scientific reports

Check for updates

OPEN

# Diagnostic performance of real-time artificial intelligence using deep learning analysis of endoscopic ultrasound videos for gallbladder polypoid lesions

Young Hoon Choi[1,7], Jun Young Park[2,7], See Young Lee[3], Jae Hee Cho[3], Young Jae Kim[4], Kwang Gi Kim[5,6,8]✉ & Sung Ill Jang[3,8]✉

**Endoscopic ultrasound (EUS) is accurate for diagnosing gallbladder (GB) polyps but is limited by subjective interpretation and operator expertise. Although artificial intelligence (AI) has been applied to still EUS images of GB polyps, its application to EUS videos, which provide richer diagnostic data, remains unexplored. This study evaluated the diagnostic performance of AI models in analyzing EUS videos for GB polyp assessment. EUS videos of patients with histologically confirmed GB polyps were divided into training and validation cohorts. Segmentation models (Attention U-Net, Residual U-Net, and deep understanding convolutional kernel [DUCK] net) identified polyp regions, followed by classification into neoplastic and non-neoplastic polyps using classification models (EfficientNet-B2, ResNet101, and vision transformer). The training cohort included 17 (11 patients) and 79 (39 patients) videos with neoplastic and non-neoplastic polyps, respectively, and the validation cohort included 11 (6 patients) and 25 (11 patients) videos, respectively. Attention U-Net (0.998) and DUCK Net (0.995) achieved the highest training cohort segmentation accuracy. EfficientNet-B2 showed the highest classification performance (accuracy 0.957, recall 0.954, F1-score 0.939, AUC 0.991) and maintained strong performance on the validation dataset (accuracy 0.879, recall 0.968, F1-score 0.917, AUC 0.861). AI demonstrated high accuracy in EUS video-based GB polyp analysis, warranting further prospective validation.**

**Keywords** Gallbladder polyp, Endoscopic ultrasound video, Artificial intelligence, Diagnostic performance

Gallbladder (GB) polyps, lesions protruding from the gallbladder wall into the lumen, are common and affect approximately 5% of the general population[1]. They can be categorized into non-neoplastic and neoplastic polyps, such as adenomas and adenocarcinomas, which is crucial in determining the need for surgical treatment[2,3]. As biopsy of GB polyps is not feasible, imaging modalities, primarily abdominal ultrasonography, are used to differentiate neoplastic polyps, with sensitivity and specificity around 68% and 79%, respectively[4].

Compared with abdominal sonography, endoscopic ultrasonography (EUS) is considered superior in differentiating neoplastic polyps, producing high-resolution images resulting from the proximity of GB visualization and utilization of high ultrasound frequencies[5,6]. Several EUS features have been suggested for distinguishing neoplastic GB polyps, such as sessile shape and hypoechoic foci, and some studies have proposed scoring systems based on these features[7–9]. However, the interpretation of these EUS features tends to be

[1]Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea. [2]Department of Translational-Clinical Medicine, Gachon University, Incheon, Republic of Korea. [3]Department of Internal Medicine, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea. [4]Gachon Biomedical & Convergence Institute, Gachon University Gil Medical Center, Incheon, Republic of Korea. [5]Medical Devices R&D Center, Gachon University Gil Medical Center, Incheon, Republic of Korea. [6]Department of Biomedical Engineering, Pre-medical Course, College of Medicine, Gachon University, Incheon, Republic of Korea. [7]Young Hoon Choi and Jun Young Park contributed equally to this study and are co-first authors. [8]Kwang Gi Kim and Sung Ill Jang contributed equally to this study and are co-corresponding authors. ✉email: kimkg@gachon.ac.kr; aerojsi@yuhs.ac

somewhat subjective, and a significant drawback is that diagnostic accuracy is influenced by the proficiency level of the EUS endoscopist. To address these limitations, our previous study utilized deep learning analysis on EUS images of GB polyps, achieving diagnostic accuracy levels comparable to those of expert endoscopists[10].

In that study, only one or a few still images per GB polyp were provided for artificial intelligence (AI) analysis, which was also the case in other studies using abdominal ultrasound data for AI analysis of GB polyps[10–12]. Still images per polyp provide substantially less diagnostic information than the dynamic evaluation performed in clinical practice, in which EUS endoscopists diagnose GB polyps by observing real-time ultrasonography videos. In this regard, we sought to determine whether employing EUS video data for AI analysis of GB polyps would provide improved diagnostic accuracy. To our knowledge, no studies have utilized EUS video data for AI analysis of GB polyps or employed abdominal ultrasound video data. Therefore, this study aimed to evaluate the accuracy of detecting GB polyps and differentiating neoplastic GB polyps using deep learning analysis of EUS videos in patients with GB polyps.

## Materials and methods

### Data collection

Preoperative EUS videos of patients with histologically confirmed GB polyps were collected from Yonsei University College of Medicine, Gangnam Severance Hospital, between April 2020 and December 2023. All EUS examinations were performed by two experienced endoscopists using echoendoscopes (GF-UCT260 or GF-UE260-AL5; Olympus, EG3870UTK; Pentax, Tokyo, Japan). Multiple EUS videos could be generated per patient due to the recording time limit of the EUS system and procedural factors. All frames from the EUS videos were used for the analysis. The videos had a width of 800, height of 600, and frame rate of 30 frames per second. Cases in which polyps appeared relatively blurred or were difficult to detect with the naked eye were excluded when constructing the training and testing datasets. In total, 15 patients in the training cohort and six patients in the validation cohort were excluded due to unclear lesion visibility. Supplementary Figure S1 shows examples of the collected EUS videos and polyp-labeling data used to train the deep learning model. The polyp-labeling data were obtained using a custom labeling tool, the Korean-Medical Imaging System. The study protocol was approved by the Institutional Review Board of Gangnam Severance Hospital (IRB No. 3-2020-0089). The Institutional Review Board of Gangnam Severance Hospital also waived the requirement for informed consent due to the retrospective nature of the study. The study was conducted in accordance with the principles of the Declaration of Helsinki.

### Research environment

The analysis in this study was conducted on a system equipped with an NVIDIA Tesla P100-SXM2 GPU (NVIDIA, Santa Clara, CA, USA), an Intel® Xeon® CPU E5-2698 (Intel, Santa Clara, CA, USA), and 32GB of RAM, running on the Ubuntu 20.04.6 LTS operating system. The libraries used in the experiment included TensorFlow (version 2.10.0), an open-source library that supports various features for designing and training deep learning models; Compute Unified Device Architecture (CUDA, version 11.8.89), developed by NVIDIA for the parallel processing of large-scale computations on GPUs; OpenCV (version 4.6.0.66), a library providing diverse image-processing capabilities; and Matplotlib (version 3.5.2), which enables the visualization of analyzed data in various forms, such as graphs and charts.

