

## BIOINFORMATICS SEMINAR

### MAHDI BELCAID

*Assistant Research Professor, Information & Computer Sciences*  
UNIVERSITY OF HAWAII AT MANOA

#### VIRAL METAGENOMICS AND DEEP LEARNING: THE PERFECT PAIRING

In viral metagenome assemblies, fragmentation and sequences with unknown functions can substantially hamper biological analysis and knowledge inference. For example, the vast majority of viral metagenome assemblies consist of relatively short unannotated contigs that are typically discarded from downstream analysis. Although various statistical and algorithmic approaches have been used to minimize fragmentation and derive contigs' functional annotations, deep learning-based methods, which have shown unprecedented results in academic and industrial applications, are still underexplored as potential solutions to these problems.

This talk will present two novel bioinformatics applications that exploit novel and divergent deep learning model architectures with viral metagenomics data. In the first part, we describe how deep contrastive learning was successfully used to mitigate fragmentation. This approach, which was implemented in a tool called CoCoNet, produces more complete or near-complete genomes in real and simulated data than other approaches. In the second part, we address ongoing research with collaborators at the University of Delaware to leverage natural language models, specifically transformers, as a meaningful encoding for protein sequences. These encodings not only represent a more efficient way to numerically represent protein sequences but can also encode functional modules in a protein leading to potentially more accurate sequence annotations.

Our results support the view that deep learning techniques are capable of overcoming the complexity of working with viral metagenomics data and facilitating accurate assessments of the diversity and role of viruses.

#### BIOGRAPHY

Dr. Belcaid earned his M.Sc. in Computer Science from the University of Quebec at Montreal and his Ph.D. in Computer Science from the University of Hawaii at Manoa. As an Assistant Research Professor at UH Manoa, he holds appointments in Information & Computer Sciences and the Hawaii Institute of Marine Biology. Dr. Belcaid previously served as Executive Director of the Hawaii Data Science Institute and as Director of the Bioinformatics Core.

Dr. Belcaid's work focuses primarily on modeling biological big data, with an emphasis on machine learning and probabilistic algorithms. In addition to bioinformatics, Dr. Belcaid is interested in data science education and integrating AI models to facilitate collaborative research and decision making.



**CBCB SEMINAR**

**5/9/2022**

**3:30-4:30PM**

**AP BioPharma**

**Room 140**

**(590 Avenue 1743)**

or via ZOOM:

[https://udel.zoom.us/j/  
93068494454](https://udel.zoom.us/j/93068494454)

**(Passcode: BINF)**

[bioinformatics.udel.edu](http://bioinformatics.udel.edu)

