Big Data Health Science Conference Participants Share Data Science Innovations

The University of South Carolina (USC) hosted the National Big Data Health Science Conference on February 13-14, 2025. The conference, which is supported in part by NIH, centers on the critical questions and emerging methods in big data health science which are contributing to the present and future of healthcare.

The purpose of the conference is to foster multidisciplinary collaboration and serve as a venue for the exchange of new concepts, methods, and findings. The conference aims to bring together researchers from academia, government, and industry to share theoretical, methodological, and substantive knowledge about big data in biomedical science and public health.

This is the sixth year that the USC's Big Data Health Science Center (BDHSC) has hosted the National Big Data Health Science Conference. The theme of the 2025 conference was "Unlocking the Power of Big Data in Health: Transforming Data into Actionable Intelligence."

Conference organizers reported a record number of more than 150 abstract submissions for the event. Over 300 participants joined the conference from 19 states across the United States and 11 foreign countries, including South Africa, Nigeria, and Australia.

Presenters discussed breakthroughs in big data science such as the utilization of big data and advanced data analytics to improve the precision diagnosis of infections and other clinical conditions (e.g., epilepsy), the use of geospatial data science for identifying and examining opioid overdose patterns in multimorbidity diseases like HIV and substance use disorder, the importance of balance between realism, security, and utility in synthetic data generation and application, and the ethical landscape to ensure fairness, transparency, accountability, privacy/data protection, human oversight on AI applied to healthcare.

Besides 6 plenary speakers and 80 poster presentations, the conference organized 7 breakout scientific sessions with more than 40 oral presentations around the themes of "Big Data Analytics and Emerging Methodology", "Big Data for Public Health and Community Health (I&II), "Big Data for Biomedical Research", "Big Data for Clinical Practice and Patient Care", "Big Data Health Science Research: Findings from Africa", and "Utilization of All of Us Research Platform for HIV Research". In addition, the conference featured a NIH Trainee Session, a Career Development Luncheon, and a formal Networking Reception. During the NIH trainee session, trainees from four NIH/NIAID funded training programs (three R25s, one T35) at USC presented their experiences in research training in the areas of Big Data and infectious diseases.

One mission of the conference is to provide a platform for students, researchers and healthcare professionals to interact and engage with the data science industries. This year, the conference featured a pre-conference SAS Viya training workshop provided by SAS, Inc and a panel presentation on the AI agents and agentic workflows by two IBM innovators/engineers. The IBM presentation described agents as an "evolution of generative AI" capable of planning, multi-step reasoning, action-taking, and self-correction. Agentic use cases for healthcare like mammography screening and medical coding were discussed. Techniques to address challenges of AI, like hallucinations, bias, regulatory compliance, resource consumption, and privacy were provided.

Industry advice on the adoption of AI agents, centered on impact on healthcare jobs and education, and the importance of continuous monitoring and evaluation.

The conference was preceded by a case competition featuring 39 teams totaling 115 students from 27 universities across the United States. This annual competition is designed to be a hands-on experience that tests undergraduate and graduate students' analytical, teamwork, communication, and presentation skills to further develop the talent pipeline in big data health science. Participating teams are given healthcare problems and datasets to solve a challenging problem using a data analytics approach. A panel of industry and academic experts judge the presentations based on teams' use of the full analytics process, from framing the problem to methodology selection, data use, model building, and innovation. This year, the teams were asked to identify the best performing antiretroviral treatment by patient cluster for people with HIV using synthetic data and advanced data analytics. The winning teams were recognized at the conference with cash prizes ($5000 first prize; $3000 second prize, $2000 third prize, and $500 for honorable mentions).

About the Big Data Health Science Center

The BDHSC, which hosts and coordinates the Conference, is an interdisciplinary Big Data health science research enterprise. Established in 2019, the center promotes the utilization of Big Data analytics in healthcare research, services improvement, and academic training.

"This recurring national conference reflects USC's leadership and commitment to understanding and utilizing the immense potential that lies within the vast datasets at our disposal. In collaboration with our communities, healthcare systems, and data science industry partners, we are poised to explore groundbreaking developments, foster collaboration, and pave the way for transformative advancements in Big Data health science research". Says the USC's Vice President for Research Julius Fridriksson.

Funding for this conference was made possible (in part) by NIH under the award R13LM014347 from the National Library of Medicine. The views expressed in written conference materials or publications and by speakers and moderators do not necessarily reflect the official policies of the Department of Health and Human Services; nor does mention by trade names, commercial practices, or organizations imply endorsement by the U.S. Government.