

**NON-LINEAR HYBRID RANDOM PROJECTION
TECHNIQUES FOR HIGH-DIMENSIONAL DATA**

BY

ADENIRAN USMAN

A Thesis Presented to the
**COLLEGE OF GRADUATE AND INTERDISCIPLINARY STUDIES
KING FAHD UNIVERSITY OF PETROLEUM & MINERALS**

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

APPLIED STATISTICS

MAY 2025

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

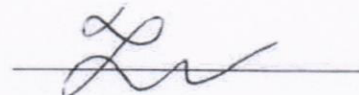
COLLEGE OF GRADUATE AND INTERDISCIPLINARY STUDIES

This thesis, written by **ADENIRAN USMAN** under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree **MASTER OF SCIENCE IN APPLIED STATISTICS**.

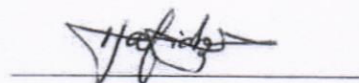
Thesis Committee



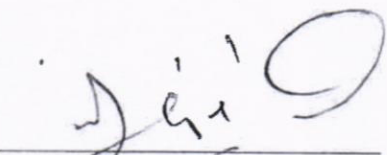
Dr. Ridwan A. Sanusi (Adviser)



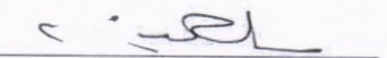
Dr. Zaid Sawlan (Member)



Dr. Mohammad Hafidz Omar (Member)



Prof. Monther Rashed Alfuraidan
Department Chairman



Prof. Suliman S. Al-Homidan
Dean, College of Graduate and Interdisciplinary
Studies

Date



©Usman Adeniran
2025

Dedication

I dedicate this thesis to Almighty Allah, whose blessings and guidance have always been with me, and to my family, whose unwavering support and encouragement have been my strength throughout this journey.

ACKNOWLEDGMENTS

I would like to extend my deepest gratitude to the Department of Mathematics, KFUPM, for granting me the invaluable opportunity to pursue my studies through their scholarship support. This generosity has been instrumental in shaping my academic journey, and I am profoundly appreciative of the resources and encouragement provided. A heartfelt thank you to my former supervisor, Dr. Jimoh Ajadi, for his unwavering guidance and mentorship. His insights, patience, and encouragement laid a strong foundation for my research. In addition, I am sincerely grateful to my current supervisor, Dr. Ridwan Sanusi, whose invaluable advice and support have been pivotal in the completion of this thesis. His expertise and commitment have continually inspired me to refine my work. I also would like to express my appreciation to the members of my thesis committee for their constructive feedback and invaluable contributions. Their thoughtful suggestions and rigorous academic discussions have enriched my research and expanded my perspective. Finally, to my family and friends, who have been my pillars of strength, offering unwavering support, encouragement, and

motivation, I am eternally grateful. Your belief in me has made this journey possible, and I dedicate this accomplishment to you all.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	v
LIST OF TABLES	x
LIST OF FIGURES	xii
ABSTRACT (ENGLISH)	xiii
ABSTRACT (ARABIC)	xv
CHAPTER 1 INTRODUCTION	1
1.1 Random Projection (RP)	2
1.2 Hybrid Random Projection (HRP)	3
1.3 Machine Learning	4
1.4 Aims and Objectives	5
1.5 Thesis Layout	7
CHAPTER 2 HYBRID DIMENSIONALITY REDUCTION TECHNIQUES BASED ON RANDOM PROJECTIONS: A LITERATURE REVIEW	8
2.1 Overview of DR Techniques	10
2.1.1 Linear Dimensionality Reduction Methods	11
2.1.2 Nonlinear DR Methods	14
2.2 Random Projection and Its Hybrid Variants	16
2.2.1 Integration of RP with Graph Embeddings	21

2.2.2	Integration of RP with Machine Learning Models	26
2.2.3	Integration of RP with Non-linear Techniques	29
2.2.4	Combined variants of RP (Type 1 and Type 2)	32
CHAPTER 3 NON-LINEAR HYBRID RANDOM PROJECTIONS		35
3.1	Objectives	35
3.2	Introduction	36
3.3	Overview of RPs	39
3.3.1	Dimensionality and Error Tolerance	40
3.3.2	Plus-Minus One Random Projection (PMRP)	42
3.3.3	Normal Random Projection	43
3.4	Formulation of the NHRP Method	44
3.4.1	Determining the Optimal Value of α	45
3.4.2	Supporting Lemmas	47
3.5	Simulation Study and Results	57
3.5.1	Simulation Study	57
3.5.2	Simulation Results	59
CHAPTER 4 APPLICATIONS OF THE DIMENSIONALITY REDUCTION METHODS ON MACHINE LEARNING DATASETS		66
4.1	Description of the Sample Datasets	66
4.2	Performance Assessment of NHRP	68
CHAPTER 5 DISCUSSION, CONCLUSION AND FUTURE PERSPECTIVES		74
5.1	Discussion	74
5.2	Conclusion and Future Perspectives	78
APPENDIX A		81
APPENDIX B		92
REFERENCES		102

LIST OF TABLES

2.1	Comparison of Dimensionality Reduction Methods	16
2.2	Taxonomy of Approaches in Implementing Hybrid DR Techniques Based on RP	18
2.3	Standard Error (SE) and Mean Distortions for each of the four Random Projection (RP) techniques utilised with dimension, $d=500$, sample size ($n=50$), across different densities (δ) and epsilon ($\epsilon = 0.8$) [1].	34
3.1	Standard Error (SE) and Mean Distortions for each of the four random projection (RP) techniques utilised with dimension, $d=500$, across various sample sizes (n), across different densities (δ) and epsilon (ϵ).	63
3.2	Standard Error (SE) and Mean Distortions for each of the four random projection (RP) techniques utilised with dimension, $d=1000$, across various sample sizes (n), across different densities (δ) and epsilon (ϵ).	64
B.1	The time taken to calculate each average distortion and the standard error for each method employed for $d = 500$	93
B.2	The time taken to calculate each average distortion and the standard error for each method employed for $d = 10000$	93
B.3	Optimal Alpha for The P_NHRP when $d = 500$	94
B.4	Optimal Alpha for The T_NHRP when $d = 500$	95
B.5	Optimal Alpha for the P_NHRP when $d=5000$	96

B.6	Optimal Alpha for the T_NHRP when $d=5000$	97
B.7	95% Confidence Intervals of Average Distortions for the five random projection (RP) methods employed with dimensions, $d = 500$ across different densities (δ) and epsilon (ϵ).	98
B.8	95% Confidence Intervals of Average Distortions for the five RP methods employed with dimensions, $d = 1000$ across dif- ferent densities (δ) and epsilon (ϵ)	99

LIST OF FIGURES

2.1	Dimensionality Reduction Techniques	11
2.2	Dimensionality in hyper-spectral images (Luo et al. 2020). . .	26
2.3	Motif Discovery Process	28
2.4	Classification Performance Using Baseline and Random Projections with LLE (Cheng et al. 2016).	31
3.1	Comparison of Methods across Different n Values for d = 500.	65
3.2	Comparison of Methods across Different n Values for d = 5000.	65
4.1	Comparison of (a) test and (b) train average distortion for NRP, PMRP, HRP and T_NHRP methods for digit mnist data.	69
4.2	Comparison of (a) test and (b) train average distortion for NRP, PMRP, HRP and T_NHRP methods for toxicity data.	70
4.3	Comparison of (a) test and (b) train average distortion for NRP, PMRP, HRP and T_NHRP methods for HAR data. . .	71
4.4	Comparison of (a) test and (b) train average distortion for NRP, PMRP, HRP and T_NHRP methods for period changer data.	72
B.1	Comparison of Methods across Different Epsilons for d = 1000	100
B.2	Comparison of Methods across Different Epsilons for d = 10000	101
B.3	Average Distortion for the Random projection (RP) methods employed with d = 1000 and n=50	101

