CMPE 492

AI Chatbot in Human Pathology

Alperen Dağı

Simar Achmet Kechagia

Yiğit Şekerci

Advisor:

Mehmet Turan

TABLE OF CONTENTS

1. INTRODUCTION

Human pathology embodies a range of diseases and disorders, and developing new outcomes for improving diagnostic accuracy and streamlining medical care strategies is critical. Because current diagnostic techniques applied in pathology are primarily based on the subjective interpretation of pathological images, it may lead in the lack of consistency in diagnosis and treatment recommendations. The growing complexity and quantity of pathologic cases increase the need for strategies in care that would be efficient and effective.

Accordingly, we propose the AI Chatbot in Human Pathology in this project. The chatbot uses multimodal models and deep learning techniques to assist in pathology diagnosis and treatment planning, which will have impacts for patients, caregivers, and the medical community.

1.1. Broad Impact

1.1.1. Healthcare Accessibility

This would indeed lead to access to pathology diagnosis and treatment planning, more so in under-served or remote areas. The chatbot could give near-instant feedback on pathological images, thus potentially reducing the time between first consultation and treatment. This way, early detection and treatment of diseases could be improved with better patient outcomes.

1.1.2. Healthcare Efficiency

AI-enabled integration into healthcare processes could also make changes to improve the efficiency of patient care and the quality of services in medicine. Automation at the early stage of diagnosis or treatment planning will enable healthcare professionals to apply more of their expert time to patient care and more complex cases.

The chatbot can play an essential role in reducing delays in patient diagnosis. Chatbots for diagnostics offer instant preliminary diagnostics. Although they provide a low rate of false alarms, they inform patients of high disease probability, hence saving time. Thus, it could enhance the performance of healthcare. It makes it possible to serve more people within a short period, which is very important because there are many patients for whom every second of delay in these results is inadmissible.

1.1.3. Educational and Research Applications

In addition to its practical application the AI chatbot can also serve the purpose of an learning tool for medical students and professionals. Such AI systems can be beneficial for students to learn about diagnostic and treatment approaches that are new to them. The use of the chatbot should result in the production of vast amounts of data about its accuracy and effectiveness. This might drive even more future research and development in the field of pathology and AI.

1.2. Ethical Considerations

1.2.1. Accuracy and Reliability

The consequences related to the misdiagnosis or inappropriate treatment recommendation of a chatbot have the potential for huge ethical risks. The second fact becomes a reason for the importance of validating the chatbot against live medical practices and monitoring any deviations in performance.

1.2.2. Human Oversight

AI recommendations should be supervised by human judgment and empathy. Healthcare professionals should take final decisions, while the recommendations given by the chatbot should be integrated into the broader picture of patient care.

2. PROJECT DEFINITION AND PLANNING

2.1. Project Definition

The "AI Chatbot in Human Pathology" project aims to use artificial intelligence to improve the precision and efficiency of pathology diagnoses. The project consists of several components, including a literature review, data collection, dataset preparation, and fine-tuning. By fine-tuning large language models (LLMs) with pathological images, this project aims to create a chatbot to augment the capabilities of healthcare experts, therefore enhancing patient care.

2.2. Project Planning

2.2.1. Project Time and Resource Estimation

<u>2.2.1.1. Weeks 1-2: Initial Planning and Data Collection.</u> These kinds of projects need a huge amount of data to show significant results. So, we have contacted several experts. The data collection requires communication with the experts and information gathering from all the sources. Approximately 450 pathological textbooks are currently collected in the cloud. These books will be used to extract image text pairs for the dataset. The sources were collected in cloud storage so all team members could easily access them.

<u>2.2.1.2. Weeks 3-4: Literature Survey.</u> The literature review was carried out for data extraction and the selection of the models for LLMs. This overview reviews existing AI applications in figure caption extraction and medical diagnostics.

2.2.1.3. Weeks 5-7: Data Extraction and Preparation. We employ AI tools to extract figures and captions from medical textbooks and then clean and organize the data. The process may have some computationally intensive resources, so the turnaround time to generate pairs of images might be slower.

