Multimodal Representation Learning for Medical Applications

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

In this project, we compare different multimodal fusion methods for decision support systems. We collect publicly available multimodal medical datasets and sytematically compare different multimodal approaches including unimodal vs. multimodal, early vs. late fusion, concatenation vs. MLP vs. attention-based features fusion, and joint embeddings (e.g., contrastive) methods. The modalities used may include radiological images, text, patient records (tabular data), laboratory test results, or genomic data. The suggested datasets include Alzheimer's Disease, COVID-19, Skin Lesions, Vertebral Fracture, and Chest X-Ray classification. Finally, novel feature fusion methods like attention-based (transformer) and optionally competing methods like contrastive learning approaches are explored. The main goal is to implement a novel transformer-based multimodal framework that can visualize the attention on which modality was used for a specific patient prediction.

3. Background and Motivation

Multimodal data are becoming increasingly available in the medical domain from sources such as radiological images, text, patient records, and laboratory test results. However, most machine learning methods in healthcare focus on single-modality data. This becomes particularly apparent within the field of radiology, which, due to its information density, accessibility, and computational interpretability, constitutes a central pillar in the healthcare data landscape and has traditionally been one of the key target areas of medically-focused machine learning. Computer-assisted diagnostic systems of the future should be capable of simultaneously processing multimodal data, thereby mimicking physicians, who also consider a multitude of resources when treating patients.

[1] Stahlschmidt, Sören Richard, Benjamin Ulfenborg, and Jane Synnergren. "Multimodal deep learning for biomedical data fusion: a review." *Briefings in Bioinformatics* 23.2 (2022): bbab569.
https://academic.oup.com/bib/article/23/2/bbab569/6516346?login=true
[2] Liang, Paul Pu, et al. "Multibench: Multiscale benchmarks for multimodal representation learning." *arXiv preprint arXiv:2107.07502* (2021). https://arxiv.org/abs/2107.07502
[3] https://multimodal-toolkit.readthedocs.io/en/latest/index.html
[4] Heiliger, Lars, et al. "Beyond Medical Imaging-A Review of Multimodal Deep Learning in Radiology." (2022). https://www.zora.uzh.ch/id/eprint/219067/
[5] Pölsterl, Sebastian, Christina Aigner, and Christian Wachinger. "Scalable, Axiomatic Explanations of Deep Alzheimer's Diagnosis from Heterogeneous Data." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2021. https://arxiv.org/abs/2107.05997
[6] Xu, Peng, Xiatian Zhu, and David A. Clifton. "Multimodal Learning with Transformers: A Survey." *arXiv preprint arXiv:2206.06488* (2022). https://arxiv.org/abs/2206.06488
[7] Zheng, Hanci, et al. "Multi-transSP: Multimodal Transformer for Survival Prediction of Nasopharyngeal Carcinoma Patients." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham, 2022. https://link.springer.com/chapter/10.1007/978-3-031-16449-1_23

4. Technical Prerequisites

- Background in deep learning
- Experienced in PyTorch

5. Benefits:

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

6. Work-packages and Time-plan:

	Description	#Students	until
WP1	Group 1: Understanding medical multimodal deep learning literature and state-of-the-art methods for feature fusion and implementing baselines for encoders and fusion methods	Group 1	intermediate presentation
WP2	Group 2: Understanding medical multimodal deep learning literature and publicly available multimodal datasets and implementing data loaders	Group 2	intermediate presentation
WP3	Group 1: Exploring novel attention-based feature fusion methods	Group 1	final presentation
WP4	Group 2: Exploring alternatives e.g. contrastive learning based feature fusion methods	Group 2	final presentation
M1	Intermediate Presentation II	all	
WP5	Group 1&2: Explore and analyze the results	all	final presentation
WP6	Documentation	all	final presentation
M2	Final Presentation	all	