

Cancer Subtype Discovery using Unsupervised Learning on Gene Expression Data

1. Introduction

This project explores cancer subtype discovery through unsupervised learning applied to gene expression data...

2. Data Loading and Preprocessing

We used a matrix of 1000 genes across 50 samples...

3. Dimensionality Reduction

Three methods were applied: PCA, t-SNE, and UMAP...

4. Clustering Analysis

Three clustering methods were used: K-Means, Hierarchical, and DBSCAN...

5. Metadata Comparison

We simulated subtype labels and evaluated clustering using ARI...

6. Conclusion

This project demonstrates how unsupervised learning can identify biologically meaningful subtypes...