THESIS ABSTRACT

NAME: ADENIRAN USMAN

TITLE OF STUDY: NON-LINEAR HYBRID RANDOM PROJECTION
TECHNIQUES FOR HIGH-DIMENSIONAL DATA

MAJOR FIELD: Applied Statistics

DATE OF DEGREE: May 2025

Dimensionality reduction (DR) is a cornerstone in managing high-dimensional data for efficient computation, analysis, and storage. DR of large datasets is essential in fields such as cloud computing, genetic distribution, image processing, and smart grids, which require high-dimensional data analysis. Linear Hybrid Random Projection (HRP) methods are used to minimize the data dimension of the data while maintaining the data structure. However, they are limited in capturing the inherent structure of complex nonlinear data. This limitation necessitates this paper to introduce a novel Nonlinear Hybrid Random Projection (NHRP) method, combining normal random projection (NRP) and plus-minus random projection (PMRP) with nonlinear transformations. This

approach aims to enhance the representation of nonlinear relationships, thereby improving the accuracy and effectiveness of DR in various applications such as cloud computing, genetic distribution, image processing, and smart grids. NHRP dynamically balances NRP and PMRP contributions using a blending parameter, effectively reducing dimensions while preserving complex relationships. Our results indicate that the NHRP method significantly improves over traditional HRP methods. The effectiveness of NHRP, compared with NRP, PMRP, and HRP, is evaluated through simulations and real-world data analyses, including datasets from human activity recognition and hyperspectral images. The findings suggest that the NHRP performs better than other methods (NRP, PMRP and HRP), indicating that the added complexity of nonlinear transformations does necessarily enhance the outcomes of dimensionality reduction. This paper discusses these findings and explores the implications for future research in dimensionality reduction techniques.

ملخص الرسالة

الاسم: اسم الطالب

عنوان الدراسة: عنوان الرسالة

التخصص: قسم الطلاب

تاريخ الدرجة العلمية: أدخل الشهر والسنة

يُعدّ تقليل الأبعاد (DR) حجر الزاوية في إدارة البيانات عالية الأبعاد لضمان كفاءة الحوسبة والتحليل والتخزين. ويُعدّ تقليل الأبعاد لمجموعات البيانات الكبيرة أمرًا بالغ الأهمية في مجالات مثل الحوسبة السحابية، والتوزيع الجيني، ومعالجة الصور، والشبكات الذكية، والتي تتطلب تحليل بيانات عالية الأبعاد. وتُستخدم أساليب الإسقاط العشوائي الهجين الخطي (HRP) لتقليل أبعاد البيانات مع الحفاظ على هيكلها. ومع ذلك، فإن قدرتها على التقاط البنية الجوهرية للبيانات غير الخطية المعقدة محدودة. ويتطلب هذا القيد من هذه الورقة تقديم طريقة جديدة للإسقاط العشوائي الهجين غير الخطي (NHRP)، تجمع بين الإسقاط العشوائي العادي (NRP) والإسقاط العشوائي الموجب والسالب (PMRP) مع التحويلات غير الخطية. ويهدف هذا النهج إلى تحسين تمثيل العلاقات غير الخطية، وبالتالي تحسين دقة وفعالية تقليل الأبعاد في تطبيقات مختلفة (xv) مثل الحوسبة السحابية، والتوزيع الجيني، ومعالجة الصور، والشبكات الذكية. يوازن NHRP ديناميكيًا مساهمات NRP و PMRP باستخدام مُعامل دمج، مما يُقلل الأبعاد بفعالية مع الحفاظ على العلاقات المُعقدة. تُشير نتائجنا إلى أن طريقة NHRP تُظهر تحسنًا ملحوظًا مقارنةً بطرق HRP التقليدية. تُقيّم فعالية NHRP، مقارنةً بطرق NRP و PMRP و HRP، من خلال عمليات المحاكاة وتحليلات البيانات الواقعية، بما في ذلك مجموعات البيانات المُستقاة من التعرف على النشاط البشري والصور الطيفية الفائقة. تُشير النتائج إلى أن أداء NHRP أفضل من أداء الطرق الأخرى (NRP و PMRP و HRP)، مما يُشير إلى أن التعقيد المُضاف للتحويلات غير الخطية يُعزز بالضرورة نتائج تقليل الأبعاد. تُناقش

هذه الورقة هذه النتائج وتستكشف آثارها على الأبحاث المستقبلية في تقنيات تقليل الأبعاد.

CHAPTER 1

INTRODUCTION

The rapid expansion of data across various domains, including machine learning, bioinformatics, and social network analysis, has introduced significant challenges related to high-dimensional data processing. As the dimensionality of data increases, traditional dimensionality reduction techniques, such as Principal Component Analysis (PCA) and Singular Value Decomposition (SVD), become computationally expensive and often require extensive parameter tuning (Johnson & Lindenstrauss, 1984) [2]. Random projections (RP) offer a scalable and efficient alternative for dimensionality reduction while preserving the geometric properties of high-dimensional data. One of the fundamental challenges in high-dimensional data processing is the curse of dimensionality, where distance metrics become less informative as dimensions increase [3]. RP mitigates this issue by mapping high-dimensional data into a lower-dimensional subspace while approximately maintaining pairwise distances, as guaranteed by the Johnson-Lindenstrauss (JL) Lemma [2]. This property is particularly useful in machine learning applications where preserving the structure of data is critical for classification, clustering, and regression tasks [4]. Taking into account these varied characteristics, this

thesis introduces innovative methods of random projection by merging the advantages of two separate random projection techniques. This integration seeks to create a balance between distance preservation and interpretability, making it suitable for various data properties and analytical tasks.

1.1 Random Projection (RP)

RP is a widely adopted technique for dimensionality reduction, particularly suited for high-dimensional data analysis. The method works by transforming data from a high-dimensional space into a lower-dimensional one through the application of a randomly generated matrix, preserving essential geometric properties of the data while significantly reducing computational complexity [4], [5]. Unlike traditional dimensionality reduction techniques such as PCA, which require eigenvalue decomposition and are computationally expensive, RP provides an efficient and scalable alternative with theoretical guarantees on distance preservation [6], [7]. The theoretical basis of RP techniques is rooted in the JL lemma, which offers a probabilistic assurance that pairwise Euclidean distances can be approximately preserved when a set of n points in a data is projected into a lower-dimensional space of dimension $k = O(\log n)$ with minimal distortion in pairwise Euclidean distances [2]. This result implies that high-dimensional structures can be approximately preserved even after projection, making RP an attractive choice for applications in machine learning, clustering, and information retrieval [8]. Several studies have explored different random projection schemes, such as Gaussian random matrices and structured sparse projection matrices, each

offering trade-offs between accuracy and computational efficiency [7], [9].

Beyond its theoretical significance, RP has demonstrated remarkable practical utility. It has been successfully applied to large-scale text processing, where dimensionality reduction is essential for handling sparse, high-dimensional feature representations [10]. Moreover, RP has gained attention in signal processing and bioinformatics, where it facilitates efficient data compression while preserving meaningful relationships within the data [11], [12]. Recent advances also explore hybrid approaches that combine RP with other dimensionality reduction techniques to enhance performance in various domains, including graph embeddings and deep learning [13], [14].

In general, RP offers an effective balance between computational efficiency and accuracy, making it a compelling choice for high-dimensional data analysis. Its theoretical underpinnings, along with extensive empirical validation, reinforce its role as a fundamental tool in modern data science and machine learning applications.

1.2 Hybrid Random Projection (HRP)

HRP has been proposed as an advanced approach to enhance the performance of traditional Random Projection (RP) methods by integrating different RP techniques or combining RP with complementary methods. One notable example is linear HRP, which integrates normal random projection (NRP) and plus-minus one random projection (PMRP) has shown promising results in various applications [15]. The HRP approach proficiently integrates NRP and PMRP using a blending parameter to optimize their contributions [15]. This parameter allows the method to balance between

accuracy and computational efficiency. By combining NRP and PMRP, HRP achieves better performance in terms of distance distortion, which signifies the preservation of the dataset’s structure following dimensionality reduction [15]. Moreover, HRP can be tailored for specific applications such as text processing, bioinformatics, and large-scale machine learning, where the balance between computational cost and data fidelity is crucial [7], [8]. However, these methods are limited by their linear nature, which may not capture the nonlinear relationships present in high-dimensional data, and despite its advantages, HRP still involves an inherent approximation error compared to exact methods [15].

