



OPEN

DATA DESCRIPTOR

# Towards Automated Reporting: A Bronchoscopy Report Dataset for Enhancing Multimodality Large Language Models

Xingjian Luo<sup>1,2,5</sup>, Xinyan Huang<sup>3,5</sup>, Xusheng Liang<sup>1,5</sup>, Jiyu Wang<sup>3</sup>, Jincui Gu<sup>3</sup>, Dong Yi<sup>1</sup>, Haohan Zhao<sup>1</sup>, Haihong Zhang<sup>1</sup>, Jinlin Wu<sup>1</sup>, Zhen Lei<sup>1</sup>, Gaofeng Meng<sup>1</sup>, Hongliang Ren<sup>2</sup>, Jiebo Luo<sup>4</sup>, Huai Liao<sup>3</sup>✉ & Hongbin Liu<sup>1</sup>✉

Bronchoscopy examination is essential for diagnosing and managing respiratory diseases. While Multimodality Large Language Models (MLLMs) can enhance the efficiency and accuracy of medical report writing, existing datasets lack descriptive and comprehensive annotations for complex cases, hindering their ability to facilitate adequate learning of image-report relationships. To address this problem, we introduce BERD, a Bronchoscopy Examination Report Dataset, which includes 3,692 bronchoscopy examination reports. Among these reports, 6,330 representative images are annotated with single-image text descriptions and classification labels. BERD emphasizes the provision of versatile and detailed descriptions of findings. All these reports and annotations were performed by experienced clinicians specializing in bronchoscopy. Furthermore, experimental results show that fine-tuning state-of-the-art MLLMs on BERD significantly improves their ability to generate accurate and comprehensive reports, advancing AI applications in bronchoscopy.

## Background & Summary

Bronchoscopy examination is a vital diagnostic and therapeutic tool in respiratory medicine<sup>1</sup>. It allows direct visualization of the tracheobronchial tree, enabling clinicians to identify abnormalities such as inflammations, infections, tumors, or structural changes<sup>2</sup>. In addition to its diagnostic utility, bronchoscopy examination is widely used for therapeutic interventions, such as foreign body removal, airway stenting, or lavage for microbiological analysis<sup>3</sup>. The findings from bronchoscopy are typically documented in detailed reports that provide crucial information for diagnosis, treatment planning, and follow-up care.

However, generating these reports is a labor-intensive task that relies heavily on the experience and expertise of clinicians<sup>4</sup>. Each report must not only accurately document the observed findings but also provide cohesive and structured descriptions to ensure effective communication among clinicians. The increasing demand for bronchoscopy examination writing in clinical workflows has amplified the need for efficient and accurate report generation methods, highlighting the potential role of artificial intelligence in automating and enhancing this process.

With recent advancements in artificial intelligence, Multimodality Large Language Models (MLLMs) have shown great promise in medical applications, especially in tasks requiring the integration of visual and textual data<sup>5</sup>. These models<sup>6–10</sup>, trained on paired image-text datasets, can analyze medical images and generate descriptive reports, offering a solution to the time and expertise constraints faced in clinical settings. For bronchoscopy examination reports, MLLMs can potentially automate the generation of structured, accurate, and comprehensive reports, reduce the workload of clinicians and improve reporting quality.

<sup>1</sup>Centre for Artificial Intelligence and Robotics (CAIR), Hong Kong Institute of Science & Innovation, Chinese Academy of Sciences, Hong Kong SAR, China. <sup>2</sup>Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China. <sup>3</sup>Department of Pulmonary and Critical Care Medicine, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong Province, P. R. China. <sup>4</sup>Hong Kong Institute of Science & Innovation, Chinese Academy of Sciences, Hong Kong SAR, China. <sup>5</sup>These authors contributed equally: Xingjian Luo, Xinyan Huang, Xusheng Liang. ✉e-mail: [liaohuai@mail.sysu.edu.cn](mailto:liaohuai@mail.sysu.edu.cn); [hongbin.liu@cair-cas.org.hk](mailto:hongbin.liu@cair-cas.org.hk)

Dataset Name	Data Source	Patient Numbers	Images	Sub-Task
BronchoLC <sup>11</sup>	Public	208	2,921	CLS, SEG
UAAL <sup>12</sup>	Public	—	3,814	CLS, SEG
BI2K <sup>14</sup>	Public	615	2,900	CLS
PKDN <sup>13</sup>	Private	200	2,029	CLS
Ours	Public	3,692	6,330	CAP, CLS

**Table 1.** Statistics comparison of existing datasets and our Broncho-R dataset, including the dataset name, dataset source, number of samples, and multiple sub-task involvement. In general, the main shortcomings of existing datasets include the following: (i) there is a lack of comprehensive and evenly distributed departmental coverage, as well as a sufficient number of patients, to prevent bias; (ii) data sources are often private and inaccessible; (iii) existing datasets usually only have single-task annotation. “CLS” refers to classification task, “SEG” refers to segmentation task, and “CAP” refers to caption task.