### Data preprocessing

The ground truth for segmentation was manually annotated by four EUS experts who had performed more than 500 EUS procedures in pancreatobiliary imaging, using Aview software (Coreline Soft, Seoul, Republic of Korea). Each annotator was assigned a distinct subset of videos for labeling, focusing on delineating the visible margins of GB polyps while excluding surrounding mucosal reflections, artifacts, and acoustic shadows. To improve consistency, representative cases were jointly reviewed and discussed among annotators.

A video preprocessing algorithm was applied to adapt the collected EUS videos to train the deep learning model. The data preprocessing procedure for segmenting the polyp regions was as follows. First, all frames from the EUS videos were extracted. Histogram equalization was performed to enhance the contrast of the EUS frames, making the polyp regions more distinct. This technique improves contrast by equalizing the brightness values of the image, effectively restoring lost contrast in the process[13]. After histogram equalization, cropping was performed by specifying fixed coordinates to exclude unnecessary areas of the video, retaining only the endoscopic video region. By removing unnecessary areas, the model was guided to ignore irrelevant video information and focus on learning significant details. To prevent the videos from becoming elongated or distorted, zero padding was applied to the top and bottom of the frames, which were filled with pixel values of 0. The frames were then resized to $512 \times 512$.

For segmentation, frames were resized to $512 \times 512$; for classification, inputs were $256 \times 256$. We used the same sizes at inference.

The data preprocessing procedure for classifying the segmented polyp regions as neoplastic or non-neoplastic was as follows. The segmented regions were cropped and resized to $256 \times 256$. To more accurately evaluate the generalized performance of the deep learning model and prevent overfitting, fivefold cross-validation was performed[14]. The cross-validation was conducted on a video-wise basis, ensuring that frames from the same video were not mixed across the training, validation, and test sets within any fold. The videos of non-neoplastic and neoplastic polyps were appropriately distributed across each fold, ensuring balance. All collected data were split at the video level into training/validation/testing to avoid cross-video leakage; no frames from a given video appeared in more than one split.

## Convolutional neural network model for deep learning

Attention U-Net, Residual U-Net, and Deep Understanding Convolutional Kernel (DUCK) Net architectures were used for polyp segmentation. Attention U-Net enhances the standard U-Net structure by incorporating attention gates, which emphasize the features necessary for segmentation while suppressing irrelevant information[15]. Residual U-Net combines the advantages of the U-Net structure with residual units, facilitating easier model training.

The skip connections transmit the information and features necessary for segmentation without degrading model performance, enabling the design of a neural network with significantly fewer parameters[16]. DUCK Net employs six convolutional blocks in parallel, allowing the model to train on the block deemed most suitable. It is designed with kernel sizes configured in three different ways to identify common regions while also capturing regions at the edges[17]. Using the segmentation model, the polyp area was predicted, cropped, and classified into neoplastic and non-neoplastic polyps using a classification model. EfficientNet-B2, ResNet101, and Vision Transformer (ViT) were employed for polyp classification in this study. EfficientNet-B2 enhances model performance efficiently by systematically scaling the width, depth, and resolution of convolutional neural network operations[18]. ResNet101 is a model designed with residual blocks to address the gradient vanishing problem that occurs as network depth increases. This design achieves a lower error rate and maximizes classification performance[19].

The ViT model incorporated Shifted Patch Tokenization (SPT) and Locality Self-Attention (LSA) to enhance locality inductive bias[20]. SPT spatially shifts the input image in multiple directions and connects it with the original input image. LSA primarily sharpens the attention score distribution by learning the temperature parameters of the softmax function. The training hyperparameters for all deep learning models were as follows: 300 epochs, a batch size of eight, a learning rate of 0.0001 with the Adam optimizer, an input size of 256 × 256, and the use of pretrained ImageNet weights. To prevent overfitting during training, the EarlyStopping mechanism was implemented[21,22]. The ReduceLROnPlateau function was added to dynamically adjust the learning rate based on continuous monitoring of the validation loss[23,24]. During the training of the classification model, EfficientNet-B2 and ResNet101 were enhanced with a global average pooling layer added to the final layer to visualize which areas of the image the model focused on when making predictions[25,26].

Unlike the other two classification models, the ViT model belongs to the Transformer family and, unlike convolutional neural network models, directly calculates gradients using self-attention[27]. Fig. 1 illustrates the workflow for polyp segmentation and classification in this study. The preprocessed data were segmented into polyp regions using Attention U-Net, Residual U-Net, and DUCK Net. The segmented regions were then cropped and classified into neoplastic and non-neoplastic polyps.

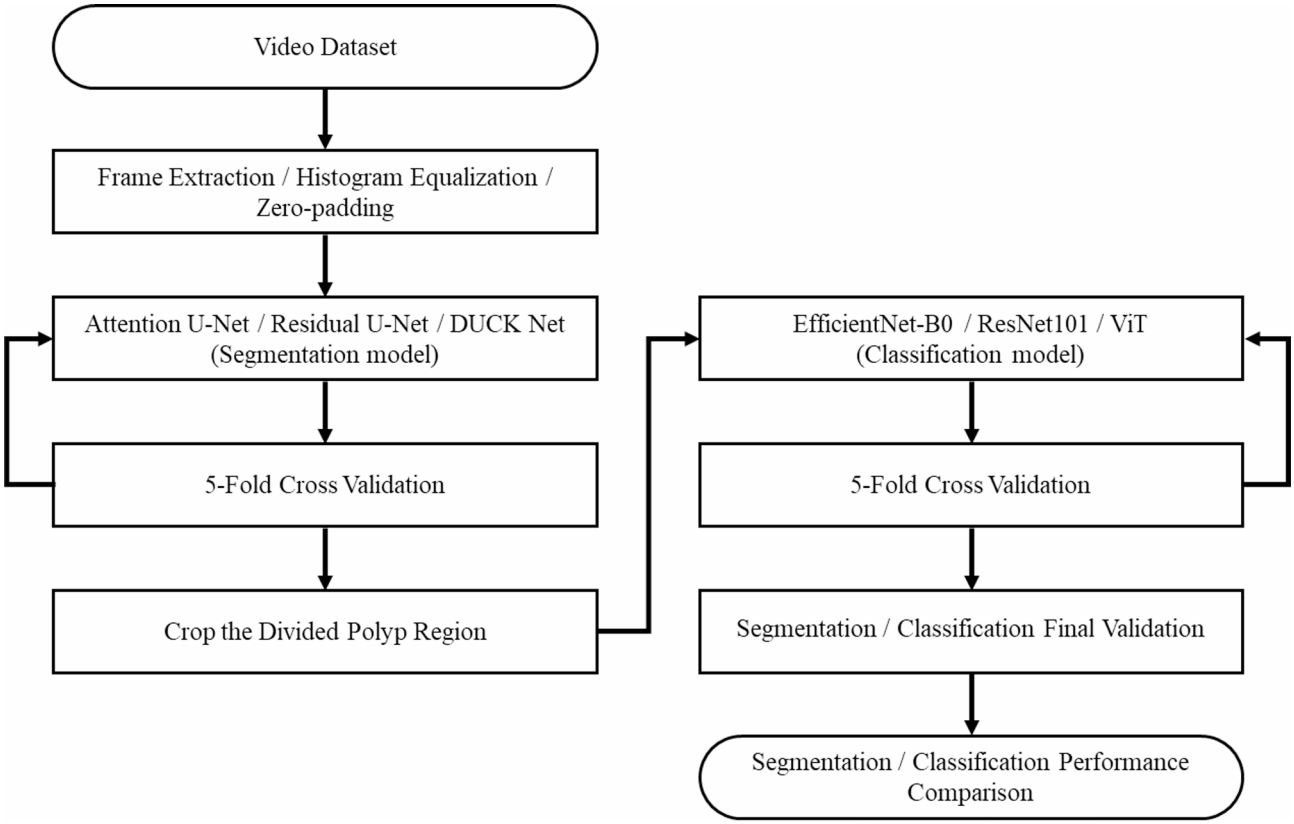


**Fig. 1**. Flowchart for segmenting and classifying polyp regions using a segmentation and classification model.

The final prediction for the video was determined by counting the number of non-neoplastic and neoplastic frames in the still images and selecting the class with the highest count as the final predicted class. The confidence level of the predicted class was assessed by examining the average probability of the classified predictions.

### Statistical analysis

The performance of polyp segmentation and classification was evaluated by comparing the deep learning model predictions with the visual analysis results from medical professionals and histologically confirmed outcomes. True positive (TP), false negative (FN), true negative (TN), and false positive (FP) values were obtained based on the comparison. Based on the obtained TP, FN, TN, and FP, the segmentation performance metrics included accuracy, precision, recall, and dice similarity coefficient (DSC), as well as intersection over union (IoU), which compares the actual polyp regions with the predicted polyp regions. The classification metrics included accuracy, precision, recall, F1-score, and area under the curve (AUC). The performance of the segmentation and classification was evaluated using these five metrics. For the ROC and AUC analyses, the decision thresholds were not arbitrarily assigned but were systematically optimized at the video level to achieve optimal discrimination performance. The 95% confidence intervals (CIs) were estimated using a bootstrap resampling approach at the video level. The final performance was obtained by conducting fivefold cross-validation and calculating the average across all folds. All statistical analyses were performed using IBM SPSS Statistics software (version 29.0; IBM Corporation, Armonk, New York, USA).

## Results

### Dataset composition and patient characteristics

Seventeen EUS videos of 11 patients with neoplastic GB polyps and 79 videos of 39 patients with non-neoplastic GB polyps were included in the training cohort. For the validation cohort, 11 EUS videos from six patients with neoplastic GB polyps and 25 videos from 11 patients with non-neoplastic GB polyps were analyzed. Across both cohorts, patients with neoplastic polyps had a higher mean age and larger mean polyp size than those in patients with non-neoplastic polyps (Table 1).

### Performance of GB polyp segmentation models on training and validation data

Table 2 shows the fivefold cross-validation results of Attention U-Net, Residual U-Net, and DUCK Net for polyp segmentation. Figure 2 shows the mean confusion matrix of the training cohort, calculated based on proportional values. Attention U-Net achieved an accuracy of 0.998, Residual U-Net 0.992, and DUCK Net 0.995. The corresponding DSC values were 0.894, 0.729, and 0.822, and IoU values were 0.818, 0.614, and 0.706, respectively.

Figure 3 shows the preprocessed EUS frames, ground truth obtained from specialists, and the segmentation prediction results of Attention U-Net, Residual U-Net, and DUCK Net. The green areas represent the actual polyp regions, whereas the red areas indicate the predicted results of the segmentation models. Attention U-Net produced relatively accurate segmentation results compared to the other two models. In contrast, Residual U-Net and DUCK Net occasionally segmented non-polyp regions but failed to correctly segment the actual polyp regions.

Table 3 presents the performance results of the segmentation models using the final validation dataset, excluding the training and testing datasets that were initially used. Figure 4 shows the mean confusion matrix

| Variables | All | Neoplastic polyp | Non-neoplastic polyp |
|---|---|---|---|
| Training cohort | | | |
| Number of patients | 50 | 11 | 39 |
| Number of videos | 96 | 17 | 79 |
| Number of video frames | 2953 | 1059 | 1894 |
| Age, years, mean ± SD | 51.3 ± 15.5 | 63.5 ± 18.6 | 47.9 ± 12.8 |
| Sex | | | |
| Male | 22 (44%) | 7 (63.6%) | 15 (38.5%) |
| Female | 28 (56%) | 4 (36.4%) | 24 (61.5%) |
| Polyp size, mm, mean ± SD | 12.9 ± 10.1 | 21.3 ± 19.1 | 10.6 ± 3.1 |
| Validation cohort | | | |
| Number of patients | 17 | 6 | 11 |
| Number of videos | 36 | 11 | 25 |
| Number of video frames | 1375 | 899 | 476 |
| Age, years, mean ± SD | 50.8 ± 15.2 | 72.7 ± 7.5 | 44.8 ± 10.2 |
| Sex | | | |
| Male | 5 (35.7%) | 2 (66.7%) | 3 (27.3%) |
| Female | 9 (64.3%) | 1 (33.3%) | 8 (72.7%) |
| Polyp size, mm, mean ± SD | 14.0 ± 7.7 | 26.7 ± 2.9 | 10.6 ± 3.7 |

