiple instance learning

Dongwoo Kim[1, #], Jongwon Lee[1, #], Minsoo Jung[1, #], Kwangil Yim[2], Gisu Hwang[3], Hongjun Yoon[3], Daeky Jeong[3], Won June Cho[3], Mohammad Rizwan Alam[2], Gyungyub Gong[4], Nam Hoon Cho[5], Chong Woo Yoo[6], Yosep Chong[2, *], Kyung Jin Seo[2, *]

[1]The Catholic University of Korea College of Medicine, Seoul, Korea; [2]Department of Hospital Pathology, The Catholic University of Korea College of Medicine, Seoul, Korea; [3]AI Research Lab, DEEPNOID Inc., Seoul, Korea; [4]Department of Pathology, Asan Medical Center, Seoul, Korea; [5]Department of Pathology, Yonsei University College of Medicine, Seoul, Korea; [6]Department of Pathology, National Cancer Center, Ilsan, Gyeonggi-do, Korea

[#]D.K., J.L., and M.J. are medical students who participated in this research as a part of the Student Research Program (SRP) of the Catholic University of Korea and contributed equally to this work as first authors.

Short title: WSI-CLAM model for Effusion Cytology.

WORD COUNTS, FIGURES AND TABLES

Word count for the abstract	: 220 words
Word count for the text	: 3,489 words
Number of Figures	: 3
Number of Tables	: 3

***Corresponding author:**

Dr. Kyung Jin Seo, MD, PhD and Dr. Yosep Chong, Department of Hospital Pathology, Uijeongbu St. Mary's Hospital, College of Medicine, The Catholic University of Korea, 271 Cheonboro, Uijeongbu 11765, Republic of Korea, Tel: +82-31-820-3158, E-mail: ywacko@catholic.ac.kr; ychong@catholic.ac.kr

Abstract

Background: Cytological diagnosis of pleural effusion plays an important role in the early detection and diagnosis of lung cancers. Recently, attempts have been made to overcome low diagnostic accuracy and interobserver variability using artificial intelligence-based image analysis. However, such analysis is primarily performed at the image-patch level and not at the whole-slide image (WSI) level. This study aims to develop a WSI-level classification of malignant effusions in metastatic lung cancer based on pleural fluid cytology using a quality-controlled, nationwide dataset.

Methods: The dataset was collected by a consortium research group that included three major university hospitals and the Committee of Quality Assurance Program Committee of the Korean Society of Cytopathology. It contains 576 normal and 309 cancer WSIs from pleural fluids. A clustering-constrained attention multiple-instance learning (CLAM) model was used for WSI-level classification.

Results: The CLAM model achieved a high accuracy of 97%, with an area under the curve of 0.97, representing a 13% improvement over the image patch classification model-based WSI classification. It also significantly reduced the analysis time and computing resources compared with those required during image patch-level classification and heat map generation on the WSIs.

Conclusion: The CLAM model successfully demonstrated high performance in differentiating malignant effusion at the WSI level using a large, quality-controlled, nationwide dataset. Further external validation is required to ensure generalizability.

Keywords: Lung Neoplasm; Malignant Effusion; Cytology; Deep Learning; Multiple Instance Learning

1. Introduction

Pleural fluid cytology is a critical first-line diagnostic test for evaluating pleural effusion, aiding in the diagnosis of inflammatory and infectious diseases, as well as respiratory malignancies, such as lung cancer, which ranks as the leading cause of cancer-related death worldwide [1, 2]. Its advantages include simplicity, cost-effectiveness, relative non-invasiveness, reusability for repeated testing, and potential therapeutic applications [3]. However, the effectiveness of the test is limited by secondary morphological alterations secondary to pleural effusion, making it insufficient for a definitive diagnosis alone; therefore, it can only be utilised as a screening test. The diagnostic sensitivity of pleural effusion cytology ranges from 40 to 90%, with significant interobserver variability that presents challenges even for experienced cytopathologists [4, 5]. Consequently, various efforts have been made to overcome these limitations and enhance the diagnostic accuracy and reliability of pleural fluid cytology.

Recently, artificial intelligence (AI) has been widely adopted in various medical fields and has shown promising results, including in histology and cytopathology [6]. Several studies have applied AI-based image analysis to pleural fluid cytology to enhance its accuracy and consistency. Barwad et al. constructed an artificial neural network model using detailed cytological features, including morphometric, densitometric, and chromatin textural data from 114 effusion cytology cases [7]. Tosun et al. applied machine learning trained with quantified chromatin distributions from 34 mesothelioma cytology cases [8]. Both studies demonstrated 100% accuracy in testing and validation; however, they were limited by small dataset sizes and an inability to use whole-slide images (WSIs) as input data. Moreover, both studies analysed data exclusively at the image patch level, rather than at the WSI level, requiring fully supervised learning.

Fully supervised learning, used in patch-level classification, requires individual cell annotations within each specimen, which poses considerable challenges to manage as dataset size increases. Weakly supervised learning, by contrast, can overcome this limitation by requiring only slide-level annotations with multiple instance learning (MIL) models [9]. The clustering-constrained attention MIL (CLAM) framework offers a novel approach to WSI-level classification in computational pathology. CLAM efficiently learns from WSIs with slide-level labels and has shown success in multiclass classification and subtyping of renal cell carcinoma, non-small cell lung cancer (NSCLC), and lymph node metastasis [10]. Ren et al. developed a model using MIL for several tasks, including differentiating benign and malignant pleural effusion, identifying the primary location of metastatic cancer, and predicting genetic alterations associated with targeted therapies, based on 1,321 effusion cytology cases [11]. However, this study was limited by its single-centre nature, the absence of external validation, and reliance solely on haematoxylin and eosin (H&E) stains.

We address these limitations by utilizing a quality-controlled, nationwide, multi-centre pleural fluid dataset containing H&E and Papanicolaou (Pap)-stained WSIs. Our objective was to develop classification models for malignant effusion in metastatic lung cancer using patch- and WSI-level classifications. This approach aims to provide a more comprehensive representation of cytopathological conditions, thereby enhancing the accuracy and reliability of AI predictions. The patch-level model employs z-stacking to enhance 3D structures and cellular focus, potentially improving algorithm performance. The WSI-level model utilises CLAM to leverage WSI-level labels, representing a substantial advancement by addressing the size constraints and other challenges faced by previous studies.

2. Methods

This study was approved by the Institutional Review Boards of the Catholic University of Korea College of Medicine (UC21SNSI0064), Yonsei University College of Medicine (4-2021-0569), and the National Cancer Center (NCC2021-0145). An overview of the method for classifying malignant and benign cells in pleural cell slide images of patients with lung cancer using a deep convolutional neural network (DCNN) model is presented in **Figure 1**.