Despite these advancements, the training of MLLMs for generating bronchoscopy examination reports is hindered by the limitations of existing datasets. Most publicly available datasets for bronchoscopy focus on narrow tasks, providing only limited support for report generation, as shown in Table 1. For instance, the BronchoLC<sup>11</sup> dataset is designed exclusively for tumor localization, offering annotations about tumor presence and its corresponding bronchial location, but neglects other common findings such as sputum, clot, or bleeding that are critical in routine bronchoscopic reports. Similarly, the UAAL<sup>12</sup> dataset, primarily developed for bronchoscopy navigation, focuses solely on the position of the bronchoscopic device relative to the airway, without capturing any pathological or descriptive information. The PKDN<sup>13</sup> dataset, while notable for its annotated bronchoscopic images, is a proprietary resource and focuses only on binary classification tasks (lesion vs. non-lesion), offering no insights into nuanced findings necessary for comprehensive report generation. The BI2K<sup>14</sup> dataset, though broader in scope, divides the data into benign lesions, malignant lesions, and normal conditions, which still falls short of the granularity required to describe routine findings such as sputum, bleeding, edema, or congestion.

The limitations of these datasets highlight their inadequacy in facilitating MLLMs for detailed and comprehensive bronchoscopy examination report generation. Unlike radiology images, such as CT or MRI, where datasets like MIMIC<sup>15</sup> and PMC<sup>16</sup> provide paired image-text report data for training models capable of generating structured reports, bronchoscopy examination reports remain a shortage. Existing datasets have constrained the field to tasks such as navigation or single-lesion segmentation, leaving the task of comprehensive report generation largely unaddressed. This gap has hindered the ability of AI systems to provide meaningful assistance to clinicians, especially in automating the time-consuming process of detailed report writing.

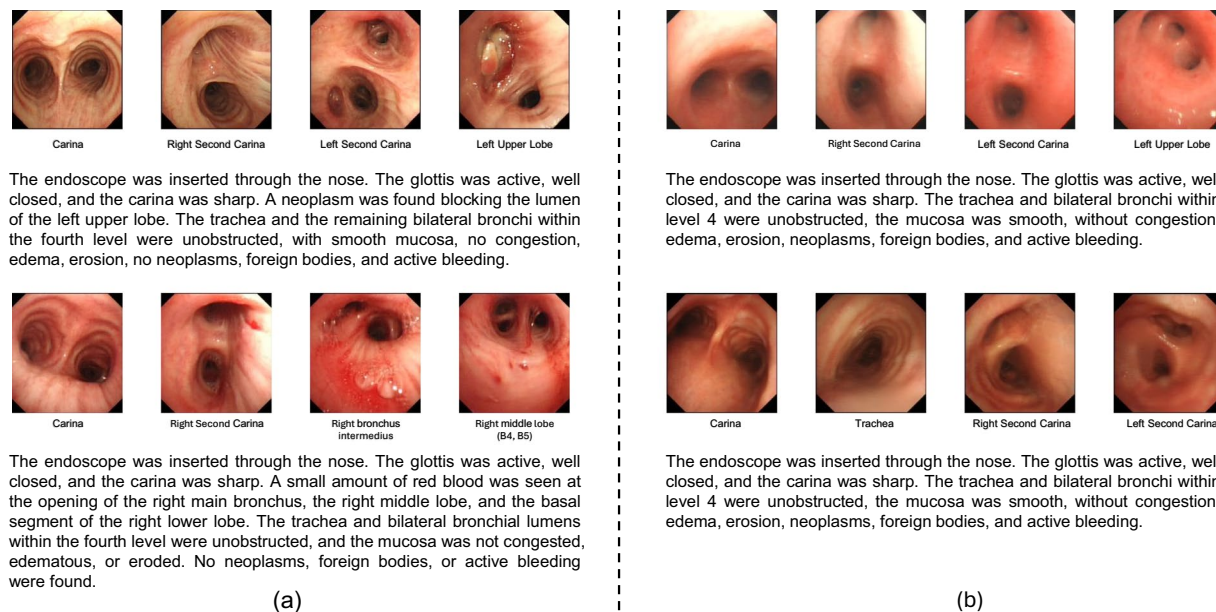
To address these challenges, our BERD dataset provides a high-quality resource for training and evaluating MLLMs. By including 3,692 bronchoscopy examination reports, with 6,330 images annotated with detailed descriptions, BERD enables MLLMs to learn holistic and nuanced representations of bronchoscopic findings. Unlike existing datasets, BERD emphasizes report-centric annotations, capturing a wide range of findings, including common yet clinically significant observations. This dataset bridges the gap between current MLLM capabilities and the demands of clinical bronchoscopy report generation, paving the way for more accurate, efficient, and clinically relevant AI applications.