**Table 1**. Baseline characteristics of the training and validation cohort patients. SD, standard deviation.

| Model | Accuracy (95% CI) | Precision (95% CI) | Recall (95% CI) | DSC (95% CI) | IoU (95% CI) | PPV (95% CI) | NPV (95% CI) |
|---|---|---|---|---|---|---|---|
| Attention U-Net | 0.998 (0.997–0.999) | 0.934 (0.904–0.964) | 0.859 (0.766–0.952) | 0.894 (0.813–0.975) | 0.818 (0.697–0.939) | 0.938 (0.717–0.989) | 0.975 (0.913–0.993) |
| Residual U-Net | 0.992 (0.981–0.999) | 0.855 (0.687–0.999) | 0.653 (0.415–0.890) | 0.729 (0.523–0.934) | 0.614 (0.370–0.858) | 0.846 (0.578–0.957) | 0.928 (0.851–0.966) |
| DUCK Net | 0.995 (0.991–0.999) | 0.900 (0.834–0.967) | 0.773 (0.630–0.915) | 0.822 (0.738–0.905) | 0.706 (0.592–0.821) | 0.867 (0.621–0.963) | 0.951 (0.880–0.981) |

**Table 2**. Attention U-Net, residual U-Net and DUCK net fivefold cross-validation results for polyp segmentation. DUCK, deep understanding convolutional kernel; CI, confidence interval; DSC, dice similarity coefficient; IoU, intersection over union.

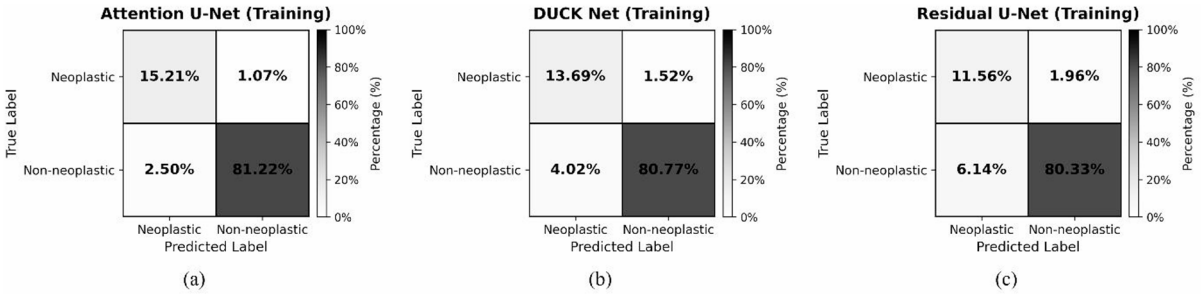


**Fig. 2**. Mean confusion matrix of fivefold cross-validation in the training cohort.

of the final validation cohort, calculated based on proportional values. Accuracy values were 0.941 for both Attention U-Net and Residual U-Net, and 0.943 for DUCK Net. DSC values were 0.612, 0.675, and 0.683 for Attention U-Net, Residual U-Net, and DUCK Net, respectively.

## Performance of GB polyp classification models on training and validation data

Table 4 presents the fivefold cross-validation results of EfficientNet-B2, ResNet101, and ViT for polyp classification. Figure 5 shows the mean confusion matrix of the training cohort, calculated based on proportional values. EfficientNet-B2 achieved an accuracy of 0.957, F1-score of 0.939, and AUC of 0.991. Accuracy for ResNet101 and ViT was 0.907 and 0.853, respectively, with corresponding F1-scores of 0.873 and 0.774. Among all the performance metrics, EfficientNet-B2 demonstrated the best performance.

Supplementary Figure S2 shows the results of the gradient-weighted class activation map (Grad-CAM) for the classification of neoplastic and non-neoplastic cases using EfficientNet-B2, ResNet101, and ViT. Grad-CAM is a visualization algorithm that highlights regions that contribute to the classification of a specific class. Regions with higher importance are displayed in red, whereas those with lower importance are displayed in blue. As shown in Supplementary Figure S2, (a) and (b) are images of histologically confirmed non-neoplastic polyps, whereas (c) and (d) are images of histologically confirmed neoplastic polyps. EfficientNet-B2 accurately highlighted the polyp regions in red and yellow, demonstrating the precise classification of neoplastic and non-neoplastic polyps compared to the other two models. In contrast, ResNet101 displayed red regions in empty spaces or black backgrounds rather than in the polyp regions. This indicates that ResNet101 failed to capture features in the EUS images, leading to a significantly lower performance in classifying neoplastic and non-neoplastic polyps. Unlike the other two classification models, the ViT model, as a transformer-based model, processes the image by dividing it into patches for training. As shown in the Grad-CAM visualization of the ViT model, small red patches were observed across the image. The ViT model highlighted the polyp regions in red and demonstrated decent classification performance for distinguishing between neoplastic and non-neoplastic polyps. Figure 6 shows the polyp contours and probability markings of non-neoplastic and neoplastic polyps displayed in the EUS video (Supplementary Videos 1 and 2). The attached video link of Fig. 6 shows the classification of non-neoplastic (yellow) and neoplastic (blue) polyps using colors. The number displayed alongside the color represents the probability of the predicted class.

Table 5 presents the performance results of the classification models using the final validatiosn dataset, excluding the initially used training and testing datasets. Figure 7 shows the mean confusion matrix of the final validation cohort, calculated based on proportional values. Accuracy was 0.879 for EfficientNet-B2, 0.871 for ResNet101, and 0.755 for ViT. The corresponding F1-scores were 0.917, 0.825, and 0.823, and the AUC values were 0.861, 0.893, and 0.794, respectively.

## Discussion

This study evaluated the performance of AI in analyzing EUS videos for GB polyp assessment using a training cohort of 96 videos from 50 patients and a validation cohort of 30 videos from 14 patients. The analysis was