1.3 Machine Learning

Machine learning has fundamentally reshaped numerous fields, driving significant advancements in various sectors since its emergence. This transformation stems from the widespread adoption of machine learning techniques, which have enabled data-driven decision making and automation in various applications. Arthur Samuel [16] defined machine learning as a discipline that allows computers to learn autonomously without being explicitly programmed. Positioned as a subset of *Artificial Intelligence (AI)*, machine learning focuses on developing algorithms and models capable of extracting patterns from data, facilitating predictive analytics and automated decision-making [17]. Unlike conventional programming, which relies on predefined rule-based instructions, machine learning leverages data-driven approaches to identify underlying trends and generalize knowledge to unseen data. The impact of machine learning spans mul-

multiple disciplines, revolutionizing research and industry applications. For example, in materials science and chemistry, machine learning has significantly improved catalyst performance prediction, as demonstrated in recent studies [18], [19]. Similarly, healthcare has witnessed transformative progress with the integration of machine learning models in disease diagnosis, drug discovery, and personalized medicine [20], [21], [22]. In bioinformatics, machine learning techniques have been widely applied in genomics, protein structure prediction, and biological data analysis [23], [24]. Furthermore, agriculture has benefited from machine learning applications in precision farming, crop yield prediction, and pest control [25], [26].

Machine learning has also been instrumental in computer vision and speech recognition, leading to breakthroughs in image classification, facial recognition, and natural language processing [27], [28]. These advancements illustrate the broad and profound influence of machine learning across disciplines, underscoring its role as a key driver of innovation in modern computational research. As machine learning techniques continue to evolve, their applications are expected to expand further, shaping the future of numerous scientific and industrial domains.

1.4 Aims and Objectives

The rapid growth of data in various domains has presented significant challenges in processing and analyzing high-dimensional datasets. Traditional dimensionality reduction techniques, such as PCA, often struggle with computational scalability and efficiency, particularly in large-scale data applications. In contrast, RP methods have

gained prominence due to their computational simplicity and ability to preserve the structure of high-dimensional data. However, linear HRP techniques frequently fail to capture the nonlinear relationships inherent in complex datasets, such as those encountered in image processing, biological data, and other structured domains.

This thesis aims to address these limitations by developing an advanced Nonlinear Hybrid Random Projection (NHRP) method that integrates NRP and PMRP. The objective is to create a scalable and computationally efficient dimensionality reduction technique capable of preserving the intrinsic geometric structure of the data while ensuring robust performance across various high-dimensional data analysis tasks. The thesis will pursue the following key objectives:

- a literature review on hybrid dimensionality reduction techniques based on RP;
- development of a novel nonlinear hybrid random projection (NHRP) method;
- performance evaluation of NHRP through simulation studies;
- to assess the performance of NHRP algorithms using machine learning datasets.

By addressing these objectives, this research aims to contribute novel insights into dimensionality reduction methodologies, enhancing their applicability in real-world high-dimensional data analysis. The proposed NHRP method is expected to offer a robust, scalable, and computationally efficient alternative to existing RP techniques, facilitating improved performance in various practical applications.

1.5 Thesis Layout

In Chapter 2, we explored the theoretical background and related works of the thesis. Chapter 3 contains details on random projections to reduce the dimensionality of high-dimensional datasets, and introduces our novel random projection technique. Chapter 4 assesses the performance of the novel algorithms using machine learning datasets. Finally, Chapter 5 synthesizes the findings of the research, providing insight into the implications of the technique developed for high-dimensional data analysis. It also discusses the limitations of the study and provides recommendations for future research directions, aiming to advance the field of hybrid dimensionality reduction and its applications in various scientific and industrial domains.

CHAPTER 2

HYBRID DIMENSIONALITY REDUCTION TECHNIQUES BASED ON RANDOM PROJECTIONS: A LITERATURE REVIEW

This chapter provides a brief literature review on hybrid dimensionality reduction techniques based on random projection. It explores the theoretical foundations, existing methodologies, and advancements in the field, highlighting their strengths, limitations, and applications. The chapter also identifies research gaps and sets the stage for the development of the proposed novel approach.

In recent years, high-dimensional data analysis has become essential across vari-

ous fields, including bioinformatics, finance, physics, and machine learning. Such data refers to datasets with an extensive number of variables (p), often surpassing the number of observations (n), (e.g., $n < p$), leading to a situation where traditional statistical methods may not be applicable due to the “curse of dimensionality” [29]. The rise in high-dimensional data is largely attributed to advancements in data collection technologies, such as high-throughput sequencing and sensor-based measurements, which generate large volumes of features per instance [29], [30]. For example, in genomics, high-throughput DNA sequencing records thousands of features per patient despite having relatively fewer patients, while in image analysis, each pixel in an image can represent an individual feature, transforming the image into high-dimensional space [31].

High-dimensional data analysis often faces challenges such as computational complexity, data sparsity, and overfitting, which make traditional approaches inadequate for such tasks [32]. Dimensionality reduction (DR) methods are widely employed to address these challenges by transforming high-dimensional data into a lower-dimensional space while retaining essential information [33], [34]. Conventional DR approaches such as PCA and Linear Discriminant Analysis (LDA) are well known and efficient for linear data, but tend to struggle with nonlinear structures and intervariable correlations [35], [36], [37]. Consequently, hybrid dimensionality reduction methods based on random projection have gained popularity for their ability to reduce computational load and maintain important data patterns through a combination of random projections and established DR techniques [34], [38].

RP has been recognized as an effective technique for DR, leveraging the JL lemma,

which asserts that data points in high-dimensional space can be projected into a lower-dimensional space while approximately preserving distances [39]. Hybrid random projection techniques have been developed by combining RP with other methods, offering improved efficiency and accuracy in various applications [40].

This review provides a brief overview of dimensionality reduction techniques based on random projection, focusing on recent trends, theoretical advancements, and application-specific insights. Specifically, this review focuses on different hybrid techniques involving RP, for DR. It also examines the interconnections between different hybrid methods, which enable efficient high-dimensional data analysis in domains such as real-time monitoring, quality control, and complex pattern recognition [41].

2.1 Overview of DR Techniques

Dimensionality reduction techniques can be broadly categorized into linear and nonlinear methods, each suited for different data structures. While linear approaches such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are computationally efficient, they struggle with complex, nonlinear relationships in data. This has led to the development of hybrid techniques that combine the strengths of different dimensionality reduction methods. The following sections provide an overview of both linear and nonlinear techniques before discussing hybrid approaches based on random projection. Figure 2.1 displays the DR techniques.

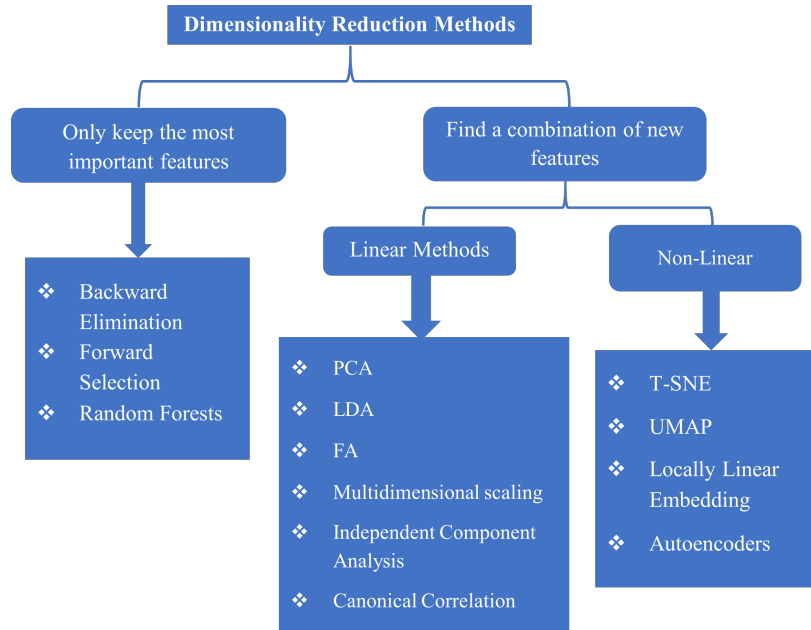


Figure 2.1: **Dimensionality Reduction Techniques**

2.1.1 Linear Dimensionality Reduction Methods

Linear techniques focus on reducing data dimensionality by identifying linear combinations of features that capture the maximum variance or structure of the data. Among these methods, Principal Component Analysis (PCA) is one of the most widely used [42], [43]. PCA identifies the directions (principal components) that explain the most variance in the dataset and projects the data onto these directions, thus reducing dimensionality while retaining the most informative features [17]. The method is particularly useful in applications such as image compression, financial modeling, and genomics, where reducing dimensionality helps improve computational efficiency while preserving essential patterns in the data [43], [44].

