HEIDELBERG
UNIVERSITY
HOSPITAL

Publication Series:
Heidelberg Spring Symposium Medical Informatics

INSTITUT FÜR
MEDIZINISCHE
INFORMATIK

Petra Knaup, Martin Dugas, Matthias Ganzinger (eds.)

3rd Heidelberg Spring Symposium Medical Informatics

18. May 2025

Abstract Volume

This publication is the third volume of a series of publications by the Institute of Medical Informatics, Heidelberg University Hospital (Germany). The series contains the abstract volumes of the annual Heidelberg Spring Symposium Medical Informatics, which has been taking place since 2023.

Acknowledgements
The editors would like to thank all the staff of the Institute who contributed to the success of the symposium with their commitment. Organizing a symposium is always a team effort, but we would especially like to thank Karin Schmid and Veronica Wilhelm for their excellent organizational support.

Cite as:
Knaup P, Dugas M, Ganzinger M (2025): 3rd Heidelberg Spring Symposium Medical Informatics. Heidelberg University Library. https://doi.org/10.11588/heidok.00036594

Contact:
Institute of Medical Informatics (Teamassistenz-Postfach.IMI@med.uni-heidelberg.de)

Published by Heidelberg University Library, 2025

The electronic version of this work is permanently available on:
https://archiv.ub.uni-heidelberg.de/volltextserver/

doi: https://doi.org/10.11588/heidok.00036594

Table of Contents

Extended Talks

Lightning Talks

Poster

Scientific Committee

Extended Talks

AI-enabled medical information systems: from scientific ideas to patient care

Hans-Ulrich Prokosch[a]

[a]*Friedrich-Alexander-Universität Erlangen-Nürnberg, Institute for Medical Informatics, Biometrics and Epidemiology, Medical Informatics*

1. Introduction

Healthcare systems worldwide face significant challenges, such as e.g. labor shortages, clinician burnout, affordability, operational inefficiencies, quality of care, and concerns about patient safety. [1] On the other side Artificial intelligence (AI) in healthcare promises transformative advancements, from enhancing diagnostics to optimizing personalized treatments [2] and many researchers believe that AI, particularly generative AI, has the potential to address the before mentioned challenges. [1] Thus, tremendous amounts of research focus on the development and improvement of AI models by using continuously emerging new and innovative methodologies. Numerous publications based on such research are published concluding that AI "holds the promise that", "has the potential to" optimize patient care, or even "may revolutionize" our healthcare systems. Nevertheless, the actual use of these technologies, even in the most promising field of diagnostic radiology, is still not widespread, [3] and it has been recently illustrated that AI adoption, effectiveness, and barriers to implementation are still not well understood. [1]

2. Reality Check

Let´s face reality.
How many of you are medical informatics, biometrics or data science researchers and pursuing research to develop new deep learning based algorithms for diagnostic or other clinical decision support? How many of you are working on the application and adaption of large language models to healthcare scenarios, e.g. for improving clinical documentation, discharge letter writing, streamlining quality control in hospital procedures or simplify billing encoding as well as other administrative hospital processes?
How many of you however are really considering how to overcome e.g. the interoperability and regulatory hurdles in order to get such algorithms integrated into real life clinical processes and an existing hospital IT infrastructure, have it certified as medical product, get clinical users trained in understanding the potential, but also limitations of your system? How many of you have thought about how to continuously monitor the quality of your model over time?
This and many more considerations are required to achieve AI-enabled trustworthy medical information systems and to move from interesting and innovative scientific ideas to real patient care improvements and finally enhancements in patient outcome. Unfortunately, only few researchers are willing to tackle those hurdles and to go ahead for the last mile, which is required to finally bring an AI model into the real life of care processes.

3. The "AI may …" Paradigm

In the last years and with the exponential increase of AI model developments for healthcare media are full of overstatements concerning the capabilities of such algorithms. I would like to illustrate this by just two small randomly selected examples. We should not blame however the writers of the subsequently cited articles and publications, since internet sources and newsletters are full of similar examples. We should however realize the danger behind, what I call the "AI may …" paradigm.

3.1. Example 1

In the category "Clinical News / Womens Imaging" of the internet site AuntMinnie.com Amerigo Allegretto on January 14th 2025 cited an article from *Academic Radiology* with the headline "CEM-, radiomics-based models **predict** breast biopsy outcomes" [4]. A little bit further down in the article this was changed to "Machine-learning models combining features from radiomics and contrast-enhanced mammography (CEM) **can predict** breast biopsy outcomes in high-risk women." [4] In AMIA´s Informatics SmartBrief from January 14th the same website was cited with the correct statement, but less fancy „AI, radiomics models **may predict** breast biopsy outcomes." [5]
In the original article from Liu et al. the authors of the publication describe „The team included 201 women in the development and training cohort and another 86 women in an internal test set. … Machine learning models combining radiomics features and clinical descriptors on CEM can predict breast biopsy outcomes on women with BI-RADS 4A/4B/4C or 5 lesions." [6] To validate these results the study authors called for larger, multi-center datasets and told that the team is working on expanding the study to verify their findings in datasets from external centers. [4]
Those three different citations of the same research nicely illustrate that one should not trust only the "eye-catching" headlines of any publication, but always read all the details, facts and original scientific statements in the underlying scientific publication.

3.2. Example 2

Just a few weeks ago, another article on the HPCLive Network about early detection of nonalcoholic steatotic liver disease and the prediction of cardiovascular mortality in metabolic dysfunction-associated steatotic liver disease (MASLD) patients catched my eyes. [7] It referred to a presentation given at the Digestive Disease Week in May 2025 and cited the abstract of this conference presentation [8] with the headline „Explainable AI Model Bests FIB-4 for Detecting MASLD Fibrosis, Predicting Outcomes". In this article one could read about the **potential utility** of the new AI model (called FibroX) and that it **outperformed** FIB-4 (the current assessment tool: fibrosis-4 index) **showing superior accuracy and interpretability** compared to FIB-4. The article even mentions, that "FibroX **could prevent 16.5 million** unnecessary VCTEs and **save $3.3 billion in US healthcare costs**." [7]
When searching for the original abstract however, I could not find the abstract online. Rather, I found an earlier publication describing research of the same team, which was probably pursued already in 2022/2023 (it was submitted in July 2023 and published in April 2024). This article presents an explainable machine learning approach (XGBoost model) to develop a clinically applicable model that **outperforms** commonly used clinical risk indices and that has **shown the ability** of machine learning to detect high-risk metabolic dysfunction-associated steatohepatitis in a more comprehensive and flexible manner. [9] Even though, this 2 year old results have shown the potential of the developed AI model, I could not find any publication about a subsequent clinical implementation trial. This was probably related to some limitations, which were also mentioned in the publication, e.g. „although the optimized XGBoost model was validated using a test (holdout) set, it would need to be further prospectively externally validated before widespread adoption." [9]

While searching for this article however, I additionally found several other publications focusing on very similar AI model developments being published between January 2024 and March 2025. Naderi et al. e.g. published a study that **demonstrated the potential** of using machine learning techniques for the early and noninvasive diagnosis of non-alcoholic steatohepatitis (NASH). [10] Their **promising results** underscored machine learning's capabilities in augmenting and potentially replacing invasive diagnostic procedures like liver biopsy that are currently considered the gold standard for confirming NASH. However, their **study had some limitations**, including the small dataset size, which might have limited the model's generalizability. Thus, they recommended **future research** focusing on expanding to multi-center data encompassing wider demographics and disease severities. [10] Further, Yu et al. presented an AI model with ten features, which could predict MASLD in internal and external validations with high discrimination and calibration performance. [11] According to the authors their results **imply the potential integration** of this model into clinical workflows. Nevertheless, they also recommended **further prospective, randomized, and controlled studies** to validate the final prediction model. [11] Finally, Long et al. "maintain that **the implementation of their newly developed DA-GAG score would significantly augment clinicians' capacity** to screen≥F2 fibrosis of MASLD without imposing additional clinical burden". [12] However, again **more**

prospective and cross-sectional studies should be performed to further confirm the robustness and accuracy of the current model.
Thus, in the last year several research teams had published AI models for early detection of steatotic liver disease with very promising results, which even – in their opinion – had **the potential for integration** into clinical patient care workflows. However, all of them retracted from the important next step to tackle the challenge of clinical implementation, referring to still existing limitations and the need for further research.

In my impression this is exemplary for many current AI model developments, which are published with "superior and very promising results", at the same time however mentioning still existing limitations and the need for further research. Reports about projects to implement any of the formerly published AI models are more than rare. If we really want to advance patient care with innovative AI models, there is however an urgent need to also set the focus on this challenge, even if it might be much harder, than just improving the accuracy of an already existing AI model, but still not implementing it.

4. Final recommendations

What needs to be done, in order to come from scientific ideas to real support for patient care based on AI-enabled medical information systems?
Elaborating this in a very comprehensive scientific publication would however go beyond the scope of this short abstract. Here I would like to close with five key recommendations:

1. Our society needs a much stronger **research focus on successful AI implementation**
    - even though the hurdles here are much higher than for AI development.
2. (University) Hospitals need to establish **AI Governance regulations** and checklists as guidance to assure future AI implementations will be trustworthy,
    - compare e.g. the AI governance framework developed within the MII Open Medical Inference project [13] and the AI regulation established in summer 2024 at Erlangen University Hospital [14].
3. (University) Hospitals need to establish and fill the position of a **Chief AI Officer** (CAIO).
4. (University) Hospitals need to implement a **multi-level teaching program** on AI (see current AI Act regulation)
5. University Hospitals need to **establish AI implementation Labs** to support validation, certification and implementation of trustworthy AI applications
    - this should be an interdisciplinary synergistic effort of academic institutes and routine IT

5. Acknowledgements

OMI (Open Medical Inference) is funded by the Federal Ministry of Education and Research (BMBF) under the FKZ 01ZZ2315C.

References /Literatur

[1] Poon, E. G., Lemak, C. H., Rojas, J. C., Guptill, J., & Classen, D. (2025). Adoption of artificial intelligence in healthcare: survey of health system priorities, successes, and challenges. Journal of the American Medical Informatics Association: JAMIA, ocaf065. Advance online publication. https://doi.org/10.1093/jamia/ocaf065
[2] Huang, X., Gu, L., Sun, J., & Eils, R. (2025). Bridging the gaps: Overcoming challenges of implementing AI in healthcare. Med (New York, N.Y.), 6(4), 100666. https://doi.org/10.1016/j.medj.2025.100666
[3] Haserück, A. (2024). Blickwinkel: Künstliche Intelligenz und Radiologie. Deutsches Ärzteblatt 21. 8626.