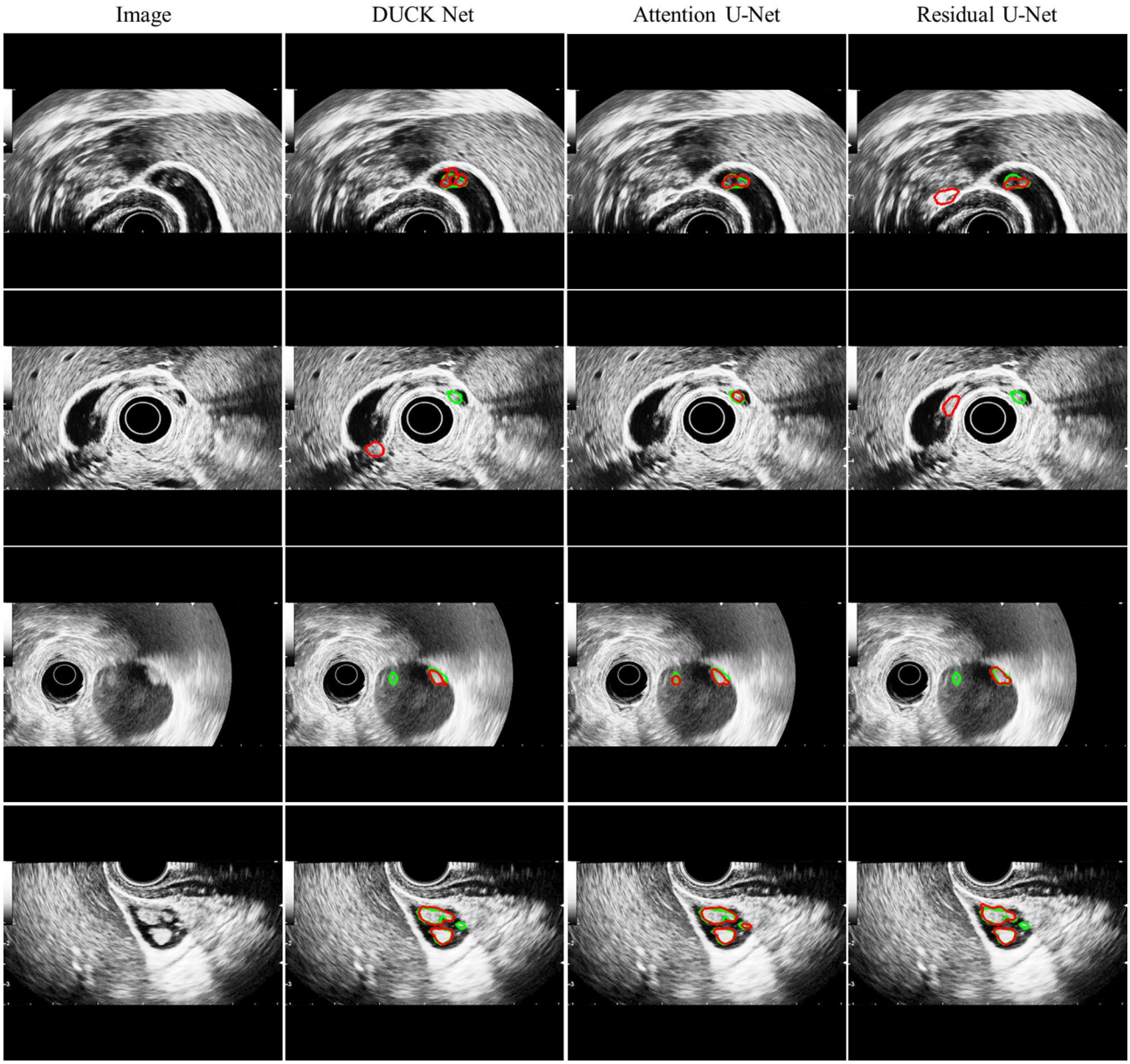


**Fig. 3**. Segmentation models predict polyp regions. The red line is the prediction of the segmentation model, and the green line is the ground truth.

| Model | Accuracy (95% CI) | Precision (95% CI) | Recall (95% CI) | DSC (95% CI) | IoU (95% CI) | PPV (95% CI) | NPV (95% CI) |
|---|---|---|---|---|---|---|---|
| Attention U-Net | 0.941 (0.940–0.942) | 0.809 (0.798–0.822) | 0.537 (0.522–0.552) | 0.612 (0.598–0.626) | 0.490 (0.476–0.505) | 0.808 (0.487–0.974) | 0.822 (0.655–0.924) |
| Residual U-Net | 0.941 (0.940–0.942) | 0.810 (0.799–0.822) | 0.607 (0.595–0.620) | 0.675 (0.664–0.686) | 0.555 (0.543–0.567) | 0.807 (0.453–0.937) | 0.845 (0.675–0.941) |
| DUCK Net | 0.943 (0.942–0.944) | 0.715 (0.700–0.729) | 0.692 (0.675–0.709) | 0.683 (0.668–0.698) | 0.574 (0.560–0.588) | 0.717 (0.434–0.903) | 0.866 (0.700–0.958) |

**Table 3**. Attention U-Net, residual U-Net, and DUCK net final validation results for polyp segmentation. DUCK, deep understanding convolutional kernel; CI, confidence interval; DSC, dice similarity coefficient; IoU, intersection over union.

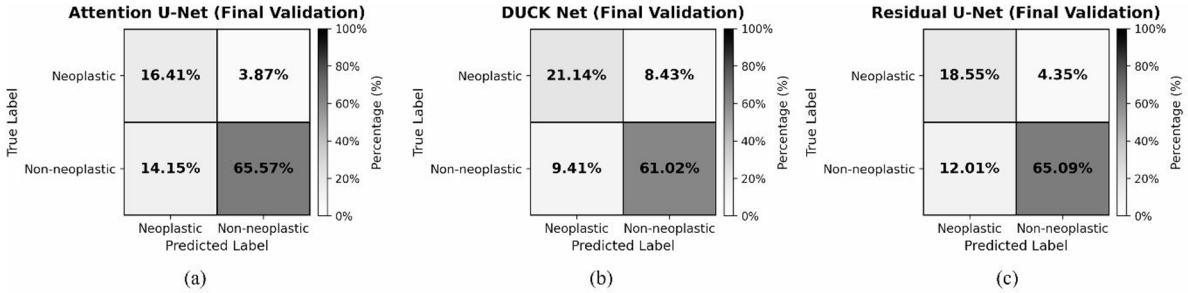


**Fig. 4**. Mean confusion matrix of fivefold cross-validation in the final validation cohort.

| Model | Accuracy (95% CI) | Precision (95% CI) | Recall (95% CI) | F1-score (95% CI) | AUC (95% CI) | PPV (95% CI) | NPV (95% CI) |
|---|---|---|---|---|---|---|---|
| EfficientNet-B2 | 0.957 (0.950–0.964) | 0.925 (0.902–0.949) | 0.954 (0.937–0.970) | 0.939 (0.929–0.949) | 0.991 (0.988–0.993) | 0.926 (0.730–0.990) | 0.990 (0.932–0.998) |
| ResNet101 | 0.907 (0.864–0.951) | 0.857 (0.766–0.947) | 0.900 (0.843–0.957) | 0.873 (0.821–0.925) | 0.965 (0.945–0.986) | 0.855 (0.608–0.942) | 0.978 (0.911–0.993) |
| ViT | 0.853 (0.834–0.873) | 0.829 (0.788–0.870) | 0.730 (0.670–0.790) | 0.774 (0.739–0.808) | 0.895 (0.872–0.917) | 0.827 (0.548–0.930) | 0.943 (0.864–0.973) |

**Table 4**. EfficientNet-B2, ResNet101, and ViT fivefold cross-validation results for polyp classification. ViT, Vision Transformer; CI, confidence interval; AUC, area under the curve.

