INNOVATING CROSS-DOMAIN SOLUTIONS TO DETECT EMERGING BIOLOGICAL THREATS

Genai Aided Synthetic Twin Of A Population For Developing Accurate Low Swap-c Multilayered Anomaly Detection

Kunal Poorey Sandia National Labs **Uma Balakrishnan** Sandia National Labs **Md Kamruzzaman** Sandia National Labs
Jorge Salinas Sandia National Labs **Hemanth Kolla** Sandia National Labs **Kenneth Sale** Sandia National Labs

Today, advancements in artificial intelligence and machine learning are revolutionizing detection, diagnostics, and intervention strategies across all sectors of science and technology. In alignment with the Department of Defense's goal to enhance biological threat detection, we propose a cutting-edge computational strategy for the early detection of emerging biothreats such as infectious outbreaks (natural or intentional). Our approach utilizes advanced anomaly detection techniques applied to a diverse array of health datasets, including wearable technology, healthcare site data, Google search terms, and Twitter feeds. These sources are analyzed at multiple scales, from individual health to broader population health dynamics, to identify and notify anomalies at multiple resolutions.

For method development we usually need large well marked datasets and no single sourced dataset existed which was suitable for direct method development. Hence, we employed Generative Adversarial Networks (GANs) to create synthetic, yet real-world inspired, datasets complete with wearable data and supporting metadata. This innovative technique addresses critical gaps in existing public datasets during SARS-CoV2 pandemic timeline by generating a data rich dynamic synthetic twin of the population that maintains the statistical signature of real-world data.

Using both synthetic and real-world dataset, we apply a co-kurtosis-based projection to detect anomalies within multivariate time-series data effectively for different population resolution. This approach enhances our early detection capability in infection timelines in individuals and locate outbreak hotspots across populations. We also conduct a thorough evaluation of uncertainties in both synthetic and real-world data, optimizing anomaly detection in physiological metrics. Our results demonstrate significant improvements in the early detection of anomalies across multivariate datasets. These findings have broad implications for emerging bio-threat detection and can be adapted for other critical applications such as climate extremes, radiation anomaly detection, and other Chemical, Biological, Radiological, Cyber and Nuclear threats.

A particularly novel aspect of our work is the application of Generative AI to mimic dynamic systems over time, offering a robust and computationally efficient alternative to traditional Agent-Based Models (ABM) for scenario testing and forecasting. These advance paves the way for data-driven ABM, further advancing forecasting capability and better decision support system development. Furthermore, our approach supports edge computing and federated learning, ensuring privacy, and providing a resilient distributed computing framework. This holistic methodology not only bolsters our capacity to tackle current and emerging bio-threats but also prepares us for future challenges through innovative, low SWaP-C technologies and enhanced data-sharing networks for early warning systems.

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.