## Methods

To facilitate the development of AI-powered automatic report generation in the Bronchoscopy field, we collected an image-caption pair dataset with high-quality complete annotations done by two professional clinicians. In the process of data collection, we removed all parts that might contain personal information about patients and clinicians, retaining only bronchoscopy images and objective descriptive reports without any private information. This retrospective study was approved by the Clinical Research and Laboratory Animal Ethics Committee of the First Affiliated Hospital of Sun Yat-sen University (Approval Number: Ethical Review No. [2024]517), permitting data collection, annotation, subsequent research, and publication. Since this study does not involve specimen collection, does not interfere with patient examination procedures, and does not include follow-up or biological samples, an application for exemption from informed consent was submitted and approved within the hospital.

**Bronchoscopy examination reports.** The dataset was collected by the Department of Pulmonary and Critical Care Medicine, First Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. Between 2022 and 2023, a total of 8,477 bronchoscopy examinations were performed by experienced clinicians in the hospital, from which we selected 3,692 representative patient cases with 6,330 images. Each original report was generated with four images that were selected by the clinicians, as shown in Fig. 1. Clinicians can take screenshots at critical moments during bronchoscopy examinations and mark their positions using a bronchoscope made by Olympus. The captured images are usually of representative significance. After the examination, the clinician writes an examination report and selects the four most representative images from all the captured images to include in the final report.

**Image-caption pair.** For each bronchoscopy report, the process of pairing images with captions was carefully conducted to ensure clinical relevance and accuracy. Clinicians first manually reviewed each report to identify the most relevant descriptions for each image in the bronchoscopy examination report. Then, the clinicians



**Fig. 1** Examples of translated original bronchoscopy reports. A report typically contains four images selected by the clinicians who conducted the examination. The four images are annotated with location. **(a)** Examples of the abnormal cases, where the lesions are depicted with their positions in the report. **(b)** Example of normal cases, where one template is used when no lesion is found.

removed the location information from the sentence and retained only the descriptive text. In addition, to ensure the robustness of the description and the accuracy of subsequent model training, text containing specific numbers, such as “3 mm” has been removed from the description. These textual descriptions were extracted directly from the examination reports, capturing details such as abnormalities (e.g., tumors, edema, or exudates) and their associated observations, including color, size, and amount. For images with no visible abnormalities, a standardized caption, such as “The lumen is unobstructed, and the mucosa is free of congestion, edema, or erosion. No neoplasms, foreign bodies, or active bleeding are found.” was given, we simplified this sentence to “It is normal.” and assigned to maintain consistency across the dataset. The caption corresponding to the image might contain lesions of more than one type. This meticulous pairing process ensures that every image is tightly linked to a meaningful and comprehensive description, providing a strong foundation for AI models to learn image-text relationships effectively. The original report and position marks were written in Chinese, after the image-caption pair annotation, we translated them into English using a locally deployed LLM (Large Language Model) Qwen3-32B<sup>17</sup> to protect data privacy.

**LLM-assisted classification annotation.** To streamline the classification of images, we integrated a locally deployed LLM into the annotation workflow. Experienced clinicians first defined a comprehensive list of disease categories from expert consensus and bronchoscopy reporting guidelines, including common terms like congestion, edema, or tumor. Using this reference, LLM was employed to extract relevant keywords and synonyms from the captions, automatically categorizing each image-caption pair into one or more predefined classes. After the initial classification, all LLM-generated labels were reviewed and refined by clinicians to ensure clinical accuracy and alignment with medical standards. This semi-automated approach significantly reduced the manual workload while preserving the overall quality and consistency of the dataset.

The whole annotation process is illustrated in Fig. 2(a), and the final annotation result is shown in Fig. 2(b).

