Novelty in Public Health and Epidemiology Informatics

Gayo Diallo[1], Georgeta Bordea[2], Section Editors for the IMIA Yearbook Section on Public Health and Epidemiology Informatics

[1] Inria SISTM, Team AHeaD - INSERM Bordeaux Population Health Research Center, Univ. Bordeaux, Bordeaux, France

[2] Team AHeaD - Inserm BPH Research Center & LaBRI UMR 5800, Univ. Bordeaux, Bordeaux, France

Summary

Objectives: To highlight novelty studies and current trends in Public Health and Epidemiology Informatics (PHEI).

Methods: Similar to last year's edition, a PubMed search of 2021 scientific publications on PHEI has been conducted. The resulting references were reviewed by the two section editors. Then, 11 candidate best papers were selected from the initial 782 references. These papers were then peer-reviewed by selected external reviewers. They included at least two senior researchers, to allow the Editorial Committee of the 2022 IMIA Yearbook edition to make an informed decision for selecting the best papers of the PHEI section.

Results: Among the 782 references retrieved from PubMed, two were selected as the best papers. The first best paper reports a study which performed a comprehensive comparison of traditional statistical approaches (e.g., Cox Proportional Hazards models) vs. machine learning techniques in a large, real-world dataset for predicting breast cancer survival, with a focus on explainability. The second paper describes the engineering of deep learning models to establish associations between ocular features and major hepatobiliary diseases and to advance automated screening and identification of hepatobiliary diseases from ocular images.

Conclusions: Overall, from this year edition, we observed that the number of studies related to PHEI has decreased. The findings of the two studies selected as best papers on the topic suggest that a significant effort is still being made by the community to compare traditional learning methods with deep learning methods. Using multimodality datasets (images, texts) could improve approaches for tackling public health issues

Keywords

Public health; epidemiology informatics; IMIA Yearbook 2022

Yearb Med Inform 2022:273-6
http://dx.doi.org/10.1055/s-0042-1742526

1 Introduction

There are various types of health informatics, which is the confluence of people, technology, and data to improve the safety and quality of patient care. Patient portals, electronic health records (EHRs), telehealth, healthcare apps, and a range of data reporting tools are all just a few examples. Nowadays, solving pressing public health needs requires access to and cross-referencing of various types of data and the use of approaches and techniques from different fields: epidemiology and population health studies, bio-statistics and machine learning, big data and heterogeneous data integration, etc. In this context, there are currently various initiatives around the world consisting in enabling health related data sharing frameworks. For instance, in Europe, the European Health Data Space (EHDS)[1] is a recent initiative which aims at providing a health-specific data sharing framework that establishes clear rules, common standards and practices, infrastructures, and a governance framework for patients to use electronic health data for research, innovation, policy-making, patient safety, statistics, and regulatory purposes. An important pillar of the EHDS is the secondary use of data for better research and policy-making. Similarly, in the United States (US), the National COVID Cohort Collaborative (N3C) is worth mentioning. N3C maintains one of the largest collections of clinical data related to Covid-19 symptoms and patient outcomes in the US[2] allowing many organisations to give clinical data in near-real time, enabling to learn more about COVID-19 and potential treatment options. Indeed, early in the current COVID-19 pandemic, issues arose that highlighted the need for public health authorities to combine diagnostic, clinical, and demographic data at scale to facilitate real-time monitoring and decision-making [1]. Overall, EHRs and databases holding other health-related data (claims, pharmacy) can help support observational research and pragmatic clinical trials, which are both valuable sources of real-world evidence [2].

The current Public Health and Epidemiology Informatics (PHEI) synthesis, which follows the last edition of the IMIA Yearbook, reviews the scientific literature published in 2021 in the domain of medical informatics and which addresses the above challenges. Targeted studies focus more specifically on the sub-field of public health and epidemiology. In comparison to past years, it tries to find new subjects and trends within this discipline. It also explains how the best articles published in 2021 were chosen through a selection procedure. To that purpose, the section editors oversaw the peer review of the most representative research papers, which were chosen for their quality and originality.

[1] https://ec.europa.eu/health/publications/proposal-regulation-european-health-data-space_en

[2] https://ncats.nih.gov/n3c/about

2 Methods

As with the previous edition of the IMIA Yearbook for the PHEI section [3], a comprehensive literature search was performed by the section editors using the PubMed/Medline database from the National Center for Biotechnology Information (NCBI). A large set of MeSH descriptors was used to retrieve relevant studies. The queries targeted public health or epidemiological journal articles over the year 2021 and that included medical informatics topics. Returned references addressing topics of the other sections of the Yearbook, e.g., those related to COVID-19 or sensors, were excluded from our search.

