UNIVERSITÄTS KLINIKUM HEIDELBERG

Publication Series:
Heidelberg Spring Symposium Medical Informatics

INSTITUT FÜR MEDIZINISCHE INFORMATIK

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

2. Heidelberger Frühjahrssymposium Medizinische Informatik

8. Mai 2024

Abstractband

This publication is the second 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 (2024): 2. Heidelberger Frühjahrssymposium Medizinische Informatik. Heidelberg University Library. https://doi.org/10.11588/heidok.00034810

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

Published by Heidelberg University Library, 2024

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

doi: https://doi.org/10.11588/heidok.00034810

Inhaltsverzeichnis

Hauptvorträge

Impulsvorträge

Poster

Hauptvorträge

Unsupervised release of clinical Information to Patients: Disaster or Blessing?

Christoph U. Lehmann [a,b]

[a] *University of Texas Southwestern Medical Center,* Dallas, USA

[b] *Institut für Medizinische Informatik, Universitätsklinikum Heidelberg, Heidelberg, Deutschland*

1. Introduction

In 2009, the US Congress passed legislation that invested US$ 36 Billion in the implementation of electronic health records (EHR). Subsequently, as of 2021, 96% of hospitals and 78% of physician practices used EHRs. In 2016, the 21st Century Cure Act introduced the concept of information blocking and in 2021 its provisions went into effect making information blocking illegal in the United States. As a result, patients now are entitled to immediate access to their health information including clinical notes, problems, procedures, medications, and laboratory and imaging results. To better understand the implications of this legislation, we conducted three studies exploring the effects of "Open Results".

2. Methods

In the first study we reviewed audit logs of patients who presented in the emergency department in the year following the 21st Century Cures Act going into effect. We explored how many patients reviewed their health information while they were still in the ED [1] Subsequently, we extended the study to and additional 11 teaching hospitals and 24 academic affiliated EDs across the US. [2] Finally, we sent a survey to patients and their caregivers, who accessed their results during the same period. We queried them regarding their perspectives regarding immediate access to their information.[3]

3. Results

Of patients already enrolled in the Patient Portal, 18.8% reviewed their personal health information while they were still in the ED at UT Southwestern in Dallas. Older age, increased social vulnerability, male sex, Hispanic ethnicity, Black race, and public or lack of insurance reduced the odds that a patient reviewed the information. When expanding the study across the US, we found that 17.4% of patients logged into the portal, 14.1% viewed test results, and 2.5% viewed clinical notes while still in the ED. Patients who were male, Black, or without commercial insurance had lower odds of logging into the portal, viewing results, and viewing clinical notes.

When surveying patients, who had reviewed their results (laboratory or imaging results were most commonly viewed) in the portal, most respondents (95.7%), including 95.3% of individuals who received nonnormal results, preferred to immediately receive test results through the portal. Few respondents (7.5%) reported that reviewing results before they were contacted by a health care practitioner increased worry, though increased worry was more common among respondents who received abnormal results (16.5%) than those whose results were normal (5.0%). The result of the pooled model for worry as a function of test result normality was statistically significant (odds ratio [OR], 2.71; 99% CI, 1.96-3.74), suggesting an association between worry and nonnormal results.

4. Discussion

A significant number of patients in the US review their results while they are still in the ED often before a provider has seen the results. Regarding patient attitudes and preferences toward receiving immediately released test results via a patient portal, most respondents preferred to receive test results via the patient portal despite viewing results prior to discussion with a health care professional. This preference persisted among patients with nonnormal results.

References

[1] Turer RW, Martin KR, Courtney DM, Diercks DB, Chu L, Willett DL, Thakur B, Hughes A, Lehmann CU, McDonald SA. Real-Time Patient Portal Use Among Emergency Department Patients: An Open Results Study. Appl Clin Inform. 2022 Oct;13(5):1123-1130. doi: 10.1055/a-1951-3268. Epub 2022 Sep 27. PMID: 36167337; PMCID: PMC9713300.

[2] Turer RW, McDonald SA, Lehmann CU, Thakur B, Dutta S, Taylor RA, Rose CC, Frisch A, Feterik K, Norquist C, Baker CK, Nielson JA, Cha D, Kwan B, Dameff C, Killeen JP, Hall MK, Doerning RC, Rosenbloom ST, Distaso C, Steitz BD. Real-Time Electronic Patient Portal Use Among Emergency Department Patients. JAMA Netw Open. 2024 May 1;7(5):e249831. doi: 10.1001/jamanetworkopen.2024.9831. PMID: 38700859; PMCID: PMC11069088.

[3] Steitz BD, Turer RW, Lin CT, MacDonald S, Salmi L, Wright A, Lehmann CU, Langford K, McDonald SA, Reese TJ, Sternberg P, Chen Q, Rosenbloom ST, DesRoches CM. Perspectives of Patients About Immediate Access to Test Results Through an Online Patient Portal. JAMA Netw Open. 2023 Mar 1;6(3):e233572. doi: 10.1001/jamanetworkopen.2023.3572. PMID: 36939703; PMCID: PMC10028486.

Bedürfnisorientierte App für ehemalige Intensivpatient/innen- Optimierung durch Feedback

Denise Molinnus [a,b,1], Angelique Kurth [a,b], Anne Mainz [a,c], Sven Meister [a,c,d], Volker Lowitsch [a], Matthias Nüchter [a,e], Gernot Marx [a,b] und Johannes Bickenbach [a,b]

[a] *DISTANCE-Konsortium der Deutschen Medizininformatik-Initiative*

[b] *Klinik für Operative Intensivmedizin und Intermediate Care, Universitätsklinikum RWTH, Aachen, Deutschland*

[c] *Gesundheitsinformatik, Fakultät für Gesundheit/Fakultät für Medizin, Universität Witten/Herdecke, Witten, Deutschland*

[d] *Abteilung Gesundheitswesen, Fraunhofer-Institut für Software- und Systemtechnik, Dortmund, Deutschland*

[e] *Universität Leipzig, LIFE Management Cluster, Leipzig, Deutschland*

[f] *Healthcare IT Solutions GmbH, Deutschland*

1. Einführung

Im Rahmen des Digitalen FortschrittsHub Gesundheit DISTANCE (Digital Smart Hub for Advanced Connected Care) wird der interoperable Datenaustausch auf medizinische Einrichtungen der regionalen Versorgung erweitert [1]. Dazu wurde im Use Case PICOS (Post Intensive Care Outcome Surveillance) eine patientenorientierte Mobile Health (mHealth)-Anwendung konzipiert und entwickelt. Diese ermöglicht erstmals die Erfassung einer Vielzahl von Vitalwerten, Aktivität sowie kognitiver und psychomentaler Einschränkungen als Längsschnittdaten bei Intensivpatient/innen, die anfällig für das Post Intensive Care Syndrom (PICS) sind [2]. Derzeit gibt es nur begrenzte Richtlinien zur Identifizierung und Behandlung von PICS, und es besteht ein dringender Bedarf an weiteren Erkenntnissen, um ein besseres konzeptionelles Verständnis zu erlangen, das zu einer Verbesserung der Langzeitbehandlung führen, und Patient/innen bewusst in ihren eigenen Krankheitsverlauf einbinden soll.

2. Methoden

Die Evaluation des Prototypen der PICOS App erfolgte im Rahmen einer Vorstudie, die den Prinzipien für interaktive Systeme des “Usability Engineerings” entspricht. Dieser Prototyp wurde der Zielgruppe vorgestellt, um zu überprüfen, ob die App nutzbar ist und wie deren Attraktivität mit dem Ziel der Nutzerakzeptanz gesteigert werden kann. In einem Zeitraum von 11 Monaten wurden insgesamt 123 Teilnehmende aus Krankenhäusern unterschiedlicher Versorgungsstufen rekrutiert. Diese wurden nach demografischen Daten, Interaktion mit Technologie und Wahrnehmung des Prototyps in Bezug auf Motivation, Benutzerfreundlichkeit und Leistungserwartung befragt.

3. Ergebnisse

Die Umfrage lieferte Einblicke in die Zielgruppe und ihre Anforderungen an die PICOS App. Die überwiegende Mehrheit der Patient/innen (92,7 % der Befragten) besaß und nutzte ein Smartphone. Allerdings würde die Hälfte (49,5 %) der Befragten bei der Nutzung der App Unterstützung von Familienmitgliedern oder Pflegepersonen suchen. Teilnehmende empfanden den Prototyp sowohl nützlich als auch einfach zu bedienen, was auf den Erfolg der finalen PICOS-App und deren tatsächliche Nutzung durch Patienten hinweist.

4. Diskussion

Diese Erkenntnisse sollen dabei helfen, das tatsächliche Roll-Out der finalen PICOS-App an 13 Standorten zu erleichtern. Über den Nutzen für Teilnehmende hinaus werden Forschende, Informatiker/innen und Fachkräfte in der Medizin an den Universitätskliniken in Zukunft die Möglichkeit haben, über die PICOS-App gesammelte Daten zu nutzen. Durch die Bereitstellung und Analyse von Langzeitbeobachtungsdaten mit Methoden der künstlichen Intelligenz können sie klinische Zustände und Behandlungsprozesse für diese spezielle Patientengruppe optimieren.

References

[4] Molinnus D, Kurth A, Lowitsch V, Marx G, Bickenbach J. Towards an Advanced Digital Infrastructure Within the Non-University Sector Demonstrated by the PICOS App. Stud Health Technol Inform. 2023 May 18; 302:366-367. doi: 10.3233/SHTI230143 PMID: 37203687.

[5] Ramnarain D, Aupers E, den Oudsten B, Oldenbeuving A, de Vries J, Pouwels S. Post Intensive Care Syndrome (PICS): an overview of the definition, etiology, risk factors, and possible counseling and treatment strategies. Expert Rev Neurother. 2021 Oct;21(10):1159-1177. doi: 10.1080/14737175.2021.1981289. Epub 2021 Sep 22. PMID: 34519235.

Advancing Personalized Medicine through N-of-1 Trials: Introducing the StudyU Platform

Johannes Vedder [a, 1], Stefan Konigorski [a, 2]

[a] *Hasso Plattner Institute, University of Potsdam, Potsdam, Germany*

1. Introduction

Traditional randomized controlled trials (RCTs) are widely considered the gold standard for clinical trials in medical therapy. However, RCTs only allow conclusions to be drawn about average effects in the population and do not provide treatment recommendations for individuals.
N-of-1 trials provide a unique approach to patient-oriented interventions in personalized medicine. Each participant completes their own study in a crossover design to ensure objectivity and draw accurate, individualized conclusions [1]. Bayesian autoregressive models are commonly used as statistical inference in N-of-1 trials [2]. Historically, the high cost and effort required to conduct N-of-1 trials has hindered their widespread use. Digitization has facilitated data collection and made analysis more cost-effective on an individual basis.