Exploring Large Language Models for Automated Gait-Analysis

Rebecca Keilhauer[a,1], Carlo Dindorf[b], Julie Pham[c], Didier Stricker[a], Michael Lorenz[c,a]

[a] *German Research Center for Artificial Intelligence, Augmented Vision Group*

[b] *Department of Sports Science, RPTU University of Kaiserslautern-Landau*

[c] *Department of Computer Science, RPTU University of Kaiserslautern-Landau*

Biomechanical gait analysis plays a crucial role in various domains, including medical diagnosis, treatment evaluation, and the customization of medical aids [1]. However, traditional gait analysis relies heavily on the expertise of practitioners, is prone to subjective interpretation biases, and is a time-intensive process [2]. In the context of increasing workloads, staffing shortages, and growing documentation demands in healthcare, the need for efficient and objective assistive technologies for gait analysis and documentation has become more pressing.

Recently various works such as MotionScript [3] demonstrated that large language models (LLMs) can be leveraged to generate natural language descriptions of 3D human motions. These descriptions capture the movement patterns with detailed non-technical textual representation.

In this work we explore the usage of LLMs such as GPT-4o [4] for automated clinical gait analysis. We target patients undergoing total knee arthroplasty. Specifically, we recorded the movement of 23 patients (average age: 67.52 with standard deviation of 6.15 years, gender: 12F, 11M) before and 6 weeks after surgery using an optical motion capture system by SIMI Motions. Additionally, we conducted interviews with practitioners such as physiotherapists to gather descriptions of gait characteristics based on visual observation. These descriptions served as the ground truth for validating the LLMs.

Based on the spatiotemporal data, parameters commonly used in gait analysis [1] such as range of motions (ROMs) are extracted and used as input for LLMs. Through optimized prompt engineering [5], it is demonstrated that LLMs are capable of generating reliable gait descriptions, which can be used in an automatic documentation procedure. The validity of these analyses was verified by comparison with the ground truth interviews.

The results indicate that the gait descriptions generated by LLMs align closely with the ground truth expert descriptions and could potentially support a documentation process, easing the daily work of practitioners.

References

[1] LUDWIG, Oliver. Ganganalyse in der Praxis: Anwendung in Prävention, Therapie und Versorgung. Maurer, 2012.

[2] HULLECK, Abdul Aziz, et al. Present and future of gait assessment in clinical practice: Towards the application of novel trends and technologies. Frontiers in medical technology, 2022, 4. Jg., S. 901331.

[3] YAZDIAN, Payam Jome, et al. Motionscript: Natural language descriptions for expressive 3d human motions. arXiv preprint arXiv:2312.12634, 2023.

[4] HURST, Aaron, et al. Gpt-4o system card. arXiv preprint arXiv:2410.21276, 2024.

[5] WANG, Jiaqi, et al. Prompt engineering for healthcare: Methodologies and applications. arXiv preprint arXiv:2304.14670, 2023.

[1] Corresponding Author: Rebecca Keilhauer, Rebecca.Keilhauer@dfki.de

The SPHN Researcher Journey: Enabling Seamless Interoperability in Personalized Health Research

Sabine Österle[a,1]

[a] *SPHN Data coordination center, SIB Swiss Institute of Bioinformatics, Basel, Switzerland*

The Swiss Personalized Health Network (SPHN) [1] is advancing personalized health research by building a comprehensive data infrastructure grounded in FAIR (Findable, Accessible, Interoperable, Reusable) principles. This ecosystem provides researchers with a suite of tools to design, integrate, and reuse health data effectively. In this presentation, we will guide you through the researcher's journey, from data exploration, schema design to dataset reuse, highlighting the key tools that support each step.

The journey begins with the SPHN Data Exploration and Analysis System (currently in pilot stage), which allows researchers to query data availability across hospitals and identify suitable cohorts for their studies. Researchers can also conduct preliminary analyses, such as exploring value distributions or survival curves in a privacy preserving way.

To initiate their research project, researchers can use SPHN Schema Forge [2], a tool that enables the design and customization of semantic data schemas in line with SPHN standards [3]. With Schema Forge, researchers can create an RDF schema, SHACL shapes for data validation, SPARQL queries for advanced analysis, and human-readable HTML documentation. These schemas lay the foundation for linking clinical, biological, and omics data in a standardized and interoperable manner. Once schemas are designed, the SPHN Connector streamlines the transformation of raw data into the defined specification, ensuring consistency with SPHN conventions at hospitals and other data generating facilities. This tool semantically enriches datasets, making them ready for integration across domains and further analysis.

After the research project, datasets are cataloged in the SPHN Metadata Catalog [4], a centralized platform for discovering and reusing health data across Switzerland. Built on the FAIR Data Point specification and utilizing the DCAT vocabulary, the catalog provides both human- and machine-readable metadata access. The Schema Scope component further enhances the catalog by enabling interactive exploration of data schemas, helping researchers identify and reuse relevant datasets more efficiently.

By offering a seamless pathway from data exploration to schema design and data reuse, SPHN equips researchers with the tools necessary to navigate the complexities of health data and accelerate progress in personalized health research.

References

[1] www.sphn.ch
[2] https://schemaforge.dcc.sib.swiss/
[3] Touré, V., Krauss, P., Gnodtke, K. et al. FAIRification of health-related data using semantic web technologies in the Swiss Personalized Health Network. Sci Data 10, 127 (2023). https://doi.org/10.1038/s41597-023-02028-y
[4] https://fdp-schemascope.dcc.sib.swiss/

[1] Corresponding Author: Sabine Österle, sabine.oesterle@sib.swiss.

PROMOP: Integrating Patient-Reported Outcomes for Metastatic Breast Cancer into the OMOP Common Data Model

Adnan Jouned[a], Florian Katsch[a,b], Marko Todorovix[a], Laura Verbeij[c], Tanja Stamm[d], Georg Duftschmid[a], Renske Los[e]

[a] *Institute of Medical Information Management, Center for Medical Data Science, Medical University of Vienna, Vienna, Austria*

[b] *Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria*

[c] *Department of Quality and Patient Care, Erasmus University Medical Center, Rotterdam, The Netherlands*

[d] *Institute for Outcomes Research, Center for Medical Data Science, Medical University of Vienna, Vienna, Austria*

[e] *Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands*

1. Introduction

The Health Outcomes Observatory (H2O) [1] project integrates patient-reported outcomes (PROs) into clinical care and aims to enable federated analyses across European sites. To facilitate multisite analyses, we adopted the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) [2], which supports federated, multi-institutional research and harmonizes healthcare data. However, because the OMOP CDM was initially not designed for questionnaire data, integrating PROs within it presents certain challenges, including distinguishing patient-reported data from clinician observations, representing dates, and accurate vocabulary mapping when standardized concepts are lacking.

2. Objectives & Methods

Our initiative ”PROMOP” aims at standardizing mapping methods for accurately incorporating PROs for metastatic breast cancer (MBC) into the OMOP CDM. Through an iterative, collaborative process involving domain experts from multiple disciplines, the following points are proposed to address the challenges:

1. Differentiating Data Sources: To clearly identify entries derived from PROs, we recommend using specific *type_concept_id* values such as "Patient self-report" (Concept ID: 32865) or "Patient filled survey" (Concept ID: 32862).
2. Representing Dates: Since PROs often provide only the date when a questionnaire was completed rather of when an event actually occurred, we suggest mapping these entries to the Observation table, using *observation_date* to record the completion date.
3. Expanding Concept Mapping: For some of our questions (e.g., from EORTC panel [3]) OMOP concepts exist, allowing a lossless representation. For others, we search for similar concepts in the OMOP CDM or propose adding new concepts to accurately capture patient responses.
4. Capturing Negative Responses: Negative responses (e.g., when a patient answers “No” to the question “Did you have an earlier diagnosis of breast cancer?”) can be challenging to map. We explore three approaches to represent negative answers:
    - **Approach 1:** Use an appropriate concept for the questions in *concept_id* and "No" (Concept ID: 45878245) in *value_as_concept_id*.
    - **Approach 2:** For observations, *qualifier_concept_id* may be used to represent negation with concepts like "No history of" (Concept ID: 4032324). However, the negation could be easily misinterpreted as a positive statement unless the *qualifier_concept_id* is considered.
    - **Approach 3:** Employ concepts that inherently express negation, e.g., “No history of malignant tumor of breast” (Concept ID: 45763684).

3. Results

For 30 out of 53 MBC-specific PRO questions, lossless mapping to standardized concepts that provided a 1:1 representation of the questions was possible. For the remaining questions, we identified concepts that capture the questions' semantics as closely as possible, accepting some degree of information loss. The methods developed can also be applied to integrate other diseases in the H2O project specifically Inflammatory Bowel Disease (IBD) and diabetes. ETL implementations are still in progress, once completed, PROMOP would enable robust, harmonized integration of PRO data into the OMOP CDM.

Keywords: Patient-Reported Outcomes, OMOP CDM, Metastatic Breast Cancer, Health Outcomes Observatory

Note: This work was supported by the Innovative Medicines Initiative (IMI) 2 Joint Undertaking under grant agreement No 945345-2. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program, The European Federation of Pharmaceutical Industries and Associations (EFPIA), Trial Nation and Juvenile Diabetes Research Foundation (JDRF). The funder had no role in the conception of this work. This manuscript reflects only the authors' view. The European Union and IMI are not responsible for any use that may be made of the information it contains.

References

[1] Stamm, Tanja, et al. "Building a value-based care infrastructure in Europe: The Health Outcomes Observatory." NEJM Catalyst. 2021;2(3). doi:10.1056/CAT.21.0146; (2021).

[2] Observational Health Data Sciences and Informatics (OHDSI). The Book of OHDSI. OHDSI; (2019).

European Organisation for Research and Treatment of Cancer (EORTC). EORTC QLQ-C30 Scoring Manual. 3rd ed. EORTC; (2001).

Lightning Talks

An Open Benchmark for Assessing Large Language Models for German Tumor Documentation

Stefan Lenz[a,1], Arsenij Ustjanzew[a], Marco Jeray[b], Torsten Panholzer[a]

[a] *Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Centre of the Johannes Gutenberg-University Mainz, Mainz, Germany*

[b] *Privacy, Compliance, and Risk Management Office, University Medical Centre of the Johannes Gutenberg-University Mainz, Mainz, Germany*

1. Introduction

Automating the extraction of structured information from medical documentation via large language models (LLMs) is an emerging approach and active area of research [1]. Due to the sensitivity of the patient data, only locally deployed open source LLMs can be used for this purpose in clinical practice. New pre-trained models are published at a high frequency and excel in benchmarks with respect to mathematics, programming and general question answering. Yet, this does not guarantee strong performance for information extraction on clinical texts in German medical language. Public benchmarks are lacking in this area and the amount of accessible data is very limited [2].

2. Methods

We evaluate three basic tumor documentation tasks: identifying tumor diagnoses in text, mapping them to ICD-10 codes, and determining the date of the first diagnosis. The LLMs are tested with various zero-shot and few-shot prompts, which guide them to extract the required information from the texts. To assess the performance, we use a dataset of 149 text snippets derived from anonymized urology doctors' notes [3]. The texts in the dataset are labeled with the diagnoses and the corresponding dates.

3. Results

The dataset is available at https://huggingface.co/datasets/stefan-m-lenz/UroLlmEvalSet, and the Python code for the benchmark can be found at https://github.com/stefan-m-lenz/UroLlmEval. Our flexible benchmark code supports locally deployable LLMs via Hugging Face's "transformers" package [4]. We tested eleven open-source models with sizes ranging from 1.7 to 70 billion model parameters [5]. Models with 7-12 billion parameters performed best and, surprisingly, were not clearly outperformed by larger models. Smaller models (1.7-3 billion parameters) were clearly inferior, however. They struggled more with following instructions and recognizing key concepts. Their incorporated knowledge was also more limited, particularly regarding ICD-10 codes. A particular challenge was to specify what constitutes a tumor diagnosis in a concise way in the prompts. Few-shot prompting improved the recognition of tumor diagnoses across all models compared to zero-shot prompting.

4. Conclusions

We present the first complete, openly accessible LLM benchmark for information extraction from German clinical texts. This benchmark facilitates systematic evaluation of new models, offering guidance towards the automation of tumor documentation tasks.

[1] Corresponding Author: Dr. Stefan Lenz, stefan.lenz@uni-mainz.de .

References

[1] Wiest, I. C., Ferber, D., Zhu, J., van Treeck, M., Meyer, S. K., Juglan, R., Carrero, Z. I., Paech, D., Kleesiek, J., Ebert, M. P., Truhn, D., & Kather, J. N. (2024). Privacy-preserving large language models for structured medical information retrieval. Npj Digital Medicine, 7(1), 1–9. https://doi.org/10.1038/s41746-024-01233-2

[2] Melnik, S., Brix, T., Storck, M., Riepenhausen, S., Varghese, J., & Rudack, C. (2024). Overview of German Clinical Text Corpora for Large Language Models – Scoping Review. GMDS Kooperationstagung "Gesundheit - Gemeinsam" 2024, DocAbstr. 743. https://doi.org/10.3205/24gmds079

[3] Lenz, S., Ustjanzew, A., Jeray, M., & Panholzer, T. (2024). Few-Shot-Prompting von Large Language Models zur Extraktion von Daten zu Tumordiagnosen aus urologischen Arztbriefen – eine Evaluation. GMDS Kooperationstagung "Gesundheit - gemeinsam" 2024, DocAbstr. 650. https://doi.org/10.3205/24gmds179

[4] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., … Rush, A. (2020). Transformers: State-of-the-Art Natural Language Processing. In Q. Liu & D. Schlangen (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (pp. 38–45). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-demos.6

[5] Lenz, S., Ustjanzew, A., Jeray, M., & Panholzer, T. (2025). Can open source large language models be used for tumor documentation in Germany? - An evaluation on urological doctors' notes (No. arXiv:2501.12106). arXiv; arXiv preprint. https://doi.org/10.48550/arXiv.2501.12106

LLM-based document classification and parameter extraction pipeline for efficient information exchange in the German healthcare

Farkas Marton Csaszar [a,1], Isabella Catharina Wiest [b], Marie-Elisabeth Leßmann [c], Thoralf Stange [d], Daniela Richter [e], Heiko Böhme [f], Marcel Susky [g]

[a] *Department of Translational Medical Oncology, National Center for Tumor Diseases (NCT/UCC) Dresden, Germany*

[b] *Else Kroener Fresenius Center for Digital Health, Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, Germany*

[c] *Else Kroener Fresenius Center for Digital Health, Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, Germany*

[d] *Core Unit for Data Management and Analytics, National Center for Tumor Diseases (NCT/UCC) Dresden, Germany*

[e] *Department of Translational Medical Oncology, National Center for Tumor Diseases (NCT/UCC) Dresden, Germany*

[f] *Core Unit for Data Management and Analytics, National Center for Tumor Diseases (NCT/UCC) Dresden, Germany*

[g] *Research Group Digital Health, Faculty of Business and Economics, TUD Dresden University of Technology, Germany*

1. Introduction

Communication between healthcare providers in Germany often takes manual effort using mail or fax. Medical documents are scanned, assigned to patient and document type, a time consuming, error prone approach [1]. Electronic data processing can accelerate this, but the medical infrastructure of the providers in Germany is not consistent, and new installations are frequently rejected due to missing infrastructure or competence [2]. A solution is required for a more efficient digital approach.

2. Methods

An integrated LLM based medical data processing system concept has been developed consisting of several components, and the text processing accuracy has been validated. A document classification system with a request database, communication module and LLM processing pipelines using prompt engineering is able to transform medical data into structured digital format and store it automatically.
A task management database requests Large Language Model (LLM) jobs to a server hosting LLM-Aix information extraction pipelines (https://medrxiv.org/content/10.1101/2024.09.02.24312917v1) using the Q4 quantised version of Llama 3.1 70B, then collects the responses of several pipelines and forwards them to the output.
Medical documents are supplied to the system input and it automatically sorts the documents into 17 predefined classes and class information are saved as tags.
A target database for document sorting tasks has been designed to assign the documents with its class tags to a personal identity. The Extract-Load-Transform (ELT) based integration process stores the source data in its original terminology, resulting in flexible data transformation to various target systems, thus supporting precise and efficient medical research [3]. For improved semantic interoperability, document classes are derived from the “Klinische Dokumentenklassen-Liste” (KDL) [4].

3. Results

The Gematik standard KIM (“Kommunikation im Medizinwesen”) is an e-mail based platform with secure end-to-end encryption, already used by hospitals and doctors. The fully automated LLM pipeline is integrated with KIM providing an easy data exchange solution based on existing

[1] Corresponding Author: Farkas Marton Csaszar, farkas.csaszar@nct-dresden.de

infrastructure, processing the documents without requiring manual intervention or data entry. The prompt engineering-based LLM processing efficiently extracts relevant parameters [5] from the attached documents without explanatory text or metadata, thus accelerating data processing significantly. The document sorting has been evaluated with a curated set of 90 documents, it demonstrates robustness and reliability. The first classification method sorting into unique document classes achieves 75% accuracy. A second method comprising a sophisticated class representation table using one-hot-encoded version of the document classes allows multiple class categories to be assigned simultaneously to the same document, which proves to be highly accurate with 94% accuracy.

4. Conclusions and outlook

The LLM-based document classification simplifies communication between providers in intersectoral healthcare scenarios. Providing the two validated methods it can be seen that a method providing multiple simultaneous classes for documents results in more accurate classification, because medical documents are often used for multiple purposes. Limitations of the system are most often due to the difficulties in the recognition of hand written information and due to the fluctuations in the source data quality regarding critical identifying personal data. Improvements are planned to provide interoperable and consistent structured data by the extraction of key clinical parameters for clinical specialists, e.g. pathologists for analyzing morphology data.

5. Acknowledgements

This work is part of the project “Medical Informatics Hub in Saxony (MiHUBx)”, funded by the German Ministry of Education and Research (Grant numbers 01ZZ2101A-E).

References

[1] Manuel Iserloh (2024). Patientenkommunikation: Persönlich, aber bitte digital. In: kma - Klinik Management aktuell 2024; 29(07/08): pp. 38-40 DOI: 10.1055/s-0044-1793883

[2] Ursula Hübner, Jan-David Liebe, Moritz Esdar, Jens Hüsers, Jens Rauch, Johannes Thye & Jan-Patrick Weiß (2019). Stand der Digitalisierung und des Technologieeinsatzes in deutschen Krankenhäusern. In: Krankenhaus-Report 2019. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-58225-1_3

[3] Edward Manopo Haryono, Fahmi, Adi Setiawan Tri W, Indra Gunawan, Achmad Nizar Hidayanto, Untung Rahardja (2020). Comparison of the E-LT vs ETL Method in Data Warehouse Implementation. In: A Qualitative Study in 2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS) pp. 115-120. 978-1-7281-9167-6/20/$31.00

[4] Annett Müller (2022). Die Klinische Dokumentenklassen-Liste (KDL) als Werkzeug für die semantische Interoperabilität. In: Digitalstrategie im Krankenhaus, 2022, pp. 333-342 DOI: 10.1007/978-3-658-36226-3_22

[5] Aishwarya Vijayan (2023). A Prompt Engineering Approach for Structured Data Extraction from Unstructured Text Using Conversational LLMs. In: ACAI '23: Proceedings of the 2023 6th International Conference on Algorithms, Computing and Artificial Intelligence pp. 183 – 189. https://doi.org/10.1145/3639631.3639663

Medication information extraction using local large language models

Phillip Richter-Pechanski[a,b,c,d,e], Dominic M. Schwab[b], Christina Kiriakou[b], Nicolas A. Geis[b,d], Anette Frank[e*], Christoph Dieterich[a,b,c,d*1]

[a] *Section of Bioinformatics and Systems Cardiology, Klaus Tschira Institute for Integrative Computational Cardiology, Heidelberg, DE*

[b] *Department of Internal Medicine III, University Hospital Heidelberg, DE*

[c] *German Center for Cardiovascular Research (DZHK) - Partner site Heidelberg/Mannheim, DE*

[d] *Informatics for Life, Heidelberg, DE*

[e] *Department of Computational Linguistics, Heidelberg University, DE*

1. Introduction

Medication information is crucial for clinical routine and research. However much of it is stored in unstructured text, such as doctoral letters, requiring time-intensive and error-prone manual information extraction [1]. Recent advances in NLP and machine learning have demonstrated the potential of pre-trained generative language models and parameter-efficient fine-tuning methods to automate medical information extraction [2][3].

2. Methods

In our study we evaluate local generative large language models (LLMs) for end-to-end extraction of nine classes of medication information (drug, adverse drug events, reason, duration, dosage, form, frequency, route and strength), combining named entity recognition and relation extraction. To address the challenge of ensuring consistent structured output from generative models, we use format-restricting instructions and propose an automatic feedback pipeline to identify semantically similar predictions for automated evaluation.
We experimented with two open-source models in different sizes -- one general, one domain-specific -- in zero-shot and supervised fine-tuning scenarios on the English i2b2 2018 clinical corpus [1] and the German CARDIO:DE corpus, containing doctoral letters from the cardiology department of the Heidelberg University Hospital [4].

3. Results

While the domain-specific model struggled to generate reliable structured outputs, the fine-tuned open-source general models outperformed the recurrent neural network–based SOTA on complex relation classes such as adverse drug events (ADE) and medication reason by up to 26 percentage points F1-score (75% for ADE; 84% for reason). Moreover, the fine-tuned open-source general models established a new SOTA baseline on the German data that outperforms traditional extraction methods using BERT by 16 percentage points micro-average F1-score (89%). We use Shapley value-based interpretability methods to attribute input token contributions to predictions. We will demonstrate how this method can enhance the transparency of our system to help physicians in making informed decisions.

