

Broad Trends in Public Health and Epidemiology Informatics

Gayo Diallo¹, Georgeta Bordea², Cécilia Samieri¹, Section Editors for the IMIA Yearbook Section on Public Health and Epidemiology Informatics

¹ Univ. Bordeaux, Inserm, BPH, U1219, F-33000 Bordeaux, France

² Univ. La Rochelle, L3i, EA 2118, F-17000, La Rochelle, France

Summary

Objectives: The objective of this study is to highlight innovative research and contemporary trends in the area of Public Health and Epidemiology Informatics (PHEI).

Methods: Following a similar approach to last year's edition, a meticulous search was conducted on PubMed (with keywords including topics related to Public Health, Epidemiological Surveillance and Medical Informatics), examining a total of 2,022 scientific publications on Public Health and Epidemiology Informatics (PHEI). The resulting references were thoroughly examined by the three section editors. Subsequently, 10 papers were chosen as potential candidates for the best paper award. These selected papers were then subjected to peer-review by six external reviewers, in addition to the section editors and two chief editors of the IMIA yearbook of medical informatics. Each paper underwent a total of five reviews.

Results: Out of the 539 references retrieved from PubMed, only two were deemed worthy of the best paper award, although four papers had the potential to qualify in total. The first best paper by pertains to a study about the need for a new annotation framework due to inadequacies in existing methods and resources. The second paper elucidates the use of Weibo data to monitor the health of Chinese urbanites. The correlation between air pollution and health sensing was measured via generalized additive models.

Conclusions: One of the primary findings of this edition is the dearth of studies identified for the PHEI section, which represents a significant decline compared to the previous edition. This is particularly surprising given that the post-COVID period should have led to an increased use of information and communication technology for public health issues.

Keywords

Public Health; Epidemiology Informatics; IMIA Yearbook 2023

Yearb Med Inform 2023;264-8

<http://dx.doi.org/10.1055/s-0043-1768754>

1 Introduction

Public Health and Epidemiology Informatics (PHEI) encompasses all approaches and methods based on the use of information and communication technology to investigate and/or propose solutions to population health issues. More specifically, the focus is on the application of informatics methods to the solution of public health problems, from research question formulation to data collection, analysis, and interpretation.

In this context, it seems useful to distinguish the PHEI section from other disciplinary sub-sections, such as those established by the International Medical Informatics Association (IMIA) working groups in general. At the same time, the panel considered the boundary of PHEI versus other sections (e.g., Pandemics) was sometimes difficult to define precisely. Consequently, in the PHEI section, we defined that the perspective or prism of studies should focus primarily on population-level health using large data sets. Another focus of interest for our section is about digital approaches to data collection, including the integration of heterogeneous health related data [12] whenever necessary, followed by analysis and interpretation of the generated results. Finally, an emphasis is placed in PHEI on the applied nature of these approaches, which we expect to be translatable into improving public health research or implementation. Applying statistical analysis and processing methods cannot be ruled out from this perspective.

Following the previous edition of the IMIA Yearbook focused on Inclusive Digital Health [8], the present synthesis in the field of PHEI examines the scientific

literature that emerged in 2022 in the field of medical informatics, with a focus on the subfield of public health and epidemiology. This analysis, unlike previous years, seeks to identify novel topics and trends in this field. In addition, it describes the meticulous procedure used to select the best articles published in 2022, which included section editors supervising the peer review of the most representative research papers chosen for their excellence and originality.

2 Methods

Similarly to the last 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 were used to retrieve relevant studies ranging from January 1, 2022 to December 31, 2022. The queries targeted public health (e.g., “Public health”, “Public health practice”, “Registries”, “Population surveillance”) or epidemiological journal articles (e.g., “Epidemiological Monitoring”, “Epidemiologic surveillance”, “Outbreak”), which included medical informatics topics (e.g., “Medical informatics”, “Expert systems”, “Neural networks”, “Social network”). Returned references addressing topics of the other sections of the Yearbook, e.g., those related to COVID-19 (including MeSH terms “COVID-19”, “Pandemics”, “SARS-CoV-2”) or sensors, covered in the Sensors, Signals, and Imaging Informatics section were excluded from our search.

3 Results and Discussion

3.1 Overall Observation

In terms of the article search and the results obtained, we observed a clear reduction in the number of papers published in 2022 compared with 2021. This may seem surprising for this PHEI theme, especially after the recovery period from the COVID-19 pandemic, when we might have expected a rise in the number of publications in the public health domain.