2. Methods

Here, we introduce the StudyU platform (https://studyu.health), which consists of the StudyU Designer and the StudyU App [3]. The platform supports scientists and physicians in conducting digital N-of-1 trials for their patients. The StudyU Designer enables the creation of N-of-1 trials, recruitment of participants, and analysis of results. It also allows sharing of studies through collaboration with other researchers. The StudyU App enables individuals to participate in digital N-of-1 trials securely and anonymously through their mobile devices. StudyU is developed open source and can be used free of charge.

3. Results

As the first scalable platform for N-of-1 trials, StudyU is currently being used in several studies in Germany [4], the United States, Australia, and Ghana. Participants include healthy individuals, as well as patients with depression, breast cancer, and chronic pain. We provide an overview of these studies, covering both inpatient and outpatient settings. The use of StudyU in an international environment has yielded valuable insights for the success of digital N-of-1 trials. Key factors include a user-friendly design, the provision of training materials, and continuous support. Additionally, various requirements for N-of-1 trial platforms are highlighted, such as data protection, security, interoperability, and adaptability in diverse contexts.

4. Discussion

StudyU is a platform that transforms the way N-of-1 trials can be conducted at the intersection of research and clinical treatment, fostering broader acceptance and utilization. By empowering patients and researchers alike, StudyU contributes to the advancement of personalized medicine and promotes a patient-centered approach to healthcare decision-making. Recent enhancements include the integration of multimodal data [5] and adaptive study designs.

[1] Johannes Vedder, johannes.vedder@student.hpi.de

[2] Corresponding Author: Stefan Konigorski, stefan.konigorski@hpi.de

References

[1] Nikles, J., & Mitchell, G. K. (2015). The Essential Guide to N-of-1 Trials in Health. In Springer eBooks. Springer Nature. https://doi.org/10.1007/978-94-017-7200-6

[2] Mirza, R., Punja, S., Vohra, S., & Guyatt, G. (2017). The history and development of N-of-1 trials. Journal of the Royal Society of Medicine, 110(8), 330–340. https://doi.org/10.1177/0141076817721131

[3] Konigorski, S., Wernicke, S., Slosarek, T., Zenner, A. M., Strelow, N., Ruether, D. F., Henschel, F., Manaswini, M., Pottbäcker, F., Edelman, J. A., Owoyele, B., Danieletto, M., Golden, E., Zweig, M., Nadkarni, G. N., & Böttinger, E. (2022). StudyU: A Platform for Designing and Conducting Innovative Digital N-of-1 Trials. Journal of Medical Internet Research, 24(7), e35884. https://doi.org/10.2196/35884

[4] Amke Müller, Konigorski, S., Meißner, C., Tahmine Fadai, Warren, C. V., Falkenberg, I., Kircher, T., & Nestoriuc, Y. (2023). Study protocol: combined N-of-1 trials to assess open-label placebo treatment for antidepressant discontinuation symptoms [FAB-study]. BMC Psychiatry, 23(1). https://doi.org/10.1186/s12888-023-05184-y

[5] Fu, J., Liu, S., Du, S., Ruan, S., Guo, X., Pan, W., Sharma, A., & Konigorski, S. (2023, February 15). Multimodal N-of-1 trials: A Novel Personalized Healthcare Design. ArXiv.org. https://doi.org/10.48550/arXiv.2302.07547

Harmonizing Real-World Data Networks with Common Data Models for Pan-Cancer Immune Checkpoint Inhibitor Insights in a Multicenter Study

Clara L. Oeste[a], Iege Bassez[a], Dries Hens[a, 1]

[a] *LynxCare Inc., Leuven, Belgium*

1. Introduction

Oncology data is highly heterogeneous among different hospitals, making data standardization a priority for retrospective analytics. Much of the information is in free-text electronic health records (EHR), warranting the use of natural language processing (NLP) to leverage patient-level insights. We have established an ongoing network of Belgian sites to assess the use of immune checkpoint inhibitors (ICI) across different cancer types in a real-world population.

2. Methods

Structured and unstructured oncology data from participating Belgian hospitals were processed with LynxCare technology, including NLP and machine learning, on over 10 data sources extracted from 4 different EHR systems. OMOP-CDM databases were generated with 597 variables that were mapped to SNOMED-CT, including demographics, comorbidities, cancer diagnosis, tumor staging, performance status, oncology treatments and procedures, anatomical pathology, and adverse events. The OHDSI OncoRegimenFinder was used to assess treatment patterns. A total of 49,364 structured mappings and 7.1 million unstructured concepts were processed. Resulting databases were validated per hospital to ensure patient privacy. Inclusion criteria were adult cancer patients receiving ICIs between March 2017 and August 2022.

3. Results

Initial results included 1724 patients with a mean (SD) age of 66 (12) years and 66.2% males. Smoking status was 35.3% current smokers, 18.0% ex-smokers, 18.2% non-smokers, and 28.5% with status unknown. Detected ICI treatments (as monotherapy or in combination with other ICIs or antineoplastic drugs) were pembrolizumab (43.9%), followed in frequency by nivolumab (29.5%), atezolizumab (10.7%), ipilimumab + nivolumab (8.4%), durvalumab (4.5%), avelumab (2.0%), cemiplimab (0.8%), and dostarlimab (0.2%). The most frequent cancer types were lung cancer (49.8%), melanoma (9.1%), renal cancer (8.6%), bladder cancer (8.4%), and head and neck cancer (7.3%). Preliminary overall survival was 8-30 months, depending on cancer type and treatment.

4. Discussion

We confirm the feasibility of using automatic, AI-driven methods to extract and validate data to generate research-grade, OMOP-CDM harmonized hospital databases. Addition of unstructured data sources through NLP enriches traditional structured data analytics and enables high granularity per patient. We are broadening the scope to include comorbidities, anatomical pathology insights, and immune-related adverse events. Ongoing steps include expanding this approach to other European hospitals, increasing the representativity of the study.

[1] Corresponding Author: Dries Hens, LynxCare Inc. Tiensevest 132, 3000 Leuven, Belgium. dries.hens@lynx.care

Impulsvorträge

Benutzer*innenzentrierter Entwicklungsprozess eines Dashboards im Kontext der Melanom Behandlung

Eva Maria Hartmann[a,1], Dr. med. Georg Lodde[b], PD Dr. med. Elisabeth Livingstone[b], Prof. Dr. med. Dirk Schadendorf[b], Prof. Dr. rer. nat. Sabine Sachweh[a]

[a] *Fachhochschule Dortmund, Fachbereich Informatik, Dortmund, Deutschland*

[b] *Universitätsklinikum Essen, Klink für Dermatologie, Essen, Deutschland*

1. Motivation

Für eine optimale Patient*innenversorgung ist es für Ärzte und Ärztinnen unabdingbar, sich einen schnellen und umfassenden Überblick über den Patient*innenstatus zu verschaffen. Medizinische Dokumentationssysteme, die diesen Prozess unterstützen, sind in der EU weiterhin auf dem Vormarsch [1]. Jedoch verbleibt die Usability dieser Systeme auf einem schlechten Niveau [2]. Dieses Projekt hat es sich daher zum Ziel gesetzt, den Entwicklungsprozess eines Dashboards, das Ärzt*innen bei der Behandlung von Melanom-Patient*innen unterstützen soll, komplett kontextorientiert und nutzer*innenzentriert zu gestalten.

2. Methodik

Grundlegend folgt das Projekt dem iterativen Ansatz des design science research process (DSRP) [3]. Um diesen nutzer*innen- und kontextorientiert zu gestalten, wird zunächst der Kontext der Melanombehandlung mit Hilfe der Think Aloud Methode (Lautes Denken) und Contextual Inquiry (Teilnehmende Beobachtung) ausgiebig erforscht. Die dabei gesammelten qualitativen Daten dienen beispielsweise als Grundlage zur Ableitung von Anforderungen und dem Entwurf des Designs. Letzteres wird im Folgenden zur Erweiterung des vorhandenen Patient Dashboards des Universitätsklinikums Essen genutzt. Zuletzt erfolgt die Evaluation der Zufriedenheit und der Integration in den Kontext in drei Zyklen. Genutzt wird dabei eine Kombination aus der Think Aloud Methode und dem Post-Study System Usability Questionnaires (PSSUQ) mit insgesamt neun Studienteilnehmer*innen. Durch regelmäßiges Feedback zu Zwischenständen und die Einarbeitung dabei gewonnener Einsichten wird das Dashboard schrittweise für den Kontext und die Nutzer*innen optimiert und die Qualität der Anwendung sichergestellt.

3. Ergebnisse

Die Erforschung des Kontexts legte eine Reihe vorhandener Usability Probleme offen. Darüber hinaus konnte explizites Wissen über Prozessabläufe, die Nutzer*innen und die aufgabenabhängige Priorisierung von Daten gewonnen werden. Auf dieser Grundlage ist ein Dashboard-Entwurf entstanden, welcher aufgabenabhängig die wesentlichsten Informationen strukturiert auf einer Seite zusammenführt. Die angezeigten Daten sind kontextspezifisch benannt und in wiederverwendbaren Modulen gruppiert. Die Evaluation ist aktuell in der letzten Iteration und deutet mit einem Gesamtwert des PSSUQ aus der ersten Iteration von 2,9 auf eine erfolgreiche Integration des Dashboards in den Arbeitsalltag der Dermatoonkologen hin. In den einzelnen Kategorien wurden Bewertungen von 2,2, sowohl auf der Nützlichkeitsskala als auch für die Qualität der Nutzeroberfläche, und 2,6 für die Informationsqualität erzielt.

[1] Corresponding Author: Eva Maria Hartmann, University of Applied Sciences and Arts Dortmund, Department of Computer Science, Otto-Hahn-Strasse 27, 44227 Dortmund, E-Mail: eva.hartmann@fh-dortmund.de

4. Ausblick

Nach dem Abschluss der Optimierung auf das Melanom soll das Dashboard zunächst auf andere Hauttumore erweitert werden. Die generische Implementierung soll darüber hinaus eine einfache Anpassung an andere Krebsarten ermöglichen.