The study was performed early March 2021, and the search returned a total of 782 references, which is lower than for the 2021 Yearbook edition. Articles were reviewed separately by the two section editors, considering the number of citations for each reference in the process. They were first classified into three categories: keep, discard, or pending using the BibReview tool [4]. Then, the two lists of references were merged, yielding 88 references that were retained by at least one section editor or classified as "pending" by both of them. The two section editors jointly reviewed these 88 references and selected a consensual list of 11 candidate best papers. These papers are depicted in Table 1.

All of these papers were then peer-reviewed by the PHEI section editors and external reviewers. Each paper was reviewed by at least five reviewers to ensure a very informed decision making. Three papers emerged at the end of this process. However, the third paper [5], which addresses the detection of medical misinformation, although very relevant to the section, was removed due to the lack of an appropriate discussion section about the ethical consideration for using social media data. The two papers that have been finally selected, respectively [6] and [7] both rely on Machine Learning (ML) approaches applied to large datasets of patients. A summary of these papers is provided in the appendix.

Table 1 List of the 11 candidate papers for the PHEI section.

Authors & References	Title
Wang et al., [5]	Detecting Medical Misinformation on Social Media Using Multimodal Deep Learning
Moncada-Torres et al., [6]	Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival
Xiao et al., [7]	Screening and identifying hepatobiliary diseases through deep learning using ocular images: a prospective, multicentre study
Herbuela et al., [8]	Early Detection of Dengue Fever Outbreaks Using a Surveillance App (Mozzify): Cross-sectional Mixed Methods Usability Study
Homer and Wakeland [9]	A dynamic model of the opioid drug epidemic with implications for policy
Kuzmenkov et al., [10]	AMRmap: An Interactive Web Platform for Analysis of Antimicrobial Resistance Surveillance Data in Russia
Razavi-Termeh et al,. [11]	Asthma-prone areas modeling using a machine learning model
Zwan et al,. [12]	Dutch Brain Research Registry for study participant recruitment: Design and first results
Köhler et al., [13]	The Human Phenotype Ontology in 2021
Huang et al., [14]	Modeling the competitive diffusions of rumor and knowledge and the impacts on epidemic spreading
Vable et al., [15]	Code Review as a Simple Trick to Enhance Reproducibility, Accelerate Learning, and Improve the Quality of Your Team's Research

3 Results

3.1 Overall Observation

Broad trends in public health and epidemiology informatics include an increased interest in surveillance methods based on mobile apps for early detection of disease outbreaks [8]. Public health issues related to the opioid drug epidemic [9], antimicrobial resistance [10], or the impact of environmental pollution [11] are more and more analysed on a wide scale using data sources such as social media, internet search data, or Wikitrends and on a variety of modalities including images, text and spatial data. The practice of recruiting participants for large-scale studies through online social media campaigns has seen an increased adoption [12]. In terms of informatics approaches, ontologies have matured to the level of worldwide standards for health data exchange [13]. The impact of misinformation and rumors during an epidemic outbreak was modeled using social networks [14]. In addition, the focus has moved from simply applying and increasing performance of deep learning approaches [7] to explainable machine learning and in depth analysis of salient features [6].

3.2 Selected Papers

Thus, among the 782 retrieved articles from the PHEI section's relevant keywords, two best papers have been selected this year (Table 2). The first study focuses on explainability in the context of ML [6] while the second study is a multi-centric study to identify hepatobiliary diseases from ocular images using deep learning techniques [7].

4 Conclusion

The 2022 edition of the PHEI section has confirmed the use of ML and deep learning approaches for data analysis to address public health issues.

A concern that has become more apparent in the wake of the recent epidemic and has yet to be addressed by the community is the ability to design and implement individualised public health approaches and tools.

Table 2 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2022 in the section 'Public Health and Epidemiology Informatics'. The articles are listed in alphabetical order of the first author's surname.