Of the 539 papers retrieved from PubMed, 10 articles were selected for a detailed evaluation. These 10 articles focused mainly on the following issues: (i) the use of social networks to study public health phenomena [1, 6], such as the impact of pollution on human health as addressed by Ji et al. [6], and parents' attitudes to vaccination as influenced by social networks [1]; (ii) the use of machine learning and deep learning, in particular for prediction strategies [4, 9, 10, 13, 14]; (iii) the coupling of a health registry and semantic approaches to identify trajectory patterns leading to certain pathologies; and finally (iv), the use of Natural Language Processing approaches over free text datasets to address public health issues [5, 11]. As in previous years, we noted a preponderance of machine learning approaches to address different PHEI questions.

3.2 Selected Papers

Among the 10 candidate papers included for this year's selection process, two best papers were selected, as depicted in Table 2. We considered the study by Valentin

and colleagues [11] to be perfectly in line with the special topic of this year's IMIA Yearbook, namely "One Digital Health". It describes a framework for the annotation of epidemiological information in animal disease-related news articles. The second best paper, authored by Ji et al. [6] focused on environmental health. The authors used Weibo data (the Chinese Twitter) to monitor the health of Chinese urbanities. The study was able to establish an association between air pollution and health sensing by using generalized additive models.

4 Conclusion

The 2022 selection process of PHEI confirmed the trend towards addressing public health and computational epidemiology issues with machine learning and deep learning techniques. In addition, although we noticed an overall decreasing trend in the number of papers published according to our usual keywords in the field of public health in comparison to the year 2021, we also found an apparent growing body of literature on environmental and exposure issues.

There is a need to keep a close eye on the trend over the next years to determine whether there is a genuine groundswell of work on the topic, or whether it is merely a matter of adapting the keywords that have been used up to now.

Acknowledgements

The PHEI section editors would like to thank the external reviewers of the section for their valuable contribution in evaluating and selecting the best papers for this

2023 IMIA Yearbook PHEI section: Marta Aavalos, David Buckeridge, Sébastien Cossin, Hossein Estiri, Aimé Patrice Koumamba, and Jean-Noël Nikiema. They also thank the two chief editors Fleur Mougín and Lina Soualmia for evaluating each of the 10 papers in the section.

References

1. Al-Regaiey KA, Alshamry WS, Alqarni RA, Albarrak MK, Alghoraiby RM, Alkadi DY, et al. Influence of social media on parents' attitudes towards vaccine administration. *Hum Vaccin Immunother* 2022 ;18(1):1872340. doi: 10.1080/21645515.2021.1872340.
2. Brakefield WS, Ammar N, Shaban-Nejad A. An Urban Population Health Observatory for Disease Causal Pathway Analysis and Decision Support: Underlying Explainable Artificial Intelligence Model. *JMIR Form Res* 2022 ;6(7):e36055. doi: 10.2196/36055.
3. Diallo G, Bordea G. Novelty in Public Health and Epidemiology Informatics. *Yearb Med Inform* 2022 ;31(1):273–5. doi: 10.1055/s-0042-1742526.
4. Fu L, Xia W, Shi W, Cao G-X, Ruan Y-T, Zhao X-Y, et al. Deep learning based cervical screening by the cross-modal integration of colposcopy, cytology, and HPV test. *Int J Med Inform* 2022;159:104675. doi: 10.1016/j.ijmed-inf.2021.104675.
5. Humbert-Droz M, Izadi Z, Schmajak G, Giranfrancesco M, Baker MC, Yazdany J, et al. Development of a natural language processing system for extracting rheumatoid arthritis outcomes from clinical notes using the national RISE registry. *Arthritis Care Res (Hoboken)* 2023;75(3):608–15. Epub 2022 Oct 31. doi: 10.1002/acr.24869.
6. Ji H, Wang J, Meng B, Cao Z, Yan T, Zhi G, et al. Research on adaption to air pollution in Chinese cities: Evidence from social media-based health sensing. *Environ Res* 2022 ;210:112762. doi: 10.1016/j.envres.2022.112762
7. Lamy J-B, Séroussi B, Griffon N, Kerdelhué G, Jaulent M-C, Bouaud J. Toward a formalization of the process to select IMIA Yearbook best papers. *Methods Inf Med* 2015 ;54(2):135–44. https://doi.org/10.3414/ME14-01-0031.
8. Mougín F, Hollis KF, Soualmia LF. Inclusive Digital Health. *Yearb Med Inform* 2022;31(1):2–6. https://doi.org/10.1055/s-0042-1742540.
9. Rieckmann A, Dworzynski P, Arras L, Lapuschkin S, Samek W, Aniweta Arah O, et al. Causes of Outcome Learning: a causal inference-inspired machine learning approach to disentangling common combinations of potential causes of a health outcome. *Int J Epidemiol* 2022;51(5):1622–36. doi: 10.1093/ije/dyac078.
10. Su P-Y, Wei Y-C, Luo H, Liu C-H, Huan W-Y, Chen K-F, et al. Machine Learning Models for Predicting Influential Factors of Early Outcomes in Acute

Table 2 Selection of best papers for the 2023 IMIA Yearbook of Medical Informatics for the Public Health and Epidemiology Informatics section. The articles are listed in alphabetical order by the first author's surname.