Literatur

[6] European Commission, RAND Europe, Open Evidence, and BDI Research, Benchmarking deployment of eHealth among general practitioners: final report: Publications Office, 2018.

[7] J. Viitanen, H. Hyppönen, T. Lääveri, J. Vänskä, J. Reponen, and I. Winblad, "National questionnaire study on clinical ICT systems proofs: physicians suffer from poor usability," *International journal of medical informatics*, vol. 80, no. 10, pp. 708–725, 2011, doi: 10.1016/j.ijmedinf.2011.06.010.

[8] K. Peffers et al., "Design Science Research Process: A Model for Producing and Presenting Information Systems Research," Proceedings of the first international conference on design science research in information systems and technology (DESRIST 2006), abs/2006.02763, pp. 83–106, 2006. [Online]. Available: https://arxiv.org/pdf/2006.02763

PathoBot – KI-gestützte Übertragungsanalyse zur Prävention von Krankenhausinfektionen

Marius Take[1], Armin Haas[2], Jan Liese[3], Christoph Becker[1]

[1] FZI Forschungszentrum Informatik, 76131 Karlsruhe, {take, christoph.becker}@fzi.de

[2] Code'n'ground AG, 89518 Heidenheim an der Brenz, armin.haas@codenground.de

[3] Universitätsklinikum Tübingen, 72016 Tübingen, jan.liese@med.uni-tuebingen.de

1. Einleitung

Übertragungen bakterieller und viraler Krankheitserreger sind in Krankenhäusern besonders gefährlich und haben weitreichende Folgen. Infektionen verursachen längere stationäre Aufenthalte von Patienten und erhöhen die Patientenmorbidität und -mortalität. In Deutschland beträgt die Fallzahl bis zu 600.000 pro Jahr – etwa 10.000 bis 20.000 Fälle enden tödlich [1].

Die Erkennung und Analyse entsprechender Infektionssituationen sind in Krankenhäusern häufig mit hohem personellem Aufwand verbunden. Komplexe Krankenhausprozesse, hunderte täglich anfallende mikrobiologische Laborbefunde und häufige Zimmer- und Stationswechsel von Patienten erschweren die Arbeit, sodass neben dem hohen Aufwand auch Übertragungen unentdeckt bleiben könnten. Im Rahmen des Forschungsverbundprojekts PathoBot werden diese Probleme adressiert und Ansätze erforscht die Abhilfe schaffen. Die resultierende Softwareanwendung wird Krankenhaushygienikern als essenzielles Hilfsmittel bei der Entdeckung, Analyse und Nachverfolgung von Erreger-Ausbrüchen unterstützen.

2. Methode

Zur Umsetzung der Zielsetzung werden vorhandene Datensätze, wie Daten aus dem Laborinformationsmanagementsystem (LIMS) oder dem Patientendatenmanagementsystem (PDMS), über standardisierte Schnittstellen regelmäßig und automatisiert abgefragt, aggregiert sowie analysiert.

Um die Untersuchung der umfangreichen Datenbezüge zu vereinfachen und insbesondere eine automatisierte Erkennung von Mustern in den Daten zu ermöglichen, sollen zwei KI-basierte Analyse-Bausteine den Kern der Softwareanwendung bilden. Aktuell wird an einem automatisierten Alarmierungssystem geforscht, welches eine frühzeitige Warnung bei ungewöhnlicher Häufung eines Erregers im Krankenhaus versendet. Vergleichbare Ansätze untersuchten Wissenschaftler unter anderem im Rahmen experimenteller Erprobungen bereits in verschiedenen Kliniken [2, 3, 4]. In PathoBot sollen entsprechende Techniken nun in eine interaktive, KI-gestützte Gesamtlösung einfließen.

Im Falle einer Alarmierung wird des Weiteren die automatisierte Ableitung geeigneter Maßnahmen angestrebt und durch ein zweites KI-basiertes System, zur Übertragungsanalyse und somit zur Identifikation von Infektionsketten, umgesetzt. Die Übertragungsanalyse soll auf automatisiert aus den Daten des PDMS erstellten Bewegungsprofilen der Patienten basieren und potenzielle Übertragungen auch jenseits eines direkten Kontakts identifizieren. Hierzu wird ein neuronales Netz trainiert, welches jeweils zwei Bewegungsprofile abhängig des Erregers auf Zusammenhänge überprüft. Durch eine iterative Durchführung dieser Prüfung und der daraus resultierenden Auflistung bezüglich des Erregerausbruchs gefährdeter Patienten, können mit geringem zeitlichem Aufwand akkurate Maßnahmen zur Einschränkung des Infektionsgeschehens, wie zusätzliche Tests, Quarantänen etc., angeordnet werden. Die KI-basierte Übertragungsanalyse soll darüber hinaus zur Aufarbeitung bereits bekannter Infektionsketten dienen und ermöglichen Verbesserungen des Hygienekonzeptes abzuleiten.

3. Ausblick

Die KI-basierten Analyse-Bausteine werden zukünftig mit weiteren Datenanalysen in einer grafischen Anwendung kombiniert, sodass diese perspektivisch die zentrale Stelle im Krankenhaus zur Beobachtung des Infektionsgeschehens bilden wird. Durch die resultierende Anwendung – und insbesondere der KI-Bausteine – wird PathoBot einen entscheidenden Beitrag zur Vermeidung von Krankenhausinfektionen und zur weiteren Verbesserung Klinik-interner Hygienesysteme leisten.

Förderhinweis

Dieses Forschungs- und Entwicklungsprojekt wird durch das Ministerium für Wirtschaft, Arbeit und Tourismus Baden-Württemberg im Rahmen des Programms „Invest BW“ gefördert.

Literatur

[1] Robert Koch Institut (2024) Antibiotic Resistance and Hospital Infections. https://www.rki.de/EN/Content/infections/antibiotic/antibiotic_node.html

[2] Leclère, B., Buckeridge, D. L., Boëlle, P. Y., Astagneau, P. & Lepelletier, D. (2017) Automated detection of hospital outbreaks: A systematic review of methods. PLoS ONE 12(4): e0176438. https://doi.org/10.1371/journal.pone.0176438

[3] Schröder, C., Peña Diaz, L. A., Rohde, A. M., Piening, B., Aghdassi, S. J. S., Pilarski, G., Thoma, N., Gastmeier, P., Leistner, R. & Behnke, M. (2020) Lean back and wait for the alarm? Testing an automated alarm system for nosocomial outbreaks to provide support for infection control professionals. PLoS ONE 15(1): e0227955. https://doi.org/10.1371/journal.pone.0227955

[4] Baker, M. A., Huang, S. S., Letourneau, A. R., Kaganov, R. E., Peeples, J. R., Drees, M., Platt, R. & Yokoe, D. S. (2016). Lack of Comprehensive Outbreak Detection in Hospitals. Infect Control Hosp Epidemiol, 37(4), 466-468. https://doi.org/10.1017/ice.2015.325

Homogene Basisdokumentation als Grundlage der Digitalen Zusammenarbeit für Menschen mit Seltenen Erkrankungen

Adam GRAEFE[a,d 1], Filip REHBURG[a], Miriam HÜBNER[a], Steffen SANDER[a], Daniel DANIS[b], Susanne WIEGAND[c], Ana GRÖNKE[d], Annic WEYERSBERG[e], Jana ZSCHÜNTZSCH[g], Elisabeth NYOUNGUI[h], Josef SCHEPERS[a], Peter KÜHNEN[c], Peter N ROBINSON[a, b], Sylvia THUN[a], Oya BEYAN[b]

[a] *Core Facility Digital Health and Interoperability, Berlin Institute of Health at Charité - University Hospital Berlin*

[b] *The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, United States*

[c] *Berlin Center for Rare Diseases - Charité University Hospital, Berlin*

[d] *Institute for Biomedical Informatics – University Hospital Cologne*

[e] *Department of Paediatrics - University Hospital Cologne*

[g] *Department of Neurology - University Medical Center Goettingen*

[h] *Department of Medical Informatics - University Medical Center Goettingen*

1. Einleitung

Seltene Erkrankungen (SE) betreffen schätzungsweise fünf Prozent der Weltbevölkerung, davon in Deutschland circa vier Millionen. SE sind durch die Europäische Union definiert als Krankheiten mit einer Punktprävalenz geringer als fünf in 10.000 Personen. Mehr als 70 % der SE sind genetisch bedingt. [1] Für viele SE-Betroffene erstreckt sich die Zeit bis zur korrekten Diagnose über mehrere Jahre mit schwerwiegenden Folgen für die Betroffenen [2]. Fehlendes Fachwissen in der nicht-SE spezifischen Routineversorgung und kaum vergleichbare einrichtungsübergreifende Daten erschweren zudem die Versorgung und Forschung der SE [4].
Eine Herausforderung der SE-Versorgung und -Forschung ist das Fehlen einer bundes- und europaweit homogenen Datenerfassung in den derzeitigen Krankenhausinformationssystemen (KIS), bei denen anstatt der Versorgungsqualität und Forschungseffizienz die Abrechnung im Vordergrund steht. Aktuell erfolgt die Dokumentation in den Zentren für Seltene Erkrankungen (ZSE) flächendeckend heterogen, durch die Nutzung von unstandardisierten Papierakten oder digitalen Erfassung. Daraus ergibt sich ein Mangel an spezialisierten SE-Daten, die den FAIR-Prinzipien (Findability, Accessability, Interoperability, Reusability) entsprechen [3].

2. Methoden

Um diesen Herausforderungen zu begegnen, wird in der vorliegenden Arbeit unser Streben nach einem FHIR-basierten deutschen Minimalbasisdatensatz für Seltene Erkrankungen (MBDS-SE.de) vorgestellt, der in der NUM-FOSA Arbeitsgemeinschaft ZSE ausgestaltet wird. Deren Taskforce für das Ergänzungsmodul SE im MII-KDS begann ihre Arbeit vor Kurzem mit Einladung zur Partizipation. Das initiale Datenmodell und FHIR-Profil des MBDS-SE.de orientiert sich an den miteinander verwandten Datensätzen der Europäischen Rare Disease Registry Infrastructure (ERDRI-CDS.eu), der französischen Banque National de Données Maladies Rares (SDM-MR.fr) und der Global Alliance for Genomics and Health (GA4GH Phenopacket Schema) [4] sowie am Informationsmodell der Medizininformatik-Initiative (MII-KDS) und dem Modellvorhaben GenomDE (§64e) [5]. Angestrebt wird die Aufbereitung für Mehrfachnutzung in den MII- Datenintegrationszentren mit einwilligungsbasierter Übermittlung an das Nationale Register für Seltene Erkrankungen (NARSE), ERN- und andere Register sowie für föderierte Auswertungen über das FDP-G sowie die VP ERDERA.