While PCA is highly effective for datasets with linear relationships, its primary limitation lies in its inability to capture complex, nonlinear relationships in high-dimensional data [17], [43]. Other linear methods, such as Linear Discriminant Anal-

ysis (LDA) and Factor Analysis (FA), similarly rely on linear assumptions but are designed for specific tasks. LDA maximizes the separation between multiple classes by finding linear combinations of features, making it particularly valuable for classification tasks [45]. Whereas FA models the underlying structure of the data through latent variables, which can help identify hidden patterns in psychometrics, social sciences, and econometrics [46]. Both methods, while useful for specific purposes, face similar challenges when applied to nonlinear data.

In addition to PCA, LDA, and FA, several other linear dimensionality reduction methods have been developed. Multidimensional Scaling (MDS) seeks to preserve pairwise distances between data points while projecting them onto a lower-dimensional space, making it useful for visualization and exploratory data analysis [47]. Independent Component Analysis (ICA) focuses on separating mixed signals into statistically independent components, making it highly applicable in biomedical signal processing, such as EEG and fMRI analysis [48]. Canonical Correlation Analysis (CCA) is another linear method that identifies relationships between two sets of variables, commonly used in multimodal data analysis, such as linking brain activity data with behavioral responses in neuroscience [49].

RP is an alternative technique that overcomes some limitations of traditional linear methods by efficiently reducing dimensionality while preserving the structure of high-dimensional data. RP relies on the JL Lemma, which states that high-dimensional data can be projected into a lower-dimensional space using a random matrix while approximately preserving pairwise distances [2]. Unlike PCA, which requires eigenvalue decomposition, RP is computationally efficient and scalable, making it suitable

for large-scale applications such as natural language processing, image compression, and large-scale clustering [5], [40]. One common variation of RP is Gaussian Random Projection, where elements of the projection matrix are drawn from a Gaussian distribution, ensuring that projections remain approximately orthogonal [8]. Another variant, Sparse Random Projection, uses sparse matrices to further optimize computational performance while maintaining high accuracy [5]. Applications of RP include large-scale text document classification, where it enables efficient feature reduction in high-dimensional term-frequency matrices, and real-time anomaly detection in cybersecurity [5], [50]. The flexibility and efficiency of RP make it a powerful tool for handling high-dimensional data in modern machine learning and data science applications.

These linear dimensionality reduction techniques provide diverse solutions for extracting meaningful structure from high-dimensional data, but their performance depends on the nature of the dataset. Although methods such as PCA, LDA, and ICA excel in structured datasets with clear linear relationships, RP offers a fast, scalable alternative that is particularly well suited for very large, high-dimensional problems. As research continues to explore hybrid techniques that combine RP with nonlinear methods, new opportunities arise to tackle the challenges of high-dimensional data analysis more effectively [5], [40].

Although PCA, LDA, and related techniques effectively reduce dimensionality in structured datasets, their reliance on linear transformations limits their applicability to datasets with complex, nonlinear structures. This limitation has motivated researchers to explore nonlinear techniques and hybrid approaches that integrate random

projections with machine learning models for improved scalability and performance.

2.1.2 Nonlinear DR Methods

Nonlinear dimensionality reduction (DR) methods have been developed to better manage complex data structures, where relationships between features are not strictly linear [51], [52], [53]. Key techniques in this category include t-distributed Stochastic Neighbor Embedding (t-SNE), Uniform Manifold Approximation and Projection (UMAP), Locally Linear Embedding (LLE), and Autoencoders.

T-SNE is widely used to visualize high-dimensional data by effectively preserving local pairwise similarities within a lower-dimensional embedding [51]. It achieves this by constructing probability distributions over pairs of data points, based on similarity, and then mapping these distributions into a low-dimensional space where similar points remain close. However, the computational complexity of t-SNE can be limiting for larger datasets, particularly those with high dimensionality, due to the need for extensive pairwise calculations [54].

UMAP, introduced in 2018, combines topological and geometry insights to represent global and local structures in the data [55]. UMAP has demonstrated better scalability on large datasets compared to t-SNE, making it a preferred method for applications that require the visualization or compression of large, high-dimensional datasets, such as genomics or image processing [56]. UMAP’s design also allows it to preserve more global structures, making it a robust alternative for a variety of high-dimensional data analysis tasks.

Another widely used nonlinear DR technique is LLE, which preserves local neigh-

neighborhood relationships by reconstructing each data point as a linear combination of its nearest neighbors before embedding the data into a lower-dimensional space [52]. LLE is particularly useful for detecting intrinsic manifolds in data, making it effective in applications such as speech processing and biological data analysis. However, LLE may struggle with noisy or sparse datasets and does not explicitly preserve global structure.

Autoencoders, a deep learning-based DR technique, have gained prominence due to their ability to learn complex, nonlinear transformations [53]. Autoencoders consist of an encoder that maps high-dimensional inputs into a compressed latent representation and a decoder that reconstructs the original input from this representation. Variational Autoencoders (VAEs) extend this approach by incorporating probabilistic modeling, allowing for smoother latent space representations [57]. Autoencoders are widely used in applications such as anomaly detection, image compression, and feature learning in large-scale machine learning tasks.

These nonlinear DR methods offer powerful tools for understanding complex datasets beyond the limitations of linear techniques. While t-SNE and UMAP excel in visualization tasks, LLE is effective for detecting local manifold structures, and Autoencoders provide a scalable approach to representation learning. The choice of method depends on the dataset characteristics, scalability requirements, and the specific analytical goals [52], [55], [53].

Table 2.1: **Comparison of Dimensionality Reduction Methods**

Method	Type	Advantages	Limitations	Applications
PCA	Linear	Efficient, preserves variance, interpretable	Fails for non-linear data	Image compression, finance, genomics
LDA	Linear	Optimized for classification tasks	Assumes normality, struggles with high-dimensional space	Face recognition, text classification
t-SNE	Nonlinear	Preserves local structures well, good for visualization	Computationally expensive, does not scale well	High-dimensional visualization
UMAP	Nonlinear	Faster and preserves both local/global structures	Can be sensitive to hyperparameters	Bioinformatics, machine learning
RP	Linear	Scalable, computationally efficient	Can lead to information loss if not optimized	Large-scale clustering, NLP

2.2 Random Projection and Its Hybrid Variants

Although RP is computationally efficient, it often falls short in capturing task-specific information due to the latent space being generated without considering the intrinsic structure of the original data. This limitation has led to the development of various methods aimed at enhancing RP’s performance.

Hybrid Random Projection (HRP) addresses these shortcomings by integrating

traditional RP techniques with additional strategies that consider the data’s inherent structure. These methods can be classified into four categories based on this review: RP integrated with ML, Graph Embedding Enhanced via RP, Combination of RP with Nonlinear Methods, and Combined variants of RP (Type 1 and Type 2), ensuring improved retention of features, better adaptability to complex datasets, and improved predictive performance.

This section presents a chronological review of related works on hybrid DR techniques based on random projection, categorized according to the four approaches. Table 2.2 shows the taxonomy approaches developed in the implementation of hybrid DR techniques to enhance the performance of RP detailing their key advantages, limitations, applications, and relevant literature.

