FAIR Development of Data-integrated AI to Detect Breathing Motion in Dynamic Lung MRI

J. Grolig[1,2], A. Koeppe[1,3], M. Selzer[3,4], S. Triphan[2], M.O. Wielpütz[2], B. Nestler[1,4]

[1]Institute for Applied Materials – Microstructure Modelling and Simulation (IAM-MMS), Karlsruhe Institute of Technology (KIT), Straße am Forum 7, 76131 Karlsruhe, Germany
[2]Department of Diagnostic and Interventional Radiology (DIR), Heidelberg University Hospital (UKHD), Im Neuenheimer Feld 130.3, 69120 Heidelberg, Germany
[3]Institute of Nanotechnology – MicroStructure Simulation (INT-MSS), Karlsruhe Institute of Technology (KIT), Straße am Forum 7, 76131 Karlsruhe, Germany
[4]Institute for Digital Materials Science (IDM), Hochschule Karlsruhe - University of Applied Sciences (HKA), Moltkestrasse 30, 76133 Karlsruhe, Germany

Introduction

Enabling Data-Driven Modeling

All research generates data, extracts knowledge, and develops models within scientific workflows. Manual knowledge extraction and execution are often implicitly used for conventional static datasets but becomes unfeasible for vast, dynamically changing datasets. However, the data-driven modeling paradigm necessitates efficient interfaces between data and models through research data management.

Aims for Artificial Intelligence (AI) in Medicine

Effective research data management, efficient interfaces and AI enable to
- Automatically extract knowledge from data (unsupervised learning),
- Train generic models to predict directly from data (supervised learning), and
- Control scientific workflows based on data (active and reinforcement learning).

The Kadi Ecosystem for Integrated AI

Kadi4Mat – Open-source platform for FAIR research data management [1]

KadiStudio – Electronic Lab notebook and scientific workflow engines [2]

KadiAI – Interface between Kadi4Mat and machine learning tools [3]

cids tools – Computational Intelligence and Data Science framework [3]

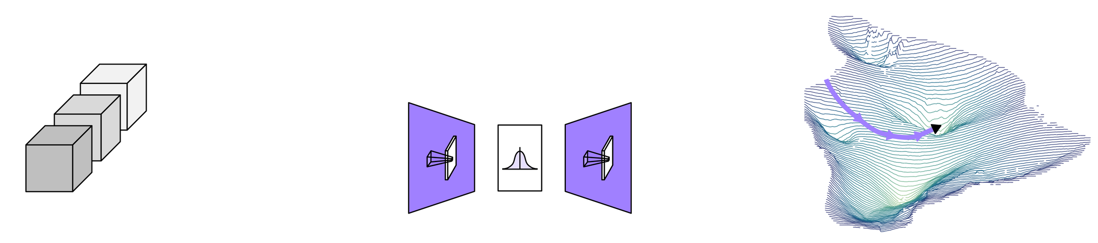
AI Models and Learning

Data-Integrated AI

Efficient **interfaces between data and AI models** are necessary to enable data-driven modeling. Data-integrated AI models directly connect with the research data management solution to **extract knowledge, learn, and control scientific workflows** based on data.