2.2.1.4. Weeks 8-10: Conversation generation. Our focus was on generating realistic and contextually relevant conversations based on the already presented figure-caption pairs. Some models we tested were GPT-4, Llama3, and Gemini 1.0 Pro. We chose Gemini 1.0 Pro, which is quite effective and has given sufficient performance in similar tasks. We created a template which will be used to create conversations, making sure they are an appropriate imitation of practical user input and medical questions. The template consisted of diagnostic questions, common follow-up inquiries, and contextual explanations. In this way, we could make a rich, highly varied data set aligned with a real-life medical setup for training our AI model.

2.2.1.5. Weeks 10-12: Fine-tuning of the model. For selecting an appropriate large multimodal model (LMM) and start the process with the prepared dataset, we replicated some of the pioneer projects on this subject. We need high-performance computing resources, pre-trained LMMs, and the development environment for example a local GPU setup or Google Cloud.

2.2.1.6. Weeks 12-14: Evaluation and Validation. Testing the chatbot's performance concerning the established metrics, expert reviews, and perspectives. We would also compare the result with existing state-of-the-art to see how well our fine-tuning process has been.

2.2.2. Success Criteria

A couple of metrics are examined to evaluate our project performance.

- Create a clean dataset from pathology books.
- High accuracy in diagnosis, with performance metrics equivalent to or better than the benchmarks set.
- Positive feedback regarding the utility and accuracy of the chatbot from the medical professionals, showing the probable value offered to professionals within a clinical setup.
- Convenient chatbot interface, easy to use, reflecting accessibility—ease of use to the general public or medical staff.

2.2.3. Risk Analysis

- This field has state-of-the-art models that use computational power and data in excess. Currently, our project is using even lower than sufficient amount of computational power and data. This might have put us behind the curve in development compared with others working in the same field. This means that it is pretty crucial that we think of getting more resources for our project work to ensure competitiveness and further progress. Lack of these resources will likely result in slow processing times and low-precision results.
- The system, even though the user knows about it, generates problems in the field of security. For future work, it should be updated with a procedure in the security area, and users should be trained to take best practices to safeguard the data.

2.2.4. Team Work

We are working on the project together, sharing responsibilities within the group. All the group members strive to complete the project on time with successful results. Each member brings their unique capability to the project, which augments the overall outcome.

3. RELATED WORK

3.1. Publicly Available Multimodal Models

3.1.1. Visual Instruction Tuning [1]:

The paper introduces LLaVA (Large Language and Vision Assistant), a large multimodal model trained through instruction tuning using language-only GPT-4 to generate multimodal language-image instruction-following dataset. This is not widely used; however, it was geared to improve zero-shot capabilities in most of the new task domains. LLaVA fuses a vision encoder with a large language model (LLM) to achieve broad visual and language understanding. Initial experiments show that LLaVA can hold its own impressive multimodal chat abilities and performs up to state-of-the-art by 85.1% compared to GPT-4 on a synthetic multimodal instruction-following dataset. This multimodal model serves as a resource for our project.

3.1.2. BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models [2]:

BLIP-2 presents a new strategy for vision-language pre-training that tackles the prohibitively expensive cost of end-to-end training large-scale vision-language models. This paper proposes BLIP-2: a more efficient and cost-effective method, with frozen pre-trained image encoders to bootstrap the vision-language pre-training in a low-resource setting. BLIP-2 leverages a lightweight querying transformer to close the modality gap, trained via a two-stage process: The first stage puts more weight on the vision-language representation learning task with a frozen image encoder; the second stage puts more weight on the vision-to-language generative learning from a frozen language model. Their use has much fewer trainable parameters compared to the existing methods and still outperforms by a significant margin on various vision language tasks. Moreover, BLIP-2 leads to state of the art in zero-shot instructed image-to-text generation and shows the potential towards being a building block in the multimodal conversational AI agent landscape.