Table 2.2: **Taxonomy of Approaches in Implementing Hybrid DR Techniques Based on RP**

Approach	Advantages	Limitations	Applications	References
RP integrated with ML	Efficient dimensionality reduction, reduced overfitting, scalable for large datasets, compatible with multiple ML models.	Potential loss of information if projection is too aggressive, may introduce variance in results, depends on ML model choice.	High-dimensional classification, clustering, preprocessing for neural networks, bioinformatics applications.	[2], [5], [7], [8], [13], [40], [58], [59], [60], [61], [62], [63], [64], [65], [66]

Approach	Advantages	Limitations	Applications	References
RP integrated with Graph Embed- dings	- Efficient for reducing graph embedding dimensions. - Scalable for large graphs and networks. - Preserves essential graph structure while reducing complexity. - Reduces storage and computational requirements for graph-based models. - Improves speed and efficiency of graph-based ML tasks.	- Potential loss of structural information during projection. - Performance highly dependent on embedding method used. - Graph topology may not always be well-represented after projection. - Limited interpretability in transformed space.	- Link prediction in social networks. - Node classification and community detection. - Recommender systems for large-scale networks. - Graph-based anomaly detection in cybersecurity. - Network compression and visualization.	[67], [68], [69], [70], [71], [72], [73], [74], [75]

Approach	Advantages	Limitations	Applications	References
Combination of RP with Nonlinear Methods	- Enhances RP by preserving local relationships and capturing non-linear structures in data. - Ideal for complex datasets, but with increased computational complexity and interpretability challenges.	- Higher computational complexity when combined with non-linear models. - Trade-off between preserving local/global structures. - Harder to interpret compared to linear projection techniques.	- Non-linearly separable data (e.g., image/speech recognition), medical imaging, biomedical analysis, and signal processing.	[76], [77], [78], [79], [80], [81], [82], [53], [83], [51]

Approach	Advantages	Limitations	Applications	References
Combined variants of RP (Type 1 and Type 2)	- Combines different RP methods for better accuracy and sparsity balance. - Can be optimized for specific dataset characteristics. - More robust against variations in data distribution.	- Trade-offs between different RP methods can affect consistency. - Potential information loss in highly structured datasets. - Requires tuning to optimize projection parameters.	- Text classification, genomic data analysis, large-scale image/video processing, and data compression.	[1], [84], [85], [86], [87]

2.2.1 Integration of RP with Graph Embeddings

The integration of RP into graph embedding techniques offers a powerful approach to reducing the dimensionality of graph-based data while preserving important structural properties. In graph-based learning, node embeddings are widely used to capture essential structural information by mapping high-dimensional adjacency matrices or feature vectors into a lower-dimensional space [88]. However, the direct application of RP to graph embeddings may

disrupt local topological structures, leading to the loss of important connectivity patterns [71]. To address this limitation, hybrid approaches such as Sparse Random Projection (SRP) and Laplacian Eigenmaps (LE) have been proposed, where RP serves as a preprocessing step to reduce the dimensionality of the adjacency matrix before applying a more structured embedding method [72], [70].

Graph embedding methods, such as Node2Vec and DeepWalk, generate vector representations of nodes, capturing information about their local and global contexts within the graph [67], [75]. However, these embeddings are often high-dimensional, making them computationally expensive for downstream tasks like link prediction, node classification, and graph clustering [89]. By applying RP, it is possible to efficiently reduce the dimensionality of these embeddings, improving computational efficiency and scalability while maintaining the structural integrity necessary for accurate predictions. RP is particularly useful in graph-related tasks where large-scale networks need to be processed. RP helps in scaling up graph-based machine learning models by transforming high-dimensional embeddings into lower-dimensional ones, which reduces memory usage and accelerates training times, thus improving the overall efficiency of graph-based learning methods [68]. In graph-based learning, node embeddings are widely used to capture essential structural information by mapping high-dimensional adjacency matrices or feature vectors into a lower-dimensional space [88]. However, the direct application of Random Projection (RP) to graph embeddings may disrupt local topological structures, leading to the loss of important connectivity patterns [71]. To address this limitation, hybrid approaches such as Sparse Random Projection (SRP) and Laplacian Eigenmaps (LE) have been proposed, where RP serves as a preprocessing step to reduce the dimensionality of the adjacency matrix before applying a more structured embedding method [72], [70].

These hybridized techniques enhance computational efficiency while preserving critical graph relationships, making them particularly suitable for applications such as fraud detection in social networks, anomaly detection in cybersecurity, and large-scale knowledge graph embeddings [75], [68]. By integrating RP with graph-based learning, researchers aim to maintain the structural integrity of the network while benefiting from the reduced computational complexity associated with lower-dimensional embeddings.

2.2.1.1 Mathematical Representation

Let a graph be represented as $G = (V, E)$, where V is the set of nodes and E is the set of edges. The structure of the graph can be captured using an adjacency matrix $A \in \mathbb{R}^{n \times n}$, where n is the number of nodes. To perform hybrid dimensionality reduction, a sparse random matrix $R \in \mathbb{R}^{n \times k}$ is generated, where $k \ll n$, ensuring a lower-dimensional representation while approximately preserving graph properties.

The transformation using RP can be mathematically formulated as:

$$A' = AR, \tag{2.1}$$

where $A' \in \mathbb{R}^{n \times k}$ represents the reduced adjacency matrix, and R is a sparse random matrix whose elements are drawn from a predefined distribution, typically $\{-1, 0, 1\}$ with equal probabilities to ensure computational efficiency [5]. This transformation significantly reduces computational complexity while maintaining essential connectivity information within the graph.

The choice of RP type (e.g., Gaussian, Sparse, or Binary) influences the trade-off between computational efficiency and accuracy in preserving graph properties [2], [7]. Empirical studies have demonstrated that applying RP before graph embedding techniques

such as node2vec, DeepWalk, or Graph Convolutional Networks (GCNs) leads to improved scalability in large-scale network analytics [67], [65].

2.2.1.2 Related Work

RP has also found wide application in graph-based learning tasks, particularly in generating efficient graph embeddings and handle large-scale and heterogeneous graph structures. Raducanu and Dornaika [72] proposed a hybrid dimensionality reduction approach that combines Sparse Random Projections (SRP) with Laplacian Eigenmaps (LE) to address the computational challenges of graph-based non-linear dimensionality reduction methods. SRP is used as a preprocessing step to project high-dimensional data into a lower-dimensional space while preserving pairwise distances, as supported by the JL lemma [72], [39]. This reduction significantly decreases the computational burden of constructing adjacency graphs and solving eigenvalue problems, which are core components of LE. The hybrid approach achieves scalability for larger datasets while maintaining fidelity to the original manifold structure, making it particularly suitable for semi-supervised learning tasks that rely on graph-based methods [72].

Extending these concepts, Zhu and Ermon [65] present a novel meta-algorithm for probabilistic inference in graphical models using random projections. The method leverages universal hash functions to generate an exponentially large number of random samples, effectively projecting the original statistical model onto a lower-dimensional subspace. This hybrid approach balances sampling-based methods and variational techniques, offering improved accuracy by managing the trade-off between approximation accuracy and variance. Empirical results demonstrate that the integration of random projections improves the performance of approximate inference algorithms, making this a significant contribution to the

field of probabilistic graphical models [65].

Efforts to enhance network embedding through sparse RPs have shown significant promise. The study by Chen et al., [71] proposes an embedding method that leverages very sparse random projections to efficiently reduce the dimensionality of graph adjacency matrices while preserving their essential structural properties. This approach ensures both computational efficiency and accuracy in representing graph topologies, making it ideal for large-scale networks. Sparse RPs not only reduce storage requirements but also facilitate faster downstream tasks, such as clustering and link prediction, which are critical for network analysis [71].

Luo et al. [73] propose an innovative method for dimensionality reduction in hyperspectral image classification, termed Enhanced Hybrid-Graph Discriminant Learning (EHGDL). The method constructs intraclass and interclass hypergraphs to capture complex high-order relationships within hyperspectral images (HSIs). In addition, a supervised locality graph and a weighted neighbourhood margin model are used to enhance the separability of different classes. The resulting dimensionality reduction model effectively improves the compactness of intraclass samples and the separability of interclass samples, leading to an optimal projection matrix for low-dimensional embedding features. The effectiveness of EHGDL is demonstrated through experiments on the Indian Pines, PaviaU, and HoustonU datasets, where it outperforms traditional methods in terms of classification accuracy. This approach highlights the potential of hybrid-graph learning in addressing the challenges of high-dimensional data in hyperspectral image classification, offering a robust solution for improving classification performance [73]. Dimensionality reduction in hyperspectral images is shown in Fig. 2.2.

