

20th International Workshop on Data Mining in Bioinformatics (BIOKDD 2021)

Da Yan
yanda@uab.edu
Department of Computer Science
The University of Alabama at
Birmingham
Birmingham, AL, USA

Steve Qin
zhaohui.qin@emory.edu
Department of Biostatistics and
Bioinformatics
Emory University
Atlanta, GA, USA

Debswapna Bhattacharya
bhattacharyad@auburn.edu
Department of Computer Science and
Software Engineering
Auburn University
Auburn, AL, USA

Jake Chen
jakechen@uab.edu
Informatics Institute
The University of Alabama at
Birmingham
Birmingham, AL, USA

Mohammed J. Zaki
zaki@cs.rpi.edu
Department of Computer Science
Rensselaer Polytechnic Institute
Troy, NY, USA

ABSTRACT

The goal of the 20th International Workshop on Data Mining in Bioinformatics (BIOKDD 2021) is to encourage KDD researchers to tackle the numerous problems and challenges in Bioinformatics using Data Mining technologies. Based on the organizers' expertise and the BIOKDD communities, BIOKDD 2021 features the theme of "Artificial Intelligence in Medicine". This topic focuses on the use of machine learning and data mining techniques for the analysis of large amounts of heterogeneous, complex, biological and medical data, with a particular focus on deep learning methods that have seen rapid advance and wider adoption in Bioinformatics (e.g., DeepVariant, AlphaFold 2). We also particularly welcome COVID-19 related research. The key goal is to accelerate the convergence between Data Mining and Bioinformatics communities to expedite discoveries in basic biology, medicine and healthcare.

CCS CONCEPTS

- **Applied computing** → **Life and medical sciences; Bioinformatics**;
- **Computing methodologies** → **Artificial intelligence**;
- **Information systems** → **Data mining**.

KEYWORDS

Bioinformatics, artificial intelligence, health informatics

ACM Reference Format:

Da Yan, Steve Qin, Debswapna Bhattacharya, Jake Chen, and Mohammed J. Zaki. 2021. 20th International Workshop on Data Mining in Bioinformatics (BIOKDD 2021). In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21), August 14–18, 2021, Virtual Event, Singapore*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3447548.3469442>

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 '21, August 14–18, 2021, Virtual Event, Singapore

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8332-5/21/08.

<https://doi.org/10.1145/3447548.3469442>



Figure 1: BIOKDD 2021 Workshop Logo

1 INTRODUCTION

Bioinformatics is the science of managing, mining, and interpreting information from biological data, and data mining plays an essential role in addressing the fundamental problems in Biomedical Informatics. The annual BIOKDD workshop is an established forum for researchers and practitioners of Biomedical Informatics and Data Mining to explore cutting-edge techniques and to exchange ideas and experiences. BIOKDD has been successfully held in conjunction with SIGKDD for 19 years. We are in our 20th anniversary this year! Figure 1 shows the logo of BIOKDD 2021.

The goal of the 20th International Workshop on Data Mining in Bioinformatics (BIOKDD 2021) is to encourage KDD researchers to work on the numerous problems and challenges in Bioinformatics using Data Mining technologies. Based on the organizers' expertise and the BIOKDD communities, BIOKDD 2021 features the theme of "Artificial Intelligence in Medicine". The emphasis is on the use of machine learning and data mining techniques for the analysis of large amounts of heterogeneous, complex, biological and medical data, with a particular focus on deep learning methods. We also welcome research related to COVID-19. The key goal is to accelerate the convergence between Data Mining and Bioinformatics communities to expedite discoveries in basic biology, medicine and healthcare.

2 TOPICS OF INTEREST

Besides workshop papers on original research contributions, we also solicit 1-page abstracts that introduce preliminary research outcomes to increase research impact and to nurture collaborations,

and late-breaking research that introduces published work with exceptionally high practical significance. We invite submissions on research contributions in (but not limited to) the areas listed below:

- Development of deep learning methods for biological and clinical data
- Novel methods and frameworks for mining and integrating big biological data
- Discovering biological networks and pathways underlying biological processes and diseases
- Analysis, discovery of biomarkers and mutations, and disease risk assessment
- Comparative genomics
- Metagenome analysis using sequencing data
- RNA-seq and microarray-based gene expression analysis
- Genome-wide analysis of non-coding RNAs
- Genome-wide regulatory motif discovery
- Structural bioinformatics
- Automated annotation of genes and proteins
- Discovery of structural variations from next-generation sequencing (NGS) data
- Correlating NGS with proteomics data analysis
- Discovery of genotype-phenotype associations
- Building predictive models for complex phenotypes
- Functional annotation of genes and proteins
- Cheminformatics
- Special biological data management techniques
- Privacy and security issues in mining genomic databases
- Predictive modeling for personalized treatment
- Semantic web and ontology-driven data integration methods
- Text mining for biomedical literature and clinical notes
- Information retrieval for healthcare and biomedical applications
- Biomedical signal analysis and processing
- Intelligent medical data management
- Collaboration technologies for biomedicine
- Social networks for biomedicine
- Bioimage analysis, single-cell analysis
- Biological network visualization
- Information visualization and visual analytics for biomedical data
- Research that battles COVID-19

3 WORKSHOP ORGANIZERS

• **Program Chairs.** **Da Yan** is an Assistant Professor at the Department of Computer Science of the University of Alabama at Birmingham (UAB). He was the sole winner of Hong Kong 2015 Young Scientist Award in Physical/Mathematical Science, and he is a senior member of ACM and IEEE. Dr. Yan's research expertise lies in developing scalable systems and algorithms for Big Data analytics. He frequently publishes in conferences such as SIGMOD, VLDB, ICDE, SIGKDD and he also regularly serves in the program committee of these conferences, and serves as reviewers of journals such as ACM TODS, VLDB Journal, IEEE TPDS, and IEEE TKDE. Dr. Yan has been organizing BOKDD workshops since 2018.

Steve Qin is a Professor of Biostatistics and Bioinformatics at Emory University. He received his Ph.D. degree in Statistics from the University of Michigan. He has extensive experience in statistical modeling and computing with applications to statistical genetics and genomics. His recent research is focused on developing Bayesian model-based methods to analyze data generated from applications of next-generation sequencing technologies, and on developing software so that the methods can be easily adopted by the research community.

Debswapna Bhattacharya is an Assistant Professor in the Department of Computer Science and Software Engineering at Auburn University. He received his Ph.D. in Computer Science from the University of Missouri-Columbia. His research focuses on computational biology, bioinformatics, machine learning, and data science. His research group has developed and freely disseminated novel computational and data-driven approaches for protein 3D modeling by combining cutting-edge machine learning and optimization algorithms. He is a recipient of NSF CAREER, an NIH Maximizing Investigators' Research Award (MIRA), and Auburn University College of Engineering's Ginn Faculty Achievement Fellowship.

• **General Chair.** **Jake Chen** is the Chief Bioinformatics Officer of Informatics Institute at the University of Alabama at Birmingham and a Professor of Genetics. Dr. Chen is President-elect of the Midsouth Computational Biology and Bioinformatics Society. He has over 25 years of R&D experience in biological data mining and systems biology with >180 peer-reviewed publications and >200 invited talks worldwide on bioinformatics methodologies and biomedical applications. He is an elected fellow of the American College of Medical Informatics (ACMI) and the American Institute of Medical and Biological Engineering (AIMBE).

• **Steering Committee.** **Mohammed J. Zaki** is a Professor of Computer Science at RPI. He is also the associate department head and the graduate program director for the CS department at RPI. He received his Ph.D. degree in computer science from the University of Rochester in 1998. His research interests focus on developing novel data mining and machine learning techniques, especially for applications in text mining, social networks, bioinformatics and personal health. He has over 250 publications (and 6 patents) and he is the founding co-chair for the BOKDD series of workshops. He was a recipient of the NSF CAREER Award and the DOE Early Career Principal Investigator Award, as well as HP Innovation Research Award, and Google Faculty Research Award. He is an ACM Distinguished Scientist and a Fellow of the IEEE.

• **Program Committee.** We are very grateful to each of our Program Committee members for their reviews and service. Each submission received at least 3–4 reviews.

The PC members of BOKDD 2021 were: Asa Ben-Hur, Rita Casadio, Zechen Chong, Ana Crisan, Alexandre P. Francisco, Mengting Gu, Ye Guan, Sumit Kumar Jha, Barbora Kozlíková, Stefan Kramer, Anshul Kundaje, Sangkeun Lee, Haichao Miao, Giri Narasimhan, William Stafford Noble, Tony Pan, Miguel Rocha, Saeed Salem, Yingfeng Wang, Liqiang Wang, Yanhong Wu, Zhaozheng Yin, Kevin Yip, Shibu Yooseph, Xiaoru Yuan, Wei Zhang, Shaojie Zhang, Zhongming Zhao, Jie Zheng, Fengfeng Zhou, Jaroslav Zola, Tianlin Zhang, Hao-Bo Guo, Mehran Ghafari, Jan Byška, Celine Sin.