3.2. Biomedical Domain-Specific Approaches

3.2.1. A Foundational Multimodal Vision Language AI Assistant for Human Pathology [3]:

This paper introduces PathChat: a multimodal vision-language AI assistant developed in the context of human pathology. Despite a great deal of progress on task-specific predictive models and self-supervised vision encoders, the development of such computational pathology assistants lags. PathChat is built on top of a foundational vision encoder pre-trained on a large dataset of histology images and pathology image-caption pairs, followed by fine-tuning using a pre-trained large language model, and finally fine-tuned on a diverse set of visual language instructions in the context of pathology. Comparative assessment against distinct multimodal vision language AI assistants and a commercial grade AI assistant (ChatGPT-4) revealed that under in-context clinical context, PathChat achieved high diagnostic accuracy with histology images and generally produced responses that were more accurate and preferable for different pathology-related queries. An interactive and flexible AI assistant, PathChat will provide novel values in the pathology domain for education, research, and clinical decisions, where expert pathologists are required.

3.2.2. LLaVA-Med: Training a Large Language-and-Vision Assistant for Biomedicine in One Day [4]:

This paper introduces LLaVA-Med, a vision-language conversational assistant pre-trained to respond to open-ended research questions regarding the content of biomedical images. Although preliminary studies have demonstrated that generative conversational AI could be potentially empowering for biomedical practitioners, current investigations are focused on unimodal text. This paper introduces LLaVA-Med, a vision–language-based conversational engine pre-trained for reply generations in tasks of open-domain research questions related to biomedical-image content. While some recent works have shown initial encouraging results for such initiatives, current empirical investigations primarily focus on unimodal text. For adjusting a large, general-domain, vision-language model with GPT-4 on self-instructed open-ended instruction-following data from the captions and a novel curriculum learning technique, the biomedical vocabulary is matched to use figure-caption pair and open-ended conversational semantics. On visual question answering on three standard biomedical datasets in LLaVA-Med, results show excellent multimodal conversational capability compared to previous models, which achieve state-of-the-art performance on some metrics. LLaVA-Med contributes to multimodal research within the biomedical field by releasing the instruction-following data and the LLaVA-Med model, enabling further work in this field.

These articles will assist in understanding how to train large language-and-medicine models using biomedical datasets, which will help develop our chatbot within the pathology domain.

3.3. Dataset Preparation

3.3.1. Figure and Caption Extraction from Biomedical Documents [5]:

This paper is based on the extraction of figures and captions from biomedical documents. This concept of extraction is fundamental because most of the information lies between these two parameters and, hence, very critical to the requirements of researchers and practitioners in this domain. Most PDF parsing tools face challenges in solving the real issue in the identification of figures and captions, faced due to complex graphical objects and non-trivial structures existing in scientific publications. The solution to that problem is what the authors do by coming up with a new system of figure and caption extraction.

PDFigCapX improves this by a new method of segregating the content into text and graphics and using layout information to identify the captions and figures. It then outputs a file with extractions of the figures and their corresponding caption.

The effectiveness of PDFigCapX is demonstrated through experiments conducted on a public dataset of computer science documents and two other newly collected sets of biomedical publications. Results show significant performance improvements, compared with all other published systems, and highlight the effectiveness and robustness of the proposed approach.

The system, PDFigCapX, is open to the public to use, and authors host the implementation at the URL provided. This allows better accessibility for researchers and practitioners who may want to extract figures and captions from biomedical documents properly. Software support for extracting such pathological relevant images or even their corresponding textual descriptions can help faster acquire such information from medical books or publications.

3.3.2. PDFFigures 2.0: Mining Figures from Research Papers [6]:

This paper presents, an algorithm conceived to extract figures, tables, and captions from scholarly documents, so as to harness the several hidden values contained in them for academic search engines. Though figures and tables are presented sources of keypoints in a research paper, academic search engines often ignore them while reading the document or summarizing it to its final users. PDF-Figures 2.0 fills in this gap by analyzing the structure of each page, detecting the captions and graphical elements along with body text chunks, and then identifying it figures and tables by using a heuristic based on the empty regions that the text leaves.

