Clinical Information Systems Research in the Pandemic Year 2020
An Overview of the CIS Section of the IMIA Yearbook of Medical Informatics

W.O. Hackl[1], A. Hoerbst[2], Section Editors of the IMIA Yearbook Section on Clinical Information Systems
[1] Institute of Medical Informatics, UMIT - Private University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria
[2] Medical Technologies Department, MCI — The Entrepreneurial School, Innsbruck, Austria

Summary

Objective: In this synopsis, we give an overview of recent research and propose a selection of best papers published in 2020 in the field of Clinical Information Systems (CIS).

Method: As CIS section editors, we annually apply a systematic process to retrieve articles for the International Medical Informatics Association Yearbook of Medical Informatics. For seven years now, we use the same query to find relevant publications in the CIS field. Each year we retrieve more than 2,400 papers which we categorize in a multi-pass review to distill a preselection of 15 candidate papers. External reviewers and yearbook editors then assess the selected candidate papers. Based on the review results, the IMIA Yearbook editorial board chooses up to four best publications for the section at a selection meeting. To get an overview of the content of the retrieved articles, we use text mining and term co-occurrence mapping techniques.

Results: We carried out the query in mid-January 2021 and retrieved a deduplicated result set of 2,787 articles from 1,135 different journals. We nominated 15 papers as candidates and finally selected four of them as the best papers in the CIS section. As in the previous years, the content analysis of the articles revealed the broad spectrum of topics covered by CIS research. Thus, this year we could observe a significant impact of COVID-19 on CIS research.

Conclusions: The trends in CIS research, as seen in recent years, continue to be observable. What was very visible was the impact of the Corona Virus Disease 2019 (COVID-19) pandemic, which has affected not only our lives but also CIS.

Keywords

Medical informatics, International Medical Informatics Association, yearbook, clinical information systems; COVID-19

1 Introduction

For seven years now, we are responsible for the clinical information systems (CIS) section of the International Medical Informatics Association Yearbook of Medical Informatics. In our search for the best papers in the field, we systematically screen more than 2,400 papers each year, retrieving from PubMed and Web of Science™ (WoS) using standardized queries. By doing so, we also get a good overview of the research activities in the CIS field in general. Additionally, every edition of the IMIA Yearbook is dedicated to a special topic that is reflected against the background of the retrieved papers.

We observed a move away from clinical documentation to patient-focused knowledge generation and support of the informed decision during the last years. Today, CIS are more than just tools or infrastructure for health care professionals and hospitals. Instead, they constitute the backbone for a very complex, trans-institutional information logistics process. The patient moves in the focus of interest, and data from the patient is used to produce value for the patients. Thus, trans-institutional information exchange, data aggregation, and data analysis are important research fields in the CIS domain [1–3].

Regarding the special topics, our observations vary each year. Last year, the focus was on "Ethics in Health Informatics". We realized that ethical aspects seem to be only a side issue as a research topic in the CIS domain [4]. This year, the special focus was on "Managing Pandemics with Health Informatics: Successes & Challenges". We were amazed at the strong influence of the Corona Virus Disease 2019 (COVID-19) on CIS research.

2 About the Paper Selection

The selection process in the CIS section is stable now for seven years. We described it in detail in [2], and the full queries are available upon request.

In mid-January 2021, we carried out the queries and retrieved 2,787 unique papers. We found 2,516 papers in PubMed and an additional set of 271 papers (deduplicated) in Web of Science. The resulting articles had been published in 1,135 different journals. Table 1 depicts the top-15-ranked journals with the highest numbers of resulting articles.

Interestingly, only 1,624 publication records included location information. Here, papers from the United States (39%, n=638), United Kingdom (27%, n=438), and the Netherlands (7%, n=121) made up the majority.

