DOI 10.5162/SMSI2023/A3.1

Making AI Measureable - Approaches within the “Metrology for Artificial Intelligence in Medicine Programme” of PTB

Hans Rabus[1]
[1] Physikalisch-Technische Bundesanstalt (PTB), Abbestrasse 2-12, 10587 Berlin, Germany
hans.rabus@ptb.de

Summary:
PTB's program "Metrology for Artificial Intelligence in Medicine" was launched in 2021 as part of the QI Digital initiative, which aims at the digital transformation of the quality infrastructure system. The program comprises thirteen projects addressing the key quality aspects of artificial intelligence, - explainability, robustness, uncertainty - in medical applications, with a focus on generic methods that can be transferred to other areas of metrology.

Keywords: Explainability, robustness, uncertainty, data quality

Background and Motivation

Artificial intelligence (AI) methods are increasingly being used in medicine with the main application Being in imaging [1,2]. Currently, AI methods are also increasingly used in other areas such as intensive care [3–5], radiotherapy [6,7], and laboratory medicine [8].

There is currently no legal framework for the certification of AI-enabled medical devices. Standards are being developed to underpin the forthcoming EU law on AI. However, the development of standards addressing concrete test specifications and criteria for conformity assessment according to ISO/IEC 17067 had to be put on hold. The reason for this is the lack of a sound and generally accepted scientific basis for the criteria and procedures to be used.

QI Digital and M4AIM

The German QI Digital initiative was launched in 2019 to pioneer the digital transformation of quality infrastructure in Europe. Funded by the German government for a period of four years starting 2021, QI Digital is now being implemented as a joint project between the key players in the German quality infrastructure to demonstrate the approach through several use cases.

Use case 3 of QI Digital concerns artificial intelligence in medical engineering and is addressed by PTB’s research programme "Metrology for AI in Medicine (M4AIM)" [9]. The program aims to build competence in AI for metrology and to provide a metrological basis for the evaluation of quantifiable performance criteria of AI methods as a prerequisite for their certification.

The program focuses on quantitative measures and criteria for the explainability and robustness of AI algorithms and the uncertainty of their predictions. As an essential prerequisite, the program also includes the development of evaluated reference datasets.

The Projects

Thirteen PhD students and post-doctoral fellows are working on research projects that address the key performance aspects of AI algorithms mentioned above. Some projects were delayed in starting due to difficulties in recruiting staff, and two projects are currently suspended due to parental leave of the investigators.

All projects address specific topics of medical relevance with the aim of developing generic methods that can also be used in other applications of AI in metrology.

One of the more fundamental projects is investigating the potential of active learning to overcome the need for large datasets in supervised machine learning by creating optimal datasets for training neural networks.

Another project addresses methods of explainable or interpretable artificial intelligence (XAI), for which theoretical verification and empirical validation have been lacking. Using a toolkit of transparently manipulated ground truth data, the reliability of existing XAI methods is assessed and quantitative metrics for explanatory performance are explored.

Two projects are looking at AI applications in critical care. One evaluates causal machine learning approaches for analysing heterogeneous datasets consisting of asynchronous time series.

The other project is exploring ways to generate realistic synthetic reference datasets for training and testing AI algorithms to circumvent the problem of patient data privacy.

Three of the projects are related to medical imaging using CT or MRI techniques. One project is investigating the suitability of deep neural networks for image optimization in CT imaging, with a focus on developing test criteria for robustness evaluation. The other two projects are investigating the benefits of physical learning for the robustness of reconstructing images from noisy (low field) or insufficiently sampled MRI measurements and the associated uncertainties.

Another three projects investigate the use of AI-based approaches for dose calculation in radiation therapy of cancer. One of the projects investigates AI-based methods in adaptive radiotherapy, including faster analysis of measurements for quality assurance and direct dose calculation. The two other projects investigate generic methods to accelerate the generation of synthetic reference by simulations based on physical models and the uncertainties associated with detailed simulation results and the derived synthetic data.

Of the remaining three projects, one is investigating the potential of invertible neural networks for dealing with measurement error, model error, hyperparameters, and multimodal posterior distributions, and for applying the network to hemodynamic problems. Another is investigating a normative modeling approach to the problem of data heterogeneity, such as in clinical databases for mental illness. The third project is investigating uncertainties in simultaneous quantitative measurement of metabolites and machine learning analysis of potential biomarkers for early diagnosis of Parkinson's disease.

Conclusions

A review day was held in October 2022, one year after the program's launch, to report on the progress of ongoing projects to a panel of external reviewers. Feedback was generally positive and helped to establish further cross-connections between several projects. For one of the projects that had stalled due to insurmountable technical difficulties, a decision was made to realign the scope of the project.

In addition to the research activities of junior scientists, some of the major investors are involved in standardization committees such as the joint AI for Health (AI4H) focus group of ITU and WHO and ISO/IEC JC1/SC42. The development of metrology services for AI quality assurance is also being explored..

Acknowledgment

The programme “Metrology for Artificial Intelligence in Medicine” (M4AIM) is funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) in the frame of the “QI-Digital” initiative.

References

[1] Z. Fang et al., Deep Learning for Fast and Spatially Constrained Tissue Quantification from Highly Accelerated Data in Magnetic Resonance Fingerprinting, *IEEE Transactions on Medical Imaging* 38 23642374 (2019); doi: 10.1109/TMI.2019.2899328

[2] L. Vandewinckele et al. Overview of artificial intelligence-based applications in radiotherapy: Recommendations for implementation and quality assurance. *Radiotherapy and Oncology* 153, 55–66 (2020); doi: 10.1016/j.radonc.2020.09.008

[3] A. Meyer et al., Machine learning for real-time prediction of complications in critical care: a retrospective study, *The Lancet Respiratory Medicine* 6, 905–914 (2018); doi: 10.1016/S2213-2600(18)30300-X

[4] N. Tomašev et al., A clinically applicable approach to continuous prediction of future acute kidney injury, *Nature* 572, 116–119 (2019); doi: 10.1038/s41586-019-1390-1

[5] G. Lichtner et al., Predicting lethal courses in critically ill COVID-19 patients using a machine learning model trained on patients with non-COVID-19 viral pneumonia *Scientific Reports* 11, 13205 (2021); doi: 10.1038/s41598-021-92475-7

[6] H. Arabi, H. Zaidi, Applications of artificial intelligence and deep learning in molecular imaging and radiotherapy, *European Journal of Hybrid Imaging* 4, 17 (2020); doi: 10.1186/s41824-020-00086-8

[7] D. Sarrut et al., Artificial Intelligence for Monte Carlo Simulation in Medical Physics, *Frontiers in Physics* 9, 738112 (2021); doi: 10.3389/fphy.2021.738112

[8] J.-P. Trezzi et al., Metabolic profiling of body fluids and multivariate data analysis, *MethodsX* 4, 95–103 (2017); doi: 10.1016/j.mex.2017.02.004

[9] https://www.m4aim.ptb.de/