[1] Corresponding authors: phillip.richter-pechanski@med.uni-heidelberg.de , christoph.dieterich@med.uni-heidelberg.de, frank@cl.uni-heidelberg.de . * CD and AF share the last authorship.

4. Discussion & Conclusion

Overall, our findings show that automated medication information extraction from unstructured text is feasible for locally implemented open-source generative LLMs. Key aspects encompass fine-tuning with parameter efficient methods and format-restricting instructions. Our approach outperforms SOTA methods for medication information extraction both on English and German data, delivering high performance in a clinical setup where constraints in time and computational resources matter, and where safety measures should be integrated wherever possible.

References /Literatur

[1] S. Henry, K. Buchan, M. Filannino, A. Stubbs, and O. Uzuner, "2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records," J. Am. Med. Inform. Assoc., vol. 27, no. 1, p. 3, Jan. 2020, doi: 10.1093/JAMIA/OCZ166.

[2] K. Singhal et al., "Large language models encode clinical knowledge," Nat. 2023 6207972, vol. 620, no. 7972, pp. 172–180, Jul. 2023, doi: 10.1038/s41586-023-06291-2.

[3] J. Dagdelen et al., "Structured information extraction from scientific text with large language models," Nat. Commun., vol. 15, no. 1, Dec. 2024, doi: 10.1038/S41467-024-45563-X.

[4] P. Richter-Pechanski et al., "A distributable German clinical corpus containing cardiovascular clinical routine doctor's letters," Sci. Data, vol. 10, no. 1, p. 207, 2023.

Enhancing Large Language Models for Structured Medical Documentation in German Healthcare

Richard Noll[a,1], Magdalena Weber[a], Marty Oelschlaeger[b], Fabian Dechent[b], Liel Glaser[b], Jana Hebeiss[c], Michael von Wagner[c], Svenja Sliwinski[d], Jannik Schaaf[a]

[a] *Goethe University Frankfurt, University Medicine Frankfurt, Institute of Medical Informatics, Frankfurt, Germany*

[b] *dida Datenschmiede GmbH, Berlin, Germany*

[c] *University Hospital Frankfurt, Goethe University, Executive Department for medical IT-Systems and digitalization, Frankfurt, Germany*

[d] *Goethe University Frankfurt, University Hospital, Department of General, Visceral, Transplant and Thoracic Surgery, Frankfurt, Germany*

1. Introduction

In Germany, there is a notable trend towards the digitisation of hospital documentation processes, offering a multitude of advantages. These include enhanced efficiency in data collection and retrieval within hospitals and improved connectivity between clinics, facilitating more straightforward access to data for research purposes. However, the standardisation and harmonisation of digital health data remain significant challenges [1].
Although clinical personnel frequently employ digital tools for text documentation, a considerable proportion of medical details, including anamnesis and risk assessments, are still recorded as unstructured free text [2]. This approach may impede the traceability of information by other hospital staff and complicate automated data analysis and research, particularly when similar medical indicators are described using different terminologies. Existing standardised terminologies, such as those for diagnoses, symptoms, and phenotypes, could address this issue by mapping free-text terms to clear, unique codes. This would not only reduce misunderstandings and errors but also enhance the comprehensibility and usability of documentation for research and subsequent clinical use.

2. Methods

The application of machine learning (ML) and natural language processing (NLP) methodologies can facilitate the automation of the categorisation and structured summarisation of such free-text documentation, while ensuring the maintenance of its individual character [3]. This includes the application of large language models (LLMs), such as GPT-4, LLaMA, and Falcon, which are highly proficient at processing natural language due to extensive training on large datasets [4]. However, generic LLMs often produce misleading or incorrect results for domain-specific tasks, rendering their direct use in healthcare impractical [5]. To address this, fine-tuning pre-trained LLMs with task-specific data is necessary.

3. Results

The objective of this study is to conduct industrial research into the fine-tuning of methods for LLMs, with the aim of enabling them to map medical history forms to medical terminology codes with greater accuracy. The goal is to facilitate the utilisation of LLMs for health data analysis and support the digital transformation of healthcare, with a particular focus on German-language datasets. The study encompasses the development of a prototype that is capable of analysing medical history forms in real time and ensuring efficient deployment on hospital servers, with a view to addressing both data privacy and resource constraints.

[1] Corresponding Author: Richard Noll, Institute of Medical Informatics, Goethe University Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany; noll@med.uni-frankfurt.de.

References

[1] Nan, Yang, et al. "Data harmonisation for information fusion in digital healthcare: A state-of-the-art systematic review, meta-analysis and future research directions." Information Fusion 82 (2022): 99-122.
[2] Adnan, Kiran, et al. "Role and challenges of unstructured big data in healthcare." Data Management, Analytics and Innovation: Proceedings of ICDMAI 2019, Volume 1 (2020): 301-323.
[3] Jiang, Min, et al. "A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries." Journal of the American Medical Informatics Association 18.5 (2011): 601-606.
[4] Thirunavukarasu, Arun James, et al. "Large language models in medicine." Nature medicine 29.8 (2023): 1930-1940.
[5] Cascella, Marco, et al. "Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios." Journal of medical systems 47.1 (2023): 33.

Current consent approaches to secondary research with patient data: Ethical challenges and outlook

Julie Schweer [a], Christoph Schickhardt [a,b,1]

[a] *National Center for Tumor Diseases (NCT), NCT Heidelberg, a partnership between DKFZ and Heidelberg University Hospital, Germany, Heidelberg University, Medical Faculty Heidelberg, Department of Medical Oncology, Section Translational Medical Ethics*

[b] *German Cancer Research Center (DKFZ) Heidelberg*

1. Introduction

Throughout their life or in the course of complex treatments, patients interact with various healthcare providers. A cancer patient, for example, may have been seeing a primary care physician long before receiving diagnosis. After diagnosis, they may visit a clinic for intensive treatment and, later, alternate between primary care physicians, specialized oncologists and other healthcare providers.
This not only implies that patient data is often stored in various medical information systems, but also at different locations and under different ethical and legal responsibilities. These factors have normative consequences for how patient data can be used for secondary research, specifically in light of consent approaches for secondary research use.

2. Methods

By looking at the technical and organizational aspects of medical information systems and research infrastructures in Germany, we identify corresponding normative challenges of consent approaches for secondary research with patient data. Considering recent developments from practice, e.g., in and around the German Medical Informatics Initiative, we illustrate how attempts to address these challenges raise new normative issues.

3. Results

Using and sharing patient data for research across multiple institutions and corresponding information systems is a scientific desideratum that induces various challenges, including dealing with various responsible entities. To address such challenges in practice, researchers and projects have responded by combining or supplementing consent forms for secondary research with patient data with various additional modules (e.g. [1]). This however, introduces normative challenges, e.g., in terms of burdens, comprehensibility and transparency, or when patients receive multiple consent forms in the course of their treatment. More generally, due to the demandingness of consent procedures in terms of personnel and time, it is difficult to scale them to recruit a majority of patients at the respective institutions – which would, however, be important for realizing the research potential of secondary research with patient data.

4. Discussion

As technical, organizational and scientific challenges induced by consent-based approaches translate into significant normative challenges, the question arises whether an opt-out approach could be a viable alternative that contributes to realizing the potential of secondary research with patient data while addressing many of the challenges induced by consent-based approaches.

[1] Christoph Schickhardt, christoph.schickhardt@med.uni-heidelberg.de

References

[1] Consent Working Group of the German Medical Informatics Initiative (2022). Handreichung zur national Handreichung zur Anwendung der national harmonisierten Patienteninformations- und Einwilligungsdokumente zur Sekundärnutzung von Patientendaten (v. 1.2.3). Retrieved April 14, 2025, from https://www.medizininformatik-initiative.de/sites/default/files/2022-10/MII_AG-Consent_Handreichung_v1.2.3.pdf

ICFx WebApp: application for recording participation disorders on an ICF basis

Eveline Prochaska[a, 1], Stephanie Boden[b], Peter Bernstein[b, c, d]

[a]*Institute for Medical Informatics and Biometry, Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, Dresden, Germany*

[b]*Johannesbad Fachklinik & Gesundheitszentrum Raupennest, Altenberg, Germany*

[c]*Wirbelsäulenzentrum der Asklepios Klinik Hohwald, Neustadt in Sachsen, Germany*

[d]*Medizinische Fakultät Carl Gustav Carus, Technische Universität Dresden*

1. Introduction

Professionals from different disciplines—physicians, psychologists, social workers, physiotherapists—often assess participation restrictions differently. While physicians and psychologists focus on symptoms, social workers emphasize context, and physiotherapists concentrate on functional limitations. For patients, however, the impact on daily life and participation is central. Communication barriers (e.g., fears, misunderstandings, mistrust) complicate the setting of treatment goals and may distort perceptions of outcomes [1]. Even general practitioners with long-standing patient relationships agree on only about 75% of participation-relevant aspects [2], with discrepancies more frequent in psychiatric comorbidities [3]. Although digital healthcare increasingly requires structured data—such as that provided by the ICF—its practical use remains limited [4]. The newly developed ICFx application aims to address this gap.

2. Methods

The ICFx application was conceptualized and implemented by a medical specialist. It is accessible as a web-based platform via browser and mobile devices, hosted on a Germany-based cloud server. The technical architecture includes a frontend built with Vue3 and TypeScript and a backend developed using Django REST Framework. Data is managed using a PostgreSQL database.
Development followed an iterative, user-centered process. Usability testing and feedback from experts in medical informatics and user experience design were integrated continuously. The aim was to create a tool that supports both patient-reported data and structured clinical documentation.

3. Results

The ICFx app is currently in pilot testing. Patients report their perspective using standardized questionnaires such as the SF-36. Clinicians can choose between two documentation modes: ICF browser or predefined ICF Core Sets (see Fig. 1).
To support onboarding and practical use, physicians developed supplementary materials, including an informational flyer and a short explanatory video [5]. The platform enables side-by-side comparison of patient and practitioner perspectives and supports structured, multi-professional documentation.
A significant portion of data entry can be delegated to patients, reducing the administrative workload for healthcare providers while increasing patient engagement.

4. Discussion

ICFx offers a digital framework for the structured assessment of functioning and participation. By integrating multiple viewpoints into the documentation process and allowing partial self-reporting by

[1] Corresponding Author: Eveline Prochaska, eveline.prochaska@tu-dresden.de.

patients, the application contributes to improved communication, goal setting, and care coordination across professional boundaries.

References

[1] Sharby, N., Martire, K., & Iversen, M. D. (2015). Decreasing Health Disparities for People with Disabilities through Improved Communication Strategies and Awareness. International Journal of Environmental Research and Public Health, 12(3), 3301. https://doi.org/10.3390/IJERPH120303301

[2] Casanova, L., Ringa, V., Chatelard, S., Paquet, S., Pendola-Luchel, I., Panjo, H., Bideau, C., Deflesselle, E., Delpech, R., Bloy, G., & Rigal, L. (2018). Level of agreement between physician and patient assessment of non-medical health factors. Family Practice, 35(4), 488‑494. https://doi.org/10.1093/FAMPRA/CMX141

[3] Schneider A, Wartner E, Schumann I, Hörlein E, Henningsen P, Linde K. The impact of psychosomatic co-morbidity on discordance with respect to reasons for encounter in general practice. J Psychosom Res. 2013 Jan;74(1):82-5. doi: 10.1016/j.jpsychores.2012.09.007.

[4] Leonardi, M., Lee, H., Kostanjsek, N., Fornari, A., Raggi, A., Martinuzzi, A., Yáñez, M., Almborg, A. H., Fresk, M., Besstrashnova, Y., Shoshmin, A., Castro, S. S., Cordeiro, E. S., Cuenot, M., Haas, C., Maart, S., Maribo, T., Miller, J., Mukaino, M., ⋯ Kraus de Camargo, O. (2022). 20 Years of ICF–International Classification of Functioning, Disability and Health: Uses and Applications around the World. International Journal of Environmental Research and Public Health, 19(18), 11321. https://doi.org/10.3390/IJERPH191811321/S1

[5] Bernstein, P., Boden S. (2025, April 4). Aspekte der Teilhabe. Erklärfilm zu den Hintergründen der klinischen Studie. https://www.youtube.com/watch?v=MiHqU0KLAXM

Integrating Patient-Reported Outcomes with FHIR: A Scalable and FAIR-Compliant Solution for EAV-Based Systems

Maryam Jafapour[a,1], Florian Katsch[b], Laura Sophia Tassatti[a], Georg Duftschmid[a]

[a] *Center for Medical Data Science, Medical University of Vienna, Vienna, Austria*

[b] *Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria*

1. Background

Patient-reported outcomes (PROs) are increasingly recognized as valuable inputs for personalized care and clinical decision-making [1]. Integrating PRO data into healthcare systems using HL7 FHIR standards enhances interoperability, yet the Entity-Attribute-Value (EAV) model, which is frequently used by legacy systems, presents unique challenges in data transformation and standardization [2]. EAV-based systems, while flexible, require scalable solutions to fully leverage PRO data for modern health applications. To address these challenges, this study presents a structured, FAIR-compliant approach for exporting PRO data from EAV-based systems into FHIR resources [3,4]. We focus on EAV-based systems since they are widely used in legacy health IT systems—especially in academic and research settings—due to their flexibility in capturing heterogeneous data. While other data models also pose transformation challenges, EAV systems are particularly difficult to standardize due to their schema-less structure, which decouples attributes from fixed schema definitions. This flexibility complicates semantic alignment and requires robust transformation logic, which has received limited attention in the context of HL7 FHIR and FAIR integration.

2. Objective

This study aims to design and implement a scalable, database-driven mapping strategy for transforming EAV-based PRO data into FHIR *Questionnaire* and *QuestionnaireResponse* resources. Our approach prioritizes maintainability, adaptability, and alignment with FAIR principles to enhance interoperability across diverse healthcare systems.

3. Methods

The proposed solution utilizes a dynamic mapping strategy with mappings managed separately from implementation code. Mappings from source EAV records to FHIR resources are stored in a dedicated database table. This modular design avoids hard-coded mappings hidden in implementation code, allows easy editing of mappings without requiring code changes, and thus reduces maintenance overhead. Our approach, which was implemented on the HAPI FHIR framework and Spring Boot platform, acts as a façade over the legacy system. As a demonstrational source system we use OpenRDA [5], an EAV-based electronic health record (EHR) system at the Medical University of Vienna. Our approach adheres to the FAIR principles as follows:

- **Findability**: A globally unique identifier was assigned to the software [2], ensuring discoverability. Well-documented mapping logic and rich metadata further support searchability and indexing.
- **Accessibility**: Standardized FHIR APIs provide secure access to PRO data. Metadata will remain accessible through registries even if the software becomes unavailable.
- **Interoperability**: The approach aligns with HL7 FHIR standards for compatibility with diverse systems and integrates community-relevant standards for PRO data representation.

[1] Corresponding Author: Maryam Jafarpour, maryam.jafarpour@meduniwien.ac.at.

[2] https://doi.org/10.6084/m9.figshare.28195922

- **Reusability**: The software is released under the Apache 2.0 open-source license, enabling reuse, modification, and integration into other systems. Its modular design supports diverse questionnaire structures and EAV implementations.

To handle edge cases, such as incomplete or inconsistent data, the system applies validation rules and default values during transformation, ensuring robust data representation.

4. Results

We tested our approach using two main questionnaires employed by the Medical University of Vienna—namely, the "PROMIS - Global Health 10" and the "Inflammatory Bowel Disease (IBD) Questionnaire". The data were stored in the OpenRDA schema. The implementation demonstrated high adaptability to diverse questionnaire structures and achieved efficient data transformation with minimal maintenance requirements. Additionally, the system's FAIR compliance—enabled by its unique identifier, rich metadata, open-source licensing, and adherence to community standards—facilitates its reusability and integration into various healthcare settings. Initial deployment at the Medical University of Vienna highlighted the system's scalability and alignment with FAIR principles, enabling seamless integration of PRO data across multiple healthcare information systems. The application of FAIR principles improved system transparency and data discoverability. For instance, integrating rich metadata allowed users to trace data provenance effectively. Moreover, the use of a persistent identifier enhanced software findability in public registries, facilitating wider reuse. These improvements significantly reduce onboarding time for new institutions and simplify integration into federated research networks.

5. Conclusion

The presented scalable, FAIR-compliant solution bridges the gap between EAV-based legacy systems and HL7 FHIR, enabling interoperable processing of PRO data in a standardized format. By incorporating globally unique identifiers, rich metadata, and open licensing, the solution ensures long-term findability, accessibility, and reusability of PRO data, thus advancing patient-centered healthcare interoperability. Future work will focus on adding further questionnaires, evaluating performance with larger datasets, enhancing data privacy measures, and extending support for AI-enabled analytics and clinical decision-making. These advancements will further drive innovation in patient-centered healthcare.

References

[1] Hermanns N, Kulzer B, Ehrmann D. Person-reported outcomes in diabetes care: What are they and why are they so important? Diabetes, Obesity and Metabolism. 2024;26(S1):30–45.

[2] El-Sappagh SH, El-Masri S, Riad AM, Elmogy M. Electronic health record data model optimized for knowledge discovery. International Journal of Computer Science. 2012;9(5):329–38.

[3] Jacobsen A, de Miranda Azevedo R, Juty N, Batista D, Coles S, Cornet R, et al. FAIR Principles: Interpretations and Implementation Considerations. Data Intelligence. 2020 Jan 1;2(1–2):10–29.

[4] Barker M, Chue Hong NP, Katz DS, Lamprecht AL, Martinez-Ortiz C, Psomopoulos F, et al. Introducing the FAIR Principles for research software. Sci Data. 2022 Oct 14;9(1):622.

[5] Rinner C, Duftschmid G, Wrba T, Gall W. Making the complex data model of a clinical research platform accessible for teaching. J Innov Health Inform. 2017;24(1).

Enhancing Molecular Tumor Boards with User-Centered Visualizations - Embedding PROMs in cBioPortal

Markus Wolfien[a,b1], Cosima Strantz[c], Phillip Unberath[c,d], Sara Bäcker[a], PM4Onco - AP3 Study Group[2]

[a] *Institute for Medical Informatics and Biometry, Faculty of Medicine Carl Gustav Carus, TUD Dresden University of Technology, Dresden, Germany*

[b] *Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Dresden/Leipzig, Dresden, Germany*

[c] *Medical Center for Information and Communication Technology, Universitätsklinikum Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany*

[d] *SRH University of Applied Sciences, Fürth, Germany*

1. Introduction

Sustainable integration of Patient-Reported Outcome Measures (PROMs) into Molecular Tumor Boards (MTBs) is essential for advancing personalized oncology care [1]. PROMs capture patients' subjective experiences, including symptoms, side effects, and quality of life, thereby offering complementary insights to molecular and clinical data. In the MTB setting, these insights can help assess therapy response, identify tolerability issues, and support more nuanced treatment decisions. Despite their potential, the routine use of PROMs in MTBs is rare, largely due to challenges in workflow integration, limited interoperability, and difficulties in visualizing longitudinal symptom data in ways that are clinically meaningful [2].

2. Methods

As part of the German PM4Onco project, we adopted a user-centered design methodology to develop an integrative PROM prototype within cBioPortal, an open-source platform that supports clinical decision-making through molecular and clinical data visualization [3]. Our development process included a literature-based requirements analysis, persona creation to reflect stakeholder needs, and a co-design workshop involving oncologists and MTB coordinators. The design emphasized the need for clear trend indicators, consistent scoring systems (e.g., changes in Visual Analogue Scale entries across PROMs), and intuitive representations of symptom trajectories.

3. Results

The resulting prototype integrates PROMs into cBioPortal through two primary components. Firstly, a timeline view that provides a compact longitudinal overview of PROMs relative to key clinical events. Secondly, a dedicated PROMs tab that offers detailed views of symptom trends, line graphs indicating clinically relevant thresholds, and customizable panels to prioritize QoL domains based on tumor type or treatment phase. This integration enables direct access to PROMs within the clinical workflow, thereby supporting informed treatment decisions while adhering to technical constraints of existing healthcare IT infrastructures, such as through standardization with FHIR.

4. Discussion

Our prototype demonstrates that integrating PROMs into MTBs is both technically feasible and clinically meaningful. In parallel with these technical advancements, our perspective paper, "The

[1] Corresponding Author: Markus Wolfien, Contact details: markus.wolfien@tu-dresden.de

[2] For more information, see https://pm4onco.de/en/ (last accessed April 10, 2025)

Promise of PROs: Patients as Equal Partners in the Decision-Making on Their Treatment," emphasizes that integrating PROs into MTBs represents both a technical objective and a moral imperative, ensuring that patient voices are central. MTBs, with their focused yet impactful scope, provide an ideal testing environment for integrating PROs, enabling us to refine and evaluate these systems in a specialized, multidisciplinary context [4, 5]. We emphasize that effective treatment is not pursued at any cost, but always with a focus on ensuring the best possible quality of life, even under the often burdensome conditions of therapy in the MTB context.

References

[1] C. Strantz et al., “Empowering personalized oncology: evolution of digital support and visualization tools for molecular tumor boards,” BMC Med Inform Decis Mak, vol. 25, no. 1, p. 29, Jan. 2025, doi: 10.1186/s12911-024-02821-8.

[2] D. Cella, K. Nolla, and J. D. Peipert, “The challenge of using patient reported outcome measures in clinical practice: how do we get there?,” J Patient Rep Outcomes, vol. 8, no. 1, p. 35, Mar. 2024, doi: 10.1186/s41687-024-00711-1.