For the multi-pass selection process of the best papers, we used RAYYAN[1] as an online systematic review tool. We both (WOH, AH) independently reviewed all 2,787 publications and excluded ineligible articles based on their titles and/or abstracts in the first pass (WOH: n=2,697; AH: n=2,733), which resulted in an agreement rate of 95.1%

---

[1] https://www.rayyan.ai

Rank	Journal	Number of papers
1	Journal of medical Internet research	79
2	PloS One	62
3	International journal of medical informatics	48
	Health communication	48
4	BMC medical informatics and decision making	41
5	International journal of environmental research and public health	39
6	Vaccine	30
	JAMA network open	30
7	Journal of the American Medical Informatics Association	29
8	Applied clinical informatics	28
	Computers informatics nursing	28
9	Health informatics journal	25
10	Patient education and counseling	19
	Drug safety	19
	BMC health services research	19
11	BMC public health	18
12	Scientific reports	16
	Journal of medical systems	16
	JMIR mHealth and uHealth	16
13	JMIR medical informatics	15
	Hospital pediatrics	15
	BMJ Open	15
14	Research in social & administrative pharmacy	13
15	Journal of evaluation in clinical practice	12
	Environmental monitoring and assessment	12

Table 2 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2021 in the section 'Clinical Information Systems'. The articles are listed in alphabetical order of the first author's surname.

Section
Clinical Information Systems

- Fabregat A, Magret M, Ferré JA, Vernet A, Guasch N, Rodríguez A, Gómez J, Bodí M. A Machine Learning decision-making tool for extubation in Intensive Care Unit patients. Comput Methods Programs Biomed 2021;200:105869.
- Kempa-Liehr AW, Lin CYC, Britten R, Armstrong D, Wallace J, Mordaunt D, O'Sullivan M. Healthcare pathway discovery and probabilistic machine learning. Int J Med Inform 2020;137:104087.
- Li Y, Nair P, Lu XH, Wen Z, Wang Y, Dehaghi AAK, Miao Y, Liu W, Ordog T, Biernacka JM, Ryu E, Olson JE, Frye MA, Liu A, Guo L, Marelli A, Ahuja Y, Davila-Velderrain J, Kellis M. Inferring multimodal latent topics from electronic health records. Nat Commun 2020;11(1):2536.
- Weemaes M, Martens S, Cuypers L, van Elslande J, Hoet K, Welkenhuysen J, Goossens R, Wouters S, Houben E, Jeuris E, Jeuris K, Laenen L, Bruyninckx K, Beuselinck K, André E, Depypere M, Desmet S, Lagrou K, Van Ranst M, Verdonck AKLC, Goveia J. Laboratory information system requirements to manage the COVID-19 pandemic: A report from the Belgian national reference testing center. J Am Med Inform Assoc 2020;27(8):1293–9.

(n=2,651 for "exclude", and n=8 for "not exclude" - i.e., include). We included the remaining papers (n=136) in the next screening round, where we selected 25 papers for full-text review on mutual consent.

The final candidate selection yielded 15 candidate papers for the CIS section 2020. For each of these candidate papers, at least five independent reviews were collected. Due to COVID-19 restrictions, the selection meeting

with the IMIA Yearbook editorial board was – as in 2020 – held as a video conference on Apr 30th, 2021. As a result, four papers [5–8] were finally selected as the best papers for the CIS section (Table 2). A content summary of these four best CIS papers can be found in the appendix of this synopsis.

3 Findings and Trends: Clinical Information Systems Research 2020

During the selection process of the best paper candidates, we get a comprehensive overview of the research field of our CIS section. We also apply our more formal text mining and bibliometric network visualizing approach [9] to summarize the articles' content and abstracts in our CIS result to avoid bias and selective perception.

As in the past year, we extracted the authors' keywords (n=16,017) from all articles and presented their frequency in a tag cloud (Figure 1). We found 4,646 different keywords, of which 3,082 were only used once. As in the previous year, most frequent keywords were "humans" (n=964) followed by "numerical data", "female" (n=435), "male" (n=419), "aged" (n=298), "adult" (n=278), and "middle-aged" (n=275).

The bibliometric network reveals more details on the content of the CIS publications. Figure 2 depicts the resulting co-occurrence map of the top-500 terms (n=518, most relevant 60% of the terms) from the abstracts of the 2,787 papers of the CIS result set.