Having compared the performance of PDFFigures 2.0 [6] with PDFigCapX [5], we have decided to use PDFigCapX in our project since it generates more pairs. While both algorithms could extract figures, captions, and other sections from scholarly documents, PDFigCapX was noted for creating a more significant quantity of figure-caption pairs. It is close to the goals set in our project since it makes the dataset deeper and more diverse in the training of our chatbot model. As a result, this PDFigCapX is the option to be applied to support our AI Chatbot concerning Human Pathology for having a better diagnostic potential.

4. METHODOLOGY

The structure across the several main stages of AI Chatbot in Human Pathology development is designed in a manner that would support the ability to achieve an accurate and effective final diagnostic tool. The plan for each stage is described here:

i. **Literature Survey:**
   - Review relevant literature on similar projects.
   - Review published research and papers on artificial intelligence usage in medical diagnostics.
   - Identify relevant methodologies, datasets, and model architectures applied in similar projects for benchmarking and reference.

ii. **Data Extraction from Medical Textbooks:**
   - Use ideas of the literature survey in data extraction.
   - By the use of AI tools and open source projects found during the literature survey, extract the data from pathology textbooks.
   - Focus on pulling out pairs of figures and captions that carry pathological images containing relevant textual descriptions.

iii. **Dataset Preparation:**
   - Curate and post-process the extracted data into a quality dataset that can be used for training and fine-tuning the model.
   - Applying any cleansing technique that is required to get rid of noise or irrelevant information from the dataset to ensure the figure-caption pairs are accurate and relevant.
   - Organize the dataset into appropriate formats for training the model,

considering factors such as image resolution and text encoding.

iv. **Conversation Generation:**

- Based on trials with various models, including GPT-4, Llama3, and Gemini 1.0 Pro, we select Gemini 1.0 Pro for conversation generation.
- Utilize the extracted figure-caption pairs to generate realistic and contextually relevant conversations.
- Employ a carefully designed template to structure these conversations, ensuring they accurately reflect potential user interactions and medical inquiries.
- Include common diagnostic questions, follow-up inquiries, and contextual explanations typically provided by pathologists.
- Generate conversations that are rich in context and aligned with real-world medical scenarios to create a comprehensive dataset for training the AI model.

v. **Model Fine-Tuning:**

- Select large multimodal models identified by the literature survey as possible candidates to be fine-tuned.
- Initialize the chosen models with pretrained weights to capitalize on existing knowledge and optimize performance.
- Fine tune the models using dataset consisting of figure-conversation pairs.
- Use transfer-learning methods to adapt the pre-trained models to specialize in the field of pathology.
- Iterate on the fine-tuning process, adjusting hyperparameters and model architecture as necessary to optimize performance.

vi. **Evaluation and Validation:**

- Evaluate the performance of the fine-tuned models using relevant metrics.

- Compare the chatbot's performance against existing diagnostic methods and expert opinions to assess the chatbot's effectiveness and reliability.

With such methodology, it is purposed to create solid and accurate AI Chatbots in human pathology capable of analyzing pathological images and generating responses.

5. REQUIREMENTS SPECIFICATION

- User Requirements
  - Users shall be able to submit pathological images related to via the chatbot interface.
  - Users shall receive a preliminary diagnostic assessment based on the submitted images.
  - Users shall be able to receive personalized analysis based on the chatbot's assessment.
- System Requirements
  - The system shall employ a large multimodal model capable of processing and understanding pathological images and textual data.
  - The system shall be trained on a comprehensive pathology dataset, including images and associated conversations.
  - The system shall provide a user-friendly interface that is accessible via web.
  - The system shall ensure accuracy and reliability by validating against current medical practices and expert opinions.

6. DESIGN

6.1. Information Structure

The information structure in our AI Chatbot in Pathology has been designed to be intuitive and efficient, ensuring that users can navigate through the interface well enough and that the system can process requests smoothly. The architecture is hierarchical, allowing easy data organization to retrieve content associated with patient questions, diagnostic information, and treatment recommendations. Because at its core, it is the framework that ensures an efficient flow of information between the different Web UI controller and worker components. It ensures that user inputs are interpreted correctly and responded to with appropriate medical advice and information.