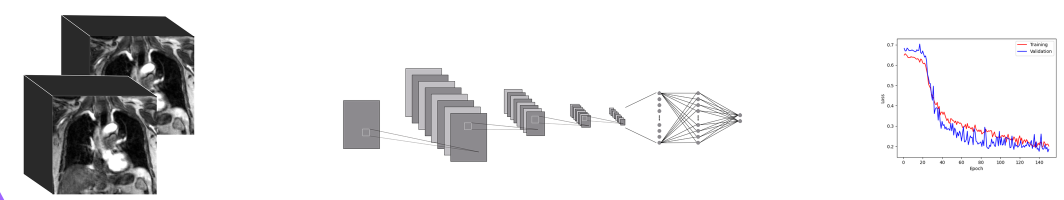
Concept Level

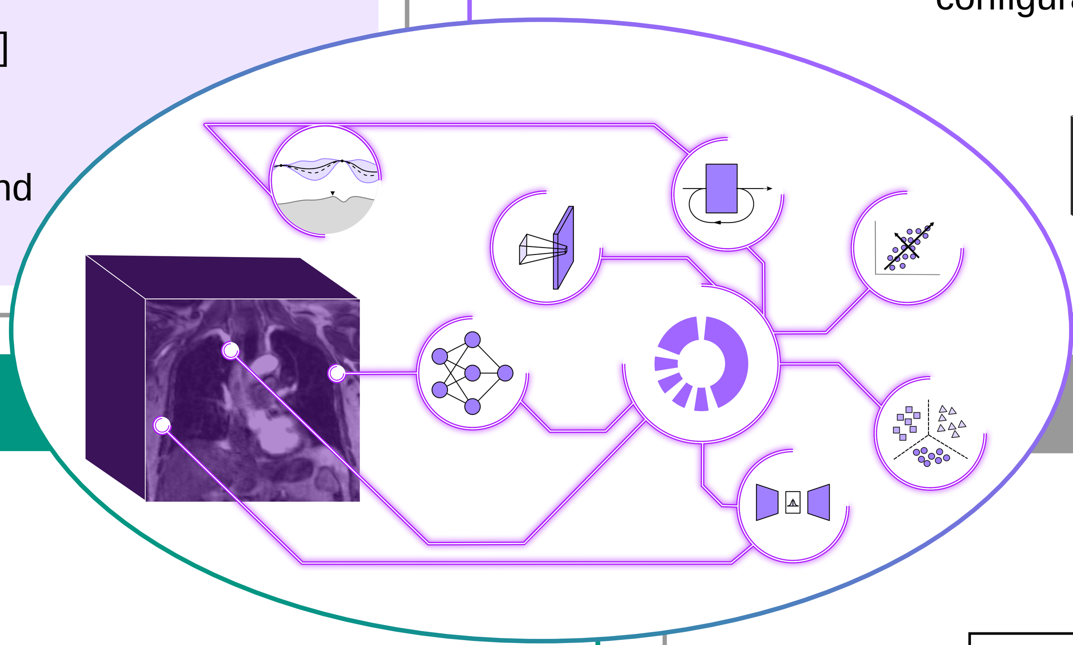
The meta information for data, models, and learning algorithms defines the boundaries in which an AI solution is investigated.



Instance Level

The individual entities of data, model, and training with fixed configurations. Leverages existing ML libraries.



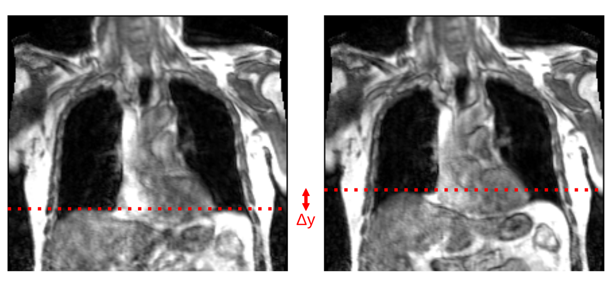


Use Case

Goal – Automatically detect breathing motion in dynamic contrast-enhanced MRIs (DCE-MRI) of the lung.

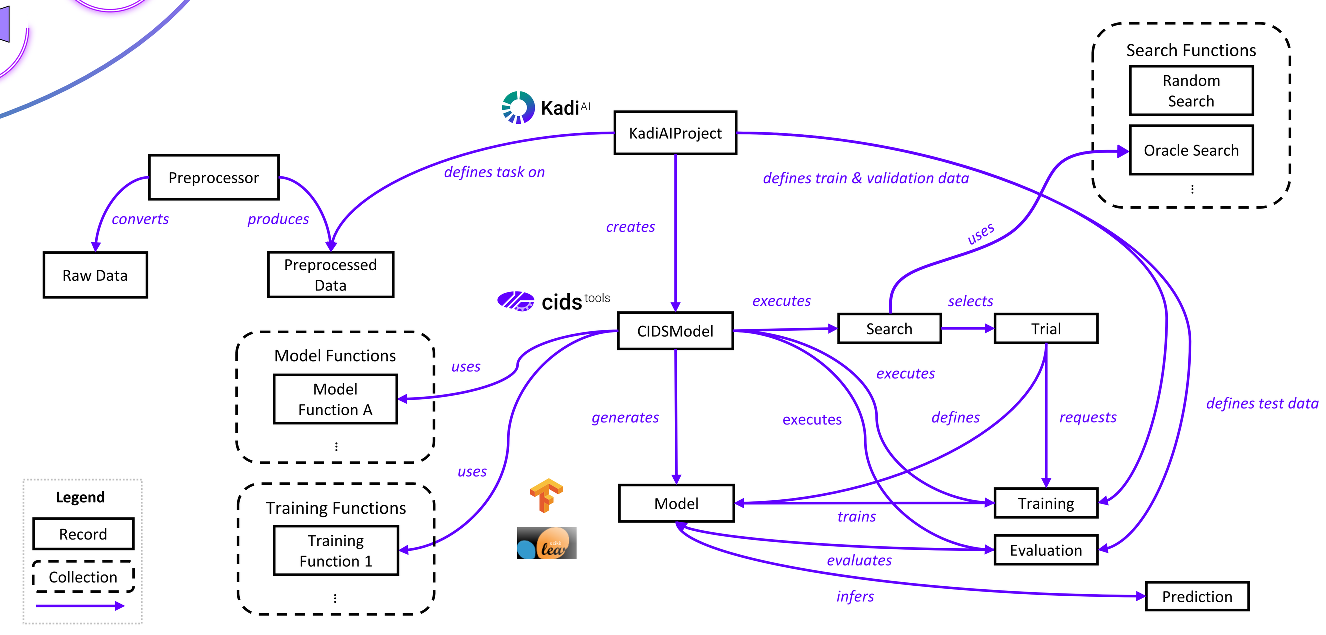
Study Data – 75 DCE-MRIs from patients with chronic obstructive pulmonary disease (COPD) of the multicenter study COSYCONET [4] resulting in 1615 pairs of adjacent image frames with manual labeled ground-truth.