The cluster analysis of the titles and abstracts yielded five clusters. The two most massive clusters, the red one on the left side (n=243 items) and the green cluster on the right (n=187 items), describe context factors, targets, and methodological aspects from the studies. The remaining three clusters are considerably smaller. The yellow one (n=28), dedicated to adverse event detection and reporting, is constantly present in our analyses during the years. Finally, the purple cluster (n=11), also constantly present, reflects location-based aspects. This seems to be an artifact caused by the query where geographic information systems are explicitly included.
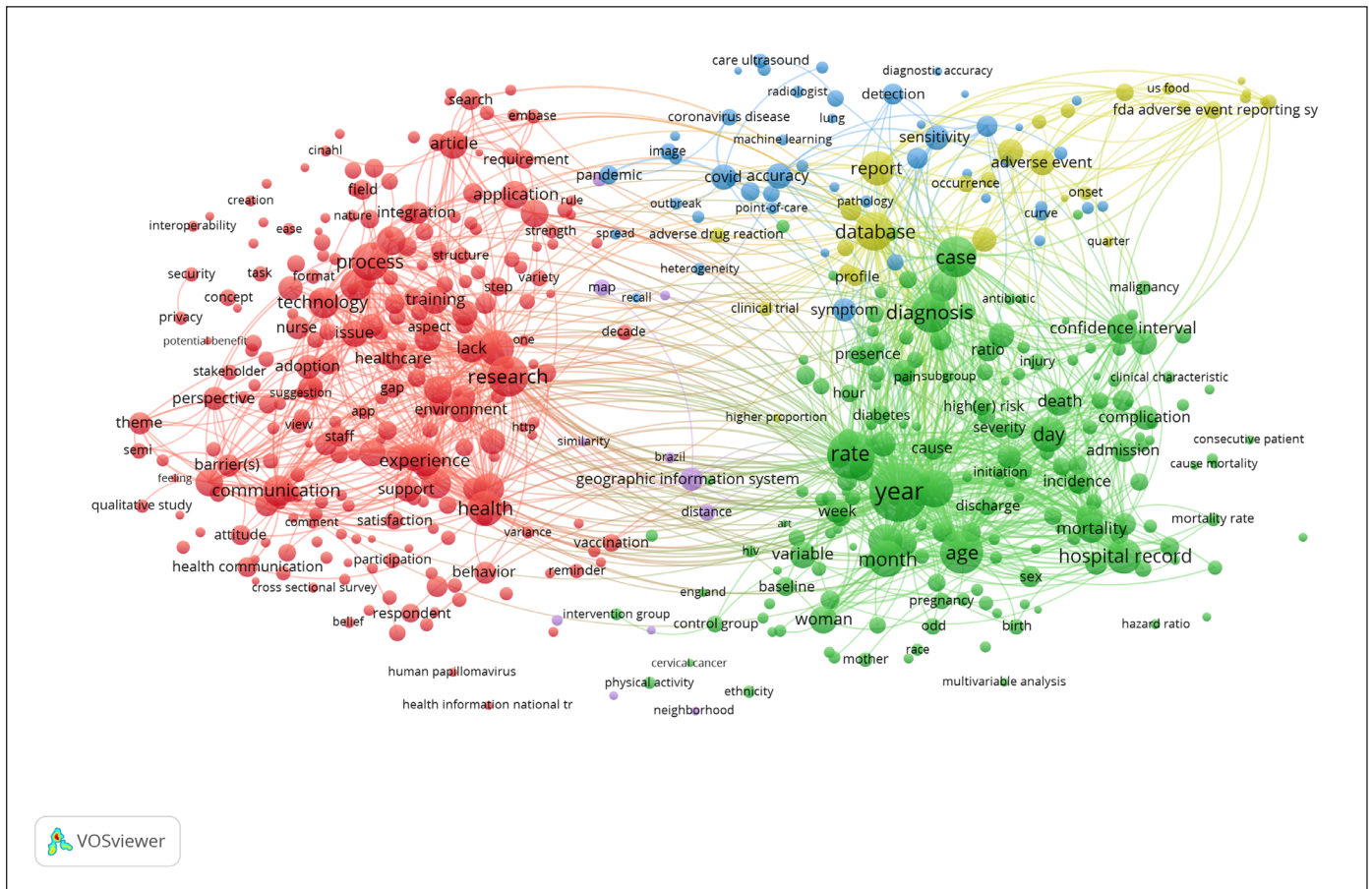
Fig. 1 Tag cloud illustrating the frequency authors' keywords (only top keywords out of n=4,646 are shown) within the 2,787 papers from the CIS query result set. Font size corresponds to frequency (the most frequent keyword was "humans" n=964).