6.2. Information Flow

The information flow in our system is designed to allow interaction with user requests made to and from our Large Language and Vision Model (LLaVA) over the web UI. First, user queries are received from the Gradio-powered chat interface, where they are formatted and sent to the our model. The our model runs complex analyses of medical data and images to respond appropriately to these requests. Finally, the response output is sent back to the Gradio interface to display the answer to the user. This bidirectional flow of information will help make dynamic interactions such that the system can handle diagnostic inquiries with high efficiency and accuracy.

6.3. System Design

Thus, we have two main components in our system explicitly designed to take care of the operations of the chatbot:

- **Web UI Controller:** This component is important in the interaction between the users and the system by providing an interface with the ease of using Gradio. The web UI controller allows for user interaction and information retrieval without pursuing any set of intricate menu or interfaces.
- **Our LLaVA Model:** This is an underpinning model for Large Language and Vision that generates all answers to user queries. The LLaVA model processes input from the web UI controller and produces diagnostics with treatment recommendations. The balance between computational cost and performance needs guides the choice of the cloud platform.

6.4. User Interface Design

The chatbot user interface will follow a straightforward model of an interface, not so different from most of the well-known chatbots, so that users can feel comfortable with it and can access it without any extra effort. The approach presented here will balance minimalism and functionality. Prioritizing clarity and intuitiveness in design, questions and answers will be neatly presented, and users will find their way through interaction with the chatbot quite naturally.

7. IMPLEMENTATION AND TESTING

7.1. Implementation

7.1.1. Figure-Caption Extraction

We primarily surveyed a few figure-caption extraction tools like PDFFigures 2.0 [6] and PDFigCapX [5]. We decided to go ahead with the implementation using PDFigCapX, as this perform better in terms of producing figure-caption pairs. Going through the source code, we tried to run the extraction on our machines. As the original code has specific dependencies, we needed to build the corresponding dependencies of the original code. The underlying extraction process indeed succeeded, although not as smooth; for example, some outputs produced were incorrect. We addressed at this issue by post processing the data using OCR libraries to detect any text on the extracted images. Also the demand for computational power was very high. We have proceeded with the extraction despite those difficulties and managed to process all the data set.

7.1.2. Conversation Generation

From our experiments from previous sections, using GPT-4, Llama3, and Gemini 1.0 Pro, we found it best to use Gemini 1.0 Pro for generating conversations since it is more powerful and performs better in tasks of this type. We used the figure-caption pairs extracted in the previous step to generate realistic and contextually relevant conversations. We used a template for conversation design based on user interaction and medical query use cases. This contained common diagnostic questions, common follow-up questions and contextual explanations that a pathologist would give. These generated conversations, along with the respective images, were used as training data. This approach made the training data highly contextual and befitting real-world medical scenarios, helping the chatbot give correct, meaningful responses.

7.1.3. Large Language and Vision Model (LLVM) Integration

Our first approach to building the chatbot involved using state-of-the-art Large Language and Visual Models, with particular focus on BLIP2 and LLAVA due to their proven effectiveness in similar applications, as observed during our literature review. Because LLAVA was a model that played such a significant role in our reference studies [1], we chose to fine-tune LLaVA. In this case, we started working on the training stage using scripts from the respective publication's repository and, where necessary, adapting to our project needs. Since the computational power need were high for the fine-tuning, we introduced the DeepSpeed and LoRA libraries, which provided efficient memory operation and maintained higher performance. The fine-tuning was done using 4 RTX 4090 24GB GPUs for about 3.5 hours, running one epoch around 120k figure-conversation pairs.

7.2. Testing Strategy

The following testing strategies were applied in the project:

- **LLaVA Model Testing:** The architecture of our system, the LLaVA model, will be benchmarked from a general performance perspective using benchmarks from the PathVQA [7] dataset. PathVQA is a question-answering dataset based on pathology images; it is designed for research in both open-ended and binary questions. Our evaluation will compare how accurate our model is in terms of the reported values from the LLaVA-Med paper [4].