2.1. Data Collection

The cytopathology images used in this study were obtained from the Open AI Dataset Project for Cytopathology in 2021 (https://www.aihub.or.kr; accessed on June 19, 2023). In collaboration with the Korean Society of Cytopathology, 13 medical institutions, including the Catholic University of Korea Uijeongbu St. Mary's Hospital, the National Cancer Center, and the Korea Cancer Center Hospital, constructed a digital cytology learning dataset. This dataset was created by refining, labelling, storing, and quality-controlling cytopathology images, and consists of 5,506 WSIs for eleven types of cancer (lung cancer, breast cancer, gastric cancer, colorectal cancer, ovarian cancer, pancreatic cancer, bladder cancer, thyroid cancer, salivary gland cancer, esophageal cancer, and lymphoma) were included, as well as benign cases. The cytological dataset collection and preparation processes are shown in **Supplementary Figures S1 and S2**. From this dataset, we used only Pap- and H&E-stained WSIs of pleural fluid specimens, each classified as benign or malignant for lung cancer.

All malignant cytopathological slides were confirmed via histological examination. The dataset for this study was collected using the Open AI Dataset and Quality Assurance Program of Korean cytopathology. In both programs, more than 10 board-certified pathologists were involved. These board-certified pathologists reviewed the cytopathological slides and corresponding histopathological slides before establishing the final diagnosis.

The collected WSIs were scanned as extended depth-of-field images by merging the multilayered z-stacked images using slide scanners to correct for image defocusing. The scanners used were AT2 (Leica Biosystems, Nussloch, Germany), Pannoramic Flash 250 III (3DHISTECH, Budapest, Hungary), and NanoZoomer S360 (Hamamatsu, Shizuoka, Japan). The scanned WSIs had Z-stacks with three layers in 70% and five layers in 30% of the WSIs. The z-layer distance was distributed as 1 µm (78%) and 0.4 µm (22%). To extract the region of interest, we used the default preprocessing settings available in CLAM. However, we made some modifications to the preset parameters of image processing. Specifically, we increased the kernel size used for the closing morphology operation because of the larger gaps between the high-value regions in cell region extraction. In addition, we lowered the minimum pixel value threshold because the preset available in CLAM was optimized for H&E-stained WSIs.

2.2. Patch classification

2.2.1. Image Preprocessing

Several processes were performed to generate patch images suitable for training the DCNN using the obtained WSIs. In the first step, the colour of each WSI was normalised to ensure colour constancy. Subsequently, we split these images into non-overlapping small patches of 1,024 × 1,024 pixels. The pathologists then classified and labelled the split images as malignant or benign. After splitting, downsampling converted the images into 256 × 256-pixel images. Finally, the image patches were augmented using horizontal, vertical, and clockwise rotations to improve the sufficiency and diversity of the training data. The image patch data were then reviewed by pathologists (**Figure 1B**).

2.2.2. Pretesting for DCNN Model Selection

We have tested six models: The models used were Inception-ResNet-v2, EfficientNet-b0, ResNet50, MobileNet-v2, DenseNet-121, and ResNeXt50.

2.2.3. Training, validation, and testing

We experimented with a 165-layer Inception-ResNet-v2 as shown in **Supplementary Figure S3A**. In this experiment, we substituted the SoftMax function in the output layer with a sigmoid function for the binary classification. Augmented patch images were randomly assigned to train, validate, and test the DCNN model. The probability of malignancy for each image patch was estimated; if the probability was 50% or higher, the image patch was classified as malignant. The dataset was divided into three parts—training, validation, and testing—in an 8:1:1 ratio, as specified.

2.2.4. Performance comparison between pathologists and AI

To compare the performance of the pathologists and AI, the sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) for the diagnosis of 1,000 patch images were measured. Four experienced cytopathologists independently made the diagnosis. After comparing the performances, the diagnoses were made once again by the cytopathologists with the assistance of the AI model to examine whether the AI model could be useful to experienced pathologists for diagnosis.

2.3. WSI-level classification by CLAM

2.3.1. Training, validation, and testing

In WSI-level classification, we substituted the SoftMax function in the output layer with a sigmoid function for binary classification. The deep learning structure of the CLAM-based model is presented in **Figure 1 C**. During the classification process, the WSIs were divided into patches, and features were extracted from these patches. The

ResNet50 model was selected owing to its suitability for feature extraction, as illustrated in **Supplementary Figure S3B**. Subsequently, the CLAM model estimated the probability of malignancy for each image patch based on these features and generated a heatmap of the entire WSI, classifying it as either benign or malignant. The dataset was divided into three parts: training, validation, and testing, in an 8:1:1 ratio.

We aimed to compare the diagnostic performance of two CLAM algorithms: single-branch (SB) and multi-branch (MB). Therefore, sensitivity, specificity, accuracy, area under the curve (AUC), and 95% confidence intervals (CI) were calculated for each algorithm during training, validation, and testing. These results were computed regardless of whether the image was stained with Pap stain or H&E.

2.3.2. External validation

To validate the performance of the model, we collected a different dataset from the archive of the Quality Assurance Program archive of the Korean Society of Cytopathology. Each WSI was labelled as either benign or malignant based on matching histological diagnoses by board-certified pathologists. The sensitivity, specificity, accuracy, AUC, and their 95% CI were calculated for CLAM-SB and CLAM-MB.

3. Results

3.1. Data collection

We acquired a total of 885 WSIs of pleural fluid specimens from the Open AI Dataset, comprising 576 WSIs confirmed to be benign and 309 WSIs confirmed to be malignant for lung cancer. These WSIs were used for the training, validation, and testing of both patch-level and WSI-level classifications. The average age of patients was 68.6 ± 14.8 years in the benign group and 68.4 ± 13.7 years in the malignant group, aligning with the typical age of patients with lung cancer. The benign group consisted of 61.6% men and 38.4% women, whereas the malignant groups consisted of 51.7% men and 47.8% women. The slides used for the benign and malignant group were collected from seven and 64 institutions, respectively. The names of the institutions are listed in **Supplementary Table 1**. Among the slides confirmed to be malignant, NSCLC accounted for the largest proportion (96.4%), with 3.5% attributed to small cell lung cancer. Among the slides confirmed to be malignant, NSCLC Subtyping information was unavailable for 40.1% of cases.