The blue cluster (n=49) is new this year. It holds items related to COVID-19 research and scientific response in the CIS field to the pandemic situation. And in fact, we found a lot of papers dealing with these topics. It was interesting to see how fast and how quickly relevant research is produced and published. In keeping with this, two of the best papers are located in this cluster.

The first of the best papers is a contribution by Weemaes *et al.* who excellently and briefly describe the development, implementation, and requirements of laboratory information system functionalities to manage test ordering, registration, sample flow, and result reporting from the Belgian national reference testing center during the COVID-19 pandemic [5]. The second one is a contribution of Fabregat *et al.* from Spain,

who developed a machine learning decision-making tool for extubation in intensive care unit patients that accurately predicts extubation outcome [6].

The third of the best papers is very interesting from a methodological point of view. Li et al present a multi-view Bayesian topic model that can reveal meaningful combinations of clinical features across highly sparse, biased, and heterogeneous electronic health record (EHR) data, and provide clinical recommendations by predicting undiagnosed patient phenotypes [7]. Finally, the fourth of the best papers also tackles an interesting and increasingly crucial methodological aspect. Kempa-Liehr *et al.* propose a pipeline for healthcare pathway discovery. In a case study, they show how to combine health-

care pathway discovery with predictive models of individualized recovery times after appendicectomy [8].

Also, among the remaining eleven candidate papers, there are very interesting contributions worth reading. For example, a systematic review on mobile health interventions in developing countries by Hoque *et al.* [11], a cluster-randomized clinical trial on mobile technology care coordination of long-term services and support by Quinn *et al.* [12], and a contribution by Weenk *et al.* who investigated positive and negative effects, barriers and facilitators for the use of wearable devices for continuous monitoring of vital signs in a randomized controlled trial [13].

Publications on the development or application of data analysis or data mining methods, machine learning, and prediction

Fig. 2 Clustered co-occurrence map of the top-500 terms (top 60% of the most relevant terms: n=518) from the titles and abstracts of the 2,787 papers in the 2021 CIS query result set. Only terms that we found in at least seven different papers were included in the analysis. Node size corresponds to the frequency of the terms (binary count, once per paper, year: n=758). Edges indicate co-occurrence (only the top 1,000 of 72,146 edges are shown). The distance of nodes corresponds to the association strength of the terms within the texts. Colors represent the five different clusters. The network was created with VOSviewer [10].

models make up a significant proportion of the papers in our query result set. Among the candidate papers, we have three representatives of this kind of research. Muñoz *et al.* developed and validated a prediction model for adverse event reporting [14], Groenhof *et al.* demonstrated how the mining of EHR data can be used to estimate patients' smoking status accurately [15], and Zhang *et al.* present a study on real-time artificial intelligence prediction for major adverse cardiac events [16].

As CIS are socio-technical systems, we have also considered the socio-organizational perspective of CIS and selected appropriate candidate papers. For example, Everson and Butler investigated hospital

adoption of multiple health information exchange approaches and information accessibility [17], Mosher *et al.* assessed the effects of patient check-in kiosks in the outpatient clinical setting [18], and Bersani *et al.* investigated use, perceived usability, and barriers to implementation of a patient safety dashboard [19]. All three are worth reading.

For those interested in the potentially seminal topic "blockchain in healthcare", we have included a scoping review by Hasselgren *et al.* in our candidate paper selection [20]. And for those interested in security, we have selected an inspiring article by Omolara *et al.* They present a prototype called HoneyDetails, a decep-

tion-based defense system against cyber-criminals seeking to steal patient data from EHR systems [21].