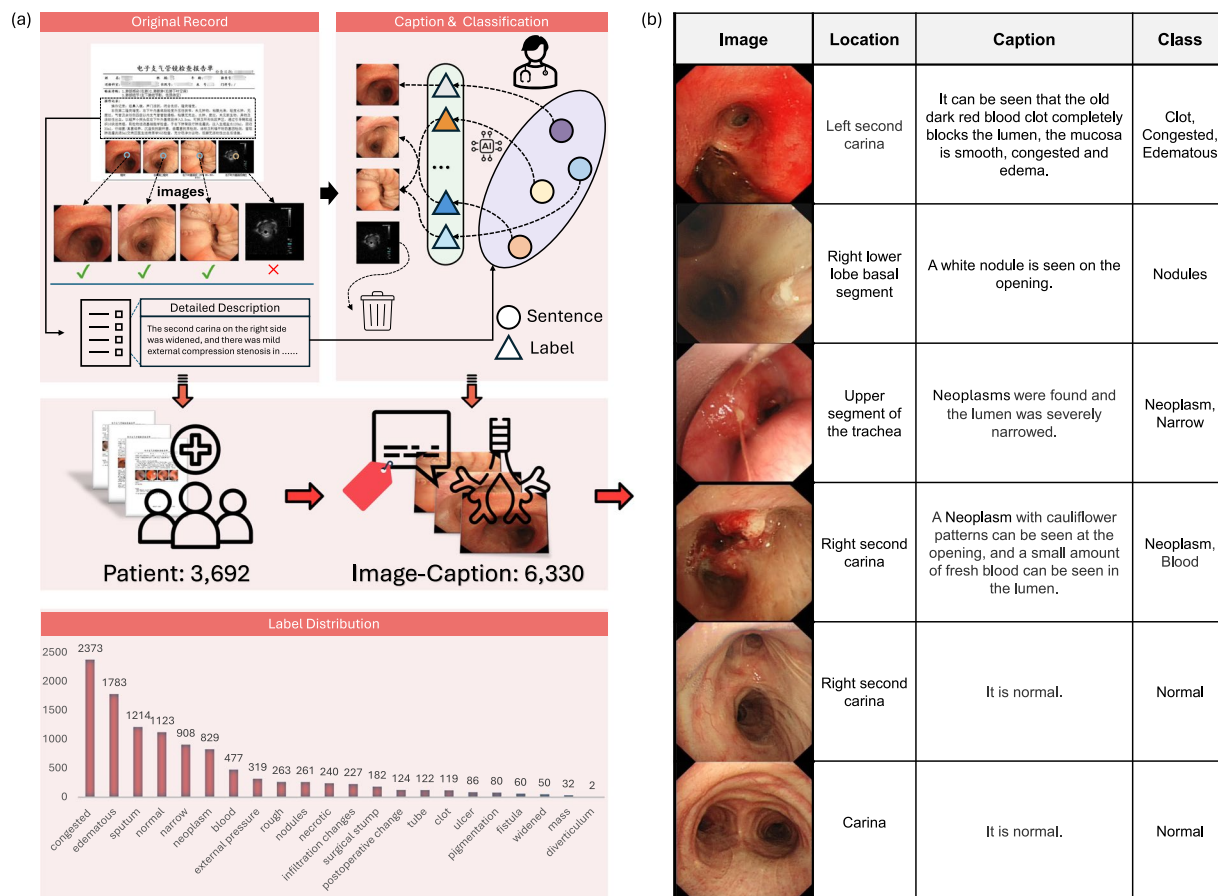
### Data Records

The dataset is available from the Science Data Bank at <https://doi.org/10.57760/sciencedb.28018><sup>18</sup>.

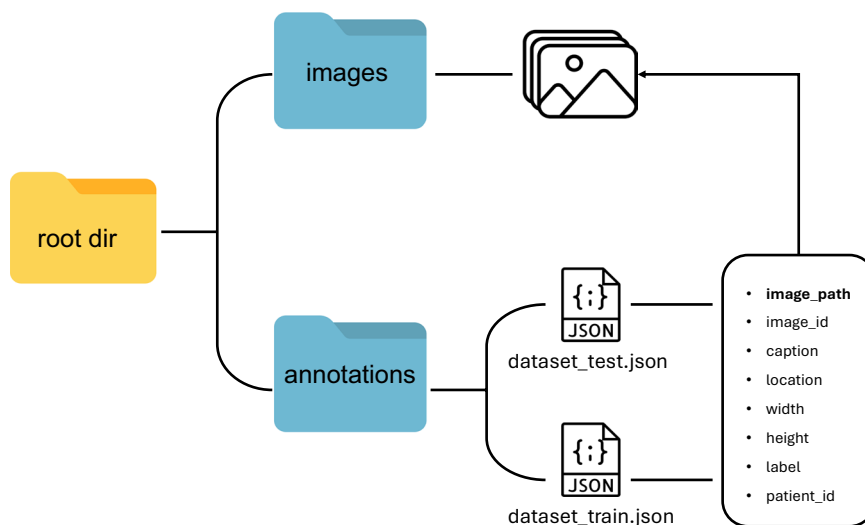
The dataset contains two folders, one of which is the annotation folder that contains annotation JSON files. The other folder contains images in PNG format.