[1] Corresponding Author: Adam Graefe, adam.graefe@charite.de

3. Ergebnisse

Als provisorisches Substitut für die SE-Datenerhebung im KIS oder KAS ist ein funktionsfähiger, auf REDCap-basierender Demonstrator mit der Bezeichnung ERKER (ERDRI-CDS Kompatible Erfassung in REDCap) entwickelt worden. Durch den Demonstrator wird die dezentrale und standardisierte Erfassung einer Modifikation des ERDRI-CDS mit noch nicht ballotierten ergänzenden Angaben zu SE-Betroffenen ermöglicht, um diese für verschiedene Nutzungen aufzubereiten. Um die Interoperabilität im REDCap Subsystem zu steigern, wurden im ERKER-Formular alle Variablen gemäß MII-KDS mit SNOMED-CT oder LOINC kodiert, alle Value Sets mit entsprechenden Ontologien versehen und die Field Annotationen für Definitionen aller Elemente verwendet. Die FHIR-basierten REDCap Integrations- und CDIS- Module ermöglichen den Import und Export der Daten im FHIR-Format, um eine Anbindung an die MII-Datenintegrationszentren zu ermöglichen. Eine beispielhafte lokale Implementierung des ERKERs ist für den MC4R Mangel realisiert worden, einer genetischen SE, die mit schwerer frühkindlicher Fettleibigkeit einhergeht.

4. Diskussion

Unser Prototyp ermöglicht die effiziente Erfassung und Standardisierung klinischer Daten und unterstützt darüber hinaus prinzipiell die Umwandlung in das Phenopacket Schema, wobei dessen Unterscheidung von Klinischer SE-Diagnose, Genetischer SE-Diagnose und sonstigen Diagnosen beachtet werden muss. Dieser Prozess soll den interoperablen Austausch und die Analyse von SE-Daten erleichtern und einen Weg zur Präzisionsmedizin bahnen helfen. Es soll zukünftig geprüft werden, ob in ausgewählten KIS- oder KAS-Systemen eine Integration mit SMARTonFHIR ermöglicht werden kann. Nach der weiteren Abstimmung und Ballotierung des Datenmodells, nach seiner Darstellung in ART-DECOR und der Publikation eines Implementation Guides wird die Integration des Europa-konformen deutschen SE-Minimalbasisdatensatzes in weiteren klinischen Informationssystemen angestrebt und an weiteren Krankheiten und Standorten erprobt werden. Durch unseren Open-Source-Ansatz wird die aktive Teilnahme und Mitgestaltung vieler Schwierigkeiten und anstehender Entwicklungen im Rahmen der verfügbaren Ressourcen ermutigt. Alle Informationen und Materialien sind in unserem GitHub-Repository verfügbar. Unsere Arbeit veranschaulicht die Notwendigkeit interoperabler und standardisierter Datenlösungen im medizinischen Bereich, insbesondere bei SE, wo detaillierte, einrichtungsübergreifend homogene Patientendaten für eine genaue Diagnose, Behandlung und Forschung entscheidend sind.

Literatur

[1] Wakap, S. N., Lambert, D., Olry, A., Rodwell, C., Gueydan, C., Lanneau, V., Murphy, D. N., Cam, Y. L. & Rath, A. (2019). Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *European Journal Of Human Genetics*, *28*(2), 165–173. https://doi.org/10.1038/s41431-019-0508-0

[2] Picci, R. L., Oliva, F., Trivelli, F., Carezana, C., Zuffranieri, M., Ostacoli, L., Furlan, P. M. & Lala, R. (2013). Emotional Burden and Coping Strategies of Parents of Children with Rare Diseases. *Journal Of Child And Family Studies*, *24*(2), 514–522. https://doi.org/10.1007/s10826-013-9864-5

[3] Lehne, M., Saß, J., Essenwanger, A., Schepers, J. & Thun, S. (2019). Why digital medicine depends on interoperability. *Npj Digital Medicine*, *2*(1). https://doi.org/10.1038/s41746-019-0158-1

[4] Schepers, J., Fleck, J. L. & Schaaf, J. (2022). Die Medizininformatik-Initiative und Seltene Erkrankungen: Routinedaten der nächsten Generation für Diagnose, Therapiewahl und Forschung. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, *65*(11), 1151–1158. https://doi.org/10.1007/s00103-022-03606-y

[5] Jacobsen, J. O. B., Baudis, M., Baynam, G., Beckmann, J. S., Beltrán, S., Buske, O. J., Callahan, T. J., Chute, C. G., Courtot, M., Daniš, D., Elemento, O., Essenwanger, A., Freimuth, R. R., Gargano, M., Groza, T., Hamosh, A., Harris, N. L., Kaliyaperumal, R., Lloyd, K. C. K., . . . Robinson, P. N. (2022). The GA4GH Phenopacket schema defines a computable representation of clinical data. *Nature Biotechnology*, *40*(6), 817–820. https://doi.org/10.1038/s41587-022-01357-4

Knowledge Connector: Decision Support System for Multi-omics-Based Precision Oncology

Daniel Hübschmann[3,8,11], Layla Tabea Riemann[9], Maximilian Schmutz[14], Simon Kreutzfeldt[2], Benjamin Roth[1], Katrin Glocker[1], Jennifer Hüllein[3], Janine Schoop[1], Lena Oeser[1], Steffen Hausmann[1], Maximilian Ataian[9], Sebastian Uhrig[3], Barbara Hutter[3], Martina Fröhlich[3], Zuguang Gu[3], Nagarajan Paramasivam[3], Malgorzata Oles[3], Christoph E. Heilig[2], Veronica Teleanu[2], Daniel B. Lipka[2,10], Irina A. Kerle[5,6,12], Ulrike Winter[2,10,11,13], Katja Beck[2,10,11,13], Christoph Heining[5,6,12], Hanno Glimm[4,5,6,12], Frank Ückert[7], Rainer Claus[14,15,16,17], Peter Horak[2,] Alexander Knurr[1], Stefan Fröhling[2,7,10,11]

[1] *Secondary Use of Data in Oncology Group, Clinical Trial Office, German Cancer Research Center (DKFZ), Heidelberg, Germany*

[2] *Division of Translational Medical Oncology, German Cancer Research Center (DKFZ), Heidelberg, Germany*

[3] *Computational Oncology Group, Molecular Precision Oncology Program, National Center for Tumor Diseases (NCT) Heidelberg and German Cancer Research Center (DKFZ), Heidelberg, Germany*

[4] *German Cancer Research Center (DKFZ) Heidelberg, Translational Functional Cancer Genomics, Germany*

[5] *Department for Translational Medical Oncology, National Center for Tumor Diseases Dresden (NCT/UCC), a partnership between DKFZ, Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, and Helmholtz-Zentrum Dresden - Rossendorf (HZDR), Germany*

[6] *Translational Medical Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, Dresden, Germany*

[7] *Institute of Human Genetics, Heidelberg University, Heidelberg, Germany*

[8] *Pattern Recognition and Digital Medicine Group, Heidelberg Institute for Stem Cell Technology and Experimental Medicine, Heidelberg, Germany*

[9] *Institute for Applied Medical Informatics, Center for Experimental Medicine, University Medical Center Hamburg-Eppendorf*

[10] *National Center for Tumor Diseases (NCT), NCT Heidelberg, a partnership between DKFZ and Heidelberg University Hospital, Heidelberg, Germany*

[11] *German Cancer Consortium (DKTK), Heidelberg, Germany*

[12] *German Cancer Consortium (DKTK), partner side Dresden*

[13] *Molecular Precision Oncology Program, NCT Heidelberg and German Cancer Research Center (DKFZ), Heidelberg, Germany*

[14] *Hematology and Oncology, Faculty of Medicine, University of Augsburg, Germany*

[15] *Pathology, Faculty of Medicine, University of Augsburg, Germany*

[16] *Bavarian Cancer Research Center (BZKF), Augsburg, Germany*

[17] *Comprehensive Cancer Center Augsburg (CCCA). University Hospital Augsburg*

1. Introduction

Precision cancer medicine seeks to improve patient outcomes by tailoring clinical management to individual molecular profiles in multidisciplinary molecular tumor boards (MTBs). The quality of MTB recommendations depends on the accurate and consistent interpretation of increasingly complex and multilayered molecular data. [1][2][3]

2. Methods

To address this challenge, we have developed the Knowledge Connector (KC), a powerful decision support system. The KC integrates individual patients' molecular and clinical data with real-world knowledge, enabling the standardized generation of MTB recommendations. It supports data curation, database integration, and case discussion based on multi-omics data. It provides an interface for generating a cross-institutional knowledge base using the newly developed concept of blocks of clinical knowledge (BoCKs). These are assembled by curators during case evaluation and subsequently reviewed and maintained by an independent team of experts. [4][5]

3. Results

The resulting and continuously growing collection of BoCKs (BoCKbase) complements external knowledge bases, allows streamlining of evidence items from different sources, and thus facilitates the reuse and storage of evidence. The BoCKbase has been successfully implemented and maintained at NCT Heidelberg and Dresden and will be rolled out to additional centers in the near future, such as University Medical Center Hamburg-Eppendorf and University Hospital Augsburg. Further work will include an NLP-based abstract classification, an NLP-based study inclusion aid, and improved visualization of the genetic pathways.

4. Discussion

In summary, the KC systematically processes relevant biomarker-drug associations and significantly enhances the efficacy of data curation with direct clinical relevance. It is a versatile tool that supports decision-making in MTBs and enables the scalability of precision cancer medicine.

References

[6] Rieke, D.T.*, et al.* Feasibility and outcome of reproducible clinical interpretation of high-dimensional molecular data: a comparison of two molecular tumor boards. *BMC Med* **20**, 367 (2022).

[7] Rieke, D.T.*, et al.* Comparison of Treatment Recommendations by Molecular Tumor Boards Worldwide. *JCO Precis Oncol* **2**, 1-14 (2018).

[8] van der Velden, D.L.*, et al.* Molecular Tumor Boards: current practice and future needs. *Ann Oncol* **28**, 3070-3075 (2017).

[9] Tamborero, D.*, et al.* The Molecular Tumor Board Portal supports clinical decisions and automated reporting for precision oncology. *Nat Cancer* **3**, 251-261 (2022).