Additionally, a new dataset of 93 pleural fluid WSIs extracted between 2019 and 2021 from the Quality Assurance Program archive of the Korean Society of Cytopathology was used to externally validate the WSI-level classification. The characteristics of the dataset are listed in **Table 1**. The average age of patients was 80.4 ± 13.1 years in the benign group and 69.6 ± 13.4 years in the malignant group, aligning with the typical age of patients with lung cancer. Benign group consisted of 65.6% men and 34.4% women, while malignant group consisted of 65.5% men and 34.4% women. Slides used for the benign and malignant groups were collected from two and 40 institutions, respectively. The names of the institutions are listed in **Supplementary Table 2**. Among the slides confirmed to be malignant in external validation , NSCLC accounted for the largest proportion (93.4%), with 6.5% attributed to small-cell lung cancer. Subtyping information for NSCLC was unavailable for 31.1% of the slides in external validation set. Detailed data characteristics of the datasets are shown in **Table 1**.

The detailed number of slides used for each patch-level and WSI-level classification is presented in **Supplementary Table 3**. For patch-level classification, we used only Pap-stained WSIs because the number of patches extracted from the H&E-stained WSIs was insufficient for valid training. In total, 583 Pap-stained WSIs, including 415 benign and 168 malignant specimens, were used to generate 30,801 patches for the training, validation, and testing process. The patch-wise distribution of benign and malignant images was well balanced (**Supplementary Table 3A**). For the WSI-level classification, 647 Pap-stained WSIs and 190 H&E-stained WSIs were used. The Pap-stained WSIs included 416 benign and 231 malignant slides, whereas the H&E-stained WSIs included 150 benign and 40 malignant slides (**Supplementary Table 3B**).

3.2. Patch-level classification

3.2.1. Pretesting for DCNN model selection

Six DCNN models (Inception-ResNet-v2, EfficientNet-b0, ResNet50, MobileNet-v2, DenseNet-121, and ResNeXt50) were trained using 24,649 patches, validated using 3,131 patches, and tested using 3,021 patches. The overall performances during training, validation, and testing were assessed to select the best model for patch-level classification. The overall test accuracy for image classification ranged from 98.3% to 99.2%, with a sensitivity of 99.4% to 99.9%, and a specificity of 97.3% to 98.7%.

The test results of Inception-ResNet-v2 showed an accuracy of 99.1%, sensitivity of 99.6%, and specificity of 98.7%. For the other models, EfficientNet-b0 showed an accuracy of 98.3%, sensitivity of 99.4%, and specificity of 97.3%; ResNet50 showed an accuracy of 99.2%, sensitivity of 99.6%, and specificity of 98.7%; MobileNet-v2 showed an accuracy of 98.3%, sensitivity of 99.4%, and specificity of 97.3%; DenseNet-121 showed an accuracy of 99.1%, sensitivity of 99.4%, and specificity of 98.7%; and ResNext50 showed an accuracy of 98.7%, sensitivity of 99.9%, and specificity of 97.6%. Based on these results, Inception-ResNet-v2 was considered the most suitable model for patch-level classification, because its results remained consistent throughout the entire process. The detailed performance data for each model in the training, validation, and test sets are presented in **Table 2**.

3.2.2. Training, validation, and testing

The Inception-ResNet-v2 model was evaluated using the results summarized in **Supplementary Table 4**. Of the 1,529 malignant patches, the AI correctly identified 1,523 as malignant, with a sensitivity of 99.6%. For benign patches, the model correctly classified 1,472 out of 1,492, resulting in a specificity of 98.7%. The false-positive and false-negative rates were 1.3% and 0.4%, respectively (**Supplementary Figure 4**)

3.2.3. Performance comparison between pathologists and AI

The performance of each pathologist, average of the pathologists, AI model, and average of the pathologists aided by the AI model are shown in **Figure 2**. The average accuracy of the four pathologists was 84.3 ± 5.5%, with a sensitivity, specificity, PPV, and NPV of 81.2 ± 15.2%, 87.4 ± 12.5%, 88.7 ± 8.7%, and 87.4 ± 12.5% NPV, respectively. The AI model achieved an accuracy, sensitivity, specificity, PPV, and NPV of 94.8%, 90.8%, 98.8%, 98.7%, and 98.8%, respectively, for classifying patch images, thus outperforming the average accuracy achieved by experienced pathologists. The average accuracy of pathologists aided by the AI model was 92.2 ± 4.2%, with a sensitivity, specificity, PPV, and NPV of 87.0 ± 8.7%, 97.5 ± 0.7%, 97.2 ± 0.7%, and 97.5 ± 0.7%, respectively. Cases with inconsistencies in the diagnosis between the AI and pathologists were re-examined, resulting in improved average accuracy, sensitivity, specificity, PPV, and NPV of 92.2, 87.0, 97.5, 97.2, and 97.5%, respectively.

Cohen’s kappa and Fleiss’ kappa scores were also calculated to assess agreement in diagnoses. Cohen's kappa scores for the cytopathologists ranged from 0.356 to 0.780. After using the AI model, these scores generally increased, ranging from 0.687 to 0.906, as depicted in **Supplementary Figure 3**. The Fleiss’ kappa score among cytopathologists also increased from 0.748 to 0.884.

Figure 3 shows examples of patch images where the pathologists and the AI model disagreed in diagnosis.

3.3. CLAM

3.3.1. Training, validation, and testing

The diagnostic performance of the CLAM-SB and CLAM-MB was assessed using a confusion matrix (**Supplementary Table 5**) and performance metrics (**Table 3**), including sensitivity, specificity, accuracy, and AUC. For CLAM-SB, 21 of the 27 malignant cases were correctly identified as malignant, resulting in a sensitivity of 77.8%. Among the 57 benign cases, 47 were correctly classified, yielding a specificity of 82.5%. The overall accuracy for CLAM-SB was 81.0% (95% CI, 72.6%–89.4%), and the AUC was 0.744 (95% CI, 0.624–0.864).

For CLAM-MB, 22 out of 27 malignant cases were correctly identified, with a sensitivity of 81.5%, whereas 54 of 57 benign cases were correctly classified, resulting in a specificity of 94.7%. The model achieved an accuracy and AUC of 90.5% (95% CI, 84.2%–96.8%) and 0.942 (95% CI, 0.879–1.000), respectively. These results demonstrate the superior performance of CLAM-MB, particularly for specificity, compared with that of CLAM-SB. An example of a heatmap image generated using the CLAM model is shown in **Supplementary Figure 6**.

3.3.2. External validation

On the external dataset, validation results for CLAM-SB showed an accuracy, sensitivity, specificity, and AUC of 85.3% (95% CI, 78.1%–92.4%), 87.3% (95% CI, 79.2%–95.5%), 81.3% (95% CI, 67.7%–94.7%), and 0.930 (95% CI, 0.880–0.980), respectively, whereas CLAM-MB showed an accuracy, sensitivity, specificity, and AUC of 90.5% (95% CI, 84.6%–96.4%), 90.5% (95% CI, 83.4%–97.6%), 90.6% (95% CI, 80.5%–100.0%), and 0.947 (95% CI, 0.904–0.989), respectively (**Supplementary Table 6**).