Methods – Two adjacent frames are used as inputs for training three model architectures: a regular CNN, a two-branch CNN, and a hybrid model consisting of a CNN followed by a LSTM network. After training the three models the whole project is pushed to Kadi4Mat by creating records for all components (e.g. model, search, ...) and linking these single records using the KadiAI ontology.



Two adjacent frames of a DCE-MRI with apparent breathing motion.

KadiAI Concept/Ontology

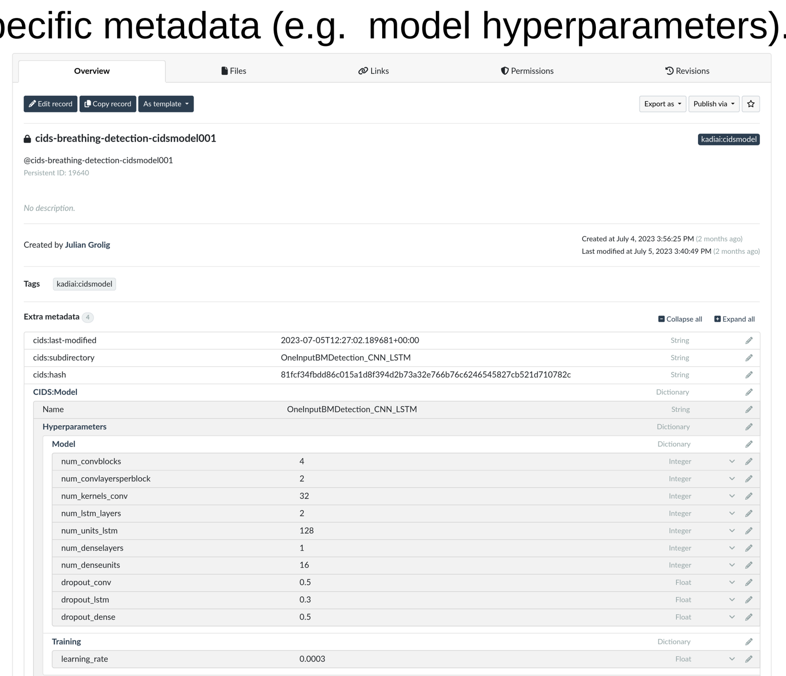


The concept for an AI work package implemented with KadiAI and CIDS

Research Data Management and Data Provenance

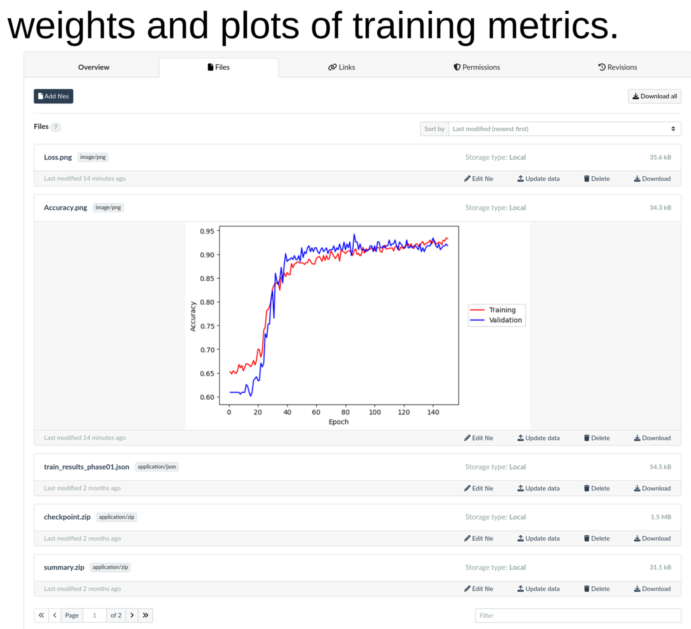
1 Metadata

The records in Kadi4Mat contain general metadata (e.g. creation date, hash values, ...) and object specific metadata (e.g. model hyperparameters).



