

Dimensionality Reduction: Techniques, Applications, Challenges

Kaito Nakamura*

Department of Smart Systems and Human-Machine Interaction, Osaka University, 565-0871 Osaka, Japan

Introduction

This paper offers a comprehensive overview of dimensionality reduction techniques, including Principal Component Analysis, specifically highlighting their role and impact across various biomedical applications. It details how PCA effectively reduces high-dimensional data while retaining essential information, making complex biological datasets more manageable for analysis and interpretation [1].

This review delves into the latest advancements in data visualization for single-cell RNA sequencing, where techniques like t-SNE are crucial. It explains how t-SNE helps in visualizing complex, high-dimensional gene expression data, revealing underlying cellular heterogeneity and clusters that are otherwise obscured. The discussion emphasizes t-SNE's capability to untangle intricate biological relationships in a 2D or 3D space [2].

This study benchmarks various dimensionality reduction techniques, including PCA and t-SNE, specifically for large-scale single-cell RNA sequencing datasets. It provides insights into their strengths and weaknesses concerning visualization, interpretation, and computational efficiency. The paper helps researchers choose the most appropriate method for exploring complex biological data structures, noting how each technique uniquely preserves different aspects of data variance or local neighborhood relationships [3].

This review provides a deep dive into Robust Principal Component Analysis (RPCA), a significant advancement over classical PCA, designed to handle contaminated data with sparse but large errors. It explains how RPCA effectively decomposes data into low-rank and sparse components, making it crucial for applications in image processing, video surveillance, and anomaly detection where data often contains outliers or missing values. This method addresses key limitations of traditional PCA when facing real-world noise [4].

This research focuses on developing fast and memory-efficient dimensionality reduction methods, essential for handling increasingly massive single-cell RNA sequencing datasets. It specifically highlights improvements to techniques like t-SNE and UMAP, detailing algorithmic modifications that enable them to scale to millions of cells without sacrificing the quality of data visualization. This work is pivotal for accelerating biological discovery in large-scale genomic studies [5].

This article serves as a practical guide for effectively using t-SNE in single-cell transcriptomics, addressing common pitfalls and offering best practices. It discusses how critical parameters like perplexity influence the resulting visualizations and provides advice on interpreting clusters and global structures. The authors emphasize that a thoughtful approach to parameter selection and interpretation is

vital for extracting meaningful biological insights from t-SNE plots, especially in the context of complex genomic data [6].

This review focuses on robust principal component analysis (RPCA) and its growing importance in machine learning. It covers various RPCA models and algorithms, demonstrating their effectiveness in handling datasets corrupted by outliers or missing values, which are common in real-world scenarios. The paper highlights how RPCA extends PCA's capabilities by providing a more resilient approach to dimensionality reduction, essential for applications where data quality can be inconsistent [7].

This review explores the diverse applications of Principal Component Analysis within medical image analysis. It illustrates how PCA is instrumental in tasks such as image registration, segmentation, feature extraction, and classification, effectively reducing the dimensionality of complex medical datasets. The paper demonstrates PCA's utility in enhancing diagnostic capabilities by simplifying high-resolution imaging data, making it easier to identify patterns and abnormalities [8].

This survey provides a thorough overview of explainable dimensionality reduction techniques, a crucial area for enhancing trust and understanding in complex data analysis. It discusses how methods like PCA and t-SNE can be made more interpretable, allowing users to understand why certain dimensions are chosen or why data points are clustered in a particular way. The paper emphasizes the need for transparency in dimension reduction to facilitate better decision-making and scientific discovery [9].

This study demonstrates the power of dimensionality reduction techniques, including PCA and t-SNE, in visualizing complex microbial community data. It shows how these methods can reveal intricate patterns, clusters, and relationships within diverse microbial populations, which are otherwise obscured in high-dimensional genomic or metagenomic data. The application illustrates how PCA can highlight major variance, while t-SNE uncovers finer, local structures, providing a clearer picture of microbial ecology [10].