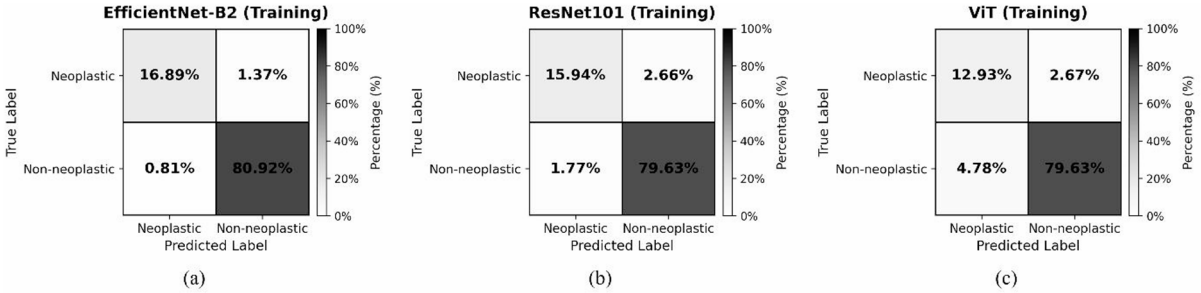


**Fig. 5**. Mean confusion matrix of fivefold cross-validation in the training cohort.

conducted in two stages: first, GB polyps were identified in the entire EUS video, and second, neoplastic and non-neoplastic polyps were differentiated.

For GB polyp segmentation, three models were tested, with the Attention U-Net model demonstrating the best overall performance, achieving an accuracy of 99.8% in the training cohort and 94.1% in the validation cohort. In the second stage of polyp classification, EfficientNet-B2 outperformed the other models, achieving 95.7% accuracy and an F1-score of 93.9% in the training cohort and 87.9% accuracy and an F1-score of 91.7% in the validation cohort. To the best of our knowledge, this is the first study to utilize AI to analyze EUS videos of GB polyps.

Few studies have applied AI to analyze gallbladder polyps, with most using still abdominal ultrasound images and only one utilizing still EUS images[10–12,28]. In these studies, the region-of-interest selection process, which corresponds to GB polyp segmentation in our study, was mostly performed manually rather than using AI[10–12,28]. One previous study using still abdominal ultrasound images attempted computer-aided segmentation, but rather than isolating only GB polyps, it segmented the entire gallbladder, distinguishing it from the background[29]. However, in this study, we successfully implemented AI-based GB polyp segmentation with high accuracy, enabling real-time EUS video analysis.

Although the DSC and IoU values for GB polyp segmentation were relatively low owing to the inherent subjectivity in manually annotated ground truth, Fig. 2 illustrates that the polyp contours generated by the Attention U-Net model closely matched the actual polyp boundaries. These minor discrepancies may have a minimal impact on polyp assessment. Nevertheless, refinement of the ground truth annotation, such as multi-expert consensus labeling, may further enhance the segmentation metrics in future studies. This automated approach demonstrates the feasibility of real-time AI-assisted EUS video analysis.

Furthermore, the accuracy of differentiating neoplastic and non-neoplastic polyps in this study exceeded that of previous studies that used still abdominal ultrasound images for AI analysis. The polyp classification

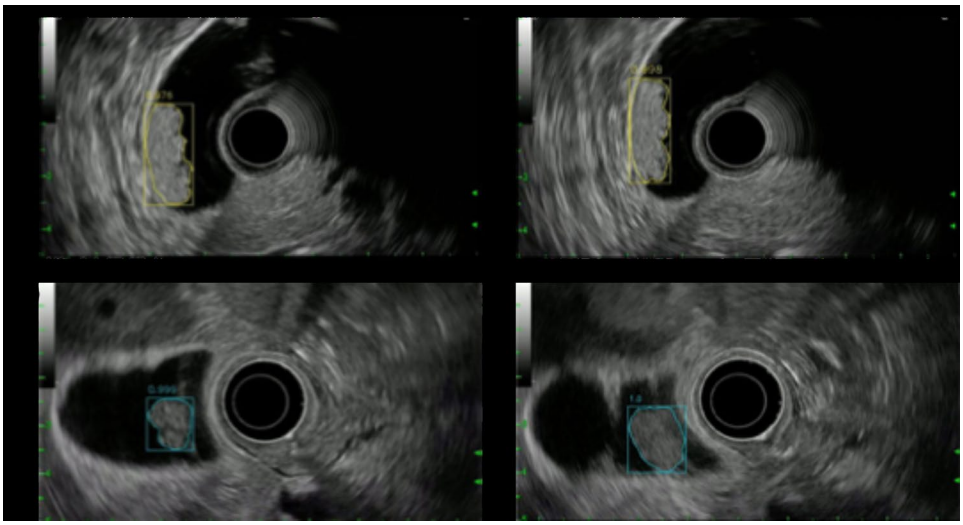

**Fig. 6**. Images captured from the EUS analysis video. (**a**,**b**) GB polyps with a higher likelihood of being non-neoplastic are outlined in yellow, with their probability displayed as a numerical value. Images were captured from a video of a single patient with a histologically confirmed non-neoplastic polyp. (**c**,**d**) GB polyps that were more likely to be neoplastic are outlined in sky blue, with their probability also displayed numerically in sky blue. Images were captured from the video of another patient with a histologically confirmed neoplastic polyp. See Supplementary Videos 1 and 2.

| Model | Accuracy (95% CI) | Precision (95% CI) | Recall (95% CI) | F1-score (95% CI) | AUC (95% CI) | PPV (95% CI) | NPV (95% CI) |
|---|---|---|---|---|---|---|---|
| EfficientNet-B2 | 0.879 (0.842–0.916) | 0.872 (0.828–0.916) | 0.968 (0.945–0.991) | 0.917 (0.889–0.943) | 0.861 (0.796–0.919) | 0.869 (0.578–0.957) | 0.983 (0.857–0.999) |
| ResNet101 | 0.871 (0.844–0.898) | 0.795 (0.742–0.847) | 0.854 (0.807–0.901) | 0.825 (0.784–0.863) | 0.893 (0.864–0.921) | 0.797 (0.523–0.949) | 0.934 (0.750–0.978) |
| ViT | 0.755 (0.706–0.803) | 0.804 (0.751–0.856) | 0.842 (0.793–0.891) | 0.823 (0.784–0.859) | 0.794 (0.744–0.841) | 0.802 (0.523–0.949) | 0.930 (0.750–0.978) |

**Table 5**. EfficientNet-B2, ResNet101, and ViT final validation results for polyp classification. ViT, Vision Transformer; CI, confidence interval; AUC, area under the curve.

