

## MIT Open Access Articles

*epiDAMIK 6.0: The 6th International Workshop on  
Epidemiology meets Data Mining and Knowledge Discovery*

The MIT Faculty has made this article openly available. **Please share**  
how this access benefits you. Your story matters.

**Citation:** Adhikari, Bijaya, Rodriguez, Alexander, Yadav, Amulya, Pei, Sen, Srivastava, Ajitesh et al. 2023. "epiDAMIK 6.0: The 6th International Workshop on Epidemiology meets Data Mining and Knowledge Discovery."

**As Published:** <https://doi.org/10.1145/3580305.3599220>

**Publisher:** ACM|Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining

**Persistent URL:** <https://hdl.handle.net/1721.1/152097>

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of Use:** Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



# epiDAMIK 6.0: The 6th International Workshop on Epidemiology meets Data Mining and Knowledge Discovery

Bijaya Adhikari  
University of Iowa  
Iowa City, IA, USA

Alexander Rodríguez  
Georgia Institute of Technology  
Atlanta, GA, USA

Amulya Yadav  
Pennsylvania State University  
State College, PA, USA

Sen Pei  
Columbia University  
New York, NY, USA

Ajitesh Srivastava  
University of Southern California  
Los Angeles, CA, USA

Marie-Laure Charpignon  
Massachusetts Institute of Technology  
Cambridge, MA, USA

Anil Vullikanti  
University of Virginia  
Charlottesville, VA, USA

B. Aditya Prakash  
Georgia Institute of Technology  
Atlanta, GA, USA

## ABSTRACT

The epiDAMIK workshop serves as a platform for advancing the utilization of data-driven methods in the fields of epidemiology and public health research. These fields have seen relatively limited exploration of data-driven approaches compared to other disciplines. Therefore, our primary objective is to foster the growth and recognition of the emerging discipline of data-driven and computational epidemiology, providing a valuable avenue for sharing state-of-the-art research and ongoing projects. The workshop also seeks to showcase results that are not typically presented at major computing conferences, including valuable insights gained from practical experiences. Our target audience encompasses researchers in AI, machine learning, and data science from both academia and industry, who have a keen interest in applying their work to epidemiological and public health contexts. Additionally, we welcome practitioners from mathematical epidemiology and public health, as their expertise and contributions greatly enrich the discussions. Homepage: <https://epidamik.github.io/>.

## CCS CONCEPTS

• **Information systems** → **Data mining**; • **Computing methodologies** → **Machine learning**; • **Applied computing** → **Epidemiology**;

## KEYWORDS

Epidemiology, public health, forecasting, AI for good

### ACM Reference Format:

Bijaya Adhikari, Alexander Rodríguez, Amulya Yadav, Sen Pei, Ajitesh Srivastava, Marie-Laure Charpignon, Anil Vullikanti, and B. Aditya Prakash. 2023. epiDAMIK 6.0: The 6th International Workshop on Epidemiology meets Data Mining and Knowledge Discovery. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '23)*, August 6–10, 2023, Long Beach, CA, USA.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

KDD '23, August 6–10, 2023, Long Beach, CA, USA

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0103-0/23/08.

<https://doi.org/10.1145/3580305.3599220>

'23), August 6–10, 2023, Long Beach, CA, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3580305.3599220>

## 1 INTRODUCTION

The devastating effects of the COVID-19 pandemic, which seems to have been elevated to an endemic [1, 2], and those of the H1N1, Zika, SARS, MERS, and Ebola outbreaks over the past few decades has sharply illustrated our enormous vulnerability to emerging infectious diseases. While the data mining research community has demonstrated increased interest in epidemiological applications, much is still left to be desired. For example, there is an urgent need to develop sound theoretical principles and transformative computational approaches that will allow us to address the escalating threat of current and future pandemics. Data mining and knowledge discovery have an important role to play in this regard. Different aspects of infectious disease modeling, analysis, and control have traditionally been studied within the confines of individual disciplines, such as mathematical epidemiology and public health, and data mining and machine learning. Coupled with increasing data generation across multiple domains/sources (e.g., wastewater surveillance, electronic medical records, and social media), there is a clear need for analyzing them to inform public health policies and outcomes timely. Recent advances in disease surveillance and forecasting, and initiatives such as the CDC Flu Challenge, CDC COVID-19 Forecasting Hub etc., have brought these disciplines closer together. On the one hand, public health practitioners seek to use novel datasets, such as Safegraph, Unacast, and Google mobility data, and techniques like Graph Neural Networks. On the other hand, data mining and machine learning researchers develop novel tools for solving many fundamental problems in the public health policy planning and decision-making process, leveraging novel datasets (e.g., COVID-19 behavioral health surveys, contact tracing trees, and satellite images of urban streets) and combining them with more traditional time series information (e.g., surveillance, hospitalization, and death records). We believe the next stage of advances will result from closer collaborations between these two groups, which is the main objective of epiDAMIK.