Section
Public Health and Epidemiology Informatics
▪ Moncada-Torres A, van Maaren MC, Hendriks MP, Siesling S, Geleijnse G. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. Sci Rep 2021 Mar 26;11(1):6968.
▪ Xiao W, Huang X, Wang JH, Lin DR, Zhu Y, Chen C, Yang YH, Xiao J, Zhao LQ, Li JO, Cheung CY, Mise Y, Guo ZY, Du YF, Chen BB, Hu JX, Zhang K, Lin XS, Wen W, Liu YZ, Chen WR, Zhong YS, Lin HT. Screening and identifying hepatobiliary diseases through deep learning using ocular images: a prospective, multicentre study. Lancet Digit Health 2021 Feb;3(2):e88-e97.

Acknowledgements

The Public Health and Epidemiology Informatics section editor would like to thank the external reviewers of the section for their valuable contribution in evaluating and selecting the best papers for this 2022 IMIA Yearbook PHEI section.

References

1. Arora J, Mega JL, Abernethy Stadtlander A. Connecting Real-World Data to Support Public Health Efforts. NEJM Catalyst Innovations in Care Delivery; May 2022.
2. National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Health Sciences Policy; Forum on Drug Discovery, Development, and Translation. Real-World Evidence Generation and Evaluation of Therapeutics: Proceedings of a Workshop. Washington (DC): National Academies Press (US); 2017 Feb 15. 2, Improving Evidence Generation for Decision Making on Approval and Use of New Treatments: Some Stakeholder Priorities.
3. Diallo G, Bordea G. Public Health and Epidemiology Informatics: Recent Research Trends. Yearb Med Inform 2021 Aug;30(1):280-2.
4. Lamy JB, Séroussi B, Griffon N, Kerdelhué G, Jaulent MC, Bouaud J. Toward a formalization of the process to select IMIA Yearbook best papers. Methods Inf Med 2015;54(2):135-44.
5. Wang Z, Yin Z, Argyris YA. Detecting Medical Misinformation on Social Media Using Multimodal Deep Learning. IEEE J Biomed Health Inform 2021 Jun;25(6):2193-203.
6. Moncada-Torres A, van Maaren MC, Hendriks MP, Siesling S, Geleijnse G. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. Sci Rep 2021 Mar 26;11(1):6968.
7. Xiao W, Huang X, Wang JH, Lin DR, Zhu Y, Chen C, et al. Screening and identifying hepatobiliary diseases through deep learning using ocular images: a prospective, multicentre study. Lancet Digit Health 2021 Feb;3(2):e88-e97.
8. Herbuela VRDM, Karita T, Carvajal TM, Ho HT, Lorena JMO, et al. Early Detection of Dengue Fever Outbreaks Using a Surveillance App (Mozzify): Cross-sectional Mixed Methods Usability Study. JMIR Public Health Surveill 2021 Mar 1;7(3):e19034. Erratum in: JMIR Public Health Surveill 2021 Apr 22;7(4):e29795.
9. Homer J, Wakeland W. A dynamic model of the opioid drug epidemic with implications for policy. Am J Drug Alcohol Abuse 2021 Jan 2;47(1):5-15.
10. Kuzmenkov AY, Trushin IV, Vinogradova AG, Avramenko AA, Sukhorukova MV, Malhotra-Kumar S, et al. AMRmap: An Interactive Web Platform for Analysis of Antimicrobial Resistance Surveillance Data in Russia. Front Microbiol. 2021 Mar 12;12:620002.
11. Razavi-Termeh SV, Sadeghi-Niaraki A, Choi SM. Asthma-prone areas modeling using a machine learning model. Sci Rep 2021(11):1912.
12. Huang H, Chen Y, Ma Y. Modeling the competitive diffusions of rumor and knowledge and the impacts on epidemic spreading. Appl Math Comput 2021 Jan 1;388:125536.
13. Zwan MD, van der Flier WM, Cleutjens S, Schouten TC, Vermunt L, Jutten RJ, et al. Dutch Brain Research Registry for study participant recruitment: Design and first results. Alzheimers Dement (N Y) 2021 Feb 14;7(1):e12132.
14. Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al. The Human Phenotype Ontology in 2021. Nucleic Acids Res 2021 Jan 8;49(D1):D1207-D1217.
15. Vable AM, Diehl SF, Glymour MM. Code Review as a Simple Trick to Enhance Reproducibility, Accelerate Learning, and Improve the Quality of Your Team's Research. Am J Epidemiol 2021 Oct;190(10):2172-7.

Correspondence to:
Gayo Diallo
Inria SISTM
Team AHeaD
INSERM Bordeaux Population Health Research Center
Univ. Bordeaux
146, rue Léo Saignat
F-33000 Bordeaux, France
E-mail: gayo.diallo@u-bordeaux.fr