For heterogeneous graph learning, Hu et al. [70] introduce an efficient method that uses

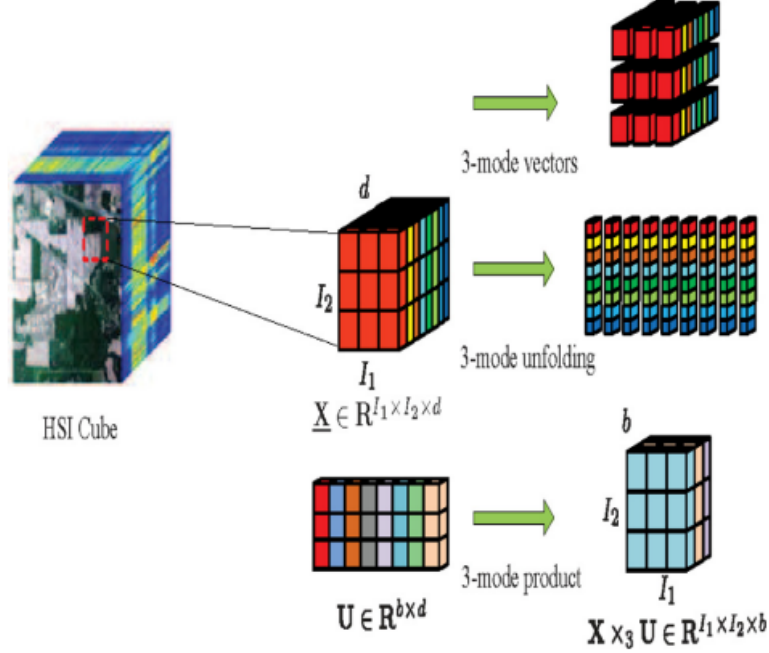


Figure 2.2: **Dimensionality in hyper-spectral images** (Luo et al. 2020).

RP to handle the diverse feature spaces of nodes and edges in multi-typed networks. By reducing the dimensionality of heterogeneous data representations, the proposed method ensures scalability while retaining the graph’s critical relational information. This approach is particularly advantageous in scenarios involving large, complex graphs, such as social networks or recommendation systems, where computational efficiency is paramount [70].

2.2.2 Integration of RP with Machine Learning Models

RP is a popular DR technique that is used to preprocess high-dimensional data, significantly reducing computational complexity and improving scalability. By projecting data onto a lower-dimensional subspace using a randomly generated matrix, RP ensures that pairwise distances between points are approximately preserved, adhering to the JL Lemma [7]. This characteristic makes it an attractive preprocessing step for machine learning models. RP is particularly beneficial in scenarios where datasets have a large number of features,

such as text, images, or genomic data [90]. This is because RP helps reduce the dimensionality of the data, making it more manageable, and improving the performance of machine learning models. For example, in text analysis, RP can help in reducing the number of features (words) while preserving the essential information [90]. By transforming the data into a lower-dimensional space, RP alleviates the curse of dimensionality, reduces storage requirements, and accelerates subsequent machine learning processes.

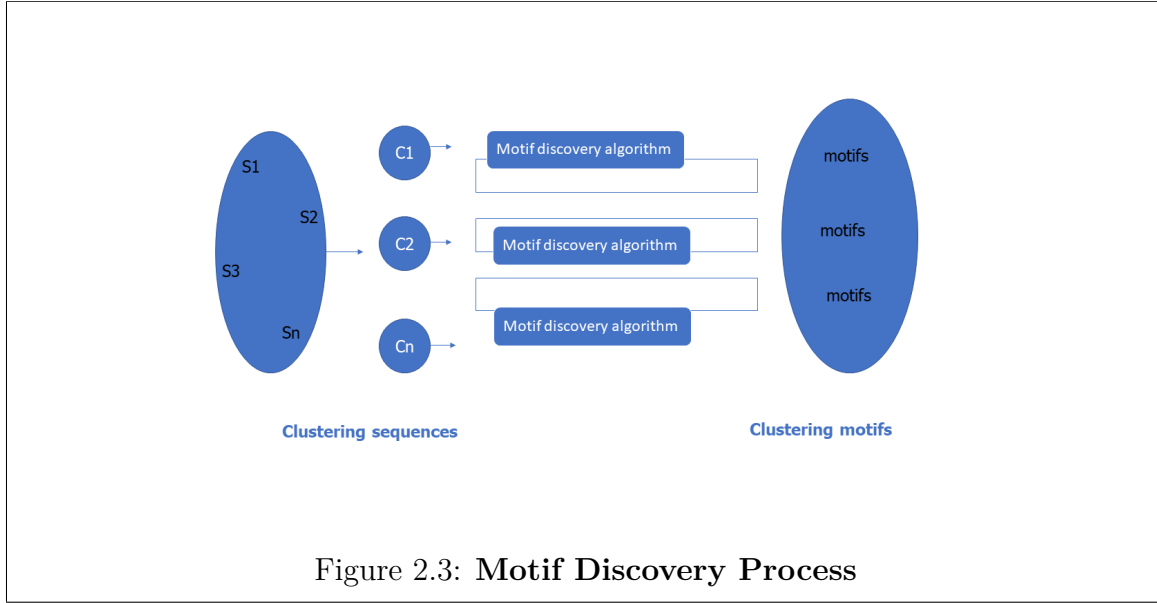
RP is commonly integrated with machine learning pipelines in the following ways:

1. **Feature Transformation:** Before applying algorithms like Support Vector Machines (SVM) or neural networks, RP reduces feature dimensions, thereby lowering training time and memory usage while maintaining model accuracy [84], [91].
2. **Clustering:** Distance-based clustering methods such as K-means benefit from RP since it preserves the geometric structure of the data, ensuring robust cluster formation [92], [93].
3. **Ensemble Learning:** RP is used in ensemble techniques to generate various subsets of characteristics, improving the overall performance of the ensemble by reducing redundancy of characteristics and overfitting [94], [95].

2.2.2.1 Related Work

The integration of RP with machine learning algorithms has led to notable advancements in various tasks, including motif discovery, time-series prediction, and speech recognition [96], [97]. By reducing the dimensionality of high-dimensional data while approximately preserving its geometric structure, RP enables efficient processing and analysis in computationally intensive applications.

One prominent application of RP is motif discovery within genomic sequence analysis. Shashidhara et al. [98] proposed a hybrid approach combining RP with the Expectation-Maximization (EM) algorithm to refine motif detection in high-dimensional biological sequences. RP reduces the computational burden by projecting the genomic data into a lower-dimensional space, allowing the EM algorithm to focus on relevant features with improved efficiency. This approach enhances motif identification accuracy while reducing processing time, a crucial factor in large-scale bioinformatics applications. However, a key limitation of this method is its sensitivity to input data quality; noisy or erroneous sequences may reduce motif detection accuracy [98]. A detailed illustration of the motif discovery process using RP and the EM algorithm can be found in the original work by Shashidhara et al., [98].



Beyond bioinformatics, RP has been widely applied in time-series prediction through reservoir computing (RC) systems. Butcher et al.[60] extended RC models by incorporating random static projections to improve both efficiency and accuracy in temporal tasks. In this approach, RP is used to reduce input dimensionality before feeding data into the RC system, leading to significant computational savings while maintaining predictive performance. The

authors further enhance the RC model by integrating Extreme Learning Machines (ELM), optimizing learning speed and prediction accuracy. This demonstrates RP’s effectiveness in accelerating real-time forecasting tasks, such as financial trend analysis and environmental monitoring [60]. However, a limitation of RP in RC models is reservoir instability in highly sparse or noisy datasets, which may degrade training effectiveness and prediction accuracy. A detailed illustration of the RP in RC models can be found in the original work by Butcher et. al., [60].

These studies highlight RP’s potential to enhance both efficiency and generalisation in machine learning models by serving as a pre-processing step for high-dimensional data. The flexibility of RP in various domains, from biological sequence analysis to real-time predictive modelling, demonstrates its versatility as a dimensionality reduction tool for modern machine learning applications.