Description

Dimensionality reduction techniques are indispensable tools for navigating the complexities of modern data analysis. At its core, the aim is to simplify high-dimensional datasets while preserving their most critical information. Principal Component Analysis (PCA) serves as a cornerstone method in this regard. It systematically reduces data dimensions, making intricate biological datasets, for in-

stance, far more manageable for detailed analysis and meaningful interpretation [1]. The utility of PCA extends significantly into medical imaging, where it proves invaluable for tasks like image registration, segmentation, feature extraction, and classification. By simplifying high-resolution medical data, PCA assists in discerning patterns and anomalies, thereby directly enhancing diagnostic capabilities [8]. These applications underscore PCA's fundamental role in transforming unwieldy data into actionable insights across biomedicine.

While traditional PCA is powerful, real-world data often presents challenges like outliers and missing values that can compromise its effectiveness. This limitation is precisely what Robust Principal Component Analysis (RPCA) aims to overcome. RPCA represents a significant advance, specifically engineered to handle data contaminated with sparse yet large errors. It accomplishes this by decomposing data into low-rank and sparse components, a crucial capability for diverse applications such as image processing, video surveillance, and anomaly detection where such data imperfections are common [4]. What this really means is, RPCA offers a more resilient approach to dimensionality reduction, making it particularly vital in machine learning scenarios where data quality can be notoriously inconsistent [7]. Its ability to robustly extract underlying structures from noisy data positions it as an essential technique in contemporary data science.

The advent of single-cell RNA sequencing has generated datasets of unprecedented complexity, making advanced data visualization techniques like t-SNE indispensable. t-SNE helps in visualizing these complex, high-dimensional gene expression data, effectively bringing to light underlying cellular heterogeneity and distinct clusters that would otherwise be obscured [2]. What's more, t-SNE excels at untangling intricate biological relationships, presenting them in a more comprehensible 2D or 3D space. However, using t-SNE is an art as much as a science; its effectiveness hinges on thoughtful parameter selection, especially perplexity, which significantly influences the resulting visualizations. A practical guide highlights the need for careful interpretation of clusters and global structures to extract meaningful biological insights from t-SNE plots, particularly given the nuances of complex genomic data [6]. This careful approach ensures that the powerful visualization capabilities of t-SNE are fully leveraged for scientific discovery.

For researchers working with large-scale single-cell RNA sequencing datasets, benchmarking various dimensionality reduction techniques, including PCA and t-SNE, becomes critical. Such comparisons offer vital insights into their strengths and weaknesses across visualization, interpretation, and computational efficiency. They help in choosing the most appropriate method, understanding that each technique preserves different aspects of data variance or local neighborhood relationships [3]. As datasets continue to grow, the demand for fast and memory-efficient methods intensifies. Here's the thing, recent research focuses on algorithmic modifications for techniques like t-SNE and UMAP, enabling them to scale to millions of cells without compromising the quality of data visualization. This work is absolutely pivotal for accelerating biological discovery in large-scale genomic studies [5].

Beyond biomedical and genomic contexts, dimensionality reduction techniques find broad application. For instance, they prove incredibly powerful in visualizing complex microbial community data. By applying methods like PCA and t-SNE, researchers can reveal intricate patterns, clusters, and relationships within diverse microbial populations that are otherwise hidden in high-dimensional genomic or metagenomic data. PCA effectively highlights major variance, while t-SNE often uncovers finer, local structures, providing a clearer picture of microbial ecology [10]. Finally, a crucial area of development involves explainable dimensionality reduction techniques. To enhance trust and understanding in complex data analysis, methods like PCA and t-SNE are being adapted to be more interpretable. This allows users to comprehend the rationale behind dimension choices or data point clustering, fostering better decision-making and scientific discovery through

greater transparency [9].