[3] C. Renner et al., “Extending cBioPortal for Therapy Recommendation Documentation in Molecular Tumor Boards: Development and Usability Study,” JMIR Med Inform, vol. 11, p. e50017, Dec. 2023, doi: 10.2196/50017.

[4] A. M. Tsimberidou, M. Kahle, H. H. Vo, M. A. Baysal, A. Johnson, and F. Meric-Bernstam, “Molecular tumour boards — current and future considerations for precision oncology,” Nat Rev Clin Oncol, vol. 20, no. 12, pp. 843–863, Dec. 2023, doi: 10.1038/s41571-023-00824-4.

[5] M. Di Maio et al., “The role of patient-reported outcome measures in the continuum of cancer clinical care: ESMO Clinical Practice Guideline,” Annals of Oncology, vol. 33, no. 9, pp. 878–892, Sep. 2022, doi: 10.1016/j.annonc.2022.04.007.

Multi-Sensor Approach for Infant Movement Classification

Tomas Kulvicius [a,b,c,1], Dajie Zhang [b,d], Luise Poustka [b], Lennart Jahn [a,c], Florentin Wörgötter [c], Peter B Marschik [a,b,d,e,1]

[a] *Child and Adolescent Psychiatry and Psychotherapy, University Medical Center Göttingen, Leibniz ScienceCampus Primate Cognition and German Center for Child and Adolescent Health (DZKJ), Göttingen, Germany*

[b] *Department of Child and Adolescent Psychiatry, University Hospital Heidelberg, Ruprecht-Karls University of Heidelberg, Heidelberg, Germany*

[c] *Department for Computational Neuroscience, Third Institute of Physics - Biophysics, Georg-August-University of Göttingen, Göttingen, Germany*

[d] *iDN – interdisciplinary Developmental Neuroscience, Division of Phoniatrics, Medical University of Graz, Graz, Austria*

[e] *Center of Neurodevelopmental Disorders (KIND), Department of Women's and Children's Health, Center for Psychiatry Research, Karolinska Institutet & Region Stockholm, Stockholm, Sweden*

1. Introduction

There is a recent boom in the development of AI solutions to facilitate and enhance diagnostic procedures for established clinical tools. The Prechtl general movement assessment (GMA) is recognized for its clinical value in diagnosing neurological impairments in early infancy [1]. GMA has been increasingly augmented through machine learning approaches, all mostly based on single sensor modalities [2, 3, 4]. These attempts are still considerably inferior to those of well-trained human assessors. Moreover, these approaches are hardly comparable as all models are designed, trained and evaluated on silo-data sets. In this study, we propose a deep learning-based multi-sensory approach for assessing infant motor patterns [5].

2. Methods

We compare three different sensor modalities (visual sensors [4], pressure, and inertial [5]) to classify fidgety movements (FMs) and evaluate whether a multi-sensor system outperforms single-modality assessments. Movements were recorded from 51 typically developing infants. We tested various sensor combinations and two different sensor fusion approaches, i.e., late- and early-sensor fusion. For movement classification, convolutional neural network (CNN) architectures were used and evaluated using 9-fold cross-validation procedure [5].

3. Results

Results show that the highest classification accuracy was obtained when using video or inertial sensors (90.7% and 90.2%, respectively) which outperformed classification accuracy using a pressure mat sensor (82.1%). Results also demonstrate that the classification accuracy of the three-sensor fusion (94.5%) was significantly higher than classification accuracy of any single sensor modality [5].

4. Conclusion

Our study suggests that the multi-sensor approach is a promising pathway for automated classification of infant motor functions. The development of a robust sensor fusion system may significantly enhance AI-based early recognition of neurofunctions and facilitate early detection of neurodevelopmental disorders [5].

[1] Corresponding Authors: Tomas Kulvicius, tomas.kulvicius@uni-goettingen.de; Peter B Marschik, peter.marschik@med.uni-heidelberg.de.

References

[1] Prechtl, H. F., Einspieler, C., Cioni, G., Bos, A. F., Ferrari, F., & Sontheimer, D. (1997). An early marker for neurological deficits after perinatal brain lesions. The Lancet, 349(9062), 1361-1363.

[2] Silva, N., Zhang, D., Kulvicius, T., Gail, A., Barreiros, C., Lindstaedt, S., ... & Marschik, P. B. (2021). The future of General Movement Assessment: The role of computer vision and machine learning – A scoping review. Research in Developmental Disabilities, 110, 103854.

[3] Reinhart, L., Bischops, A. C., Kerth, J. L., Hagemeister, M., Heinrichs, B., Eickhoff, S. B., ... & Meissner, T. (2024). Artificial intelligence in child development monitoring: A systematic review on usage, outcomes and acceptance. Intelligence-Based Medicine, 9, 100134.

[4] Reich, S., Zhang, D., Kulvicius, T., Bölte, S., Nielsen-Saines, K., Pokorny, F. B., ... & Marschik, P. B. (2021). Novel AI driven approach to classify infant motor functions. Scientific Reports, 11(1), 9888.

[5] Kulvicius, T., Zhang, D., Poustka, L., Bölte, S., Jahn, L., Flügge, S., ... & Marschik, P. B. (2025). Deep learning empowered sensor fusion boosts infant movement classification. Communications Medicine, 5(1), 16.

Poster

LSTM-based Emotion Recognition: A Comparative Study of Single-Signal and Multi-Sensor Fusion Approaches

Mostafa Haghi[a,1], Soheil Khooyooz[a], Yiting Wei[a], and Nima Taherinejad[a]

[a] *Eclectx Team, Institute for Computer Engineering, Heidelberg University, Heidelberg*

1. Introduction

Emotion recognition contributes to advancing human-computer interaction and understanding human affective states, with applications in mental health monitoring and personalized healthcare. A key challenge lies in balancing the reliable, unobtrusive signals and designing models that effectively capture the complexities of emotional states. We propose using temporal neural networks, long short-term memory (LSTM) and gated recurrent unit (GRU) models, to classify emotional states into four arousal-valence categories and analyze the correlation of various signals with these categories, both individually and through multimodal fusion [1].

2. Methods

We used the ASCERTAIN dataset, a comprehensive multimodal resource that provides a connection between emotional states and physiological responses. The dataset consists of physiological signals, including Electroencephalogram (EEG), Electrocardiogram (ECG), and Galvanic Skin Response (GSR), from 58 adult participants who viewed 36 effective short movie clips. Participants provided self-reported ratings of arousal and valence for each video [2]. We preprocessed the data. Emotional states were labeled using arousal and valence dimensions, resulting in four categories: happy, angry, calm, and sad. We trained LSTM and GRU models (each with 64 units) for classification, using categorical cross-entropy loss and the Adam optimizer. We performed all the experiments on a local server with 2x Intel Xeon Gold 6526Y 16-Core CPU.

3. Results

LSTM outperformed GRU, particularly for high-valence emotions like happy/calm, with the ECG & GSR fusion yielding peak performance (96% accuracy, F1 = 0.97). GRU showed better balance for low-arousal low-valence states, outperforming LSTM in recognizing sad using GSR (F1 = 0.83 vs. 0.42). GSR was the most effective individual signal, reaching 95% accuracy with LSTM and 89% with GRU. ECG also demonstrated high reliability, especially with LSTM (F1 = 0.89 for angry with LSTM, 0.83 with GRU). Signal fusion enhanced performance, with ECG & GSR emerging as the best for both models.

4. Discussion

This study demonstrated that the LSTM model, paired with signals such as GSR and ECG, delivers superior performance over GRU models. GSR emerged as a simpler, lower-power, and more effective alternative for detecting medium-arousal emotions. Although Photoplethysmography (PPG) was not directly tested in this study, it presents an alternative for ECG in wearable emotion recognition applications due to its ease of integration and ability to capture comparable physiological features for high-arousal emotions [3].

[1] Corresponding Author: Mostafa Haghi, Mostafa.haghi@ziti.uni-heidelberg.de.

References

[1] King RC, Villeneuve E, White RJ, Sherratt RS, Holderbaum W, Harwin WS. Application of data fusion techniques and technologies for wearable health monitoring. Medical engineering & physics. 2017 Apr 1;42:1-2.

[2] Subramanian R, Wache J, Abadi MK, Vieriu RL, Winkler S, Sebe N. ASCERTAIN: Emotion and personality recognition using commercial sensors. IEEE Transactions on Affective Computing. 2016 Nov 4;9(2):147-60.

[3] Ismail SN, Aziz NA, Ibrahim SZ. A comparison of emotion recognition system using electrocardiogram (ECG) and photoplethysmogram (PPG). Journal of King Saud University-Computer and Information Sciences. 2022 Jun 1;34(6):3539-58.

Evaluating Transformer Models for ICD Code Embeddings in Predicting Clinical Outcomes

Jonas Harriehausen[a,d], Jacqueline Michelle Metsch[a,d], Zully Ritter[a], Miriam Cindy Maurer[a,d], Lisa Weller[b], Thorsten Pollmann[b], Matthias Kretzler[c], Thomas Grobe[b], Anne-Christin Hauschild[a,d,1]

[a] *Department of Medical Informatics, University Medical Center Goettingen*

[b] *aQua-Institute for Applied Quality Improvement and Research in Health Care, Goettingen*

[c] *BKK Dachverband*

[d] *Department of Predictive Deep Learning in Medicine and Healthcare, Justus-Liebig University Gießen*

1. Introduction

Over the last decade, Artificial Intelligence (AI) has achieved remarkable progress, revolutionizing medical research and applications, particularly in well-structured data modalities such as medical imaging. Another information-rich data resource is patients' medical history, which is typically encoded in the form of internationally standardized diagnosis and procedure codes, also called International Code Diseases (ICD) codes. Such diagnosis codes comprise a large part of the information in electronic health records (EHR) and medical insurance claims data, making these a crucial resource for data-driven decision-making in healthcare. However, challenges such as high variability and gaps in patient profiles, as well as the resulting data sparsity and high dimensionality, hinder the effective representations required in AI. Therefore, in contrast to the standard encoding by code or clinically defined categories, novel approaches that leverage Large Language Models (LLM) show promising results, for instance, Pat2Vec[5], trained on large-scale claims data, or BEHRT[4] in public EHR data.

2. Methods

In this study, we adapt a state-of-the-art transformer architecture called Bidirectional Encoder Representations from Transformers (BERT) [1] and tailor it from Language towards ICD codes to obtain a dense representation of a sequence of medical history events. We compare the novel BERT encodings to the Pat2Vec encodings and 241 one-hot encoded clinical categories in three prediction tasks: (1) mortality, (2) rehospitalization, and (3) future disease. Therefore, we train and evaluate these models on a publicly available EHR dataset called MIMIC-IV [3] and the large-scale KI-THRUST dataset comprising insurance claims from German matching hospital and general practitioner records. The KI-THRUST [2] project integrated approximately 1.4 million routine insurance patient records from four insurance agencies to evaluate the potential use for healthcare management and AI applications. At first, we train and apply the three encodings to generate a representation for each training set. Subsequently, logistic regression and random forest are trained on the encoded training datasets to predict the desired outcomes. Finally, all models are validated and tested on the KI-THRUST and mimic-IV test datasets to provide a comprehensive overview of the generalization abilities of all models.