[10] Wagner, A.H.*, et al.* A harmonized meta-knowledgebase of clinical interpretations of somatic genomic variants in cancer. *Nat Genet* **52**, 448-457 (2020).

Das Data Sharing Framework als agnostische Infrastruktur: Aktuelle und zukünftige Anwendungsfälle

Reto Wettstein[a,1], Maximilian Kurscheidt[b], Hauke Hund[b], Simon Tobias Schweizer[b], Martin Dugas[a], Christian Fegeler[b]

[a] *Institut für Medizinische Informatik, Universitätsklinikum Heidelberg, Heidelberg, Deutschland*

[b] *GECKO Institut für Medizin, Informatik und Ökonomie, Hochschule Heilbronn, Heilbronn, Deutschland*

In der medizinischen Forschung spielen organisationsübergreifende Infrastrukturen zur Umsetzung datengetriebener Prozesse eine zunehmend zentrale Rolle. Diese Infrastrukturen werden in der Regel für spezifische Nutzungsszenarien und darauf aufbauenden Anwendungsfällen als geschlossene Netzwerke konzipiert und entwickelt. Ein Beispiel einer solchen organisationsübergreifenden Infrastruktur wird derzeit im Rahmen der Medizininformatik Initiative (MII) für das Nutzungsszenario der sekundären Routinedatennutzung entwickelt. Mit der Entstehung des Netzwerk Universitätsmedizin (NUM) als Reaktion auf die COVID-Pandemie wurden die Synergien mit der MII erkannt und die in der MII entstandene Infrastruktur gemeinschaftlich weiterentwickelt.

Ein zentraler Aspekt von MII und NUM ist die Einrichtung von Datenintegrationszentren (DIZ) an Universitätskliniken. In diesen werden klinische Routinedaten gemäß dem HL7-FHIR-Standard sowohl syntaktisch standardisiert als auch semantisch annotiert, um Interoperabilität zu gewährleisten. Für eine effektive Nutzung dieser Daten ist außerdem die Digitalisierung und Automatisierung organisationsübergreifender Prozesse erforderlich, welche miteinander kommunizieren können. Zu diesem Zweck wurde das Data Sharing Framework (DSF) [1] entwickelt. Es ermöglicht die föderierte Nutzung der HL7-FHIR-Daten der DIZe mithilfe von digitalisierten Geschäftsprozessen basierend auf dem Standard BPMN-2.0.

Anwendungsfälle, welche basierend auf DSF Prozessen für die MII und NUM entwickelt wurden, sind unter anderem zentralisierte [2] und SMPC-basierte [3] Machbarkeitsanfragen, Datentransfers mit Record-Linkage über eine föderierte Treuhandstelle [4] oder verteilte Datenextraktions- und Datenzusammenführungsprozesse. Über diese parallellaufenden Anwendungsfälle hat sich das DSF in deutschen Unikliniken etabliert und kann perspektivisch für neue Anwendungsfälle über die MII und NUM hinaus genutzt werden.

Ein solcher perspektivischer Anwendungsfall könnten Prozesse in multizentrischen Studien sein, welche die von Ganzinger et al. [5] beschriebene Architektur einer föderierten Datenerfassung (fEDC) unterstützen. Am Symposium präsentieren wir dafür einen neuen DSF Prozess zum organisationsübergreifenden Management von gemeinsam genutzten electronic Case Report Forms und der föderierter Datenhaltung sowie der Datenzusammenführung zu Analysezwecken.

Ein weiterer Anwendungsfall könnte die Digitalisierung von Prozessen zur automatisierten Meldung von Patientendaten an Krankheitsregister sein, die zukünftig in standardisierter Form aus Routinedaten generiert und übertragen werden könnten.

Die umgesetzten und skizzierten Anwendungsfälle zeigen auf, dass basierend auf dem DSF als Prozessautomatisierungs- und Kommunikationsinfrastruktur ein anwendungsfall-agnostischer Datenraum entstanden ist, der sich über die deutschen Universitätskliniken spannt und es ermöglicht sensible Datensätze organisationübergreifend datenschutzkonform zu verarbeiten.

[1] Korrespondierender Autor: Reto Wettstein, Institut für Medizinische Informatik, Universitätsklinikum Heidelberg, Im Neuenheimer Feld 130.3, 69120 Heidelberg, Deutschland; E-Mail: reto.wettstein@med.uni-heidelberg.de.

Literatur

[1] Hund, H., Wettstein, R., Kurscheidt, M., Schweizer, S. T., Zilske, C., & Fegeler, C. (2024). Interoperability Is a Process - The Data Sharing Framework. Studies in Health Technology and Informatics, 310. https://doi.org/10.3233/SHTI230921

[2] Gruendner, J., Deppenwiese, N., Folz, M., Köhler, T., Kroll, B., Prokosch, H. U., Rosenau, L., Rühle, M., Scheidl, M. A., Schüttler, C., Sedlmayr, B., Twrdik, A., Kiel, A., & Majeed, R. W. (2022). The Architecture of a Feasibility Query Portal for Distributed COVID-19 Fast Healthcare Interoperability Resources (FHIR) Patient Data Repositories: Design and Implementation Study. JMIR Medical Informatics, 10(5). https://doi.org/10.2196/36709

[3] Wettstein, R., Kussel, T., Hund, H., Fegeler, C., Dugas, M., & Hamacher, K. (2022). Secure Multi-Party Computation Based Distributed Feasibility Queries - A HiGHmed Use Case. Studies in Health Technology and Informatics, 296, 41–49. https://doi.org/10.3233/SHTI220802

[4] Hund, H., Wettstein, R., Hampf, C., Bialke, M., Kurscheidt, M., Schweizer, S. T., Zilske, C., Mödinger, S., & Fegeler, C. (2023). No Transfer Without Validation: A Data Sharing Framework Use Case. Studies in Health Technology and Informatics, 302, 68–72. https://doi.org/10.3233/SHTI230066

[5] Ganzinger, M., Blumenstock, M., Fürstberger, A., Greulich, L., Kestler, H. A., Marschollek, M., Niklas, C., Schneider, T., Spreckelsen, C., Tute, E., Varghese, J., & Dugas, M. (2023). Federated electronic data capture (fEDC): Architecture and prototype. Journal of Biomedical Informatics, 138. https://doi.org/10.1016/J.JBI.2023.104280

Umsetzung eines nicht-universitären Datenintegrationszentrums für den DigiHub DISTANCE

Antonia Schmidt [a,1], Hong Diem Duong [a], Volker Lowitsch [b], Denise Molinnus [c] und Marcel Klötgen [a]

[a] *Fraunhofer-Institut für Software- und Systemtechnik ISST, Dortmund, Deutschland*

[b] *Healthcare IT Solutions GmbH, Deutschland*

[c] *Operative Intensivmedizin und Intermediate Care, Universitätsklinikum RWTH, Aachen, Deutschland*

Die Datenbasis der Gesundheitsforschung zu verbessern ist eines der zentralen Ziele der sechs BMBF geförderten „Digitalen FortschrittsHubs Gesundheit“ (DigiHubs) in Deutschland. Die Förderlinie konzentriert sich darauf, die Aktivitäten der Medizininformatik-Initiative (MII) an Universitätskliniken in Bezug auf Digitalisierung und Datenaustausch auf die regionale Versorgung auszudehnen. [1]. Grundlage hierfür bieten die von der MII geschaffenen Strukturen und Rollen, insbesondere die Datenintegrationszentren (DIZ). Sie dienen der Aggregation von Daten aus der medizinischen Versorgung, die anschließend über die digitale Infrastruktur der MII den Forschenden zur Verfügung gestellt werden können. Für die DigiHubs sollen die bereits existenten DIZ an Universitätskliniken genutzt werden. Vor der Förderlinie waren die universitären DIZ ausschließlich zuständig für die Aggregation und Bereitstellung der universitätsklinischen Daten für die Gesundheitsforschung. Mit den DigiHubs werden diese bestehenden Strukturen für den Anschluss der regionalen Gesundheitsversorger an die MII-Infrastruktur weiterverwendet.

DISTANCE, der DigiHub mit Fokus auf präzisere Therapien und bessere Prognosen für Krankheitsverläufe nach intensivmedizinischer Betreuung, wählt einen anderen Ansatz und erforscht die Option, sich als eigenständiges DIZ zu etablieren. Wie ein solches nicht-universitäres DIZ aufzubauen ist und ein Verständnis für die damit verbundenen Aufwände, Ressourcen und zu beachtenden Anforderungen aufzubringen, ist eine wesentliche Forschungsaufgabe des Projekts.
Erste Erfolge dessen lagen darin, die im Rahmen der Anbindung regionaler Versorger geschaffenen technischen Komponenten so zu erweitert, dass sie mit den von der MII vorgegebenen Strukturen kompatibel sind. Die damit zusammenhängenden Tätigkeiten in DISTANCE zeigten gleichzeitig, dass noch einige Herausforderungen zu lösen sind. So besteht beispielsweise bei den technischen Dokumentationen der MII Verbesserungsbedarf, da Informationen partiell unzulänglich oder nicht mehr aktuell sind. Als weitere Herausforderung konnten die administrativen und (daten-) qualitätssichernden Tätigkeiten eines DIZ identifiziert werden. Hierbei besteht für nicht-universitäre DIZ eine Schwierigkeit darin, den Bedarf an Personal mit Entscheidungskompetenz und -autorität zu decken.

Entsprechend stellt die Forschung im Rahmen des DISTANCE DigiHubs heraus, dass hinsichtlich der Rollen- und Aufgabendefinition der MII noch Anpassungsbedarf besteht, wenn diese auch universitätsklinikumsunabhängig anwendbar sein sollen. Parallel erarbeitet und realisiert das Projekt die hierfür notwendige technische Architektur.

Literatur

[1] BMBF (2023). Digitale FortschrittsHubs Gesundheit - DLR Gesundheitsforschung. Online verfügbar unter https://www.gesundheitsforschung-bmbf.de/de/Digitale-FortschrittsHubs-Gesundheit.php. Zuletzt geprüft am 09.01.2024.