As every year, at the very end of our review of findings and trends for the CIS section, we want to recommend a reading of this year's survey article in the CIS section by Jeffery Reeves, Natalie Pageler, Elizabeth Wick, Genevieve Melton, Gamaliel Tan, Brian Clay, and Chris Longhurst. The objective of their article was to review the areas in which CIS can be and have been utilized to support and enhance the response of healthcare systems to pandemics, focusing on COVID-19. And they came up with a genuinely comprehensive analysis [22].

4 Conclusions and Outlook

Trends in CIS research observed in previous years will continue. Patient-centeredness, trans-institutional information sharing, intelligent clinical data analytics capabilities, artificial intelligence, machine learning, and decision support are on the rise. Telehealth services and networked, integrated care are other vital topics for CIS research.

Very evident this year was the impact of the COVID-19 pandemic on CIS research. We found a number of publications that addressed problems of information logistics for the management of the pandemic, some of which offered interesting approaches and solutions.

References

1. Hackl WO, Ganslandt T. Clinical Information Systems as the Backbone of a Complex Information Logistics Process: Findings from the Clinical Information Systems Perspective for 2016. Yearb Med Inform 2017;26(1):103–9.
2. Hackl WO, Hoerbst A. On the Way to Close the Loop in Information Logistics: Data from the Patient - Value for the Patient. Yearb Med Inform 2018;27(1): 91–7.
3. Hackl WO, Hoerbst A. Managing Complexity. From Documentation to Knowledge Integration and Informed Decision Findings from the Clinical Information Systems Perspective for 2018. Yearb Med Inform 2019 Aug 16;28(1):95–100.
4. Hackl WO, Hoerbst A. Trends in Clinical Information Systems Research in 2019. Yearb Med Inform 2020;29(1):121–8.
5. Weemaes M, Martens S, Cuypers L, van Elslande J, Hoet K, Welkenhuysen J, et al. Laboratory information system requirements to manage the COVID-19 pandemic: A report from the Belgian national reference testing center. J Am Med Inform Assoc 2020;27(8):1293–9.
6. Fabregat A, Magret M, Ferré JA, Vernet A, Guasch N, Rodríguez A, et al. A Machine Learning decision-making tool for extubation in Intensive Care Unit patients. Comput Methods Programs Biomed 2021;200:105869.
7. Li Y, Nair P, Lu XH, Wen Z, Wang Y, Dehaghi AAK, et al. Inferring multimodal latent topics from electronic health records. Nat Commun 2020;11(1):2536.
8. Kempa-Liehr AW, Lin CYC, Britten R, Armstrong D, Wallace J, Mordaunt D, et al. Healthcare pathway discovery and probabilistic machine learning. Int J Med Inform 2020 May;137:104087.
9. Waltman L, van Eck NJ, Noyons ECM. A unified approach to mapping and clustering of bibliometric networks. J Informetr 2010;4(4):629–35.
10. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010;84(2):523–38.
11. Hoque MR, Rahman MS, Nipa NJ, Hasan MR. Mobile health interventions in developing countries: A systematic review. Health Informatics J 2020;26(4):2792–810.
12. Quinn CC, Knopp KL, Marsiglia CA, Bearch RC, LeFever A, Barr EA, et al. Mobile Technology Care Coordination of Long-Term Services and Support: Cluster Randomized Clinical Trial. J Appl Gerontol 2021;40(5):529–35.
13. Weenk M, Bredie SJ, Koeneman M, Hesselink G, Van Goor H, Van De Belt TH. Continuous monitoring of vital signs in the general ward using wearable devices: Randomized controlled trial. J Med Internet Res 2020;22(6):1–11.
14. Muñoz MA, Dal Pan GJ, Wei YJJ, Delcher C, Xiao H, Kortepeter CM, et al. Towards Automating Adverse Event Review: A Prediction Model for Case Report Utility. Drug Saf 2020;43(4):329–38.
15. Groenhof TKJ, Koers LR, Blasse E, de Groot M, Grobbee DE, Bots ML, et al. Data mining information from electronic health records produced high yield and accuracy for current smoking status. J Clin Epidemiol 2020;118:100–6.
16. Zhang PI, Hsu CC, Kao Y, Chen CJ, Kuo YW, Hsu SL, et al. Real-time AI prediction for major adverse cardiac events in emergency department patients with chest pain. Scand J Trauma Resusc Emerg Med 2020;28(1):1–7.
17. Everson J, Butler E. Hospital adoption of multiple health information exchange approaches and information accessibility. J Am Med Inform Assoc 2020;27(4):577–83.
18. Mosher ZA, Hudson PW, Lee SR, Perez JL, Arguello AM, McGwin G, et al. Check-in Kiosks in the Outpatient Clinical Setting: Fad or the Future? South Med J 2020;113(3):134–9.
19. Bersani K, Fuller TE, Garabedian P, Espares J, Mlaver E, Businger A, et al. Use, Perceived Usability, and Barriers to Implementation of a Patient Safety Dashboard Integrated within a Vendor EHR. Appl Clin Inform 2020;11(1):34–45.
20. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A. Blockchain in healthcare and health sciences—A scoping review. Int J Med Inform 2020;134(November 2019):104040.
21. Esther Omolara A, Jantan A, Abiodun OI, Arshad H, Dada KV, Emmanuel E. HoneyDetails: A prototype for ensuring patient's information privacy and thwarting electronic health record threats based on decoys. Health Informatics J 2020;26(3):2083–104.
22. Reeves JJ, Pageler NM, Wick E, Melton GB, Tan G, Clay BJ, et al. The Clinical Information Systems Response to the COVID-19 Pandemic. Yearb Med Inform 2021:105-25.