The annotation JSON files contain one training annotation file and a testing annotation file. The annotation file includes seven elements:

*image\_path*, *image\_id*, *caption*, *location*, *width*, *height*, *label*, *patient\_id*. The *image\_path* is the relative path of the image, *image\_id* is the unique ID of the image, which should be the same as its image name. The *caption* is the caption annotation. The *location* is the anatomical location of the image. The *height* and *width* are the size of the image. And the *label* is the classification result of the image. The *patient\_id* is the ID of patients. Different images might come from the same patient. Both training and testing images are in the images folder. The dataset folder structure is shown in Fig. 3.



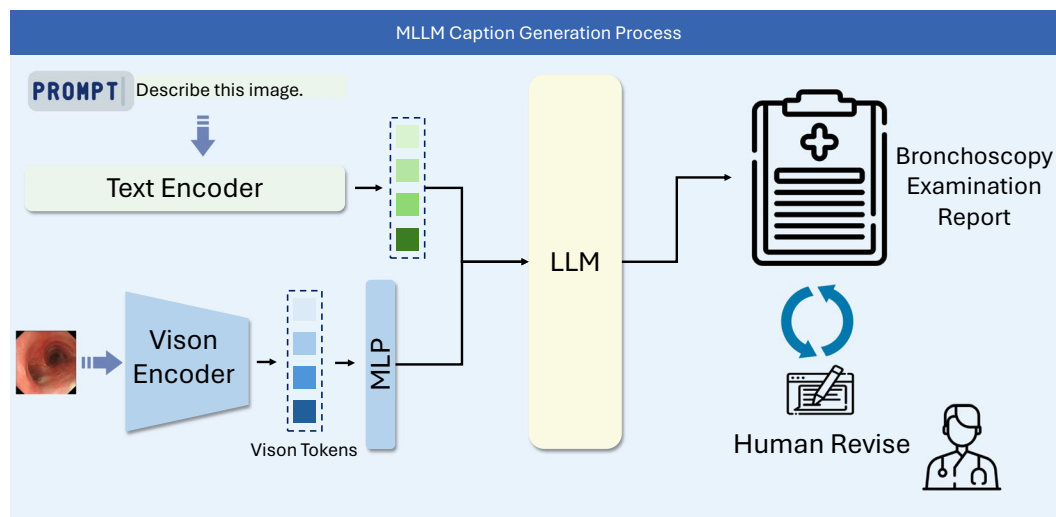
**Fig. 2** (a) The process of annotation: we first extract the detailed descriptions from the original report and then distribute corresponding descriptions to each image while removing the unrelated images, and then annotate the labels based on the sentences. (b) Examples of the annotated dataset. The dataset comprises images along with their corresponding locations, descriptions, and labels.



**Fig. 3** The structure of the dataset folder.

### Technical Validation

**Experience of the operators.** The department where the bronchoscopy examination reports are collected specializes in minimally invasive diagnosis and treatment of respiratory diseases, with an average annual



**Fig. 4** The process of caption generation and report writing. We take the image and default prompt as input to the MLLM, and the output is the corresponding caption of that image. In real practice, the MLLM-generated captions can be revised and utilized by the clinicians while writing the bronchoscopy examination report.

completion of over 5,000 bronchoscopic procedures, including diagnostic bronchoscopy and complex interventions such as tumor resection, airway stenting, and bronchial fistula occlusion. Therefore, the quality and diversity of the bronchoscopy examination report can be guaranteed.

**Experience of the annotators.** The annotation for this study was carried out by two bronchoscopists, each with 5 more years of specialized experience in bronchoscopy, supported by two standardized-trained resident clinicians. All annotations were supervised and verified by a senior expert with more than 10 years of bronchoscopic practice, who has performed over 10,000 bronchoscopic examinations and leads technical innovations in navigation-guided biopsies. Referring to clinical atlas standards, bounding boxes and labels for anatomical landmarks and airway lesions were independently annotated by the two experienced bronchoscopists, followed by final review by the senior expert to ensure annotation accuracy.

**Analysis of the dataset and annotations.** To validate the effectiveness of our dataset across various tasks and demonstrate its superiority over current state-of-the-art (SOTA) closed-source models, we conducted comprehensive experiments. These experiments primarily focused on generating bronchoscopic reports and evaluating the performance of leading closed-source General MLLMs, Medical MLLMs, and those fine-tuned on our dataset. Because the bronchoscopic images contain bloody elements, they cannot pass the image review mechanisms of most closed-source MLLMs. Therefore, closed-source MLLMs were not considered in this evaluation. The goal is to verify the utility of our dataset in this domain and to show that current SOTA models, having no prior exposure to bronchoscopic data, perform poorly on such tasks. The procession of caption generation is shown in Fig. 4. The image passes through a vision encoder and an MLP alignment layer, while the textual input passes through the text input module. The visual input is aligned with textual space. Two inputs are then sent to the LLM to generate the final textual output.