[1] Corresponding Author: Antonia Schmidt, antonia.schmidt@isst.fraunhofer.de

Factors Associated with the Risk of Patient Re-identification Using Clinical Metadata

Caroline Bönisch[a, b 1], Sebastian Behre[a], Dorothea Kesztyüs[a], Tibor Kesztyüs[a]

[a] *Medical Data Integration Center, University Medical Center Göttingen, Robert-Koch-Straße 40, 37075 Göttingen, Germany*

[b] *University of Applied Sciences Stralsund, Zur Schwedenschanze 15, 18435 Stralsund, Germany.*

1. Introduction

Medical data integration centers (MeDIC) play a crucial role in aggregating and storing heterogeneous healthcare data [1], making it interoperable while safeguarding sensitive patient information. When considering secondary data usage, it is essential to assess and categorize potential privacy risks associated with metadata, as they are the initial information requested by researchers in data usage inquiries [2]. Based on the findings of Rocher et al. [3], it can be deduced that metadata, even when partially complete, are potentially susceptible to re-identification.

2. Methodology

This study aims to deduce factors contributing to an increased identification risk based on metadata and evaluate how this risk can be assessed and visualized for researchers. A systematic literature review was conducted using specific search strategies in PubMed, Embase and JSTOR. Inclusion criteria encompassed publications with re-identification analyses of data - including image, signal and other data, as well as those estimating re-identification risks of data from a legal point of view. 65 results were obtained for Pubmed, while Embase provided 16 distinct results and JSTOR 15. The abstracts of the studies were then assessed and yielded to 20 results for Pubmed, 5 results for Embase and four results for JSTOR. in the final step, the full texts of the search results were examined in detail using predefined selection criteria. The final result was a total of 27 publications, where 19 papers described real attacks, of which 11 were directed against medical data and eight against personal data in general. 8 papers dealt with risk assessment methods for medical data.

3. Results

From the literature analysis, it was observed that the risk of re-identification increases with the quantity of available (in)direct data. Effective re-identification often involves narrowing the search space through external data source linkage (e.g., death registers, voter lists, social media). Attacks have primarily targeted relational organized data, while non-relational structures have led to graph-theoretical matching approaches and machine learning utilization. Risk assessment involves evaluating dataset uniqueness and the degree of feature intersection when combining two data sources. A larger intersection increases the risk of re-identification. The analysis provided risk assessments exclusively for data, but their applicability to metadata is plausible due to metadata serving as supplementary information about the data.

4. Discussion

This work examines the re-identification risk of clinical metadata and finds that this risk remains largely unexplored in the current literature and only risk regarding data is explored [4]. This study emphasizes the importance of researching the re-identification risks associated with medical metadata, because metadata are equally, if not more, critical than the corresponding medical data.

[1] Corresponding Author: Caroline Bönisch, University of Applied Sciences Stralsund, caroline.boenisch@hochschule-stralsund.de

The protection and risk analysis of medical metadata require further attention to ensure the security and privacy of sensitive patient information.

References /Literatur

[1] Gaddale J. R. (2015). Clinical Data Acquisition Standards Harmonization importance and benefits in clinical data management. *Perspectives in clinical research*, *6*(4), 179–183. https://doi.org/10.4103/2229-3485.167101

[9] Ulrich, H., Kock-Schoppenhauer, A., Deppenwiese, N., Gött, R., Kern, J., Lablans, M., Majeed, R. W., Stöhr, M. R., Stausberg, J., Varghese, J., Dugas, M., & Ingenerf, J. (2022). Understanding the Nature of Metadata: Systematic review. Journal of Medical Internet Research, 24(1), e25440. https://doi.org/10.2196/25440

[10] Rocher, L., Hendrickx, J. M., & De Montjoye, Y. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. Nature Communications, 10(1). https://doi.org/10.1038/s41467-019-10933-3

[11] Xia, W., Liu, Y., Wan, Z., Vorobeychik, Y., Kantacioglu, M., Nyemba, S., Clayton, E. W., & Malin, B. A. (2021). Enabling realistic health data re-identification risk assessment through adversarial modeling. Journal of the American Medical Informatics Association : JAMIA, 28(4), 744–752. https://doi.org/10.1093/jamia/ocaa327

Multiverse analysis on depression biomarkers from EEG

Yasmin HOLLENBENDERS [a,b], Christoph MAIER [c], Alexandra REICHENBACH [a,b,1]

[a] *Center for Machine Learning, Heilbronn University, Heilbronn, Germany*

[b] *Medical Faculty Heidelberg, University of Heidelberg, Heidelberg, Germany*

[c] *Medical Informatics, Heilbronn University, Heilbronn, Germany*

1. Introduction

Major depressive disorder (MDD) is one of the most common mental disorders but diagnosis is rather subjective (Cai et al., 2018). Objective EEG biomarkers can distinguish between MDD patients and healthy controls (HC) but discriminatory features are controversial (De Aguiar Neto & Rosa, 2019). To support decision-making in clinics, robust biomarkers are essential. One substantial problem is non-standardized data acquisition and processing (Kołodziej et al., 2021). Our proposed multiverse analysis (Steegen et al., 2016) aims to resolve these contradictions exemplarily on α-band biomarkers.

2. Methods

We used a public dataset with 5 min resting-state eyes-closed EEG (HC:28/MDD:30) (Mumtaz, 2016). Data was re-referenced, filtered, and artifacts were automatically removed. Afterward, 720 paths with combinations from each processing step were constructed for the multiverse analysis: A) Normalization: None, subject-wise z-normalization, channel-wise z-normalization. B) Window length: non-overlapping windows of 5-, 10-, 15-, and 20-sec. C) Feature extraction: absolute and relative α-band power, α-peak-frequency, and from the upper signal envelope kurtosis, skewness, median, interquartile range, variance, range, and their combination. D) Aggregation: 10 individual values and median. E) Classification algorithm: Logistic Regression, Random Forest, and Support Vector Machine. Classification models were trained with six-fold cross-validation. An ANOVA for comparing processing steps and t-tests against chance-level for statistical robustness were conducted with resulting accuracies

3. Results

All processing steps but normalization have a sign. influence on accuracy (all $p<.001$). Aggregating to the median (63.4±17.0%) achieves higher accuracy yet individual values yield more statistically robust models (15 vs 7). 15-sec windows (65.7±13.5%) achieve sign. higher accuracies (all $t(2158)>5.288$, $p<.001$), while 10-sec windows yield most statistically robust models (8). Random Forest performs sign. better (all $t(2878)>3.243$, $p<.001$), yet Logistic Regression yields most statistically robust models (8). Envelope skewness scores the highest accuracies (8/9, $t(862)>3.737$, $p<.001$). However, the most statistically robust marker is envelope kurtosis (7).

4. Discussion

We demonstrate that the processing steps investigated explain the contradictory results. However, this study is restricted by the choice of biomarkers, achieving neither high diagnostic value nor statistical robustness. Furthermore, none of the variations of the processing steps yield a clear advantage. Nonetheless, this approach contributes to finding robust biomarkers for clinical decision-support in MDD.

[1] Corresponding Author: Alexandra Reichenbach, alexandra.reichenbach@hs-heilbronn.de.

References /Literatur

[1] Cai, H., Han, J., Chen, Y., Sha, X., Wang, Z., Hu, B., Yang, J., Feng, L., Ding, Z., Chen, Y., & Gutknecht, J. (2018). A Pervasive Approach to EEG-Based Depression Detection. Complexity, 2018, 1–13. https://doi.org/10.1155/2018/5238028

[2] De Aguiar Neto, F. S., & Rosa, J. L. G. (2019). Depression biomarkers using non-invasive EEG: A review. Neuroscience & Biobehavioral Reviews, 105, 83-93. https://doi.org/10.1016/j.neubiorev.2019.07.021

[3] Kołodziej, A., Magnuski, M., Ruban, A., & Brzezicka, A. (2021). No relationship between frontal alpha asymmetry and depressive disorders in a multiverse analysis of five studies. eLife, 10. https://doi.org/10.7554/eLife.60595

[4] Mumtaz, W. (2016). MDD Patients and Healthy Controls EEG Data (New). figshare. Dataset. MDD Patients and Healthy Controls EEG Data generated by https://doi.org/10.6084/m9.figshare, 4244171, v2.

[5] Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. Perspectives on Psychological Science, 11(5), 702–712. https://doi.org/10.1177/1745691616658637

How to do graph-based retrieval for the Medical Data Models (MDM) Portal – a concept

Lea Gütebier[a,1], Max Blumenstock[b], Dagmar Waltemath[a], Christian Niklas[b], Ron Henkel[a]

[a] *Medical Informatics Laboratory, Institute for Community Medicine, University Medicine Greifswald, Germany*

[b] *Institute of Medical Informatics, Heidelberg University Hospital, Germany*

The MDM Portal has been established as a significant metadata repository of case report forms (CRFs) and is an important information system in medical research [1]. The portal's emphasis on data standardisation in the CDISC ODM format and the semantic encoding of medical forms enables data analytics and comparison of CRFs, thereby facilitating their reuse.
Here, we present our concept to extend the MDM Portal with a new graph-based information retrieval system. A graph-based information retrieval system will significantly extend the possibilities for study discovery, exploration, and comparison.

The goal is to build an automated pipeline that integrates the existing MDM data into a knowledge graph. A prominent example is the CovidGraph, a labelled property graph to integrate COVID-19 data [2]. The concept for a graph integration of MDM aims at incorporating nodes, edges, labels, and properties to build a detailed graph representation of the MDM data [3]. Conceptually, within the graph, each study will be represented with its core features, reflecting the internal structure of the study's forms, connecting items, and annotations. The integration of domain-specific ontologies, such as UMLS, LOINC, SNOMED, and the subsequent creation of mappings to already existing semantic encodings in the MDM Portal will allow to automate the comparison of studies and enhance the UMLS term encoding.

At its core, the concept uses graph-based similarity measures, which will identify overlap of annotations between studies, depth and distance of the ontological annotations. To strengthen and enhance the usability of our concept we develop use cases. These use cases are based on clinical studies in the field of pharmacology, and genetic phenotype studies.

Ultimately, we will extend the MDM Portal's infrastructure with a graph-based approach, creating a robust and user-friendly retrieval system that enhances metadata management and analysis.

References

[1] Riepenhausen et al. (2019). Portal of Medical Data Models: Status 2018. Stud Health Technol Inform., 258:239-240. doi: 10.3233/978-1-61499-959-1-239

[2] Gütebier et al. (2022). CovidGraph: a graph to fight COVID-19. Bioinformatics, 38(20):4843-4845. doi: 10.1093/bioinformatics/btac592.

[3] Robinson et al. (2015). Graph databases: new opportunities for connected data. O'Reilly Media, Inc.

[1] Corresponding Author: Lea Gütebier, Medical Informatics Laboratory, Institute for Community Medicine, University Medicine Greifswald, lea.guetebier@uni-greifswald.de

Poster

Consolidation of clinical data: Harmonizing EEG data of major depressive disorder patients and healthy controls

Friedrich Philipp Carrle[a,b] , Alexandra Reichenbach [a,b,1]

[a] *Center for Machine Learning, Heilbronn University, Heilbronn, Germany*

[b] *Medical Faculty Heidelberg, University of Heidelberg, Heidelberg, Germany*

1. Introduction

Major depressive disorder (MDD) is the most common mental disorder worldwide, leading to impairment in quality and independence of life [1]. Electroencephalography (EEG) biomarkers processed with machine learning (ML) algorithms have been explored for objective diagnoses with promising results [2]. However, the generalizability of those models, a prerequisite for clinical application, is restricted by small datasets [3].
Harmonizing several datasets is challenging. EEGs have a high number of degrees of freedom in data acquisition and processing. Individual research groups have their own methods and use different equipment [4]. Additionally, there is a lack of comprehensive and standardised data processing tools.

2. Methods

I will analyse existing EEG datasets and develop a harmonisation method, enabling the combination of several datasets into larger ones for unified use. To this end, an EEG data structure is designed that stores the data in a way the EEG data and selected features/biomarkers can be read out efficiently. The database will contain data from people with MDD as well as healthy controls and meet the FAIR principles. The data is initially fed in with minimal loss using ETL (Extract, Transform and Load) processes. The data schema takes the diverse data sources into account, which often contain further information about the test subjects in addition to EEG data, and is also intended for storing intermediate results after various processing steps. Common standards of data storage and processing, such as EEG-BIDS, are taken into account and included.

3. Results

A process to compile a consolidated, quality-assured, and expandable EEG database for clinical research purposes is proposed. I aim to equip scientists with the necessary tools to replicate the harmonization process with their own datasets. By providing a comprehensive framework and methodology, researchers can adapt their data to fit the proposed EEG database schema, ensuring compatibility and standardization across studies.

4. Discussion

Current literature shows conflicting results regarding which biomarkers have diagnostic properties [5]. Those contradiction can result from the aforementioned high number of degrees of freedom in data acquisition and processing as well as the small size of available datasets. Robust biomarkers for use in the diagnosis of depression is vital, since they need to be broadly applicable in the clinic and not prone to change with small variations in data processing. With this in mind, a quality assurance measure is still needed, so it can be verified, that the data used meets certain criteria and can be compared to data collecting from other research facilities.
This work contributes to the increased availability of large data sets that can be used for research by providing an analysis-centred database.

References

[1] Otte, C., Gold, S. M., Penninx, B. W., Pariante, C. M., Etkin, A., Fava, M., ... & Schatzberg, A. F. (2016). Major depressive disorder. *Nature reviews Disease primers*, *2*(1), 1-20.

[2] Yasin, S., Hussain, S. A., Aslan, S., Raza, I., Muzammel, M., & Othmani, A. (2021). EEG based Major Depressive disorder and Bipolar disorder detection using Neural Networks: A review. *Computer Methods and Programs in Biomedicine*, *202*, 106007.

[3] Rakić, M., Cabezas, M., Kushibar, K., Oliver, A., & Lladó, X. (2020). Improving the detection of autism spectrum disorder by combining structural and functional MRI information. *NeuroImage: Clinical*, *25*, 102181.

[4] Melnik, A., Legkov, P., Izdebski, K., Kärcher, S. M., Hairston, W. D., Ferris, D. P., & König, P. (2017). Systems, subjects, sessions: To what extent do these factors influence EEG data?. *Frontiers in human neuroscience*, *11*, 150.

[5] Kołodziej, A., Magnuski, M., Ruban, A., & Brzezicka, A. (2021). No relationship between frontal alpha asymmetry and depressive disorders in a multiverse analysis of five studies. *Elife*, *10*, e60595.

Artifact Detection in DCE Breast MRI: Transition from ROIs to Whole-Breast-MIPs

Lorenz A. Kapsner[a,b,1], Eva L. Balbach[a], Andrzej Liebert[a], Sabine Ohlmeyer[a], Evelyn Wenkel[c], Michael Uder[a], and Sebastian Bickelhaupt[a]
[a] Institute of Radiology, Uniklinikum Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany.
[b] Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany.
[c] Radiologie München, München, Germany.

1. Objective

To apply an image quality assessment convolutional neural network (CNN) that was originally trained on regions of interest (ROIs) – namely the left and right breasts cropped from MRI-derived dynamically contrast enhanced (DCE) subtraction maximum-intensity-projections (MIPs) – to whole-breast (WB) MIPs. DCE subtraction sequences in breast MRI allow the visualization of contrast-enhanced areas, with breast tissue being the main ROI in which suspect lesions are enriched by the contrast agent. However, in WB-MIPs also the thorax containing the heart appears contrast-enriched, potentially hampering the application of the neural network.

2. Methods

This IRB-approved retrospective study included n=2523 clinically indicated breast MRI examinations (one second post-contrast WB-MIP each) of n=1793 patients, acquired between the years 2015 and 2020 at the Institute of Radiology of the University Hospital Erlangen. An ensemble of five CNN classifier (DenseNet [1]) was applied to WB-MIPs to detect artifacts. These CNNs were trained in a previous study using the above mentioned ROI images as inputs [2]. The performance of the artifact detection on WB-MIPs was evaluated by comparing the ensemble's prediction with artifact labels that were previously determined as a ground truth by three independent raters in a another study [3].

3. Results

Before providing the WB-MIPs as inputs to the CNNs, the potentially confounding thorax region was masked with the following approach: a) sternum detection (windowing to the ten central pixel columns, remove background noise, apply Wiener filter + K-means clustering [both steps adapted from Pandey et al. [4]], Gaussian-mixture model, and binary masking), b) draw ellipsis thorax mask to cover the area below the detected sternum. On WB-MIPs, the CNN ensemble identified artifacts with an AUROC of 0.79 and a positive predictive value of 0.85. The code is publicly available at GitHub: https://github.com/kapsner/ce_mip_artifact_detection.

4. Conclusion

For translation into a diagnostic setting, artifact detection on WB-MIPs is favored over detection on ROIs. The results demonstrate that the application of the previously ROI-trained CNNs to WB-MIPs might be a resource-saving starting point for future efforts in this field. A potential limitation is the overlapping cohort of the ROI-CNNs' training data and the whole-breast MIPs, which, however, might play a subordinate role due to the significantly differing image appearance of the input data.

[1] Corresponding Author: Lorenz A. Kapsner, email: lorenz.kapsner@uk-erlangen.de

5. References

[1] G. Huang, Z. Liu, L. Van Der Maaten, and K.Q. Weinberger, Densely connected convolutional networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, 2017: pp. 2261–2269. doi:10.1109/CVPR.2017.243.

[2] L.A. Kapsner, S. Ohlmeyer, L. Folle, F.B. Laun, A.M. Nagel, A. Liebert, H. Schreiter, M.W. Beckmann, M. Uder, E. Wenkel, and S. Bickelhaupt, Automated artifact detection in abbreviated dynamic contrast-enhanced (DCE) MRI-derived maximum intensity projections (MIPs) of the breast, *Eur Radiol*. (2022). doi:10.1007/s00330-022-08626-5.

[3] L.A. Kapsner, E.L. Balbach, F.B. Laun, L. Baumann, S. Ohlmeyer, M. Uder, S. Bickelhaupt, and E. Wenkel, Prevalence and influencing factors for artifact development in breast MRI-derived maximum intensity projections, *Acta Radiologica*. (2023). doi:10.1177/02841851231198349.

[4] D. Pandey, X. Yin, H. Wang, M.-Y. Su, J.-H. Chen, J. Wu, and Y. Zhang, Automatic and fast segmentation of breast region-of-interest (ROI) and density in MRIs, *Heliyon*. **4** (2018) e01042. doi:10.1016/j.heliyon.2018.e01042.

Better Sandbox: a comprehensive platform for quality-assured healthcare software development based on openEHR

Miguel Pedrera-Jimenez[a,1], Lukasz Popowicz[a], Jovan Pavićević[a], Benjamin Muhič[a], Borut Fabjan[a], Andraž Koželj[a]

[a] *Better, Ljubljana, Slovenia*

1. Introduction

Electronic Health Record (EHR) is a collection of information from heterogeneous systems that are often based on local models of health domain concepts, hindering data interoperability and reuse [1, 2]. However, the evolution towards ecosystems based on the openEHR standard is making possible to develop digital health applications through a common, open, and stable framework for modeling, persistence and management of clinical knowledge and information [3].
The objective of this study was the implementation of "Better Sandbox", a comprehensive platform for the design and development of healthcare software through the registration, query, retrieval, persistence, and interoperability mechanisms defined by the openEHR specification [4, 5].

2. Methods

Better Sandbox is based on the openEHR specification, a health information standard that follows the paradigm of Detailed Clinical Models (DCM). This paradigm provides a set of stable and generic components for building EHRs, along with an archetype model that formalizes health domain concepts with full meaning and flexibility through the reference model and terminology bindings [3].

3. Results

The main result of this study was the design and implementation of a sandbox platform based on openEHR, consisting of four modules [5]:

- Archetype Designer: web application for the design of archetypes and openEHR templates, using the ADL language, and the generation of operational (OPT) and JSON web templates.
- Standardized Repositories: repositories for information resources (models and terminology), demographic and administrative data (based on HL7 FHIR) and health data (based on openEHR).
- EHR Studio: web tools for creating, editing, and validating clinical forms from openEHR templates, querying health data through AQL, and designing and implementing graphical portals for data entry and visualization.
- API suite: set of REST API services for interaction with archetype and terminology servers, demographic data repository, and health data repository.

This platform is currently being used by various healthcare, research, and teaching organizations, as well as software vendors, to learn about openEHR and develop their future digital health applications.

4. Discussion

Better Sandbox platform provided a comprehensive platform for quality-assured healthcare software development based on openEHR: (1) design and implementation of clinical archetypes and templates;

[1] Corresponding Author: Miguel Pedrera-Jiménez, Better, Štukljeva cesta 48, 1000 Ljubljana, Slovenia, miguel.pedrera@better.care.

(2) standardized persistence for information resources, demographic and administrative data, and health data; (3) building clinical forms, performing AQL queries, and composing graphical portals; and (4) interacting with information resources and data from external systems through API services.

