Invitation to the Oral Examination – Department CS

For the occasion of his examination for a Doctoral Degree,

Mr. Matthias Fabian Keicher

will present his dissertation entitled

Multimodal Deep Learning for Holistic Clinical Decision and Reasoning Support

on **16 December 2024** at **15:00 h**

Attendance to the presentation is open to the public. The presentation will be in English.

The candidate, all members of the Examination Committee, and authorized examiners of the TUM School of CIT are invited to the presentation and subsequent oral examination.

The presentation and subsequent examination will take place online via Zoom:

Link:

https://tum-conf.zoom-x.de/j/67814868754?pwd=4jC10y2oh12lPuOA6iDZCA5YMD6kbT.1

Meeting ID:

678 1486 8754

Passcode:

578904

as well as in presence in:

Room 22.1.27; TranslaTUM; Einsteinstraße 25, 81675 Munich.

Examination committee:

Chair: **Prof. Daniel Rückert; TUM**

First Examiner: **Prof. Nassir Navab; TUM**

Second Examiner: **Prof. Andreas Maier; Fraunhofer Institute**

Third Examiner: **Prof. Pranav Rajpurkar; Harvard Medical School**

Munich/Garching, the **14th** of **November 2024**

Mailing list:

Members of the examination committee

Doctoral candidate

Abstract:

In clinical decision-making, medical doctors rely not only on a multitude of information about a patient, including patient history, vital signs, blood markers, and imaging data, but also on their extensive knowledge gained through formal education and experience with previously treated patients. To effectively support this complex process, clinical decision support systems should be able to integrate these different data modalities and incorporate relevant medical knowledge to make accurate diagnostic predictions. Furthermore, if these systems could provide insights into their reasoning process, they would not only support healthcare professionals more effectively in reasoning but also help build trust in the system's outputs, detect flaws in its reasoning process, and thereby enable continuous improvement.

This thesis explores clinical decision support systems based on deep learning that integrate multimodal knowledge about a patient with formal and exemplar clinical knowledge while providing insight into their reasoning.

To learn from and adapt to the clinical decision-making process of medical doctors, we first detail the clinical reasoning process from a medical education and cognitive psychology perspective. We then give an overview of multimodal deep learning and how the heterogeneous modalities involved in the clinical decision-making process can be integrated into such models. The second part demonstrates how the exemplar knowledge about previously treated patients can be modeled with a population graph. Using multimodal patient data, we model the inter-patient relationships and the underlying patient characteristics. We leverage this advanced approach for toxin prediction at a poison control center, demonstrating that the system achieves superior performance compared to clinicians by additional fusion with textbook knowledge about symptoms. Next, we extend this approach with features extracted from computer tomography images and propose a novel method for multimodal population graph construction. We apply this to the outcome prediction of COVID-19 and show that the attention on relevant patients in the graph can be interpreted as mimicking the memory-retrieval reasoning of clinicians. In the third part, we explore how self-supervised pretraining on large amounts of unlabelled data can be used to extract structured knowledge from images and subsequently be used for transparent reasoning. In the first work, we predict structured radiology reporting elements for chest X-rays using only few annotated samples. We then introduce a novel zero-shot method, where instead of training, we make use of prior knowledge about disease manifestations and use this for transparent reasoning. Finally, we investigate extracting fine-grained semantic concepts from the neural activations of a deep learning model only trained on detecting vertebral fractures, assessing their radiological meaningfulness for potential utility in decision support.

We conclude by providing an outlook on how the findings in this thesis could impact the rapidly growing field of multimodal large language models.