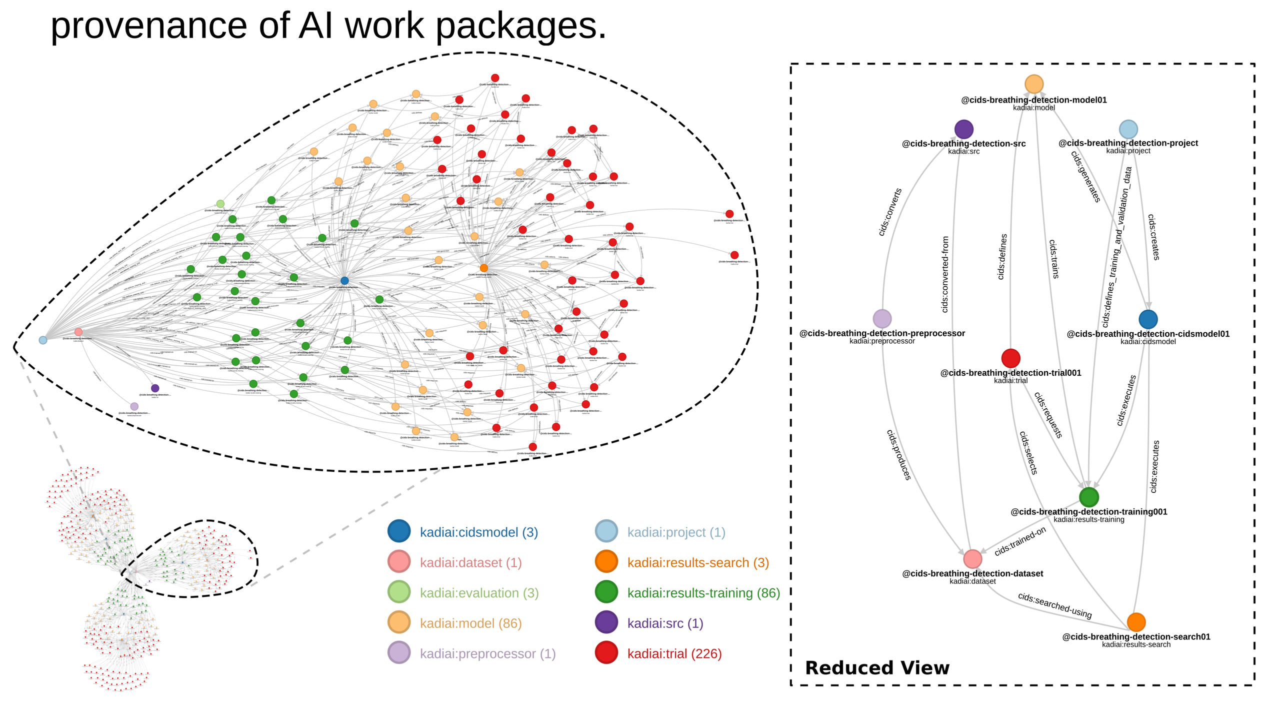
2 Files & Plots

All files of the AI project are added to their respective records. The training record shown below contains files with the model weights and plots of training metrics.



3 Data Perspective: KadiAI Bottom-Level Ontology

Kadi4Mat helps the experts to track scientific progress through records and links. KadiAI automatically creates this knowledge graph for data provenance of AI work packages.



Acknowledgement and references.

[1] N. Brandt et al., 'Kadi4Mat: A Research Data Infrastructure for Materials Science', Data Science Journal, vol. 20, no. 1, Art. no. 1, Feb. 2021, doi: 10.5334/dsj-2021-008.
[2] L. Griem et al., 'KadiStudio: FAIR Modelling of Scientific Research Processes', Data Science Journal, vol. 21, no. 1, Art. no. 1, Sep. 2022, doi: 10.5334/dsj-2022-016.
[3] A. Koeppe and The CIDS Team, 'cids: 3.1'. Zenodo, Jan. 11, 2023. doi: 10.5281/zenodo.7524476.
[4] A. Karch et al., 'The German COPD cohort COSYCONET: Aims, methods and descriptive analysis of the study population at baseline," Respiratory Medicine, vol. 114, 2016, doi: https://doi.org/10.1016/j.rmed.2016.03.008.

A special thanks to Jana Holland-Cunz for the design consultation.

This work is funded by the Ministry of Science, Research and the Arts Baden-Württemberg (MWK-BW), in the project MoMaF--Science Data Center (grant number: 34-7547.222).
This work is funded by the BMBF and MWK-BW as part of the Excellence Strategy of the German Federal and State Governments in the project Kadi4X.
This work was [partly] carried out with the support of the Karlsruhe Nano Micro Facility (KNMFi, www.knmf.kit.edu), a Helmholtz Research Infrastructure at Karlsruhe Institute of Technology (KIT, www.kit.edu).
This work was [partly] carried out with the support of the German Center for Lung Research (DZL).

julian.grolig@kit.edu
www.iam.kit.edu/mms