3. Results & Discussion

Our preliminary results indicate that the novel BERT encodings performed equal to or significantly better than the clinical categories in the (1) mortality and (2) rehospitalization prediction tasks. In contrast, in the (3) future disease prediction tasks, BERT performed significantly worse than the clinical categories. Finally, the comparison with the Pat2Vec encodings showed significant superiority of the BERT and categorical encodings in all tasks except the (2) mortality prediction task on MIMIC-IV, where the Pat2Vec performed better than the categorical encoding. Overall, the BERT encoding tends to show higher performance than the clinical categories, and both show higher performance than the Pat2Vec encoding. Moreover, the higher performance of the models based on the clinical

categories encodings for the (3) future disease prediction task might point to a simplified classification of the future diseases based on the strong correlation to the prior diseases. In summary, we conclude that the choice of model for future tasks strongly depends on a trade-off between maximum performance from the BERT encoder and the generalizibility and interpretability of the encoding based on clinical categories.

References /Literatur

[1] J. Devlin. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[2] KI-THRUST – Potenziale KI-gestützter Vorhersageverfahren auf Basis von Routinedaten. G-BA Innovationsfonds, Nr. 01VSF20014.

[3] A. E. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, S. Hao, B. Moody, B. Gow, et al. Mimic-iv, a freely accessible electronic health record dataset. Scientific data, 10(1):1, 2023.1

[4] Y. Li, S. Rao, J. R. A. Solares, A. Hassaine, R. Ramakrishnan, D. Canoy, Y. Zhu, K. Rahimi, and G. Salimi-Khorshidi. Behrt: transformer for electronic health records. Scientific Reports, 10(1):7155, 2020.

[5] E. Steiger and L. E. Kroll. Patient embeddings from diagnosis codes for health care prediction tasks: Pat2vec machine learning framework. JMIR AI, 2:e40755, 2023.

Federated documentation and data use within NUM and EU project Screen4Care

Miriam Hübner [a,1], Urban Geisthoff [b,g], Jürgen Schäfer[c], Romina Blasini[d], Richard Röttger [e], Jana Zschüntzsch [f], Simone Kesten [g], Steffen Sander [a], Anne Schwerk [a,h], Marie Gebhardt [i], Josef Schepers [a]

[a] Berlin Institute of Health (BIH) @ Charité, [b] Universitätsklinikum Gießen-Marburg; HNO Marburg and VASCERN HHT Reference Center, [c] UKGM Zentrum für Unerkannte und Seltene Erkrankungen Marburg, [d] UK GM DIZ Gießen, [e] University of Southern Denmark, Department of Mathematics and Computer Science, [f] Universitätsmedizin Göttingen, Klinik für Neurologie, [g] Morbus Osler Selbsthilfe e.V., [h] IU Internationale Hochschule GmbH, [i] FDPG c/o TMF e.V.

1. Introduction

Using the example of the often unrecognised and underdiagnosed rare diseases (RDs) Hereditary Hemorrhagic Telangiectasia (HHT) and Hypophosphatasia (HPP), with a diagnostic latency that often lasts several years, we will show how the secondary use of routine data can be of benefit to people with RDs within the Network University Medicine (NUM). This benefit shall be increased by the dissemination of the European 'Set of Common Data Elements for Rare Diseases' in routine EHR and with the support of the German Portal for Medical Research Data (FDPG) as a pillar of the federated NUM4Rare registry-system. The collaborative, cross-institutional and data protection-compliant use of RD data from care processes for clinical epidemiology, diagnostic support, therapy communication and study development is continuously being expanded.

1.1. Use case and aim

For the prototype studies Screen-for-HHT (Orpha:774; ICD:I78.0) and Screen-for-HPP (Orpha:436), the networked data repositories of the data integration centres (DIC) of several MII institutions (Medical Informatics Initiative) are used as they are currently available with coded diagnoses, symptoms, phenotypes and laboratory values. The studies investigate the **potential for identifying clues to undetected rare diseases in routine data** (e.g. based on combinations of symptoms or laboratory values such as nosebleeds, neurological deficits, pain, anemia, low alkaline phosphatase, …) through the assistance of the FDPG.

1.2. Gradual Realisation

In a resource-saving, continuous improvement process, the data level of the future integrated NUM4Rare registry system in the DICs for all RD patients will be raised from:

- Level 0 (without ORPHAcodes) to
- Level 1 through the nationwide mandatory Orpha coding of RD diagnoses.
- Level 2 is aimed for by collecting an EU-compliant minimum data set for RD in many departments and for numerous RD patients.
- Level 3 is created by implementing hybrid documentation with the inclusion of more sophisticated registry elements in clinical workplace systems.

The RD module in the MII core dataset with basic functions and extension options serves as a guideline for data collection. We are investigating whether the uniform collection of data on 'maladies rares' at clinical workplaces ('dossier patient informatisé') in France, which are merged into the central BNDMR registry via the 'mode connecté', can be reproduced in Germany by means of an EU- and FAIR-compatible federated multiple use.

[1] Miriam Hübner, miriam-rebecca.huebner@charite.de

References

[1] Schepers, J., Hübner, M., Zoch, M., Rehburg, F., Finis, C., Zschüntzsch J., Mimouni Y., & Rosenbichler B. (2024). Vernetzte medizinische Forschung für Seltene Erkrankungen. Workshop 69. GMDS Jahrestagung Dresden. doi: 10.3205/24gmds227. https://www.egms.de/static/en/meetings/gmds2024/24gmds227.shtml

[2] Geisthoff, U. W., Mahnken, A. H., Denzer, U. W., Kemmling, A., Nimsky, C., & Stuck, B. A. (2024). Hereditary Hemorrhagic Telangiectasia (Osler's Disease): Systemic, Interdisciplinary, Relatively Common—and Often Missed. *Deutsches Arzteblatt international*, *121*(18), 601–607. https://doi.org/10.3238/arztebl.m2024.0111

[3] Schepers, J., Fleck, J., & Schaaf, J. (2022). Die Medizininformatik-Initiative und Seltene Erkrankungen: Routinedaten der nächsten Generation für Diagnose, Therapiewahl und Forschung [The medical informatics initiative and rare diseases: next-generation routine data for diagnosis, therapy selection and research]. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, *65*(11), 1151–1158. https://doi.org/10.1007/s00103-022-03606-y

[4] Prokosch, H. U., Gebhardt, M., Gruendner, J., Kleinert, P., Buckow, K., Rosenau, L., & Semler, S. C. (2023). Towards a National Portal for Medical Research Data (FDPG): Vision, Status, and Lessons Learned. *Studies in health technology and informatics*, *302*, 307–311. https://doi.org/10.3233/SHTI230124

[5] Ferlini, A., Gross, E. S., Garnier, N., & Screen4Care consortium (2023). Rare diseases' genetic newborn screening as the gateway to future genomic medicine: the Screen4Care EU-IMI project. *Orphanet journal of rare diseases*, *18*(1), 310. https://doi.org/10.1186/s13023-023-02916-x

[6] Schepers, J., Geisthoff, U., Finis, C., Kühnen, P., Mai, K., Schwerk, A., Zschüntzsch, J., & Zoch, M. (2025). Integrierte, standardisierte, verteilte Dokumentationen und föderierte Auswertungen im NUM am Beispiel der Seltenen Erkrankungen HHT und HPP. Workshop 1-1-1 NUM Convention 2025. https://num.invitepeople.com/public/events/bec0c66aa7/seminars/cd4582b995

Benchmarking methods for 2D infant pose estimation

Lennart JAHN[a,b,1], Sarah FLÜGGE[a], Dajie ZHANG[c,d], Luise POUSTKA[c], Sven BÖLTE[e,f,g], Florentin WÖRGÖTTER[b], Peter B. MARSCHIK[a,c,d,e], Tomas KULVICIUS[a,b,c]

[a] *Child and Adolescent Psychiatry and Psychotherapy, University Medical Center Göttingen; German Center for Child and Adolescent Health (DZKJ), Leibniz ScienceCampus Göttingen, Germany*

[b] *University of Göttingen, III Institute of Physics – Biophysics, Göttingen, Germany*

[c] *Department of Child and Adolescent Psychiatry, University Hospital Heidelberg, Heidelberg University, Heidelberg, Germany*

[d] *iDN – interdisciplinary Developmental Neuroscience, Division of Phoniatrics, Medical University of Graz, Graz, Austria*

[e] *Center of Neurodevelopmental Disorders (KIND), Department of Women's and Children's Health, Center for Psychiatry Research, Karolinska Institutet & Region Stockholm, Stockholm, Sweden*

[f] *Child and Adolescent Psychiatry, Stockholm Health Care Services, Region Stockholm, Stockholm, Sweden*

[g] *Curtin Autism Research Group, Curtin School of Allied Health, Curtin University, Perth, Australia*

1. Introduction

We observe ever increasing efforts to automate clinical methods for early diagnosis of developmental disorders, like General Movement Assessment (GMA) [1], a video-based tool to classify infant motor functioning. Optimal pose estimation, is a crucial part of the automated pipeline.
To find the best method for infant pose estimation, we compare the performance of readily available generic pose estimators and specialized infant-pose models, including one retrained on our own data. Furthermore, we evaluate the best choice of viewing angle for optimal recordings, i.e., the conventional diagonal view as used in GMA vs. a top-down view.

2. Methods

We used 4500 annotated video-frames selected from 75 recordings of infant spontaneous motor functions from 4 to 16 weeks.
On those, we performed pose estimation using generic pose estimators, namely OpenPose [2], MediaPipe, HRNet and ViTPose [3] as well as specifically retrained models, AggPose [4], AGMA-HRNet48 [5] and ViTPose retrained on our dataset.
To determine which pose estimation method and camera angle yield the best pose estimation accuracy, we computed and compared two measures: the error with respect to human annotations and the percentage of correct key-points (PCK).

3. Results

The results show that the best performing generic model trained on adults, ViTPose, also performs best on infants. We see no improvement from using specific infant-pose estimators over the generic pose estimators on our infant dataset. However, when retraining a generic model on our data, there is a significant improvement in pose estimation accuracy.
The pose estimation accuracy obtained from the top-down view is significantly higher than that obtained from the diagonal view (the standard view for GMA) regardless of model. Retraining ViTPose specifically on this view alone does not lead to higher accuracy.

[1] Corresponding Author: Lennart Jahn, lennart.jahn@phys.uni-goettingen.de.

4. Conclusions

The results indicate limited generalization capabilities of infant-pose estimators to other infant datasets.
This means that one should be careful when choosing infant pose estimators and using them on infant datasets which they were not trained on. It might be better to just use generic pose estimators if retraining on an own dataset is not possible or feasible.
The camera angle comparison suggests that an additional, non-standard, top-down view should be included in recording setups for automated GMA research for more accurate pose estimation.