4. Discussion

The use of AI to diagnose pathological WSIs across various organs has led to considerable progress in the medical field [12-17]. Few studies have used artificial neural networks in effusion cytology [18]. Consequently, additional research is required to explore the potential of neural networks in aiding in diagnostic cytology. In this study, we successfully demonstrated for the first time that AI exhibited remarkable accuracy in diagnosing lung cancer pleural effusion cytopathology. To our knowledge, this is the first study to use a large dataset that includes z-stacking and the first to attempt WSI-level classification of malignant effusions using CLAM.

Recent efforts have aimed to enhance diagnostic accuracy and reduce inter-observer variation in cytology through AI-based image analysis. Barwad et al. (2012) developed an artificial neural network model utilizing detailed cytological features, image morphometry, densitometric data, and chromatin textures from 114 effusion cytology cases [7]. Similarly, Tosun et al. (2015) employed machine learning to quantify chromatin distribution in 34 mesothelioma cytology cases [8]. Both studies achieved 100% accuracy in testing and validation but were limited by small datasets and the inability to use the actual images as input data. Ren et al. (2023) developed models using MIL for several tasks, including diagnosing benign and malignant pleural effusion, identifying the primary location of metastatic cancer, and predicting genetic alterations associated with targeted therapy, based on 1,321 effusion cytology cases [11]. However, the study was limited by its single-centre nature, the absence of external validation, and reliance solely on H&E-stained slides.

In the present study, the sample size was sufficiently large and exhibited a high level of demographic diversity and heterogeneity, exceeding previous studies' findings. Two AI models were used to diagnose lung cancer pleural effusion cytopathology. For the first model, used for patch-level classification, we employed merged z-stacked images, resulting in improved focus. For the second model, used for WSI-level classification, we employed CLAM. This model demonstrated classification ability even more accurately than the first model, indicating the potential for WSI-level classification in cytopathology. Furthermore, when cases in which the AI and pathologists disagreed on the diagnosis were re-evaluated, the pathologists' diagnostic accuracy improved. This suggests that AI assistance can enhance cytopathologists' interpretive capabilities by balancing sensitivity and specificity, and standardizing diagnostic criteria, and mitigating the variability that may arise from human subjectivity. This improvement in diagnostic concordance not only enhances the overall accuracy of pathology assessments but also underscores the potential for AI to serve as a valuable collaborator in medical decision-making processes.

Although our AI model demonstrated better overall diagnostic performance than that of human pathologists, it is not meant to replace them but to serve as a powerful tool to enhance their efficiency and precision, ultimately improving medical treatment. Our model misdiagnosed 18 malignant cases as false negatives, which were correctly identified by all pathologists. This discrepancy, which is difficult to account for owing to the model's lack of interpretability, may be attributable to several factors. Pleural fluid acts as a culture medium for floating cells, leading to a reduced nuclear-to-cytoplasmic ratio that mimics normal cells

[19]. Additionally, mesothelial cells can become activated and exhibit traits similar to those of malignant cells, particularly when exposed to inflammatory stimuli, such as pneumonia [20]. Macrophages activated in response to pleural effusion or inflammation can also resemble malignant cells [21]. These nuances, which occasionally mislead AI models, highlight the essential role of human pathologists in correcting AI's errors to ensure accurate diagnoses. Significant improvements in medical diagnostics and treatment can be achieved by integrating AI with human expertise.

Since the introduction of CLAM, Bobowicz et al. (2023) utilised it for diagnosing breast cancer [22], and Geijs et al. (2024) applied it to diagnose basal cell carcinoma [23]. CLAM can be used in various medical fields beyond cancer diagnosis. Kim et al. (2023) compared and analysed tau immunohistochemistry WSIs of five types of tauopathies with a control group, aiming to overcome the limitations of traditional methods that can only analyze images dichotomously using CLAM [24]. King et al. (2024) aimed to implement a model to predict the recurrence of Chiari malformations by combining whole volumetric magnetic resonance imaging and clinical data through CLAM [25].

This study has few limitations. First, the lack of ethnic diversity, as the dataset consists solely of Asian individuals, may limit the generalizability of the findings to other healthcare settings. Second, we excluded cases of abnormal mesothelial proliferation, including mesothelioma, due to their low incidence in our dataset. Developing an AI model to differentiate mesothelioma from benign or metastatic conditions would require a balanced dataset with roughly 33% representation from each category. However, assembling such a dataset was not feasible for this study due to the rarity of mesothelioma in our population.

Third, the AI model used in this study is not explainable, thereby having a limited ability to elucidate the reasons behind its decisions or predictions in a way that humans can understand. Further efforts are needed to address these limitations.

Future studies should focus on using AI to determine more details regarding different types of diseases and predict their behavior at the molecular level. This can help us understand specific variants of diseases and find better ways to treat them [26, 27]. Using AI to classify subtypes and predict molecular changes, personalized treatments that match each patient's unique illness characteristics can be developed. This not only improves how AI’s role in diagnosing diseases but also has the potential to transform clinical decision-making regarding optimal treatments for patients.

5. Conclusion

This study demonstrated that patch image-level classification can provide significant accuracy in diagnosing lung cancer pleural effusion cytopathology, enhancing the interpretative ability and reducing diagnostic discrepancies among cytopathologists. This is the first study to classify lung cancer by pleural effusion at the WSI level using the CLAM model. It successfully demonstrated extraordinary performance in differentiating malignant effusions using a large, quality-controlled nationwide dataset. Future studies should focus on subtype classification and molecular predictions. Efforts should be directed towards developing methods to develop a more accurate AI model that requires fewer computing resources for model training, thereby increasing its accessibility.

ABBREVIATIONS

WSI: Whole-slide image

CLAM: Clustering-constrained attention multiple-instance learning

AI: Artificial intelligence

MIL: Multiple instance learning

NSCLC: Non-small cell lung cancer

H&E: Haematoxylin and eosin

Pap: Papanicolaou

DCNN: Deep convolutional neural network

PPV: Positive predictive value

NPV: Negative predictive value

SB: Single branch

MB: Multi-branch

AUC: Area under the curve

CI: Confidence interval

AUTHOR CONTRIBUTION

Dongwoo Kim: Data curation; formal analysis; investigation; methodology; validation; visualization; writing – original draft; writing – review and editing.

Jongwon Lee: Data curation; formal analysis; investigation; methodology; validation; visualization; writing – original draft; writing – review and editing.

Minsoo Jung: Data curation; formal analysis; investigation; methodology; validation; visualization; writing – original draft; writing – review and editing.

