



COLLEGE of ENGINEERING AND PHYSICAL SCIENCES

SCHOOL OF COMPUTER SCIENCE

PhD Seminar 1

Friday June 28, 2019 at 9AM in Reynolds, Room 2224

Classification of Pathogenicity in Genomic Sequence of Viruses

Akshay Chadha

Advisor: Dr. Rozita Dara

Co-Advisor: Dr. Zvonimir Poljak [Population Medicine]

Advisory Committee: Dr. David Pearl [Population Medicine]

Advisory Committee: Dr. Daniel Gillis

ABSTRACT:

Viral Avian Influenza (AI) outbreaks often make top place on the news articles around the globe. Incursion of a novel AI strain into susceptible poultry population could result into varying degree of clinical signs and mortality rates. AI viruses of H5 and H7 subtypes have been classified into Low Pathogenic (LP) and Highly Pathogenic (HP) categories. Due to its significance to poultry, human health and economy at large in Canada, both H5 and H7 subtypes were classified as a reportable disease to the Canadian Food Inspection Agency. Various genomic sequence collection initiatives over the last decade kick-started the usage of modern state of the art machine learning and deep learning algorithms to discover underlying relationships present in genomic sequences.

Current literature largely focuses on use of aggregating methods, classification of host species distribution and classification of antigenic subtypes based on genomic sequences. Convolutional Neural Network (CNN) is a deep learning architecture popularly used to identify hidden discriminative patterns present in genomic sequences. Our research firstly aims to classify pathogenicity of genomic strains and subsequently identify key positions of genomic sequences which are supposed to play a role in pathogenicity classification. As a preliminary study, we collected nucleotide sequences of Haemagglutinin (HA) gene for 1147 LP and 990 HP H5 strains from genomic data banks and applied CNN for classification of pathogenicity of H5 strains. We subsequently used CNN to successfully identify the position of different motifs including the one present at cleavage site of HA gene which is known to play a role in pathogenicity determination of H5 strains.