**Evaluation metrics.** We employed a combination of standard natural language processing (NLP) metrics and expert evaluations to ensure a robust assessment of the generated reports. **BLEU:** Measures the precision of n-grams in the generated text compared to the reference text. It evaluates how closely the generated reports match the ground truth at the word and phrase levels. BLEU@1 to BLEU@4 represent BLEU scores calculated using 1-gram to 4-gram precision, respectively, where higher n-gram values provide more stringent evaluation of text fluency and coherence. **ROUGE-L:** Focuses on the recall of sequences between the generated report and the reference, emphasizing the overlap of the longest matching subsequences. **METEOR:** Considers both precision and recall by aligning words and phrases semantically, using synonyms and stemming to capture meaning. **CIDEr:** Evaluates the consensus between the generated text and the reference text based on term frequency-inverse document frequency (TF-IDF), ensuring relevance and informativeness in the generated reports. **Accuracy:** To achieve a more intuitive expression while aligning with the cognition of clinicians when composing bronchoscopy examination reports, we asked clinicians to rate the generated captions. The scoring results were binary, with “1” indicating acceptance, meaning the caption could be directly included as part of the report, and “0” indicating rejection, suggesting that the report contained unreasonable elements. Rejection could arise from various reasons, such as missing content or incorrect descriptions. In such cases, clinicians deemed the results unacceptable and required modifications to be made before they were included in the report.

**Experimental results.** We randomly extracted 6,014 images for training and 316 images for testing with no overlap in patients between two datasets. First, we evaluated the performance of current general and medical

## Prompt of report generation evaluation

**Prompt** = ""Describe the bronchoscopy image in detail, focusing on the appearance and condition of the airway. Include observations such as the presence of neoplasm, narrowing, blood clots, sputum, rough surfaces, or normal structures. Note any signs of infiltration changes, congestion, or edema. Mention if the airways appear widened or show signs of external pressure. Describe any surgical stumps, nodules, fistulas, postoperative changes, granulation tissue, necrosis, or masses. Include details of pigmentation, Y-type bifurcations, or ulcers, if present. If there is no abnormality in this image, reply with "It is normal. " Provide a comprehensive description of any abnormalities or normal findings observed in the image. Here are some examples:  
 A small amount of white mucus sputum can be seen in the lumen  
 A cauliflower-like neoplasm was found  
 See a metal coated bronchial support  
 The mucosa can be seen in yellow-white membranes, and the mucosa is slightly congested and edema  
 Dark red blood marks can be seen in the lumen  
 Remember, only output the report, no other words needed.""

**Fig. 5** Prompt used in the MLLM report generation. We use this prompt to make generated reports closer to our caption style.

Model Type	Model	BLEU@1	BLEU@2	BLEU@3	BLEU@4	ROUGE-L	METEOR	CIDER	ACC
General MLLM	Qwen2.5VL-7B <sup>19</sup>	4.05	1.36	0.51	0.29	3.52	13.21	1.82	—
	InternVL-3-8B <sup>20</sup>	3.62	1.33	0.53	0.27	3.92	12.08	1.95	—
Medical MLLM	LLaVA-Med-v1.5-mistral-7B <sup>8</sup>	6.35	1.84	0.85	0.53	6.38	12.45	1.61	—
	Huatuo-GPT-Vision-7B <sup>7</sup>	4.61	1.95	0.98	0.58	4.85	14.92	2.08	—
	UniBioMed <sup>10</sup>	4.02	1.64	0.86	0.54	5.08	7.12	0.62	—
	Lingshu-7B <sup>21</sup>	4.27	1.65	0.70	0.39	3.90	13.58	2.00	—
Open Resource with BR Finetune	Qwen2.5VL-2B	27.68	24.63	22.42	21.29	27.98	29.93	21.88	—
	InternVL-3-3B	35.05	29.63	26.28	23.21	31.40	32.03	24.99	—
	Qwen2.5VL-7B	33.41	30.24	27.28	20.31	35.66	36.58	14.05	81.33
	InternVL-3-8B	35.06	30.50	27.70	25.83	36.29	38.42	27.71	82.91


**Table 2.** Validation of the Caption Generation Task on different MLLMs in percentage.

MLLMs on the test set. To align the model outputs more closely with the style of our caption dataset, we utilized prompts and few-shot examples as shown in Fig. 5. The outputs of both general models and medical domain MLLMs were inferior, failing to accurately describe the bronchoscopy images. This highlights that these MLLMs have not undergone pre-training or fine-tuning in the bronchoscopy domain, likely due to the lack of publicly available datasets in this field. To address this, we fine-tuned general models, specifically Qwen2.5VL<sup>19</sup> (2B and 7B) and InternVL-3<sup>20</sup> (3B and 8B) and tested their performance. The results demonstrate that our fine-tuned models achieved significant improvements across all metrics. The best-performing model, InternVL3-8B, achieved BLEU@1 to BLEU@4 scores of 35.06%, 30.50%, 27.70%, and 25.83%, respectively. ROUGE-L reached 36.29%, METEOR reached 38.42%, and CIDEr scored 27.71%. Additionally, clinicians assess the binary classification accuracy of the generated captions. InternVL-3-8B achieved the highest score, with an accuracy of 82.91%, outperforming the second-best model, Qwen2.5VL-7B, by 1.58% as shown in Table 2, and one result example is shown in Fig. 6.