Kwangil Yim: Data curation; formal analysis; investigation; methodology; writing – review and editing.

Gisu Hwang: Data curation; formal analysis; investigation; methodology; writing – review and editing.

Hongjun Yoon: Data curation; formal analysis; investigation; methodology; writing – review and editing.

Deaky Jeong: Data curation; formal analysis; investigation; methodology; writing – review and editing.

Won June Cho: Data curation; formal analysis; investigation; methodology; writing – review and editing.

Mohammad Rizwan Alam: Data curation; formal analysis; investigation; methodology; writing – review and editing.

Gyungyub Gong: Data curation; formal analysis; investigation; methodology; writing – review and editing.

Nam Hoon Cho: Data curation; formal analysis; investigation; methodology; writing – review and editing.

Chong Woo Yoo: Data curation; formal analysis; investigation; methodology; writing – review and editing.

Yosep Chong: Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; software; supervision; validation; visualization; writing – original draft; writing – review and editing.

Kyung Jin Seo: Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; software; supervision; validation; visualization; writing – original draft; writing – review and editing.

ACKNOWLEDGEMENT

We thank Muhammad Joan Ailia (The Catholic University of Korea (CUK)), Jamshid Abdul-Ghafar (CUK), In Park (CUK), Ah Reum Kim (CUK), Seona Shin (National Cancer Center (NCC)), Na Young Han (NCC), Joon Young Shin (Asan Medical Center), and Sook Hee Kang (Yonsei University College of Medicine) for their assistance in data collection, retrieval, management, annotation, and quality checks. We used datasets from The Open AI Dataset Project (AI-Hub, South Korea). All dataset information can be accessed through AI-Hub (www.aihub.or.kr, accessed on June 19, 2023).

FUNDING INFORMATION

This work was partially supported by a National Research Foundation of Korea (NRF, IRIS) grant funded by the Korean Government (MSIT) (2021R1A2C2013630, RS-2021-NR059550). None of the sponsors had any role in study design, data collection, data analysis or manuscript preparation.

CONFLICT OF INTEREST STATEMENT

All authors have no competing interests to disclose.

DATA AVAILABILITY

Data is available upon request from the corresponding author.

HUMAN ETHICS APPROVAL DECLARATION

This study was performed in accordance with the Declaration of Helsinki. This study was approved by the Institutional Review Boards of the Catholic University of Korea College of Medicine (UC21SNSI0064), Yonsei University College of Medicine (4-2021-0569), and the National Cancer Center (NCC2021-0145).

REFERENCES

1. Bray, F., et al., *Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA Cancer J Clin, 2024. **74**(3): p. 229-263.

2. Pairman, L., et al., *Evaluation of pleural fluid cytology for the diagnosis of malignant pleural effusion: a retrospective cohort study.* Intern Med J, 2022. **52**(7): p. 1154-1159.

3. Gayen, S., *Malignant Pleural Effusion: Presentation, Diagnosis, and Management.* Am J Med, 2022. **135**(10): p. 1188-1192.

4. Metzgeroth, G., et al., *Diagnostic accuracy of cytology and immunocytology in carcinomatous effusions.* Cytopathology, 2008. **19**(4): p. 205-11.

5. Loveland, P., et al., *Diagnostic yield of pleural fluid cytology in malignant effusions: an Australian tertiary centre experience.* Intern Med J, 2018. **48**(11): p. 1318-1324.

6. Ailia, M.J., et al., *Current Trend of Artificial Intelligence Patents in Digital Pathology: A Systematic Evaluation of the Patent Landscape.* Cancers (Basel), 2022. **14**(10).

7. Barwad, A., P. Dey, and S. Susheilia, *Artificial neural network in diagnosis of metastatic carcinoma in effusion cytology.* Cytometry B Clin Cytom, 2012. **82**(2): p. 107-11.

8. Tosun, A.B., et al., *Detection of malignant mesothelioma using nuclear structure of mesothelial cells in effusion cytology specimens.* Cytometry A, 2015. **87**(4): p. 326-33.

9. Dimitriou, N., O. Arandjelović, and P.D. Caie, *Deep Learning for Whole Slide Image Analysis: An Overview.* Front Med (Lausanne), 2019. **6**: p. 264.

10. Lu, M.Y., et al., *Data-efficient and weakly supervised computational pathology on whole-slide images.* Nat Biomed Eng, 2021. **5**(6): p. 555-570.

11. Ren, W., et al., *Deep Learning-Based Classification and Targeted Gene Alteration Prediction from Pleural Effusion Cell Block Whole-Slide Images.* Cancers (Basel), 2023. **15**(3).

12. Vaickus, L.J., et al., *Automating the Paris System for urine cytopathology-A hybrid deep-learning and morphometric approach.* Cancer Cytopathol, 2019. **127**(2): p. 98-115.

13. Elliott Range, D.D., et al., *Application of a machine learning algorithm to predict malignancy in thyroid cytopathology.* Cancer Cytopathol, 2020. **128**(4): p. 287-295.

14. Thakur, N., H. Yoon, and Y. Chong, *Current Trends of Artificial Intelligence for Colorectal Cancer Pathology Image Analysis: A Systematic Review.* Cancers (Basel), 2020. **12**(7).

15. Kim, H.K., et al., *Artificial-Intelligence-Assisted Detection of Metastatic Colorectal Cancer Cells in Ascitic Fluid.* Cancers (Basel), 2024. **16**(5).

16. Kim, T., et al., *Deep learning-based diagnosis of lung cancer using a nationwide respiratory cytology image set: improving accuracy and inter-observer variability.* Am J Cancer Res, 2023. **13**(11): p. 5493-5503.

17. Lee, Y., et al., *Improved Diagnostic Accuracy of Thyroid Fine-Needle Aspiration Cytology with Artificial Intelligence Technology.* Thyroid, 2024. **34**(6): p. 723-734.

18. Thakur, N., et al., *Recent Application of Artificial Intelligence in Non-Gynecological Cancer Cytopathology: A Systematic Review.* Cancers (Basel), 2022. **14**(14).

19. Park, H.S., et al., *Deep Learning-Based Computational Cytopathologic Diagnosis of Metastatic Breast Carcinoma in Pleural Fluid.* Cells, 2023. **12**(14).

20. Chowdhuri, S.R., et al., *Adenocarcinoma cells in effusion cytology as a diagnostic pitfall with potential impact on clinical management: a case report with brief review of immunomarkers.* Diagn Cytopathol, 2014. **42**(3): p. 253-8.