## 2 WORKSHOP RELEVANCE

Despite the end of the emergency status of COVID-19 in April 2023, state and federal public health agencies are becoming more invested in data-driven approaches for infectious disease surveillance, forecasting, and control (e.g., the CDC's new Center for Forecasting and Outbreak Analytics), and funders are committing more resources to this important area (e.g., the NSF Incorporating Human Behavior in Epidemiological Models Program). Further, the researchers in the health and data science community who have been dedicated to the emergency response over the last three years, are finally able to dedicate more time to retrospective analyses and pandemic preparedness efforts in partnership with policy stakeholders. Therefore, we are well-positioned to further attract many researchers, including those who may not attend SIGKDD otherwise. Finally, we want to highlight a new pressing issue. The quality of COVID-19 case and death data has been deteriorating over the last year, and there are plans for stopping epidemiological data collection other than hospitalizations, like in the case of Influenza. After ensuring the availability and reliability of pandemic data for three years, Johns Hopkins University plans to cease the Coronavirus Resource Center's ongoing collection and reporting of COVID-19 data on March 10, 2023. Thus, the timing of our proposed workshop is ideal for innovations in data mining methods for situational awareness, outbreak detection, and forecasting using alternative data sources.

## 3 TOPICS OF INTEREST

Topics of interest of epiDAMIK include, but are not limited to:

- Epidemiologically-relevant data collection
- Advances in modeling, simulation, and calibration of disease spread models
- Syndromic surveillance using social media, search, and other data sources
- Challenges in model validation against ground truth
- Outbreak detection and inference
- Visualization of epidemiological data
- Planning for public health policy
- Role of open source data and community in epidemiology
- Data-driven advances in control and optimization (like immunization)
- Forecasting disease outcomes including projections
- Graph mining and network science approaches to epidemiology
- Crowdsourced methods for detection and forecasting
- Use of novel datasets for prediction and analysis (including EHR records)
- Data mining data for hospital-acquired infections like C.diff, MRSA, etc.
- Identifying health behaviors
- Handling missing and noisy data
- Disease forecasting challenge (like the CDC FluSight) experiences
- Interpretable and expert-driven AI for public health
- Experiences of real-time forecasting
- Infodemic, misinformation, and disinformation
- Fairness in resource allocation and surveillance

## 4 PARTICIPATION AND REVIEW PROCESS

Each submitted paper received at least two blind reviews. The final acceptance/rejection decision is made by program chairs based on the reviews each paper received. Final decision also includes the information on whether each paper is accepted as a oral or poster paper. The accepted papers must be formatted according to the ACM SIG Proceedings template. All accepted papers will be hosted at <sup>1</sup> and featured on the workshop's website. At least one author from each accepted paper must register, attend, and present their work at the workshop.

## 5 PROGRAM OUTLINE

epiDAMIK @ KDD 2023 is a full-day workshop. We are aiming to present a balanced program consisting of the following elements, allowing for ample opportunity for discussion and networking:

- Invited Talks (about 3, 45 minutes each, including questions)
- Panel (60 minutes).  
Topic: Enhancing Data Availability and Quality in the Post-Emergency Era of COVID-19: Unlocking New Opportunities for Data-Driven Approaches
- Oral Papers (15 minutes for presentation, 5 for discussion)
- Poster Papers (45 minute poster and networking session)

## 6 EXPECTED OUTCOME

The expected outcome of the workshop are new insights into how data mining can play a bigger role in epidemiology and public health research. We want to further raise attention to the emerging research area of data-driven and computational epidemiology, and create a venue for presenting state-of-the-art and in-progress results that would benefit from peer feedback.

## 7 ORGANIZING COMMITTEE

The workshop organizing committee composed of the following:

### Program Committee Chairs

- Bijaya Adhikari, UIowa
- Alexander Rodríguez, Georgia Tech
- Amulya Yadav, Penn State
- Sen Pei, Columbia
- Ajitesh Srivastava, USC
- Marie-Laure Charpignon, MIT

### Steering Committee

- B Aditya Prakash, Georgia Tech
- Anil Vullikanti, UVA

## REFERENCES

- [1] Paul Adepoju. 2022. Africa prepares for endemic COVID-19. *Nature Medicine* 28 (2022), 876–878.
- [2] Manish Sagar, Katherine Reifler, Michael Rossi, Nancy S Miller, Pranay Sinha, Laura F White, Joseph P Mizgerd, et al. 2021. Recent endemic coronavirus infection is associated with less-severe COVID-19. *The Journal of clinical investigation* 131, 1 (2021).

<sup>1</sup>Accepted papers: [openreview.net/group?id=KDD.org/2023/Workshop/epiDAMIK](https://openreview.net/group?id=KDD.org/2023/Workshop/epiDAMIK)