References

[1] Kalra D, Blobel BG. Semantic interoperability of EHR systems. Stud Health Technol Inform. 2007;127:231-245.
[2] Parra-Calderón CL, Sanz F, McIntosh LD. The Challenge of the Effective Implementation of FAIR Principles in Biomedical Research. Methods Inf Med. 2020;59(4-05):117-118. doi:10.1055/s-0040-1721726.
[3] Beale T. Archetypes: Constraint-based Domain Models for Future- proof Information Systems. OOPSLA 2002 Work Behav Semant 2001;:1–69. doi:10.1.1.147.8835.
[4] OpenEHR Specification. URL: https://specifications.openehr.org/releases/RM/latest/ehr.html. Accessed 2024-04-15.
[5] Better Sandbox. URL: https://www.better.care/try-sandbox/. Accessed 2024-04-15.

Synthesizing tabular health data in cancer registries using generative machine learning approaches

Philipp RÖCHNER[a,b,1], Louisa SCHWARZ[b]

[a] *Chair of Information Systems and Business Administration, Johannes Gutenberg University, Mainz, Germany*

[b] *Cancer Registry, Institute for Digital Health Data Rhineland-Palatinate, Mainz, Germany*

1. Introduction

Sharing personal health data collected by medical patient registries, such as cancer registries, is highly regulated due to privacy and security concerns. Instead of sharing real patient data, synthetic data can be generated based on real patient data [1]. To generate synthetic patient data, generative machine learning approaches learn a model of the real patient data structure [2]. Ideally, the synthetic data can be shared more easily than the real data because it is impossible to reconstruct real personal data from the synthetic data (privacy). At the same time, the synthetic data should be as similar as possible to the real data to ensure that conclusions drawn from the synthetic data are also valid for the real data (utility). In this study, we compare the privacy and utility of synthetic cancer registry records generated by machine learning approaches.

2. Materials and Methods

We compared variational autoencoders (VAEs), generative adversarial networks (GANs), and diffusion models (DDPMs) to generate synthetic cancer registry records. We studied 1,218,802 electronic health records consisting of nine categorical variables describing patients with breast cancer diagnosed between 2000 and 2020. The dataset was collected and provided by the SEER program [3]. We evaluated the privacy and utility of the synthetic data.

3. Results

The structure of the records synthesized by the DDPM is closer to the structure of the real records than the VAE and GAN records. A binary classifier that separates real from synthetic data had an accuracy of 0.888 for the DDPM records, 0.952 for the VAE records, and 0.956 for the GAN records. The classifier's F_1 score was 0.891 for the DDPM records, 0.951 for the VAE records, and 0.956 for the GAN records.
At the same time, synthetic DDPM records pose a greater privacy risk because the records are more likely to reveal information from real patients than synthetic VAE and GAN records. Of the synthetic DDPM records, 83.8% share eight out of nine values with the most similar real record, where the number of identical values measures the similarity of two records. For the VAE and GAN records, this decreases to 42.1% and 34.1%, respectively.

4. Discussion

The results of this study highlight the trade-off between privacy and utility in the synthesis of electronic health records [4]: Although newer DDPMs generate records that are more similar in structure to the real records than established VAEs and GANs, DDPM records pose a greater privacy risk.
In future work, we plan to investigate further technical and legal aspects of privacy risk assessment approaches for synthetic health records.

[1] Corresponding Author: Philipp Röchner, roechner@uni-mainz.de.

References

[1] Yale, A., Dash, S., Dutta, R., Guyon, I., Pavao, A., & Bennett, K. P. (2020). Generation and evaluation of privacy preserving synthetic health data. Neurocomputing, 416, 244-255.

[2] Hernandez, M., Epelde, G., Alberdi, A., Cilla, R., & Rankin, D. (2022). Synthetic data generation for tabular health records: A systematic review. Neurocomputing, 493, 28-45.

[3] Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER Research Data, 17 Registries, Nov 2022 Sub (2000-2020) - Linked To County Attributes - Time Dependent (1990-2021) Income/Rurality, 1969-2021 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2023, based on the November 2022 submission.

[4] Yan, C., Yan, Y., Wan, Z., Zhang, Z., Omberg, L., Guinney, J., ... & Malin, B. A. (2022). A multifaceted benchmarking of synthetic electronic health record generation models. Nature communications, 13(1), 7609.

The three-year evolution of Germany's digital therapeutics reimbursement program and its path forward

Linea Schmidt[a,b,c], Marc Pawlitzki[d], Bernhard Y. Renard[a,b,c], Sven G. Meuth[d], Lars Masanneck[d,e1]

[a] *Hasso Plattner Institute, Digital Engineering Faculty, University of Potsdam, Potsdam, Germany*
[b] *Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, USA*
[c] *Windreich Dept. of Artificial Intelligence & Human Health, Icahn School of Medicine at Mount Sinai, New York, USA*
[d] *Heinrich-Heine-University Düsseldorf, Medical Faculty and University Hospital Düsseldorf, Department of Neurology, Germany*
[e] *German Society of Digital Medicine e.V. (DGDM)*

1. Introduction

The 2019 German Digital Healthcare Act introduced the Digital Health Application program, known in German as 'Digitale Gesundheitsanwendungen' (DiGA). The program has established a pioneering model for integrating Digital Therapeutics (DTx) into a healthcare system with scalable and effective reimbursement strategies.

2. Methods

This perspective provides a synthesis of the DiGA program's evolution since its inception three years ago, based on a comprehensive analysis of the BfArM DiGA registry [1], taking into account all DiGAs listed until January 16th 2024. Additionally, we assessed legislative changes with their potential influence on the DiGA program.

3. Results

A total of 53 DiGAs are analyzed (29 permanently listed and 24 provisionally listed). All but two DiGAs are classified as medical devices of risk class I, the remaining DiGAs are classified as class IIa medical devices. The mean cost for manufacturer-set prices for initial prescription is 465.42€, while for negotiated final prices it is 220.79€. Although those two prices are positively correlated (Pearson correlation coefficient of 0.79, P value < 0.001), the absolute difference in negotiated prices is relatively modest (189.00€ - 243.00€).

4. Discussion

The Digital Act [2] and the Health Data Use Act [3] increase the scope of the DiGA program by allowing, with stricter evidence requirements, medical devices of risk class IIb to be reimbursed as a DiGA. This potentially paves the way for more complex products like remote monitoring systems getting integrated into these primarily patient-facing applications. Importantly, as from 2026 onwards at least 20% of a DiGA's reimbursement price will be contingent upon success measures, enforcing a higher usage of real-world data and real-world evidence.
Despite encountering challenges related to effectiveness, evidence requirements, and integration within the healthcare system, the DiGA program continues to evolve and serves as a seminal example for the integration of DTx, offering valuable insights for healthcare systems globally.

[1] Corresponding Author: Lars Masanneck (lars.masanneck@med.uni-duesseldorf.de)

References

[1] DiGA-Verzeichnis. https://diga.bfarm.de/de/verzeichnis (Accessed January 25, 2024).

[2] Deutscher Bundestag. Gesetz Zur Beschleunigung Der Digitalisierung Des Gesundheitswesens (Digital-Gesetz - DigiG) [Digital Act]. Drucksache 20/9048 (2023).

[3] Deutscher Bundestag. Gesetz Zur Verbesserten Nutzung von Gesundheitsdaten [Health Data Use Act]. Drucksache 20/9046 (2023).

Setting the Stage for Interoperable, Interinstitutional Application of (AI) Algorithms and Services for Radiology Use Cases: the OMI Protocol

Stefan Sigle[a,1], Patrick Werner[a], Christian Fegeler[a,b]

[a] *MOLIT Institute, Heilbronn, Germany*

[b] *University of Heilbronn, Heilbronn, Germany*

1. Introduction

The Open Medical Inference (OMI) project, as part of the Medical Informatics Initiative (MII) [1], brings together thirteen university hospitals [2] and addresses the need for infrastructure for current and upcoming questions in research and clinical care settings that require algorithm driven approaches. Algorithms, often based on Artificial Intelligence (AI), are developed, and used locally in research scenarios without considering the scientific community outside of a developer's own organization. This stands in direct contrast to FAIR Guiding Principles [3] and efforts of building networks and digital ecosystems.

2. Methods

As part of OMI our work packages specifies i) a registry for inference services and ii) a protocol to invoke these services remotely. As part of the registry inference services, their in- and output parameters as well as endpoints are represented to be findable, accessible, interoperable and reusable for the scientific community beyond organizational borders. Based on business process modelling for common use cases the specification orchestrates artefacts like FHIR core dataset definitions by the MII and uses established data formats like DICOM to secure interoperability using national efforts like the Data Sharing Framework [4] for the data transport layer.

3. Results

After reviewing ongoing national efforts and considering stakeholders within the OMI project we described and modeled use cases, generated an abstract information model and profiled the FHIR (Fast Healthcare Interoperability Resource) Standard by creating a FHIR implementation Guide (IG) including examples of existing AI algorithms.

4. Discussion and Conclusion

The OMI protocol IG specifies data structure of common in- and output parameters and provides guidelines on how to represent algorithms. Specific requirements can only be modelled and profiled by the algorithm provider and made available to the OMI registry. OMI has the potential to democratize access to (AI) services for the scientific community in Germany.

[1] Corresponding Author: Stefan Sigle, stefan.sigle@molit.eu

5. Acknowledgement

This work was funded in part by the German Federal Ministry of Education and Research (BMBF), Funding reference number: 01ZZ2315J. The funding body did not play any role in the design of the study and collection, analysis, and interpretation of data as well as writing of the manuscript.

6. References

[1] S. Semler, F. Wissing, and R. Heyder, "German Medical Informatics Initiative: A National Approach to Integrating Health Data from Patient Care and Medical Research," *Methods Inf. Med.*, vol. 57, no. S 01, pp. e50–e56, Jul. 2018, doi: 10.3414/ME18-03-0003.

[2] "OMI - Medizininformatik-Plattform 'Open Medical Inference.'" Federal Ministry of Education and Research, Apr. 10, 2024. [Online]. Available: https://www.gesundheitsforschung-bmbf.de/de/omi-medizininformatik-plattform-open-medical-inference-16916.php

[3] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, no. 1, p. 160018, Mar. 2016, doi: 10.1038/sdata.2016.18.

[4] H. Hund, R. Wettstein, C. M. Heidt, and C. Fegeler, "Executing Distributed Healthcare and Research Processes – The HiGHmed Data Sharing Framework," in *Studies in Health Technology and Informatics*, R. Röhrig, T. Beißbarth, W. Brannath, H.-U. Prokosch, I. Schmidtmann, S. Stolpe, and A. Zapf, Eds., IOS Press, 2021. doi: 10.3233/SHTI210060.