Correspondence to:
Dr. Werner O Hackl
Institute of Medical Informatics
UMIT – Private University for Health Sciences, Medical Informatics, and Technology
Eduard-Wallnoefer-Zentrum 1
6060 Hall in Tirol, Austria
Tel: +43 50 8648 3806
E-mail: werner.hackl@umit.at

Dr. Alexander Hörbst
Medical Technologies Department
MCI - THE ENTREPRENEURIAL SCHOOL
Universitätsstrasse 15
6020 Innsbruck
Austria
Tel: +43 512 2070 4410
E-mail: alexander.hoerbst@mci.edu

Appendix: Content Summaries of Selected Best Papers for the IMIA Yearbook 2021 Section "Clinical Information Systems"

Fabregat A, Magret M, Ferré JA, Vernet A, Guasch N, Rodríguez A, Gómez J, Bodí M

A Machine Learning decision-making tool for extubation in Intensive Care Unit patients

Invasive Mechanical Ventilation (IMV) is central to treating patients who are unable to maintain adequate pulmonary ventilation and oxygenation to allow patients to recover. Although IMV can be a life-saving procedure, it also bears significant risks such as ventilator-induced lung injuries or infections as well as long-term problems after recovery. One of the critical decisions regarding IMV is weaning. This includes, amongst other steps, the removal of the endotracheal tube. Patients that need to be reintubated bear several risks and problems associated, including increased mortality (25%-50%). The goal of the current work was to create a machine learning (ML) model that can increase the successful extubation rate in adult Intensive Care Unit (ICU) patients. The model is based on data routinely collected from patients' health record data. Patients included were admitted to an ICU in a Spanish hospital between 2015 and 2019, and received at least 12 consecutive hours of IMV. Variables used for prediction were categorized in four types: T1: time series data (averaged over 20 minutes) (e.g., heart rate); T2: derived variables from T1 (e.g., respiratory rate); T3: discrete event information, (e.g., Glasgow coma scale (GCS)); T4: demographics and admission information (e.g., gender). In total, 20 predictors were used. The resulting dataset had a strong imbalance with regard to successful extubations (1,108 versus 100). Therefore, randomly selected data points of the most frequent class (successful extubation) were removed for the training data set

