Postgraduate Seminar Series

Venue: Graduate Seminar Room
Date & Time: March 07, 2026 at 2:15 PM

Speaker Information

Md. Shariful Islam Bhuyan (Std No. 0418054001) is a part time Ph.D. student in the department of CSE, BUET. He completed his B.Sc. Engg. in CSE in 2006 and M.Sc. Engg. in CSE in 2009, both from Bangladesh University of Engineering and Technology (BUET). His research interest lies in the field(s) of Bioinformatics, Machine Learning, and Deep Learning. He is currently doing his postgraduate thesis under the supervision of Dr. Mohammad Saifur Rahman. He will be speaking about his ongoing research in this talk.

SICaRiO: A Gradient Boosting Framework for Reliable Indel Detection Using Pipeline-Agnostic Genomic Features

Despite impressive improvement in the next-generation sequencing technology, reliable detection of indels is still a difficult endeavor. Recognition of true indels is of prime importance in many applications, such as personalized health care, disease genomics and population genetics. Recently, advanced machine learning techniques have been successfully applied to classification problems with large-scale data. In this talk, we present SICaRiO, a gradient boosting classifier for the reliable detection of true indels, trained with the gold-standard dataset from "Genome in a Bottle" (GIAB) consortium. Our filtering scheme significantly improves the performance of each variant calling pipeline used in GIAB and beyond. SICaRiO uses genomic features that can be computed from publicly available resources, i.e., it does not require sequencing pipeline-specific information (e.g., read depth). This study also sheds lights on prior genomic contexts responsible for the erroneous calling of indels made by sequencing pipelines. We have compared prediction difficulty for three categories of indels over different sequencing pipelines. We have also ranked genomic features according to their predictivity in determining false positives.