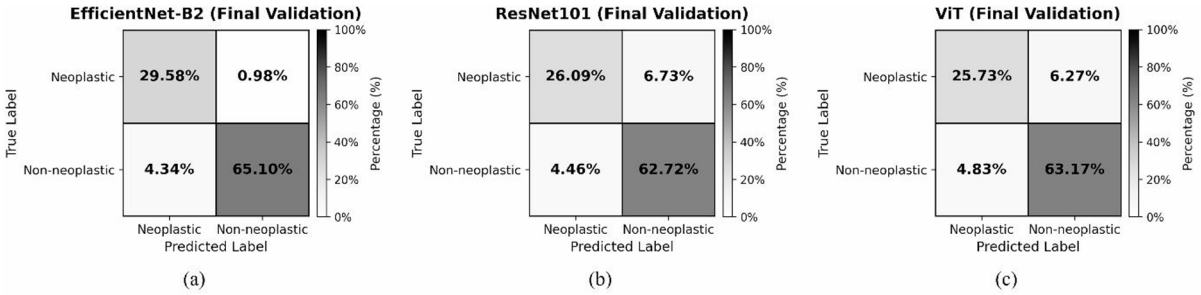


**Fig. 7**. Mean confusion matrix of fivefold cross-validation in the final validation cohort.

accuracy of the EfficientNet-B2 model in our study reached 95.7% in the training cohort and 87.9% in the validation cohort, which is higher than the reported accuracy range of 83.63–87.54% in three prior studies[11,12,28]. Additionally, our F1-score was 0.939 in the training cohort and 0.917 in the validation cohort, both higher than the 0.788 reported in a previous study[12], confirming the robust performance of our model. We excluded cases with unclear GB polyp visibility to ensure reliable annotation and data consistency. This process was not intended to select cases that were easier to differentiate but rather to maintain the overall quality and interpretability of the dataset. Such careful dataset management may have contributed to improved model accuracy.

Although previous studies using still abdominal ultrasound images included 224–535 patients, a larger sample size than that in our study, the superior accuracy in our results can be attributed to the use of the EUS modality, which has been reported to provide higher accuracy in differentiating gallbladder polyps[5,6]. Moreover, the use of

video-based AI analysis, rather than still images, likely contributed to the improved classification performance despite the relatively small sample size. In our study, although the total number of patients was relatively small ($n$ = 67), the dataset comprised 4,328 video frames, which exceeded those used in prior studies based on still images (501 images[11] and 3,118 images[28] and was only slightly fewer than in another study using 6,056 still images[12]. Thus, despite the small sample size, the relatively large amount of video-derived data available for analysis may have been one factor contributing to the favorable performance observed in our study. Our previous study utilizing still EUS images reported a polyp classification accuracy of 89.8%, which was also higher than that of prior studies using abdominal ultrasound[10]. This further supports the notion that EUS modality enhances classification accuracy. However, considering that the classification accuracy in our validation cohort was lower than that reported in a study using still EUS images from 753 patients[10], future large-scale prospective studies utilizing EUS video data are necessary to further improve accuracy.

This study had some limitations. First, it was a retrospective study with a relatively small sample size. However, despite the limited number of patients, the use of video data comprising approximately 4,000 video frames allowed for a more comprehensive analysis, leading to a diagnostic performance that surpasses that of previous AI-based studies on GB polyps. In a prior study[10], an EUS-AI system with ResNet50 architecture was trained using 1,039 still-cut images, whereas our video-based dataset provided about 3,000 frames for training, offering a richer resource for training and validation. To mitigate potential overfitting risks from the limited sample size, we implemented fivefold cross-validation at the video level, ensuring that frames from the same video were not simultaneously included in both training and validation sets. This design minimized data leakage and allowed us to confirm consistent performance across folds, suggesting that the model achieved a reasonable degree of generalizability despite the sample size limitation. Second, selection bias was inevitable because the study population was limited to patients whose diagnoses were pathologically confirmed, and consequently all subjects had undergone cholecystectomy. Currently, there is no gold standard diagnostic method that can replace pathological confirmation for GB polyps. Therefore, future prospective studies including non-surgical follow-up cohorts with long-term observation are needed to overcome this limitation and to ensure the generalizability of AI-based GB polyp diagnosis.

## Conclusions

In conclusion, this is the first study to analyze EUS videos using AI for GB polyp assessment. Our AI model for EUS video-based GB polyp segmentation and classification demonstrated strong diagnostic performance. Further large-scale prospective studies are essential to validate its clinical utility as a real-time diagnostic tool for EUS-based GB polyp evaluation.

## Data availability

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request. We have uploaded the inference code and trained model weights to GitHub (https://github.com/user-dynamite/EUS-GB-poylp).

Received: 27 April 2025; Accepted: 14 November 2025
Published online: 08 December 2025