21. Shidham, V.B., *The panorama of different faces of mesothelial cells.* Cytojournal, 2021. **18**: p. 31.

22. Bobowicz, M., et al., *Attention-Based Deep Learning System for Classification of Breast Lesions-Multimodal, Weakly Supervised Approach.* Cancers (Basel), 2023. **15**(10).

23. Geijs, D.J., et al., *Detection and subtyping of basal cell carcinoma in whole-slide histopathology using weakly-supervised learning.* Med Image Anal, 2024. **93**: p. 103063.

24. Kim, M., et al., *Diagnosis of Alzheimer Disease and Tauopathies on Whole-Slide Histopathology Images Using a Weakly Supervised Deep Learning Algorithm.* Lab Invest, 2023. **103**(6): p. 100127.

25. King, V., et al., *Use of Artificial Intelligence in the Prediction of Chiari Malformation Type 1 Recurrence After Posterior Fossa Decompressive Surgery.* Cureus, 2024. **16**(5): p. e60879.

26. Alam, M.R., et al., *Recent application of artificial intelligence on histopathologic image-based prediction of gene mutation in solid cancers.* Brief Bioinform, 2023. **24**(3).

27. Alam, M.R., et al., *Recent Applications of Artificial Intelligence from Histopathologic Image-Based Prediction of Microsatellite Instability in Solid Cancers: A Systematic Review.* Cancers (Basel), 2022. **14**(11).

FIGURE LEGENDS

Figure 1.

Overview of the study: This study consists of three steps: **(A)** data collection, **(B)** patch-level classification, and **(C)** WSI-level classification. CLAM, clustering-constrained attention multiple-instance learning; H&E, haematoxylin and eosin; Pap, Papanicolaou; WSI, whole slide image.

Figure 2.

Performance comparison between pathologists and AI. AI, artificial intelligence; NPV, negative predictive value; PPV, positive predictive value.

Figure 3.

Patch images that were diagnosed differently by pathologists and the artificial intelligence (AI) model. The model's misdiagnosis of 18 malignant cases as false negatives, despite pathologists' accuracy, may be owing to the pleural fluid's effect on cell appearance and the similarity of activated mesothelial cells and macrophages to malignant cells [19-21].

Tables

Table 1. Characteristics of data obtained from the Open AI Dataset (left) and the archive of the Quality Assurance Program of the Korean Society of Cytopathology (right). *

	Open AI Dataset (n = 885)		**Quality Assurance Program (n = 93)**	
	Benign (*n* = 576)	**Malignant (*n* = 309)**	**Benign (*n* = 32)**	**Malignant (*n* = 61)**

Age – yr	68.6 ± 14.8	68.4 ± 13.7	80.4 ± 13.1	69.6 ± 13.4
Sex – no. (%)				
Men	355 (61.6)	160 (51.7)	17 (65.6)	40 (65.5)
Women	221 (38.4)	148 (47.8)	15 (34.3)	21 (34.4)
Not reported	0 (0)	1 (0.3)	0 (0)	0 (0)
Number of institutions	7	64	2	40
Subtype – no. (%)				
NSCLC		298 (96.4)		57 (93.4)
Adenocarcinoma		167		38
Squamous cell carcinoma		7		0
NSCLC, subtype not specified (Including not reported)		124 (40.1)		19 (31.1)
SCLC		11 (3.5)		4 (6.5)

* Plus-minus values are presented as means ± standard deviation (SD). The percentage may not total 100 owing to rounding. NSCLC, non-small cell lung cancer; SCLC, small-cell lung cancer.

Table 2. Performance of deep convolutional neural network models. *

Model	Performance metric	Training	Validation	Test
Inception-ResNet-v2	Accuracy	0.988 (0.987 – 0.989)	0.975 (0.970 – 0.980)	0.991 (0.988 – 0.994)
	Sensitivity	0.987 (0.986 – 0.988)	0.974 (0.968 – 0.980)	0.996 (0.994 – 0.998)
	Specificity	0.990 (0.989 – 0.991)	0.976 (0.971 – 0.981)	0.987 (0.983 – 0.991)
Efficientnet-b0	Accuracy	0.988 (0.987 – 0.989)	0.948 (0.940 – 0.956)	0.983 (0.978 – 0.988)
	Sensitivity	0.986 (0.985 – 0.987)	0.945 (0.937 – 0.953)	0.994 (0.991 – 0.997)
	Specificity	0.990 (0.989 – 0.991)	0.951 (0.943 – 0.959)	0.973 (0.967 – 0.979)
ResNet50	Accuracy	0.985 (0.983 – 0.987)	0.968 (0.962 – 0.974)	0.992 (0.989 – 0.995)
	Sensitivity	0.984 (0.982 – 0.986)	0.970 (0.964 – 0.976)	0.996 (0.994 – 0.998)
	Specificity	0.987 (0.986 – 0.988)	0.967 (0.961 – 0.973)	0.987 (0.983 – 0.991)
Mobilenet-v2	Accuracy	0.985 (0.983 – 0.987)	0.959 (0.952 – 0.966)	0.983 (0.978 – 0.988)

	Sensitivity	0.984 (0.982 – 0.986)	0.965 (0.959 – 0.971)	0.994 (0.991 – 0.997)
	Specificity	0.986 (0.985 – 0.987)	0.954 (0.947 – 0.961)	0.973 (0.967 – 0.979)
Densenet-121	**Accuracy**	0.991 (0.990 – 0.992)	0.978 (0.973 – 0.983)	0.991 (0.988 – 0.994)
	Sensitivity	0.990 (0.989 – 0.991)	0.986 (0.982 – 0.990)	0.994 (0.991 – 0.997)
	Specificity	0.992 (0.991 – 0.993)	0.970 (0.964 – 0.976)	0.987 (0.983 – 0.991)
ResNext50	**Accuracy**	0.990 (0.989 – 0.991)	0.964 (0.957 – 0.971)	0.987 (0.983 – 0.991)
	Sensitivity	0.989 (0.988 – 0.990)	0.994 (0.991 – 0.997)	0.999 (0.998 – 1.000)
	Specificity	0.992 (0.991 – 0.993)	0.935 (0.926 – 0.944)	0.976 (0.971 – 0.981)

* Values are presented as means (95% confidence intervals).

Table 3. Performance of CLAM model. *

		Validation	**Test**
CLAM-SB	Accuracy	0.905 (0.842 – 0.968)	0.810 (0.726 – 0.894)
	Sensitivity	0.815 (0.668 – 0.961)	0.778 (0.621 – 0.935)
	Specificity	0.947 (0.889 – 1.000)	0.825 (0.726 – 0.923)

	AUC	0.968 (0.921 – 1.000)	0.744 (0.624 – 0.864)
CLAM-MB	Accuracy	0.929 (0.874 – 0.984)	0.905 (0.842 – 0.968)
	Sensitivity	0.852 (0.718 – 0.986)	0.815 (0.668 – 0.961)
	Specificity	0.965 (0.917 – 1.000)	0.947 (0.889 – 1.000)
	AUC	0.984 (0.951 – 1.000)	0.942 (0.879 – 1.000)

* Values are presented as means (95% confidence intervals). AUC, area under the curve; CLAM, clustering-constrained attention multiple-instance learning; MB, multi-branch; SB, single-branch.

