Early Detection of Critical Salmonellosis Cases for Pre-Emptive Treatment via Novel Biomarkers Discovered Through Protein Domain Analysis and Machine Learning

Bhattacharyya, Aakash (School: Little Rock Central High School)

Salmonella is a common pathogen, infecting more than a hundred million people yearly. About 8% of these cases worldwide result in life-threatening conditions around 4-7 days after exposure. The case fatality ratios of these critical conditions can be as high as 30%; however, with pre-emptive treatment, it can be reduced to lower than 1%. Thus, rapid assessment of clinical case severity is essential for improving patient outcomes and optimizing healthcare resources. Advancements in genome sequencing technologies have enabled the analysis of bacterial genomes from many clinical cases, opening new opportunities for precise and timely diagnosis. This study proposes a genome-based framework for identifying critical Salmonella cases before the onset of critical symptoms and facilitating early medical intervention. By leveraging the novel approach of Protein Family (Pfam) Domains as the representation for genomic data, the complex genetic profiles of Salmonella cases are simplified into interpretable features. The severity levels of cases were investigated through rigorous data analysis, resulting in a set of 70 novel PFAM biomarkers for severity. Using the XGBoost algorithm, a machine learning model trained on these biomarkers achieved ~93% accuracy in predicting clinical severity. The results demonstrate the efficacy of the proposed approach and biomarkers. This framework holds strong potential for integration into clinical decision-making, paving the way for timely and personalized Salmonella infection management.

Awards Won: