# Postgraduate Seminar Series

*Venue: Graduate Seminar Room*
*Date & Time: June 22, 2024 at 2:30 PM*

## Speaker Information

Tasniah Mohiuddin (Std No. 0419052101) is a part time M.Sc. student in the department of CSE, BUET. She completed her undergraduate studies from Military Institute of Science and Technology (MIST) in 2019. Her research interest lies in the field(s) of Bioinformatics, Machine Learning, and Deep Learning. She is currently doing her postgraduate thesis under the supervision of Dr. Md. Shamsuzzoha Bayzid. She will be speaking about her ongoing research in this talk.

## Correcting single cell RNA-seq data for batch effects using a deep autoencoder with adversarial learning

Recent technological advances in high-throughput single-cell profiling have facilitated the compilation of large-scale single-cell datasets, enhancing our understanding of cell types across various biological and clinical conditions. However, integrating and analyzing these datasets is challenging due to batch-specific systematic differences, known as batch effects. As both biological and batch-specific differences coexist together, correcting batch effects in scRNA-seq datasets is essential for successful data integration. But the major challenge is correcting batch effects while accurately detecting the number of cell types, which is crucial in human studies. Most scRNA-seq algorithms are designed to first remove batch effects and then perform clustering, a process that may overlook some rare cell types. This study introduces BEENE-Cor, a semi-supervised data-integration framework that utilizes a deep autoencoder model guided by a cell-type classifier network and adversarial learning. BEENE-Cor effectively integrates multiple batches while preserving biological variability. Comprehensive evaluations on seven datasets demonstrate that BEENE-Cor successfully removes complex batch effects, improves cell type clustering, accurately recovers cell types, and outperforms popular methods such as ComBat-seq, Harmony, Seurat 3, and AD-AE. Notably, BEENE-Cor retains subtle cell type information in raw data, enabling the discovery of new cell types that are difficult to identify through individual batch analysis.