- **UI Testing:** Using Gradio for the chatbot front end is simple and practical because it is easy to develop with speed. Testing will be concentrated in the area of usability and interaction workflows, supported by the embedded potency in Gradio to ensure that it is user-friendly use.

7.3. Deployment Considerations

The deployment of the Large Language and Vision Model (LLVM) presents a complex challenge, primarily due to its computational and resource intensiveness. For this project, our model is deployed on the same machine that the training was done.

8. RESULTS

We tested our model performance on the PathVQA dataset - question/answer pairs from pathology images. The data set contains open-ended questions and binary yes/no questions. From our side, we checked the model accuracy on binary questions, testing it on more than 3,250 such questions. We compared the results of our model with those reported in the LLaVA-Med paper [4] for the LLaVA and LLaVA-Med models. Our fine-tuned LLaVA model achieved an accuracy of 60.82%, outperforming the LLaVA-Med model's 59.66% and the baseline LLaVA model's 45.65%. The results showed that our model can provide accurate and meaningful answers to a wide range of pathological questions, possibly indicating its usefulness in real-world medical applications. The use of DeepSpeed and LoRA for fine-tuning made sure that it was memory-efficient, and performance was maintained. This provided the ability to achieve these results within our available computational resources. Additionally, Gradio for the web UI helped us set an intuitive and efficient interface to interact with the model. It made access to effective and fast medical service to users.

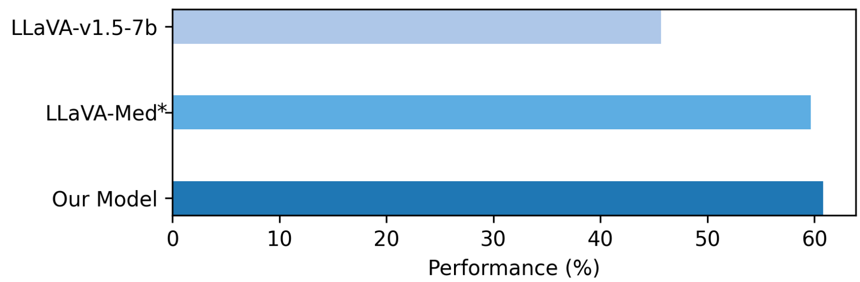


Figure 8.1. Accuracy comparison across models

*The LLaVA-Med variant was trained on 600K samples for 1 epoch (stage 1) and 60K samples for 3 epochs (stage 2).

9. CONCLUSION

The development of our AI Chatbot in Human Pathology was a step toward using artificial intelligence to medical diagnostics. This approach, coupling LLaVA with a web-based interface empowered using Gradio, has proven pretty effective for delivering accurate diagnostic information and treatment recommendations. The results of the PathVQA dataset evaluation highlight the accuracy of the model and the potential impact of the work on improving the diagnostic process in pathology. Further options for work include improvement of the training from more diverse datasets and work on the remaining inaccuracies, along with enhancing computational efficiency towards further development of this technology. Besides, if the content of the chatbot is enriched with more extensive coverage of all medical fields, it will increase its utility and purpose. Together, our project suggests feasibility and affirms the potential implementation of AI-driven chatbots in change-making pathology diagnostics; there are new perspectives for improvements in treatment quality and support of medical research.

REFERENCES

1. Liu, H., C. Li, Q. Wu and Y. J. Lee, "Visual Instruction Tuning", *arXiv e-prints*, p. arXiv:2304.08485, Apr. 2023.

2. Li, J., D. Li, S. Savarese and S. Hoi, "BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models", *arXiv e-prints*, p. arXiv:2301.12597, Jan. 2023.

3. Lu, M. Y., B. Chen, D. F. K. Williamson, R. J. Chen, K. Ikamura, G. Gerber, I. Liang, L. Phi Le, T. Ding, A. V. Parwani and F. Mahmood, "A Foundational Multimodal Vision Language AI Assistant for Human Pathology", *arXiv e-prints*, p. arXiv:2312.07814, Dec. 2023.