SUPPLEMENTARY FIGURES LEGEND

Supplementary Figure S1.

The NIA project for the cytopathology dataset.

Supplementary Figure S2.

The NIA project for the cytopathology dataset preparation process.

Supplementary Figure S3.

Scheme of (A) Inception-ResNet-v2 and (B) ResNet50 network used for binary classification.

Supplementary Figure S4.

Examples of patch images that AI diagnosed correctly or wrongly. AI denotes artificial intelligence.

Supplementary Figures S5.

Cohen's and Fleiss' kappa scores of pathologists before (left) and after (right) the assistance of AI. AI denotes artificial intelligence; Patho., pathologist.

Supplementary Figures S6.

Heatmap of a whole slide image generated by CLAM. CLAM denotes clustering-constrained attention multiple-instance learning.

SUPPLEMENTARY TABLES

Supplementary Table 1. Sources of whole slide images extracted the from OpenAI Dataset.

Name of Institutions	Number of Slides
Benign	
Bongseng memorial hospital	1
Bumin hospital	1
Catholic University of Korea, College of Medicine	362
Cheong Ju Medical Center	1
Metro hospital	1
National Cancer Center	60
Yonsei University, College of Medicine	150
Malignant	
Ajou University Hospital	1
Andong Medical Group Hospital	3
Busan Medical Center	2
Busan St. Mary's Hospital	1

Catholic University of Korea, College of Medicine	154
CHA Bundang Medical Center	2
Chinju Jeil Hospital	2
Chung-Ang University Hospital	1
Chungbuk National University Hospital	5
Daegu Catholic University Medical Center	1
Daegu Fatima Hospital	1
Daejeon Eulji Medical Center, Eulji University	1
Daerim St. Mary's Hospital	2
Dankook University Hospital	1
Dong-A University Hospital	4
Dongguk University Gyeonju Hospital	1
Dongnam Institute of Radiological & Medical Sciences	1
Good Gangan Hospital	4
GS Medical Center	1
Gunsan Medical Center	1

H Plus Yangji Hospital	1
Hallym University Dongtan Sacred Heart Hospital	1
Hallym University Kangnam Sacred Heart Hospital	1
Hanil General Hospital	1
Hanyang University Guri Hospital	3
Hanyang University Seoul Hospital	1
Hongik Hospital	2
Incheon Sarang Hospital	1
Inha University Hospital	1
Inje University Busan Paik Hospital	1
Inje University Haeundae Paik Hospital	1
Inje University Sanggye Paik Hospital	2
Inje University Seoul Paik Hospital	2
Jesus Hospital	1
Kangdong Sacred Heart Hospital	1
Korea University Anam Hospital	1

Korea University Ansan Hospital	3
Kyung Hee University Hospital at Gangdong	1
Kyung Hee University Medical Center	1
Kyungpook National University Chilgok Hospital	3
Kyungpook National University Hospital	2
Myongji Hospital	1
National Health Insurance Service Ilsan Hospital	3
National Medical Center	1
Sahmyook Seoul Medical Center	2
Samsung Medical Center	2
SNUH SMG-SNU Boramae Medical Center	2
SoonChunHyang University Bucheon Hospital	4
SoonChunHyang University Hospital Gumi	1
St. Carollo Hospital	3
Sun Medical Center	2
The Catholic University of Korea St. Vincent's Hospital	1

The Catholic University of Korea, Eunpyeong ST. Mary's Hospital	1
The Catholic University of Korea, Incheon ST. Mary's Hospital	2
The Catholic University of Korea, Seoul ST. Mary's Hospital	2
The Catholic University of Korea, Seoul ST. Paul's Hospital	2
The Catholic University of Korea, Yeouido ST. Mary's Hospital	1
Ulsan University Hospital	2
Wonju Severance Christian Hospital	2
Wonkwang University Hospital	3
Yeungnam University Hospital	2
Yonsei University College of Medicine	50
Yonsei University Severance Hospital	2
Not reported	1

Supplementary Table 2. Sources of whole slide images used for external validation of clustering-constrained attention multiple-instance learning (CLAM).

Name of institutions	Number of slides
Benign	
Changwon Hanmaeum Hospital	2
The Catholic University of Korea, Uijeongbu ST. Mary's Hospital	30
Malignant	
Ajou University Hospital	1
Andong Medical Group Hospital	1
Best Han Seo Hospital	2
Busan St. Mary's Hospital	2
Cheju Halla General Hospital	1
Cheongju St. Mary's Hospital	1
Chonan Chungmu Hospital	1
Chungbuk National University Hospital	3
Daejeon Eulji Medical Center, Eulji University	1
Digital Pathology Histology Center	2

Dongkang Hospital	3
H Plus Yangji Hospital	2
Hallym University Kangnam Sacred Heart Hospital	2
Hanil General Hospital	1
Hanyang University Seoul Hospital	1
Hongik Hospital	3
Inje University Busan Paik Hospital	1
Inseong Medical Foundation Hallym Hospital	1
Jeonbuk National University Hospital	1
Kangbuk samsung hospital	2
Kangdong Sacred Heart Hospital	2
Konyang University Hospital	2
Korea Institute of Radiological & Medical Sciences, Korea Cancer Center Hospital	1
Korea University Anam Hospital	2
Kyung Hee University Hospital at Gangdong	1
Kyungpook National University Hospital	2

National Cancer Center	1
Nowon Eulji Medical Center, Eulji University	1
Prebyterian Medical Center	1
Pusan National University Hospital	2
Samkwang Medical Laboratories	1
Samsung Medical Center	2
Samyook Medical Center	1
Seoul Clinical Laboratories	2
Seoul National University Bundang Hospital	1
Seoul National University Hospital	1
Soonchunhyang University Hospital	2
SQ Lab	1
The Catholic University of Korea, Eunpyeong ST. Mary's Hospital	1
The Catholic University of Korea, Incheon ST. Mary's Hospital	1
The Catholic University of Korea, ST. Vincent's Hospital	1
The Catholic University of Korea, Uijeongbu ST. Mary's Hospital	2

Supplementary Table 3. Dataset details used in this study. (A) Number of whole slide images (WSIs) and image patches used for patch-level classification, and (B) number of WSIs used for WSI-level classification. *

(A)		Number of whole slide images (patch images) for patch-level classification			
		Training	Validation	Testing	Total
Pap stained	Benign	329 (12,348)	45 (1,582)	41 (1,492)	415 (15,422)
	Malignant	134 (12,301)	21 (1,549)	13 (1,529)	168 (15,379)
	Total	463 (24,649)	66 (3,131)	54 (3,021)	583 (30,801)

(B)		Number of whole slide images for whole-slide image-level classification			
		Training	Validation	Testing	Total
Pap stained	Benign	332	42	42	416
	Malignant	185	23	23	231
	Total	517	65	65	647
H&E stained	Benign	120	15	15	150
	Malignant	32	4	4	40
	Total	152	19	19	190

* H&E, haematoxylin and eosin; Pap, Papanicolaou.