### Usage Notes

To facilitate the use of the dataset, we offer public access to both the database and all related code. We have provided evaluation metrics for each task and divided the dataset into training and testing sets to ensure a fair comparison. Therefore, we believe this dataset should serve as an excellent benchmark for these relevant tasks and pave the way for report generation tasks. As the details for the data digitization process and codes for pre-processing are provided.

**Limitations.** Despite the high potential in developing report generation models, the proposed dataset has certain limitations. First, all data was collected from a single hospital using the Olympus bronchoscope. This may limit the generalizability of the dataset to other institutions or equipment types. However, since the Olympus bronchoscope is widely used in clinical practice, the dataset maintains a certain level of standardization and remains broadly applicable to similar clinical settings employing the same bronchoscopic technology. Second, specific numerical values, such as lesion sizes (e.g., “3 mm”), were removed from the reports. This decision was made because such quantitative details cannot be directly observed from the corresponding images and could

Image	Ground Truth
	Neoplasms blocking the lumen can be seen at the opening.
Model	Caption
Qwen2.5VL-7B	The bronchoscopy image shows a narrowed segment of the airway with a visible mass obstructing the lumen. The mass appears to have a rough, irregular surface with a yellowish tint, suggesting possible necrosis or inflammation. There is no evidence of neoplasm, blood clots, or sputum within the lumen.
InternVL-3-8B	The bronchoscopy image shows a section of the airway with notable features. There is a visible presence of sputum within the lumen, characterized by a white, frothy appearance. The mucosal surface appears slightly erythematous, indicating possible inflammation.
LLaVA-Med-v1.5-mistral-7B	The bronchoscopy image shows a large, well-defined, and homogenous mass in the right middle lobe of the lung. The mass appears to be protruding into the right main bronchus, which is the airway that branches off from the trachea and leads to the right lung.
Huatuo-GPT-Vision-7B	The bronchoscopy image shows a narrowed airway with signs of inflammation. There is visible mucus and some areas appear slightly congested, indicating potential edema or irritation in the mucosal lining. No neoplasm are observed; however, there may be remnants from previous infections that have since cleared up.
UniBioMed	Air densities. there are air densities in the lumen in the esophagus.
Lingshu-7B	Significant narrowing and partial occlusion of the airway lumen is observed, likely due to external pressure rather than intrinsic airway pathology. The airway walls appear smooth without visible nodules, fistulas, or masses. ....
Qwen2.5VL-7B-FT	Neoplasms were seen blocking the lumen, and the surface was covered with white matter.
InternVL-3-8B-FT	Neoplasms were completely blocked from the lumen when the opening was opened.

**Fig. 6** Caption results by different MLLM models. General MLLMs and Medical MLLMs can not recognize the lesion in the image in the test dataset. After fine-tuning on the training dataset, the lesion can be recognized with a concise description as the clinician desired.

potentially introduce hallucinations in multimodal large language models during report generation. Third, the reports were written by different clinicians, leading to subtle variations in descriptive style and interpretation.

### Data availability

The dataset is available from the Science Data Bank at <https://doi.org/10.57760/sciencedb.28018><sup>18</sup>. The name of the repository is “BERD: Fine-Grained Bronchoscopy Examination Report Dataset”.

### Code availability

The code for data processing, dataset statistical analysis, and MLLM validation can be found at <https://github.com/lxj22/BERD>.

Received: 5 August 2025; Accepted: 22 January 2026;

Published online: 03 February 2026

### References

- Fulkerson, W. Current Concepts - Fiberoptic Bronchoscopy. *N. Engl. J. Med.* **311**, 511–515, <https://doi.org/10.1056/nejm198408233110806> (1984).
- Ernst, A., Silvestri, G., Johnstone, D. & Diagnost, A. I. C. Interventional pulmonary procedures: Guidelines from the American College of. *Chest Physicians. Chest* **123**, 1693–1717, <https://doi.org/10.1378/chest.123.5.1693> (2003).
- Criner, G. *et al.* Interventional Bronchoscopy. *American Journal of Respiratory and Critical Care Medicine* **202**, 29–50, <https://doi.org/10.1164/rccm.201907-1292SO> (2020).
- Zhang, J. *et al.* AI co-pilot bronchoscope robot. *Nature Communications* **15**, <https://doi.org/10.1038/s41467-023-44385-7> (2024).
- Xiao, H. *et al.* A comprehensive survey of large language models and multimodal large models in medicine. *Information Fusion* **117**, <https://doi.org/10.1016/j.inffus.2024.102888> (2025).
- Bannur, S. *et al.* MAIRA-2: Grounded Radiology Report Generation. Preprint at <https://arxiv.org/abs/2406.04449> (2024).

