ENBIS-25 Conference

European Network for Business and Industrial Statistics

Contribution ID: **74**

Type: **not specified**

Comparative Assessment of Feature Selection Techniques for Single-Cell RNA-Sequencing in Multi-Cancer Data

Single-cell RNA sequencing (scRNA-seq) enables detailed exploration of cellular heterogeneity,Yet its high dimensionality requires efficient feature selection for robust downstream analysis.This study evaluates five feature selection methods—Triku, Scanpy, Seurat, Variance Threshold, and Pearson Residual—on a multi-cancer scRNA-seq dataset comprising 801 cells from breast, colon, kidney, lung, and prostate cancers. Following standardized preprocessing, including quality control, normalization, and PCA-based dimensionality reduction, each method was assessed for clustering performance using Normalized Mutual Information (NMI), Adjusted Rand Index (ARI), and Silhouette Score, alongside Gene Ontology (GO) enrichment for biological relevance and ribosomal gene fraction for technical robustness.

Triku consistently outperformed other methods, achieving the highest Silhouette Score (0.6501), robust NMI (0.9781), and ARI (0.9734), while identifying biologically meaningful genes with low ribosomal contamination (0.006).Although Pearson Residual and Variance Threshold provided adequate clustering results, in terms of accuracy, robustness, and biological relevance, Triku was still the best method. Seurat and Scanpy showed moderate performance across metrics.

These results highlight the importance of selecting appropriate feature selection strategies in scRNA-seq workflows and identify Triku as a robust and biologically meaningful approach for analyzing heterogeneous cancer datasets. This work contributes to optimizing single-cell RNA sequencing pipelines for precision oncology research.

Special/ Invited session

Classification

Mainly application

Keywords

Triku, Normalized Mutual Information, Adjusted Rand Index, Silhouette Score ,Clustering , precision oncology

Primary author: Ms WIJEKOON, W.A.D.M. (Rajarata University of Sri Lanka)

Co-author: Mr DISSANAYAKE, R.B.N. (Rajarata University of Sri Lanka)

Presenter: Ms WIJEKOON, W.A.D.M. (Rajarata University of Sri Lanka)

Track Classification: Statistics in Pharma and Health Care