Technische Universität München – Faculty of Informatics
Chair for Computer Aided Medical Procedures (Prof. Nassir Navab)
**Practical Course: Machine Learning in Medical Imaging**

# Explaining Medical Image Classifiers with Visual Question Answering Models

## 1. General Info

Contact Person: Matthias Keicher

Contact Email: matthias.keicher@tum.de

Outcome: The result of the project will potentially be published in MIDL 2023 or at a similar venue.

## 2. Project Abstract

Flamingo[1] is a new family of Visual Language Models (VLM) that are designed for few-shot learning. It provides key architectural innovations to (i) bridge powerful pretrained vision-only and language-only models, (ii) handle sequences of arbitrarily interleaved visual and textual data, and (iii) seamlessly ingest images or videos as inputs. In this project, we propose to adopt the Flamingo architecture[2] for the diagnosis of medical images. In addition to giving a diagnosis, the visual question answering (VQA) capabilities of the model will be leveraged to provide a textual explanation for the radiological findings. The results of our evaluation on the MIMIC-CXR[3] and an in-house vertebral body fracture dataset[4] will help to determine whether Flamingo's few-shot learning capabilities can leverage domain-specific pretrained models and provide meaningful explanations in a medical setting.

## 3. Background and Motivation

Deep learning models have demonstrated promising potential in diagnosing pathologies on Chest X-rays and other medical images. However, the black-box nature of deep learning methods raises concerns regarding their reliability. For their adoption in the clinical routine, it is required to know what patterns the models rely on for the diagnosis. Another limitation in applying deep learning models for medical applications is the limited availability of annotated data. Therefore, few-shot models that provide explanations for their decision-making and achieve good performance with a few annotated samples are highly desired. In our previous work[2,] we used contrastive language image pretraining (CLIP) to improve the few-shot performance. However, the approach lacked explainability and the pretraining suffered from overfitting on limited training data. In contrast, Flamingo can incorporate independently trained image and text encoders, and its VQA capabilities can give textual explanations of its decision process.

## 4. Technical Prerequisites

- Background in deep learning and ideally NLP models
- Experienced in PyTorch

## 5. Benefits:

- Working on a state-of-the-art deep learning explanation approach
- Working on the largest multimodal medical datasets available (MIMIC-CXR)
- Scientific contribution towards reliable and explainable deep learning models for medical applications

[1] https://www.deepmind.com/blog/tackling-multiple-tasks-with-a-single-visual-language-model
[2] https://github.com/lucidrains/flamingo-pytorch
[3] Keicher, M., Mullakaeva, K., Czempiel, T., Mach, K., Khakzar, A., & Navab, N. (2022). Few-shot Structured Radiology Report Generation Using Natural Language Prompts. *arXiv preprint arXiv:2203.15723*.
[4] Engstler, P., Keicher, M., Schinz, D., Mach, K., Gersing, A. S., Foreman, S. C., ... & Navab, N. (2022). Interpretable Vertebral Fracture Diagnosis. *arXiv preprint arXiv:2203.16273*.

## 6. Work-packages and Time-plan:

| | Description | #Students | From | To |
|---|---|---|---|---|
| **WP1** | Group 1: Understanding chest X-ray models' literature and modeling image classification as a multiple-choice VQA task | Group 1 | | 01.06 |
| **WP2** | Group 2: Understanding the background of using VQA models for a textual explanation of classifiers | Group 2 | | 01.06 |
| **WP3** | Group 1: Modeling and implementing the classification problem on the MIMIC-CXR as a VQA task in Flamingo | Group 1 | 01.06 | intermediate presentation |
| **WP4** | Group 2: Implement explanation baselines, evaluate VQA explanations with clinicians, visualize attentions | Group 2 | 01.06 | intermediate presentation |
| **M1** | Intermediate Presentation II | all | | |
| **WP5** | Group 1&2: Explore and analyze the results | all | | |
| **WP7** | (Optional) Evaluate results on a second small data regime task: vertebral body fracture diagnosis | all | | |
| **WP8** | Documentation | all | | |
| **M2** | Final Presentation | all | | |