and/or a weight was assigned to data points. Seven-fold cross-validation was determined appropriate. Extubation/Reintubation was basically identified by finding gaps on the IMV monitor signal larger than 48 hours. As several possible errors have an influence on this gap, a comparison with medical records was necessary to correct the numbers (final dataset: 647 successful and 50 failed). Three different ML classifiers were compared: support vector machine (SVM) with radial basis, gradient boosting machine (GBM) with Bernoulli loss, and Linear Discriminant Analysis (LDA). Mean Accuracy and AU-ROC were used to determine performance. The following scores were achieved: SVM 94.6% and 98.3%; GBM 87% and 96%; LDA 72% and 79%. The results suggest that the top five predictors in descending order of importance are time, GCS, body mass index, respiratory rate-oxygenation index, and plateau pressure. On the other hand, the least relevant predictors in descending order of importance are Spanish Society of Intensive, Critical Medicine and Coronary Units classification code for ICU admission reason, gender, total cumulative dose, total given dose, and ventilation mode. The models should not be applied as a general-purpose predictor of success for programmed extubations or as a monitoring alarm system but as a support tool to validate the medical staff's decision. With the predictive accuracy achieved, the rate of failed extubation (currently 9%) could be reduced to a theoretical 1%. The results suggest that ML tools are especially well suited to support the decision-making protocol based on spontaneous breathing trials to decide about extubation.

Kempa-Liehr AW, Lin CYC, Britten R, Armstrong D, Wallace J, Mordaunt D, O'Sullivan M

Healthcare pathway discovery and probabilistic machine learning

The success of electronic health records has also driven several other research areas such as knowledge management in healthcare, which basically involves four steps: (1) data access; (2) knowledge discovery; (3) knowledge translation and interpreta-

tion, as well as (4) knowledge description, integration and sharing. An important role hereby is played by healthcare pathways that incorporate the operational knowledge of a healthcare organization by defining the execution sequence of clinical activities as patients move through a treatment process. In many cases, these pathways result from clinician-led practice rather than explicit design, which leads to several problems (e.g., lack of update). The study aims to combine healthcare pathway discovery with predictive models of individualized recovery times after appendicectomy. Particular emphasis is set on easy to interpret models for clinicians. The predictive model takes the stochastic volatility of pathway performance indicators into account and can replicate the dominant mode as well as the fat tail of the empirical recovery time distribution. To mine the pathways, the ProM software was used. First, healthcare pathway variations were discovered and then reduced (clustering, merging consecutive activities, condense repetitive patterns) to meaningful models. In a second step, conformance of these models with actual patient traces is evaluated, including new findings into the model leads to an iterative approach between pathway discovery and conformance analysis. The third step involves data enrichment, which comprises two stages: healthcare pathway performance evaluation and healthcare pathway performance analysis. The main objectives of evaluating healthcare pathway performance are to understand the strengths and weaknesses of the current pathway design. Analyzing the performance of healthcare pathways with respect to pathway variants and other possible influencing factors like demographics or patient-specific pathway characteristics (e.g., surgery duration) is the final step of the proposed process mining pipeline. For the appendicitis model, 13 pathway variants were discovered, whereas the top four variants accounted for approximately 88% of the patient traces. In a next step, it was analyzed if the variants are relevant features or covariates for explaining the stochastic volatility of postoperative length of stay. To build two probabilistic machine learning models, 415 individual patient traces were used. The two models showed promising results to explain the length of stay. Summarizing, the pro-

posed process mining pipeline successfully constructed concise pathway models for the appendicitis case study and, therefore, supported the generation of probabilistic machine learning models.

Li Y, Nair P, Lu XH, Wen Z, Wang Y, Dehaghi AAK, Miao Y, Liu W, Ordog T, Biernacka JM, Ryu E, Olson JE, Frye MA, Liu A, Guo L, Marelli A, Ahuja Y, Davila-Velderrain J, Kellis M

