AI for infectious disease modelling and therapeutics

Gil Alterovitz
Brigham and Women's Hospital / Harvard Medical School, Boston, MA, 02115, USA.
National Artificial Intelligence Institute, Department of Veterans Affairs, Washington, DC, 20005, USA.
Email: ga@alum.mit.edu

Wei-Lun Alterovitz
Center for Biologics Evaluation and Research (CBER), U.S. Food and Drug Administration, Silver Spring, MD, 20993, USA
Argentys Informatics, Gaithersburg, MD, 20877, USA

Gail H. Cassell
Department of Global Health and Social Medicine, Harvard Medical School Professor and Chair, Emeritus, Department of Microbiology, University of Alabama at Birmingham
Scientific Affairs (ret), Eli Lilly and Company

Lixin Zhang
State Key Laboratory of Bioreactor Engineering East China University of Science and Technology Shanghai, 200237, China

A. Keith Dunker
Center for Computational Biology and Bioinformatics (Emeritus), Department of Biochemistry and Molecular Biology Indiana University School of Medicine, Indianapolis, IN, 46202, USA

AI for infectious disease modelling and therapeutics is an emerging area that leverages new computational approaches and data in this area. Genomics, proteomics, biomedical literature, social media, and other resources are proving to be critical tools to help understand and solve complicated issues ranging from understanding the process of infection, diagnosis and discovery of the precise molecular details, to developing possible interventions and safety profiling of possible treatments.

Keywords: Artificial intelligence, infectious diseases, COVID-19, 2019-nCoV, modelling, therapeutics

1. Background

Back in the 19th century, physicians and scientists used to think "bad air" was the source of infection and disease. This miasma theory was ultimately replaced by the germ theory with the advance of the microscope and the discovery of microorganisms. This switch dramatically changed our understanding of infectious disease and started the new era of public health. This year again, the outbreak of novel coronavirus 2019-nCoV has turned people's attention to the importance of surveillance, prevention, diagnosis and treatment of infectious disease.

Besides harmful viruses like the coronavirus (e.g. 2019-nCoV, SARS, MERS), HIV, Zika, Ebola virus, some bacteria (1%) cause diseases in people such as tuberculosis (*Mycobacterium tuberculosis*) and pertussis (*Bordetella pertussis*). While most antibiotic drugs were developed for bacteria-based infection, antibiotic resistance has become a growing challenge because of antibiotic misuse and poor stewardship. On the other hand, adopting new microbiome-based therapeutics is another potential risk delivering antimicrobial resistance genes to the human body and intestinal microtome via mobile genetic elements (or the other way around). For example, an important safety alert has been issued for use of a recent successful FDA-approved microbiome-based intervention, fecal microbiota transplantation (FMT) due to transmission of multi-drug resistant organisms. Disordered protein modelling is playing an important role in understanding microorganism structure and function as well.

2. *Introduction*

New computational approaches are leading to new possibilities for AI for infectious disease modelling and therapeutics, leveraging new resources such as the 2019-nCoV-released genomic sequences along with protein-protein interactions, among others. The large number of bacteria, viruses, fungi, and other microorganism genomes that are available along with clinical implications of observed mutations, make these particularly amenable to development of novel computational methods. By combining information at multiple scales, new insights have arisen via the tools, pipelines, and associated algorithm development as well. This session has a number of areas that integrated AI for infectious disease modelling and therapeutics in the proceedings:

3. *Social Media and COVID-19*

In "Characterization of Anonymous Physician Perspectives on COVID-19 Using Social Media Data" by K. J. Sullivan, et. al. [1] explored using Twitter to characterize different perspectives on COVID-19. Specifically, physician direct messages were compared to general public tweaks. The work analyzed over 513 million tweets in the process. Sentiment and n-gram analysis revealed patterns within physician vs public discourse regarding COVID-19[1].

4. *Biomedical literature and COVID-19 plus neglected tropical diseases*

Work by B. Dinakar, et. al. [2] explored the biomedical literature using an algorithm to find novel directions in disease research, with focus on COVID-19 and also neglected tropical diseases. The paper "Semantic Changepoint Detection for Finding Potentially Novel Research Publications"

analyzed publications over time to find patterns where there are significant changes in direction such that a semantic changepoint can be defined by the algorithm. The software is also released via link in the paper.

5. *Genomics and HCV*

S. Sledzieski, et. al. [3] describe a new method for reconstructing transmission phylogenies that increases accuracy while maintaining scalability. The paper "TreeFix-TP: Phylogenetic Error-Correction for Infectious Disease Transmission Network Inference" applied the method to HCV outbreaks. It also released the software via link in the paper.

6. *Protein intrinsically disordered regions and SARS-CoV-2*

A. Mudide, et. al. [4] analyzed SARS-CoV-2 for certain protein regions, known intrinsically disordered regions, that have additional flexibility, that may be targets for drug candidates. The paper, "SARS-CoV-2 Drug Discovery Based On Intrinsically Disordered Regions," also leveraged different docking approaches to model this flexibility and prioritize potential drug candidates, analyzing over 290 thousand compounds.

G. Goh, et. al. [5] analyzed the shell disorder of SARS-CoV-2 and other viruses to establish shell disorder as a proxy for vaccine development feasibility. It characterized SARS-CoV-2 as having an exceptionally hard outer shell, suggesting that vaccine development for SARS-CoV-2 is likely feasible and may be easier than for several other viruses such as HIV, HSV and HCV. The work in the paper "Feasibility study of vaccine development for SARS-CoV-2 and other viruses using shell disorder analysis" also presented several ideas on how shell disorder can be leveraged to characterize virulence, immune system evasion, and potential animal hosts.

7. *Protein-protein interactions and SARS-CoV-2*

M. Kshirsagar, et. al. [6] examined protein-protein interactions between viruses and their host. By comparing SARS-CoV-2 and host interactions with other virus-host interactions a number of motifs and themes emerged. The paper, entitled, "Functional comparison of virus-host pathogen communication using sequence-feature based SARS-CoV-2 protein interaction prediction" also specifically created SARS-CoV-2-human protein-protein interaction predictor based on sequence information and validated it with an independent dataset.

References

1. Sullivan, Katherine J., et al, "Characterization of Anonymous Physician Perspectives on COVID-19 Using Social Media Data", *Pac Symp Biocomput, 2020.*
2. Dinakar, Bhavish, et al. "Semantic Changepoint Detection for Finding Potentially Novel Research Publications", *Pac Symp Biocomput, 2020.*
3. Sledzieski, Samuel, et al. "TreeFix-TP: Phylogenetic Error-Correction for Infectious Disease Transmission Network Inference", *Pac Symp Biocomput, 2020.*
4. Mudide, Anish. et al. "SARS-CoV-2 Drug Discovery Based On Intrinsically Disordered Regions", *Pac Symp Biocomput, 2020.*
5. Goh, Gerard Kian-Meng, et al. "Feasibility study of vaccine development for SARS-CoV-2 and other viruses using shell disorder analysis", *Pac Symp Biocomput, 2020.*
6. Kshirsagar, Meghana, et al. "Functional comparison of virus-host pathogen communication using sequence-feature based SARS-CoV-2 protein interaction prediction", *Pac Symp Biocomput, 2020.*