A Secure Implementation of Mendelian Randomization via Multi-Party Computation

Amirtharaj, Divya (School: Westview High School)

Finding correlation and causation between lifestyle habits and disease remains a major public health goal of our society. With the usage of genetic, lifestyle, and epidemiological data, we are able to use the concept of Mendelian Randomization (MR) to infer causality between human habits and potential disease. However, in order to use MR, a massive amount of private medical and genetic data is required. With data privacy and security being an utmost concern, hospitals, bio-banks, and patients are unwilling to share data for research purposes as they are vulnerable to attack or misuse. Because of this, a secure method of computation is necessary to protect patient privacy without sacrificing epidemiological advancement. I developed a web application utilizing secure multi-party computation to securely and anonymously execute Mendelian Randomization via the two stage least squares machine learning protocol. Data is encrypted using ciphers and asymmetric encryption to protect privacy, and then shared between nodes. Computation is performed on the encrypted data using the garbled circuit protocol-- coded using open source libsodium cryptographic libraries. The application was tested using data from the open source genome-wide association studies database. Mendelian Randomization results outputted by a plain-text R Studio model were compared to the secure multi-party computation environment to ensure accuracy, which was determined to be between 95% to 99%. Future steps include improving scalability through incorporation of beaver computation, and developing further computation capabilities.

Awards Won:
Fourth Award of $500
National Security Agency Research Directorate : Second Place Award "Science Security" of $1,000
King Abdulaziz &amp
his Companions Foundation for Giftedness and Creativity: $21000 Scholarship for Intelligent-Based Solutions in Cyber-security