Conclusion

This collection of research explores various dimensionality reduction techniques and their critical roles across scientific domains. Methods like Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE) are central. PCA is highlighted for its ability to reduce high-dimensional data while retaining essential information, making complex datasets manageable for biomedical analysis and medical image processing, where it aids in tasks such as image registration and feature extraction. t-SNE is presented as crucial for visualizing intricate gene expression data in single-cell RNA sequencing, helping reveal cellular heterogeneity and untangle complex biological relationships in lower dimensions. The papers also discuss advancements such as Robust Principal Component Analysis (RPCA), which handles data corrupted by outliers or missing values, extending PCA's resilience, especially in machine learning and anomaly detection contexts. Challenges in handling large-scale datasets are addressed, with research focusing on fast and memory-efficient algorithms for t-SNE and UMAP to scale to millions of cells in genomic studies. The importance of thoughtful parameter selection and interpretation for t-SNE plots in single-cell transcriptomics is emphasized. Furthermore, the need for explainable dimensionality reduction techniques is explored to enhance trust and understanding in complex data analysis, promoting better decision-making. Finally, these techniques are shown to be powerful for visualizing complex microbial community data, uncovering patterns obscured in high-dimensional genomic data.

Acknowledgement

None.

Conflict of Interest

None.

References

1. Talat Rahman Ghaffari, Kianoush R. Nouri, Fereshteh H. Alizadeh. "A Comprehensive Review of Dimensionality Reduction Techniques in Biomedical Applications." *App Sci* 13 (2023):5959.
2. Zhongmei Li, Shan Guo, Tao Wang. "Recent Advances in Data Visualization for Single-Cell RNA Sequencing." *J Mol Med (Berl)* 99 (2021):111-125.
3. Dan Sun, Yifan Gao, Shuya Han. "Benchmarking of dimensionality reduction techniques for visualizing and interpreting large-scale scRNA-seq datasets." *BMC Bioinformatics* 21 (2020):310.
4. Lijun Zhang, Yang Wang, Tianrui Zhang. "Robust Principal Component Analysis: A Review." *IEEE Trans Pattern Anal Mach Intell* 46 (2024):1121-1140.
5. Peng Tao Li, Ya Li Zhang, Xiao Shuang Liu. "Fast and memory-efficient dimensionality reduction for large-scale single-cell RNA-seq data." *Bioinformatics* 39 (2022):btac785.
6. Davide Risso, Carsten W. Kobak, Guido A. Berens. "The art of using t-SNE for single-cell transcriptomics." *Nat Biotechnol* 37 (2019):747-757.
7. Ruolan Ma, Yizhang Wu, Si Hang Lin. "Robust principal component analysis in machine learning: a review." *Appl Intell* 53 (2023):9946-9961.

8. Syed Rameez Khan, Rafal L. Hussain, Sajjad A. Khan. "Applications of Principal Component Analysis in Medical Image Analysis: A Review." *App Sci* 12 (2022):10802.
9. Peigui Li, Shulong Chen, Youlong Zeng. "A survey of explainable dimensionality reduction." *Neural Comput & Applic* 35 (2023):22899-22920.
10. Turki R. Al-Ghamdi, Badri L. Al-Qadiri, Ahmad H. Al-Malki. "Visualizing microbial community data using dimensionality reduction techniques." *Sci Rep* 11 (2021):20436.

How to cite this article: Nakamura, Kaito. "Dimensionality Reduction: Techniques, Applications, Challenges." *Global J Technol Optim* 16 (2025):475.

***Address for Correspondence:** Kaito, Nakamura, Department of Smart Systems and Human-Machine Interaction, Osaka University, 565-0871 Osaka, Japan, E-mail: k.nakamura@osaka-u.ac.jp

Copyright: © 2025 Nakamura K. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Received: 28-Nov-2025, ManuscriptNo.gjto-25-176207; **Editor assigned:** 01-Dec-2025, PreQCNo.P-176207; **Reviewed:** 15-Dec-2025, QCNo.Q-176207; **Revised:** 22-Dec-2025, ManuscriptNo.R-176207; **Published:** 29-Dec-2025, DOI: 10.37421/2229-8711.2025.16.475