7. Chen, J. *et al.* HuatuoGPT-Vision, Towards Injecting Medical Visual Knowledge into Multimodal LLMs at Scale. Preprint at <https://arxiv.org/abs/2406.19280> (2024).
8. Li, C. *et al.* LLaVA-med: training a large language-and-vision assistant for biomedicine in one day. *Advances in Neural Information Processing Systems* **36**, 28541–28564, <https://dl.acm.org/doi/10.5555/3666122.3667362> (2023).
9. Wu, C. *et al.* Towards generalist foundation model for radiology by leveraging web-scale 2D&3D medical data. *Nature Communications* **16**, <https://doi.org/10.1038/s41467-025-62385-7> (2025).
10. Wu, L. *et al.* UniBiomed: A Universal Foundation Model for Grounded Biomedical Image Interpretation. Preprint at <https://arxiv.org/abs/2504.21336> (2025).
11. Vu, V. *et al.* BM-BronchoLC - A rich bronchoscopy dataset for anatomical landmarks and lung cancer lesion recognition. *Scientific Data* **11**, <https://doi.org/10.1038/s41597-024-03145-y> (2024).
12. Hao, R. *et al.* Upper Airway Anatomical Landmark Dataset for Automated Bronchoscopy and Intubation. *Scientific Data* **12**, 1907, <https://doi.org/10.1038/s41597-025-06169-0> (2025).
13. Yan, P. *et al.* PKDN: Prior Knowledge Distillation Network for bronchoscopy diagnosis. *Computers in Biology and Medicine* **166**, <https://doi.org/10.1016/j.compbiomed.2023.107486> (2023).
14. Sun, W. *et al.* An accurate prediction for respiratory diseases using deep learning on bronchoscopy diagnosis images. *Journal of Advanced Research* **76**, 423–438, <https://doi.org/10.1016/j.jare.2024.11.023> (2025).
15. Johnson, A. *et al.* MIMIC-III, a freely accessible critical care database. *Scientific Data* **3**, <https://doi.org/10.1038/sdata.2016.35> (2016).
16. Roberts, R. PubMed Central: The GenBank of the published literature. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 381–382, <https://doi.org/10.1073/pnas.98.2.381> (2001).
17. Yang, A. *et al.* Qwen3 Technical Report. Preprint at <https://arxiv.org/abs/2505.09388> (2025).
18. LUO, X. *et al.* BERD: Fine-Grained Bronchoscopy Examination Report Dataset. *Science Data Bank* <https://doi.org/10.57760/sciencedb.28018> (2025).
19. Bai, S. *et al.* Qwen2.5-VL Technical Report. Preprint at <https://arxiv.org/abs/2502.13923> (2025).
20. Zhu, J. *et al.* InternVL3: Exploring Advanced Training and Test-Time Recipes for Open-Source Multimodal Models. Preprint at <https://arxiv.org/abs/2504.10479> (2025).
21. Xu, W. *et al.* Lingshu: A Generalist Foundation Model for Unified Multimodal Medical Understanding and Reasoning. Preprint at <https://arxiv.org/abs/2506.07044> (2025).

## Acknowledgements

This research was supported by the InnoHK initiative of the Government of the Hong Kong Special Administrative Region.

## Author contributions

Xingjian Luo: Methodology, Experiment, Writing (Draft, Graph and Visualization), Review & Revising. Xinyan Huang: Conceptualization, Supervision, Data Collection, Annotation, Review & Revising. Xusheng Liang: Experiment, Review & Revising. Jiyu Wang: Data Collection, Annotation, Review. Jincui Gu: Data Collection, Annotation, Review. Dong Yi: Conceptualization, Project Administration, Review & Revising. Haohan Zhao: Review & Revising. Haihong Zhang: Conceptualization, Project Administration. Jinlin Wu: Review & Revising. Zhen Lei: Supervision, Review & Revising. Gaofeng Meng: Supervision, Review & Revising. Hongliang Ren: Supervision, Review & Revising. Jiebo Luo: Supervision, Review & Revising. Huai Liao: Conceptualization, Review & Revising, Data Collection, Project Administration. Hongbin Liu: Conceptualization, Supervision, Review & Revising, Project Administration.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to H.L. or H.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026