4. Li, C., C. Wong, S. Zhang, N. Usuyama, H. Liu, J. Yang, T. Naumann, H. Poon and J. Gao, "LLaVA-Med: Training a Large Language-and-Vision Assistant for Biomedicine in One Day", *arXiv e-prints*, p. arXiv:2306.00890, Jun. 2023.

5. Li, P., X. Jiang and H. Shatkay, "Figure and caption extraction from biomedical documents", *Bioinformatics*, Vol. 35, No. 21, pp. 4381–4388, 04 2019, `https://doi.org/10.1093/bioinformatics/btz228`.

6. Clark, C. and S. Divvala, "PDFFigures 2.0: Mining Figures from Research Papers", *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries*, JCDL '16, p. 143–152, Association for Computing Machinery, New York, NY, USA, 2016, `https://doi.org/10.1145/2910896.2910904`.

7. He, X., Y. Zhang, L. Mou, E. Xing and P. Xie, "PathVQA: 30000+ Questions for Medical Visual Question Answering", *arXiv preprint arXiv:2003.10286*, 2020.

APPENDIX A: Github Repository

https://github.com/simarahmtkhy/492_project

APPENDIX B: Conversation Generation

You are an AI assistant specialized in biomedical topics.
You are provided with a text description (Figure Caption) of a figure image from a biomedical research paper. In some cases, you may have additional text (Figure Context) that mentions the image. Unfortunately, you don't have access to the actual image.

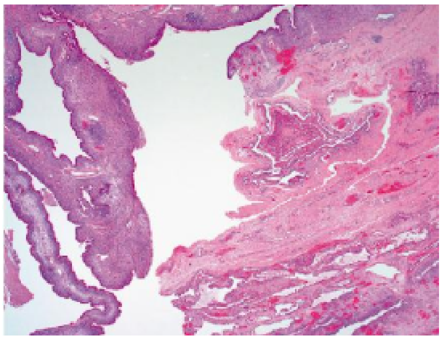
Your task is to generate a conversation between a person (User) inquiring about the image and you (Assistant) responding to their questions. The conversation should proceed as though both the User and Assistant are viewing the image, while not referring to the text information (Figure Caption and Figure Context).

Below are requirements for generating the questions and answers in the conversation:
 - Avoid quoting or referring to specific facts, terms, abbreviations, dates, numbers, or names, as these may reveal the conversation is based on the text information, rather than the image itself. Focus on the visual aspects of the image that can be inferred without the text information.
 - Do not use phrases like "mentioned", "caption", "context" in the conversation. Instead, refer to the information as being "in the image."
 - Ensure that questions are diverse and cover a range of visual aspects of the image.
 - The conversation should include at least 2-3 turns of questions and answers about the visual aspects of the image.
 - Answer responsibly, avoiding overconfidence, and do not provide medical advice or diagnostic information. Encourage the user to consult a healthcare professional for advice.

Figure B.1. Conversation generation template

Caption:
Figure 6-35: Low-power view of a type II pleuropulmonary blastoma. The spindle cell component is readily apparent.

(Image is shown only for reference)



Conversation:

Human: *What type of cells are shown in the image?*

Assistant: *The image shows a low-power view of a type II pleuropulmonary blastoma. The spindle cell component is readily apparent.*

Human: *What is the significance of the spindle cell component?*

Assistant: *The spindle cell component is a characteristic feature of type II pleuropulmonary blastoma, a rare and aggressive type of lung cancer. Spindle cells are elongated, fibroblast-like cells that are often found in mesenchymal tissues, such as muscle, connective tissue, and blood vessels. In the context of pleuropulmonary blastoma, the presence of spindle cells suggests that the tumor may have a more aggressive behavior and a poorer prognosis.*

Human: *Can you describe the other components of the tumor?*

Assistant: *In addition to the spindle cell component, the tumor may also contain other components, such as epithelial cells, blastema, and cartilage.*

Figure B.2. Example figure conversation pair