## References

1. Jørgensen, T. & Jensen, K. H. Polyps in the gallbladder. A prevalence study. *Scand. J. Gastroenterol.* **25**, 281–286. https://doi.org/10.1080/00365521.1990.12067104 (1990).
2. Ito, H. et al. Polypoid lesions of the gallbladder: diagnosis and followup. *J. Am. Coll. Surg.* **208**, 570–575. https://doi.org/10.1016/j.jamcollsurg.2009.01.011 (2009).
3. Riddell, Z. C., Corallo, C., Albazaz, R. & Foley, K. G. Gallbladder polyps and adenomyomatosis. *Br. J. Radiol.* **96**, 20220115. https://doi.org/10.1259/bjr.20220115 (2023).
4. Wennmacker, S. Z. et al. Transabdominal ultrasound and endoscopic ultrasound for diagnosis of gallbladder polyps. *Cochrane Database Syst. Reviews*. https://doi.org/10.1002/14651858.CD012233.pub2 (2018).
5. Azuma, T., Yoshikawa, T., Araida, T. & Takasaki, K. Differential diagnosis of polypoid lesions of the gallbladder by endoscopic ultrasonography. *Am. J. Surg.* **181**, 65–70. https://doi.org/10.1016/S0002-9610(00)00526-2 (2001).
6. Sugiyama, M., Xie, X. Y., Atomi, Y. & Saito, M. Differential diagnosis of small polypoid lesions of the gallbladder: the value of endoscopic ultrasonography. *Ann. Surg.* **229**, 498–504. https://doi.org/10.1097/00000658-199904000-00008 (1999).
7. Cho, J. H. et al. Hypoechoic foci on EUS are simple and strong predictive factors for neoplastic gallbladder polyps. *Gastrointest. Endosc.* **69**, 1244–1250. https://doi.org/10.1016/j.gie.2008.10.017 (2009).
8. Kim, S. Y. et al. The efficacy of real-time colour doppler flow imaging on endoscopic ultrasonography for differential diagnosis between neoplastic and non-neoplastic gallbladder polyps. *Eur. Radiol.* **28**, 1994–2002. https://doi.org/10.1007/s00330-017-5175-3 (2018).
9. Sadamoto, Y. et al. A useful approach to the differential diagnosis of small polypoid lesions of the gallbladder, utilizing an endoscopic ultrasound scoring system. *Endoscopy* **34**, 959–965. https://doi.org/10.1055/s-2002-35859 (2002).
10. Jang, S. I. et al. Diagnostic performance of endoscopic ultrasound-artificial intelligence using deep learning analysis of gallbladder polypoid lesions. *J. Gastroenterol. Hepatol.* **36**, 3548–3555. https://doi.org/10.1111/jgh.15673 (2021).
11. Kim, T. et al. Gallbladder polyp classification in ultrasound images using an ensemble convolutional neural network model. *J. Clin. Med.* **10** https://doi.org/10.3390/jcm10163585 (2021).
12. Jeong, Y. et al. Deep learning-based decision support system for the diagnosis of neoplastic gallbladder polyps on ultrasonography: preliminary results. *Sci. Rep.* **10**, 7700. https://doi.org/10.1038/s41598-020-64205-y (2020).
13. Lu, L., Zhou, Y., Panetta, K. & Agaian, S. Comparative study of histogram equalization algorithms for image enhancement. *Mob. Multimedia/Image Process. Secur. Appl.* **2010 7708**, 337–347. https://doi.org/10.1117/12.853502 (2010).
14. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proc. Int. Joint Conf. Artif. Intell.* 1137–1145 http://robotics.stanford.edu/~ronnyk/accEst.pdf (1995).
15. Oktay, O. et al. Attention u-net: learning where to look for the pancreas. *ArXiv Preprint arXiv:1804 03999*. https://doi.org/10.48550/arXiv.1804.03999 (2018).

16. Zhang, Z., Liu, Q. & Wang, Y. Road extraction by deep residual u-net. *IEEE Geosci. Remote Sens. Lett.* **15**, 749–753. https://doi.org/10.1109/LGRS.2018.2802944 (2018).
17. Dumitru, R. G., Peteleaza, D. & Craciun, C. Using DUCK-Net for polyp image segmentation. *Sci. Rep.* **13**, 9803. https://doi.org/10.1038/s41598-023-36940-5 (2023).
18. Tan, M. & Le, Q. EfficientNet: rethinking model scaling for convolutional neural networks. *Proc. Int. Conf. Mach. Learn.* 6105-6114 https://doi.org/10.48550/arXiv.1905.11946 (2019).
19. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 770–778 https://doi.org/10.1109/CVPR.2016.90 (2016).
20. Lee, S. H., Lee, S. & Song, B. C. Vision transformer for small-size datasets. *arXiv* https://doi.org/10.48550/arXiv.2112.13492 (2021).
21. Prechelt, L. Early stopping—but when? *Neural Netw. Tricks Trade*. 55–69. https://doi.org/10.1007/3-540-49430-8_3 (2002).
22. Yao, Y., Rosasco, L. & Caponnetto, A. On early stopping in gradient descent learning. *Constructive Approximation*. **26**, 289–315. https://doi.org/10.1007/s00365-006-0663-2 (2007).
23. D Zeiler, M. Adadelta: an adaptive learning rate method. *ArXiv Preprint arXiv:1212 5701*. https://doi.org/10.48550/arXiv.1212.5701 (2012).
24. N Smith, L. Cyclical learning rates for training neural networks. *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)* 464–472. https://doi.org/10.1109/WACV.2017.58 (2017).
25. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. & Torralba, A. Learning deep features for discriminative localization. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 2921–2929 https://doi.org/10.1109/CVPR.2016.319 (2016).
26. Selvaraju, R. R. et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proc. IEEE Int. Conf. Comput. Vis.* 618–626 https://doi.org/10.1109/ICCV.2017.74 (2017).
27. Leem, S. & Seo, H. Attention guided CAM: visual explanations of vision transformer guided by self-attention. *Proc. AAAI Conf. Artif. Intell.* 2956-2964 https://doi.org/10.1609/aaai.v38i4.28077 (2024).
28. Choi, J. H. et al. Analysis of ultrasonographic images using a deep learning-based model as ancillary diagnostic tool for diagnosing gallbladder polyps. *Dig. Liver Disease*. **55**, 1705–1711. https://doi.org/10.1016/j.dld.2023.06.023 (2023).
29. Chen, T. et al. Computer-aided diagnosis of gallbladder polyps based on high resolution ultrasonography. *Comput. Methods Programs Biomed.* **185**, 105118. https://doi.org/10.1016/j.cmpb.2019.105118 (2020).

## Acknowledgements

This work was supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety; Project Number: 1711196789, RS-2023-00252804), and a faculty research grant from Yonsei University College of Medicine (Sung Ill Jang, 6-2023-0209).

## Author contributions

Y.H.C.: drafting of the article; technical and material support; critical revision of the article for important intellectual content; analysis and interpretation of dataJ.Y.P.: drafting of the article; technical and material support; critical revision of the article for important intellectual content; analysis and interpretation of dataS.Y.L.: technical and material support; analysis and interpretation of the dataJ.H.J.: technical and material support; analysis and interpretation of the dataY.J.K.: analysis and interpretation of the data; critical revision of the article for important intellectual contentK.G.K.: conception and design; case collection; critical revision of the article for important intellectual content; final approval of the articleS.I.J.: conception and design; case collection; critical revision of the article for important intellectual content; final approval of the article.

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-29179-9.

**Correspondence** and requests for materials should be addressed to K.G.K. or S.I.J.