2.2.3 Integration of RP with Non-linear Techniques

The combination of RP with non-linear methods has gained attention in dimensionality reduction tasks, particularly when the data exhibits complex, non-linear relationships [99]. While RP is effective for linearly reducing dimensions, it may not capture intricate, non-linear structures inherent in many datasets, such as images, text, or complex physical systems [99]. By integrating RP with non-linear techniques like t-SNE (t-Distributed Stochastic Neighbor Embedding), autoencoders, and kernel-based methods, it becomes possible to address these challenges and enhance the performance of machine learning models [51].

- **t-SNE:** This technique is widely used for visualizing high-dimensional data by projecting it onto two or three dimensions. RP can be used as a preprocessing step to reduce the initial dimensionality, making the subsequent application of t-SNE more

computationally efficient without significantly compromising its ability to preserve the local structure of the data [51].

- **Autoencoders:** They are neural networks trained to learn a compressed representation of data through an encoder-decoder structure. RP can be applied to reduce the input dimensionality before feeding it into an autoencoder, thereby improving training efficiency and reducing the computational burden, particularly in large-scale datasets. This combination preserves the data’s essential features while reducing dimensionality [76].
- **Kernel-based Methods:** Non-linear dimensionality reduction methods like kernel PCA (Principal Component Analysis) and locally linear embedding (LLE) rely on a non-linear transformation of data to capture more complex structures. RP can be used as a preprocessing technique to first reduce the dimensionality, and then these kernel-based methods can be applied to uncover intricate patterns that linear methods may miss [100].

2.2.3.1 Related Work

Building on these ideas, RP has been widely applied in tasks requiring nonlinear dimensionality reduction. By combining RP with non-linear techniques, these hybrid methods enable efficient processing of high-dimensional data while preserving complex, non-linear relationships within the dataset.

Cheng et al.[77] introduce a hybrid nonlinear dimensionality reduction framework that combines random projections with non-linear dimensionality reduction techniques such as Locally Linear Embedding (LLE), Isometric Embedding (Isomap), and Semidefinite Embedding (SDE). This innovative framework aims to enhance computational speed and reduce

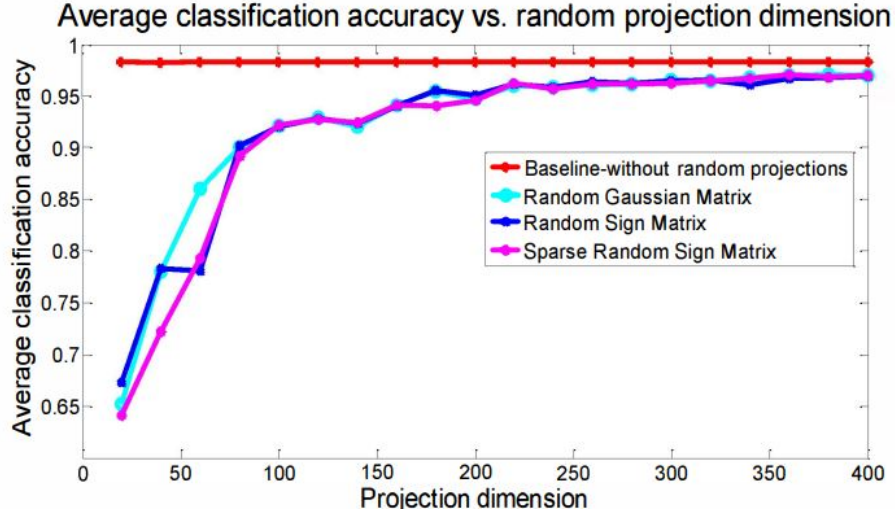


Figure 2.4: **Classification Performance Using Baseline and Random Projections with LLE (Cheng et al. 2016).**

memory usage while preserving the intrinsic non-linear geometry of the data. The authors validate the framework through experiments on a handwritten digits dataset, demonstrating its efficiency and effectiveness in maintaining data structure. The results indicate that this hybrid approach outperforms traditional linear dimensionality reduction methods in terms of both computational efficiency and accuracy in data geometry preservation [77]. A comparison of the baseline classification accuracy obtained by the combination of RPs with LLE is shown in Fig. 2.4.

Similarly, Mohan & Venkatesan [78] proposed an innovative approach to address the challenges of analyzing hyperspectral image (HSI) data, which are characterized by high dimensionality and rich spectral information. The hybrid approach integrates the strengths of RP and manifold learning to achieve efficient and effective DR for HSI. In this framework, RP serves as a preprocessing step, reducing the dimensionality of the data while preserving its global structure. The reduced data are then processed using a manifold learning algorithm to uncover its intrinsic nonlinear structure[78]. The combination results in a more effective and efficient dimensionality reduction process, suitable for various HSI tasks

such as classification, anomaly detection, and target detection. The synergy between RP and ML allows for the preservation of both global and local structures, leading to better performance in practical applications [101], [81]. The paper highlights the effectiveness of the method through case studies on hyperspectral datasets such as Indian Pines and Pavia University, showing improvements in classification accuracy compared to standalone DR techniques [102]. It also addresses challenges such as noise sensitivity and the need for adaptive projection matrices, offering directions for future research to enhance the robustness and scalability of the hybrid approach. This hybrid technique represents a promising solution for dimensionality reduction in hyperspectral imaging, leveraging the strengths of RP and ML to balance computational efficiency with data integrity.

Similarly, kernelized random projections have been employed to improve knowledge graph embeddings. In Goyal et al., [69], the authors combine RP with kernel methods to capture the non-linear relationships inherent in knowledge graphs. By applying kernelized RPs, the approach projects graph features into a low-dimensional space where complex relational patterns can be more effectively preserved. This hybrid technique bridges the gap between dimensionality reduction and advanced relational modelling, resulting in efficient and robust knowledge graph embeddings suitable for large datasets [69].

2.2.4 Combined variants of RP (Type 1 and Type 2)

The implementation of this approach is new and has only one paper published. The integration of Normal Random Projection (NRP) and Plus-Minus One Random Projection (PMRP) leverages the strengths of both methods to balance accuracy and computational efficiency in dimensionality reduction tasks using linear combination [1]. NRP, which uses a dense random matrix with elements sampled from a standard normal distribution $N(0,1)$, is

well-suited for preserving the geometric properties of data and ensures strong adherence to the Johnson-Lindenstrauss lemma. However, it can be computationally expensive for large datasets due to the need to generate and store a dense matrix [2], [39].

PMRP, on the other hand, employs a sparse random matrix with elements limited to $\{-1, 0, 1\}$, reducing memory requirements and speeding up matrix-vector multiplications. Although this sparsity offers significant computational advantages, it may lead to a slight degradation in the preservation of pairwise distances compared to NRP. By integrating these methods, one can design hybrid systems that utilize NRP for tasks requiring higher precision and PMRP for efficiency in large-scale or resource-constrained environments [1]. This integration could involve applying NRP to critical subsets of data or specific layers of a model where precision is essential, while PMRP handles broader datasets or non-critical computations. The combination allows for scalable and efficient dimensionality reduction without significant loss in accuracy [1].

The paper by Yahaya et al. [1] titled "Hybrid Random Projection (HRP) for Enhanced Representation in High-Dimensional Data" introduces a novel approach to dimensionality reduction that combines the strengths of Normal Random Projection (NRP) and Plus-Minus One Random Projection (PMRP). By incorporating a blending parameter, the HRP method optimizes the contributions of NRP and PMRP, reducing the data's dimensions while preserving its structure. Key contributions include improved computational efficiency, improved distance preservation, and application versatility. The HRP method's performance is evaluated through comprehensive simulations and analyses of real-world datasets, considering factors such as sample size, data dimensions, and sparsity. The results showed that HRP consistently exhibits the least distance distortion compared to NRP and PMRP, effectively minimizing information loss during the dimensionality reduction process Table 2.3

[1]. This positions HRP as a powerful tool for managing complex high-dimensional data, offering improved efficiency and precision in data analysis [1]. Nevertheless, they often fail to preserve the inherent geometry of data in a nonlinear manifold. This limitation becomes particularly evident in applications that involve image processing, biological data, and other domains where the data structure is inherently nonlinear [103]. The comparison of the average distortion ratios across the three methods is shown in Fig. ??.