References

[1] Prechtl, H. F., Einspieler, C., Cioni, G., Bos, A. F., Ferrari, F., & Sontheimer, D. (1997). An early marker for neurological deficits after perinatal brain lesions. *Lancet* (London, England), 349(9062), 1361–1363. https://doi.org/10.1016/S0140-6736(96)10182-3

[2] Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2021). OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE transactions on pattern analysis and machine intelligence*, 43(1), 172–186. https://doi.org/10.1109/TPAMI.2019.2929257

[3] Xu, Y., Zhang, J., Zhang, Q., & Tao, D. (2022). ViTPose: Simple vision transformer baselines for human pose estimation. In *Proceedings of the 36th International Conference on Neural Information Processing Systems* (pp. 38571–38584). Curran Associates Inc.

[4] Cao, X., Li, X., Ma, L., Huang, Y., Feng, X., Chen, Z., … Cao, J. (2022). AggPose: Deep Aggregation Vision Transformer for Infant Pose Estimation. In *International Joint Conferences on Artificial Intelligence* (pp. 5045–5051). https://doi.org/10.24963/ijcai.2022/700

[5] Soualmi, A., Ducottet, C., Patural, H., Giraud, A., & Alata, O. (2024). A 3D pose estimation framework for preterm infants hospitalized in the Neonatal Unit. *Multimedia Tools and Applications*, 83(8), 24383-24400. https://doi.org/10.1007/s11042-023-16333-6

“Now, I understand!”: Digital support in 24-hour care – learnings from co-creating and friendly user testing

Cornelia Schneider [a,b,1], Birgit Trukeschitz [c]

[a] *University of Applied Sciences Wiener Neustadt, Wiener Neustadt, Austria*

[b] *UMIT TIROL – Private University for Health Sciences and Health Technology, Hall in Tirol, Austria*

[c] *WU Vienna University of Economics and Business, Vienna, Austria*

1. Introduction

Demographic changes will challenge health and long-term care systems in coming decades (EC et al., 2021). Many countries prioritize home care over institutional care [2], including 24-hour care [3]. 24-hour care is challenged by language barriers, remote problem-solving, and climate-related issues like heat waves. A study emphasizes the potential of digital tools in 24-hour care [4]. Our project aims to develop a communication app, called *24/7 Dialog*, with two core functionalities: (1) GDPR-compliant chats and video calls (partly with AR support), and (2) translation services, both stand-alone and integrated into the chat and video call functions.

2. Methods

The *24/7 Dialog* app was developed using the “evidence-based and user-centred innovation process for AAL projects” [5]. Three co-creation workshops using the Walt Disney method with 12 participants (relatives, nurses, and 24h carers) guided the initial design. Flutter and Kotlin were used for the agile development (Scrum) of the app. The first prototype was tested within the consortium and in a friendly user test with 15 participants. Tests were conducted in the household of the care dependent person using interactive scenarios embedded in LimeSurvey®.

3. Results

Co-creation workshops identified key features for the three main functions of the *24/7 Dialog* app. Potential users placed an emphasis on data protection and simple user interfaces. The friendly user test highlighted rollout issues, such as unread SMS activations, forgotten passwords, and overlooked installation manuals. Usability challenges included navigation confusion and varying performance of language services across Eastern European languages. Insights from these tests informed the refinement of technical solutions for further testing this year.

4. Discussion

The two main functions of the *24/7 Dialog* app have proved useful in all tests so far. However, not only the technical development but also the onboarding (download of app and installation) and the app’s implementation in the 24h care workflows are key. Available language services showed considerable limitations for Eastern European languages such as Romanian, Bulgarian, and Slovak. Future development iterations will focus on improving the usability, the requirements for successful implementation to ensure accessibility and ease of use for 24-hour care networks.

[1] Corresponding Author: Cornelia Schneider, University of Applied Sciences Wiener Neustadt, Wiener Neustadt, Austria, E-Mail: cornelia.schneider@fhwn.ac.at.

References

[1] European Commission, Directorate-General for Employment, and Social Affairs and Inclusion, "Long-term care report – Trends, challenges and opportunities in an ageing society," Luxembourg, Jun. 2021. doi: 10.2767/677726.

[2] S. Spasova, R. Baeten, S. Coster, D. Ghailani, R. Peña-Casas, and B. Vanhercke, "Challenges in long-term care in Europe: A study of national policies," 2018. doi: 10.2767/84573.

[3] U. Famira-Mühlberger, "Die Bedeutung der 24-Stunden-Betreuung für die Altenbetreuung in Österreich," Wien, 2017. [Online]. Available: http://www.wifo.ac.at/wwa/pubid/60718%0AAbgerufen am 1.10.2021

[4] C. Hauser, E. Kupka-Klepsch, E. Haslinger-Baumann, and F. Werner, "Designing digital tools for quality assurance in 24-hour home-care in Austria," *Human Affairs*, vol. 32, no. 2, pp. 213–227, 2022, doi: 10.1515/humaff-2022-0016.

[5] C. Schneider, V. Venek, H. Rieser, S. Jungreitmayr, and B. Trukeschitz, "„Fit-mit-ILSE" für Senior*innen: User-Centred Design Prozess und Prototyp des Active and Assisted Living Systems," in *Aktives Altern im digitalen Zeitalter*, S. Ring-Dimitriou and M. Dimitriou, Eds., Wiesbaden: Springer Fachmedien Wiesbaden, 2022, pp. 105–132. doi: 10.1007/978-3-658-34970-7_5.

Unlocking German Clinical Text Data: Advanced De-Identification for LLM Training

Marvin Seiferling [a], Christina Lohr [b], Christoph Dieterich [a]

[a] *Klaus Tschira Institute for Computational Cardiology, Universitätsklinik Heidelberg, Heidelberg, Germany*

[b] *Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE), Universität Leipzig, Leipzig, Germany*

1. Introduction

The Medical Informatics Initiative (MII) [1] represents the most comprehensive research program in Germany aimed at integrating clinical patient data across multiple institutions. While initial efforts primarily focused on structured data, the GeMTeX (German Medical Text Project) [2] initiative extends this work to unstructured free-text patient documents. Unstructured clinical free text has remained largely untapped due to the significant challenges posed by data protection regulations. GeMTeX addresses this gap by creating a publicly accessible corpus of German clinical texts to enable the training of large language models tailored to the healthcare domain.
This cross-site initiative spans six major clinical locations, each contributing patient histories to construct a composite corpus. As no real German patient documents have ever been publicly released at this scale, stringent measures are necessary to comply with data protection regulations and safeguard patient privacy. Central to this effort is the process of de-identification, wherein sensitive information—such as names, addresses, and patient identifiers—is removed to protect patient privacy [3,4].

2. Methods

To achieve this, GeMTeX employs a manual annotation process supported by pre-annotation NLP software tools to detect the sensitive patient information embedded in the text. These annotated documents then undergo a de-identification process in which the sensitive information is removed. To accomplish this, we adapted an existing pipeline [5], to meet the specific requirements of the GeMTeX project. This ensures that no sensitive patient information is inadvertently leaked.

3. Results

We developed a de-identification pipeline tailored for GeMTeX, capable of effectively removing privacy-sensitive information from the corpus while providing multiple formats for masking the identified data. This enables participating sites to continue the semantic annotation process, ultimately supporting the successful release of the corpus.

4. Discussion

Simple PHI removal can diminish the corpus's utility for downstream NLP tasks. To address this, we are working on extending the current de-identification pipeline to generate contextually appropriate surrogates. By preserving the statistical and linguistic characteristics of the original data, realistic surrogates enhance the effectiveness of LLM training while masking potential residual sensitive information that may have been overlooked during the manual annotation process.
-
Funding: BMBF, GeMTeX (01ZZ2314I, 01ZZ2314B).

References

[1] S. Semler, F. Wissing, und R. Heyder, „German Medical Informatics Initiative: A National Approach to Integrating Health Data from Patient Care and Medical Research“, Methods Inf Med, Bd. 57, Nr. S 01, S. e50–e56, Juli 2018, doi: 10.3414/ME18-03-0003.

[2] F. Meineke, L. Modersohn, M. Loeffler, und M. Boeker, „Announcement of the German Medical Text Corpus Project (GeMTeX)“, in Studies in Health Technology and Informatics, M. Hägglund, M. Blusi, S. Bonacina, L. Nilsson, I. Cort Madsen, S. Pelayo, A. Moen, A. Benis, L. Lindsköld, und P. Gallos, Hrsg., IOS Press, 2023. doi: 10.3233/SHTI230283.

[3] P. Richter-Pechanski, S. Riezler, and C. Dieterich, "De-Identification of German Medical Admission Notes," in Studies in Health Technology and Informatics, vol. 253, "German Medical Data Sciences: A Learning Healthcare System," pp. 165–169, 2018. doi: 10.3233/978-1-61499-896-9-165.

[4] C. Lohr u. a., „De-Identifying GRASCCO – A Pilot Study for the De-Identification of the German Medical Text Project (GeMTeX) Corpus“, in Studies in Health Technology and Informatics, R. Röhrig, N. Grabe, U. H. Hübner, K. Jung, U. Sax, C. O. Schmidt, M. Sedlmayr, und A. Zapf, Hrsg., IOS Press, 2024. doi: 10.3233/SHTI240853.

C. Lohr, E. Eder, und U. Hahn, „Pseudonymization of PHI Items in German Clinical Reports“, in Studies in Health Technology and Informatics, J. Mantas, L. Stoicu-Tivadar, C. Chronaki, A. Hasman, P. Weber, P. Gallos, M. Crişan-Vida, E. Zoulias, und O. S. Chirila, Hrsg., IOS Press, 2021. doi: 10.3233/SHTI210163.

Prof. Dr. Martin Dugas, Heidelberg (Chair)
Prof. Dr. Petra Knaup-Gregori, Heidelberg (Co-Chair)
Prof. Dr. Felix Balzer, Berlin
Prof. Dr. Jörg Becker, Münster
Prof. Dr. Kerstin Denecke, Bern
Prof. Dr. Toralf Kirsten, Leipzig
Prof. Dr. Dr. Michael Marschollek, Hannover
Prof. Dr. Benjamin Meder, Heidelberg
Prof. Dr. Uta Merle, Heidelberg
Prof. Dr. Carsten Müller-Tidow, Heidelberg
Prof. Dr. Christoph Michalski, Heidelberg
Prof. Dr. Fabian Prasser, Berlin
Prof. Dr. Rüdiger Pryss, Würzburg
Prof. Dr. Ulrich Sax, Göttingen
Prof. Dr. Martin Sedlmayr, Dresden
Prof. Dr. Hanna Seidling, Heidelberg
Prof. Dr. Julian Varghese, Magdeburg
Prof. Dr. Markus Weigand, Heidelberg

Scientific Committee