Supplementary Table 4. Confusion matrix of Inception-ResNet-v2 in patch-level classification. *

		AI Prediction	
	Total (n = 3,021)	Malignant (1,543)	Benign (1,478)
Ground Truth	Malignant (1,529)	1,523 (99.6%)	6 (0.4%)
	Benign (1,492)	20 (1.3%)	1,472 (98.7%)

* AI, artificial intelligence.

Supplementary Table 5. Confusion matrix of CLAM in whole slide image-level

classification. *

		CLAM-SB Prediction		CLAM-MB Prediction	
	Total (*n* = 84)	Malignant (31)	Benign (53)	Malignant (25)	Benign (59)
Ground Truth	Malignant (27)	21 (77.8%)	6 (22.2%)	22 (81.5%)	5 (18.5%)
	Benign (57)	10 (17.5%)	47 (82.5%)	3 (5.3%)	54 (94.7%)

* CLAM, clustering-constrained attention multiple-instance learning; MB, multi-branch; SB, single-branch.

Supplementary Table 6. External validation of CLAM models. *

	CLAM-SB	CLAM-MB
Accuracy	0.853 (0.781 - 0.924)	0.905 (0.846 - 0.964)
Sensitivity	0.873 (0.792 - 0.955)	0.905 (0.834 - 0.976)
Specificity	0.813 (0.677 – 0.948)	0.906 (0.805 – 1.00)
AUC	0.930 (0.880 - 0.980)	0.947 (0.904 - 0.989)

* Values are presented as means (95% confidence intervals). AUC, area under the curve; CLAM, clustering-constrained attention multiple-instance learning; MB, multi-branch; SB, single-branch.

ICMJE DISCLOSURE FORM

Date:	12/16/2024
Your Name:	Kyung Jin Seo
Manuscript Title:	Whole slide image-level classification of malignant effusion cytology using clustering-constrained attention multiple instance learning
Manuscript Number (if known):	Click or tap here to enter text.

In the interest of transparency, we ask you to disclose all relationships/activities/interests listed below that are related to the content of your manuscript. "Related" means any relation with for-profit or not-for-profit third parties whose interests may be affected by the content of the manuscript. Disclosure represents a commitment to transparency and does not necessarily indicate a bias. If you are in doubt about whether to list a relationship/activity/interest, it is preferable that you do so.

The author's relationships/activities/interests should be defined broadly. For example, if your manuscript pertains to the epidemiology of hypertension, you should declare all relationships with manufacturers of antihypertensive medication, even if that medication is not mentioned in the manuscript.

In item #1 below, report all support for the work reported in this manuscript without time limit. For all other items, the time frame for disclosure is the past 36 months.

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	**Specifications/Comments (e.g., if payments were made to you or to your institution)**
		Time frame: Since the initial planning of the work	
1	All support for the present manuscript (e.g., funding, provision	☐ **None** This work was partially supported by a National Research Foundation of Korea (NRF,IRIS) grant funded by the Korean Government (MSIT)	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
	of study materials, medical writing, article processing charges, etc.) **No time limit for this item.**	(2021R1A2C2013630, RS-2021-NR059550)	
			Click the tab key to add additional rows.
Time frame: past 36 months			
2	Grants or contracts from any entity (if not indicated in item #1 above).	☒ **None**	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
3	Royalties or licenses	☒ None	
4	Consulting fees	☒ None	
5	Payment or honoraria for	☒ None	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
	lectures, presentations, speakers bureaus, manuscript writing or educational events		
6	Payment for expert testimony	☒ None	
7	Support for attending meetings and/or	☒ None	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
	travel		
8	Patents planned, issued or pending	☒ None	
9	Participation on a Data Safety Monitoring Board or Advisory Board	☒ None	
10	Leadership or	☒ None	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
	fiduciary role in other board, society, committee or advocacy group, paid or unpaid		
11	Stock or stock options	☒ None	
12	Receipt of equipment, materials,	☒ None	

		Name all entities with whom you have this relationship or indicate none (add rows as needed)	Specifications/Comments (e.g., if payments were made to you or to your institution)
	drugs, medical writing, gifts or other services		
13	Other financial or non-financial interests	☒ **None**	

Please place an "X" next to the following statement to indicate your agreement:

☒ I certify that I have answered every question and have not altered the wording of any of the questions on this form.

[15 April 2025]

Prof. Dr. Solange Peters, MD, PhD

Editor-in-Chief – *Lung Cancer*

Dear Editor,

Declaration of Interest Statement. We wish to submit a research article for publication in *Lung Cancer*, titled "**Whole slide image-level classification of malignant effusion cytology using clustering-constrained attention multiple instance learning**". We have read and understood your journal's policies, and we believe that neither the manuscript nor the study violates any of these.

The authors declare that they have no conflict interests.

Sincerely,

Yosep Chong, MD., Ph.D.

Department of Hospital Pathology, Uijeongbu St. Mary's Hospital, College of Medicine, The Catholic University of Korea, #271 Cheonbo-ro, Uijeongbu 11765, Gyeonggi-do, Republic of Korea. Tel: +82-031-820-3160, Fax: +82-031-820-3877

e-mail: ychong@catholic.ac.kr

Dr. Kyung Jin Seo, MD., Ph.D .

Department of Hospital Pathology, Uijeongbu St. Mary's Hospital, College of Medicine, The Catholic University of Korea, 271 Cheonboro, Uijeongbu 11765, Republic of Korea, Tel: +82-31-820-3158, E-mail: ywacko@catholic.ac.kr

Highlights

- Cytological diagnosis of pleural fluid has primarily been explored at the image-patch level using artificial intelligence.
- This is the first study to classify lung cancer from pleural effusion at the whole-slide image level.

- The study demonstrated outstanding performance in identifying malignant effusions using a large, quality-controlled dataset.
- It achieved 97% accuracy and an AUC of 0.97, representing a 13% improvement over patch-based methods.