Section
Public Health and Epidemiology Informatics
<ul style="list-style-type: none"> ▪ Ji H, Wang J, Meng B, Cao Z, Yang T, Zhi G, Chen S, Wang S, Zhang J. Research on adaption to air pollution in Chinese cities: Evidence from social media-based health sensing. <i>Environ Res</i> 2022 Jul;210:112762. doi: 10.1016/j.envres.2022.112762. ▪ Valentin S, Arsevska E, Vilain A, De Waele V, Lancelot R, Roche M. Elaboration of a new framework for fine-grained epidemiological annotation. <i>Sci Data</i> 2022 Oct 26;9(1):655. doi: 10.1038/s41597-022-01743-2.

- Ischemic Stroke: Registry-Based Study. *JMIR Med Inform* 2022;10(3):e32508. doi: 10.2196/32508.
11. Valentin S, Arsevska E, Vilain A, De Waele V, Lancelot R, Roche M. Elaboration of a new framework for fine-grained epidemiological annotation. *Sci Data* 2022;9(1):655. doi: 10.1038/s41597-022-01743-2
 12. Wang S, Celebi ME, Zhang Y-D, Hu J, Dong S. Advances in Data Preprocessing for Biomedical Data Fusion: An Overview of the Methods, Challenges, and Prospects. *Information Fusion* 2021;76:376–421. doi: 10.1016/j.inffus.2021.07.001.
 13. Yadgir SR, Engstrom CJ, Jacobsohn GC, Green RK, Jones CMC, Cushman JT, et al. Machine learning-assisted screening for cognitive impairment in the emergency department. *J Am Geriatr Soc* 2022;70(3):831-7. doi: 10.1111/jgs.17491.
 14. Yu H, Huang T, Feng B, Lyu, J. Deep-learning model for predicting the survival of rectal adenocarcinoma patients based on a surveillance, epidemiology, and end results analysis. *BMC Cancer* 2022;22(1):210. doi: 10.1186/s12885-022-09217-9.

Correspondence to:

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

Appendix: Content Summary of Selected Best Papers for the IMIA Yearbook 2023, section Public Health and Epidemiology Informatics (PHEI)

Valentin S, Arsevska E, Vilain A, De Waele V, Lancelot R, Roche M

Elaboration of a new framework for fine-grained epidemiological annotation

Sci Data 2022 Oct 26;9(1):655. doi: 10.1038/s41597-022-01743-2

This paper introduces a novel annotation framework for epidemiological surveillance of newspaper articles by proposing a sentence-level annotation approach that deviates from the traditional document-level annotation. By categorizing sentences based on their epidemiological types, the framework enables precise identification and extraction of critical information, ultimately facilitating effective outbreak detection. The authors' focus on meticulous annotation guidelines has

significantly improved accuracy and agreement among annotators in four progressive steps. Rigorous evaluation of inter-annotator agreement at each stage further enhances the reliability and robustness of the final annotations. The resulting annotated corpus offers high-quality annotations and specializes specifically in animal health and online news articles, providing a superior alternative to existing benchmark corpora in this domain.

Ji H, Wang J, Meng B, Cao Z, Yang T, Zhi G, Chen S, Wang S, Zhang J

Research on adaption to air pollution in Chinese cities: Evidence from social media-based health sensing

Environ Res 2022 Jul;210:112762. doi: 10.1016/j.envres.2022.112762

Pollution has been associated with a variety of diseases, including respiratory diseases, cardiovascular diseases, and mental illness. Geographical and socioeconomic factors affect the air quality and health effects in cities differently. Using social media data, researchers can now examine the health impacts of air pollution. It has been demonstrated that user-generated data can

be used to track diseases. This is exactly what the study in this paper accomplished. The authors mine health sensing from social media data and examine its relationship to air pollution in 70 Chinese cities in order to fill the aforementioned research gaps. The objectives of this study are: 1) to quantitatively correlate health sensing with air pollution utilizing the generalized additive model, and 2) to examine air pollution sensitivities across 70 Chinese cities. Weibo, the Chinese equivalent of Twitter, is the source of the used social data. Weibo data can detect the real-time health of urban residents. In terms of methodology, a generalized additive model was used to quantify the impact of air pollution on public health sensing from Weibo data. Resulting from the integration of geographical conditions, China is divided into seven distinct regions. In these seven regions, the sensitivity to air pollution varied. Cities with good air quality in the south and east of China were more sensitive to air pollution than those with poor air quality in the northwest and north. The study provides new perspectives and methodologies for monitoring health and the health effects of air pollution.