Inferring multimodal latent topics from electronic health records

Nat Commun 2020;11(1):2536

Electronic health records (EHRs) are heterogeneous collections of patient health information that would support multiple uses such as risk prediction, clinical recommendations, or individual therapeutic concepts. However, raw data in EHRs is in many cases not directly processable, especially when building formal models. Different challenges such as non-standardized clinical notes, heterogeneous data types, missing standardization, or diagnosis-driven lab tests pose challenges. Appropriate and effective computational methods have the potential to overcome those challenges and provide access to an encyclopedia of diseases, disorders, injuries, and other related health conditions, uncovering a modular phenotypic network. The paper introduces MixEHR to: (1) distill meaningful disease topics from otherwise highly sparse, biased, and heterogeneous EHR data; and (2) provide clinical recommendations by predicting undiagnosed patient phenotypes based on their disease mixture membership. MixEHR builds on collaborative filtering and latent topic modeling and can model various EHR categories with separate discrete distributions. A variational inference algorithm that scales to large-scale EHR data was created. The model was applied to three EHR datasets: (1) Medical Information Mart for Intensive Care (MIMIC)-III (50,000 intensive care unit admissions); (2) Mayo Clinic EHR dataset containing 187 patients, including with 93 bipolar disorders and 94 controls; (3) The Régie de l'assurance maladie du Québec Congenital Heart Disease Dataset (Quebec CHD Database; more than

80,000 patients with congenital heart disease). The authors followed a probabilistic joint matrix factorization approach. The high dimensional and heterogeneous clinical record was projected onto a low dimension probabilistic meta-phenotype signature, reflecting the patient's mixed memberships across diverse latent disease topics. Factorization is carried out at two levels. At the lower level, data-type-specific topic models, learning a set of basis matrices for each data type, were applied. A common loading matrix that connects the multiple data types for each patient was used at the higher level. The approach was used, among others, to define a disease comorbidity network, create patient risk prioritization, EHR code predictions, or mortality predictions from the given datasets. Overall, the MixEHR approach's accuracy scores top compared to other existing approaches. MixEHR can infer expected phenotypes of a patient conditioned only on a subset of clinical variables that are perhaps easier and cheaper to measure. Currently, data are a set of two-dimensional matrices of patients by measurements in the model. To model higher dimensional objects such as patient by lab test by diagnoses, MixEHR could be extended to a probabilistic tensor-decomposition framework.

Weemaes M, Martens S, Cuypers L, van Elslande J, Hoet K, Welkenhuysen J, Goossens R, Wouters S, Houben E, Jeuris E, Jeuris K, Laenen L, Bruyninckx K, Beuselinck K, André E, Depypere M, Desmet S, Lagrou K, Van Ranst M, Verdonck AKLC, Goveia J

Laboratory information system requirements to manage the COVID-19 pandemic: A report from the Belgian national reference testing center

J Am Med Inform Assoc 2020;27(8):1293–9

The paper describes the challenges faced by the Belgian National Reference Center for COVID-19 testing at the University Hospitals Leuven, when demand passed allocated surge capacity during the initial phases of the COVID-19 pandemic. This includes the design, implementation and requirements of laboratory information system (LIS) functionality related to managing increased test demand during the COVID-19 crisis. In particular, all phases in laboratory testing were streamlined: the pre-laboratory phase (test ordering, sample packaging, and shipping); the pre-analytical phase (sample registration, tracking, and test prioritization); and the post-analytical phase (automated reporting and facilitating data-driven policy-making). Apart from COVID-19 testing, the laboratory concerned performs more than 12,000,000 lab tests a year. The LIS is in-house developed and maintained by a dedicated team. The system includes a computerized physician order entry (CPOE) module for in-house test ordering, which is fully integrated into the electronic health record (EHR). All external orders were initially paper-based and required that request forms accompany the sample. In the course of the analysis, 17 major challenges were identified in the different phases of the testing process. Selected solutions included: a COVID-19 specific CPOE module was linked to both the LIS and EHR, allowing to automatically retrieve demographic information, which dramatically improved metadata completeness; a "COVID-19 status" button on the main page of the EHR of each patient was displayed, showing in real-time the results of SARS-CoV-2 laboratory testing; a database with contact details and preferred reporting methods (e.g., fax, email, electronic mailbox system) of every laboratory in Belgium was compiled, to enable automated test reporting (resulted in more than 98% automated reporting). To successfully implement such changes in a short time, several prerequisites apply. The authors, therefore, recommend that crisis management teams not only consist of staff focused on increasing analytical capacity but also information technology-staff and to apply change management frameworks. To summarize, the most effective solutions reported were to streamline sample ordering through a CPOE system and reporting by developing a database with contact details of all laboratories in Belgium. In addition, the implementation of R/Shiny-based statistical tools facilitated epidemiological reporting and enabled explorative data mining.