Table 2.3: **Standard Error (SE) and Mean Distortions for each of the four Random Projection (RP) techniques utilised with dimension, $d=500$, sample size ($n=50$), across different densities (δ) and epsilon ($\epsilon = 0.8$) [1].**

Method	δ			
	0.2	0.4	0.6	0.8
	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)
NRP	12.86 (0.108)	12.88 (0.099)	12.89 (0.117)	12.87 (0.1)
PMRP	12.88 (0.109)	12.88 (0.088)	12.86 (0.103)	12.87 (0.094)
HRP	9.09 (0.077)	9.1 (0.073)	9.09 (0.089)	9.1 (0.072)

CHAPTER 3

NON-LINEAR HYBRID RANDOM PROJECTIONS

3.1 Objectives

The objective of this chapter is to introduce a novel nonlinear hybrid random projection framework (NHRP), which integrates the precision of NRP with the computational efficiency of PMRP through a nonlinear transformation. The NHRP approach aims to overcome the individual limitations of these techniques, providing a scalable and robust solution for dimensionality reduction across diverse datasets and applications. The NHRP method seeks to balance these trade-offs by designing a new projection matrix that blends the statistical properties of NRP and PMRP. This nonlinear combination leverages the flexibility of PMRP's sparsity with the precision of NRP's Gaussian sampling, ensuring robust performance across a wide range of application and preservation of the data structure.

3.2 Introduction

The complexity and volume of data have increased rapidly in the last few years, making it difficult to analyze and process the raw data directly. High-dimensional representation is indispensable, especially in areas such as statistical learning [103], biological data [104], imaging [105], and smart grid [106]. Since these high-level representations often contain ambiguous information and have little relationship with each other, it is necessary to represent the first high-level information in 3D while providing good information and preserving its essential properties [107]. As a result, reducing the dimensionality while adequately representing the original high-dimensional data without compromising the large amount of data received is of great interest. Due to the necessity of reduction, various techniques have been developed to solve this challenge. An important way to reduce the dimension is random projection, which is valuable due to its simplicity and efficiency in storing large data patterns [108].

Random projection (RP) techniques, grounded in the JL Lemma, facilitate the conversion of high-dimensional data into a lower-dimensional space while maintaining the pairwise distances between data points with high probability [2]. This approach is particularly good because it does not require the computation of eigenvalues or numerical value decomposition operations, which are computationally expensive for large datasets [2]. This effectiveness stems from the basic idea of JL lemma, which was first presented in their groundbreaking work in 1984 [2]. They showed that the pairwise distances between points are roughly maintained when high-dimensional data is projected onto a random subspace with high enough dimensions, reducing information loss [2], [71]. This discovery serves as the foundation for the use of RP in a variety of fields. There is literature that provides a more thorough mathematical explanation of this lemma (Frankl and Maehara, 1988; Dasgupta and Gupta,

2003). The JL lemma was succinctly proved by Frankl and Maehara (1988), who used it to set upper limits on the sphericity of graphs. Additionally, [8] used basic probabilistic methods to provide an alternate proof of the JL lemma.

RP has gained recognition as a computationally efficient technique for dimensionality reduction. It involves transforming high-dimensional data into a lower-dimensional space by applying a randomly generated transformation matrix [109], [110]. The fundamental aim of this method is to retain the essential geometric structure of the data while significantly reducing its dimensionality [5]. Despite its conceptual simplicity, RP is effective and computationally lightweight, offering accurate results in reducing dimensions of complex datasets.

This capability is supported by the JL lemma [2], introduced in 1984, which provides a theoretical foundation for RP. According to this lemma, it is possible to project high-dimensional data into a randomly chosen lower-dimensional subspace such that the pairwise Euclidean distances between data points are approximately maintained, thereby minimizing information loss. This concept forms the backbone of RP’s reliability across a wide range of applications

Among RP variants, normal random projection (NRP) and plus-minus one random projection (PMRP) are widely used for their simplicity and effectiveness. However, both methods have limitations in capturing the diverse structures of high-dimensional datasets [70], [71], [99]. NRP constructs the projection matrix by sampling its entries from a Gaussian distribution, ensuring strong theoretical guarantees of distance preservation as established by the JL lemma [2]. This method is particularly effective in preserving pairwise distances in high-dimensional spaces, making it suitable for tasks requiring precise geometric structure retention [71], [110]. On the other hand, PMRP generates sparse projection matrices

with entries of -1, 0, or 1, distributed with specific probabilities, such as $1/6$, $2/3$, and $1/6$, respectively [109]. This sparsity significantly reduces computational costs while maintaining acceptable performance and preserving data integrity during dimensionality reduction [109]. Despite their effectiveness, both methods face challenges when applied to complex datasets with intrinsic properties such as sparsity, nonlinearity, or intricate feature correlations [99], [109]. These limitations arise because NRP’s dense matrix construction can be computationally prohibitive for extensive datasets, while PMRP’s sparsity may lead to insufficient representation of intricate data structures [99], [71], [103].

Consequently, there is a growing interest in hybrid approaches that combine the strengths of both techniques to better handle the diverse challenges of high-dimensional data representation. Hybrid random projection (HRP) has been proposed to integrate different types of RP techniques or combine RP with other methods to enhance the performance of traditional RP techniques. For instance, linear HRP that integrates NRP and PMRP have shown promising results in various applications [15]. The HRP approach proficiently integrates NRP and PMRP by using a blending parameter to optimize their contributions [15]. This parameter allows the method to balance between accuracy and computational efficiency. By combining NRP and PMRP, HRP achieves better performance in terms of distance distortion, which signifies the preservation of the dataset’s structure following dimensionality reduction [15]. However, these methods are limited by their linear nature, which may not capture the nonlinear relationships present in high-dimensional data, and despite its advantages, HRP still involves an inherent approximation error compared to exact methods [15].

Due to these limitations, this study introduces a novel nonlinear hybrid random projection framework (NHRP), which integrates the precision of NRP with the computational

efficiency of PMRP through a nonlinear transformation.

The chapter is organized as follows: Section 2 presents an overview of different RP techniques utilized in the study. Section 3 introduces the NHRP method and its theoretical basis. In addition to introducing the NHRP approach, the section emphasizes its original synthesis of the benefits of current RP methods and provides a clear theoretical rationale for its development. The JL lemma and other well-known dimensionality reduction theories, which seek to maintain pairwise distances in high-dimensional data, serve as the theoretical cornerstone of NHRP. Simulation experiments are presented in Section 4 to illustrate the NHRP method’s applicability and robustness. The NHRP method’s application to sample datasets is covered in Section 5. Lastly, the chapter summary is included in Section 6.

3.3 Overview of RPs

Random Projections (RPs) are widely utilized in dimensionality reduction, a process designed to transform high-dimensional data

$$X = \{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^{n \times d}$$

be the original high-dimensional dataset. It is mapped into a lower-dimensional embedding

$$Y \in \mathbb{R}^{n \times k},$$

where $k < d$, where d is the original dimensionality, n is the number of data points, and k represents the reduced dimensionality. This transformation is particularly valuable for handling large datasets, enabling efficient computation while preserving the intrinsic structure

of the data [2].

The theoretical foundation of RPs is rooted in the Johnson-Lindenstrauss (JL) Lemma, which provides probabilistic guarantees for preserving pairwise Euclidean distances (l_2 norm) when projecting from a high-dimensional space to a lower-dimensional one [2]. The JL Lemma ensures that for a given error tolerance ϵ (where $0 < \epsilon < 1$) and a number of samples n , there exists a linear transformation $f : \mathbb{R}^{n \times d} \rightarrow \mathbb{R}^{n \times k}$ such that the pairwise distances between points in the original and reduced spaces remain approximately the same. Formally, this relationship can be expressed as:

$$(1 - \epsilon)\|x_i - x_l\|_2^2 \leq \|R(x_i - x_l)\|_2^2 \leq (1 + \epsilon)\|x_i - x_l\|_2^2 \quad (3.1)$$

where x_i and x_l are data points in the original space, $i \neq l$, $\|x_i - x_l\|_2^2$ is the pairwise distance between the i^{th} and l^{th} columns of $\mathbf{X}$, $\|\cdot\|_2$ denotes the Euclidean norm, and $R(x_i - x_l)$ represents the random transformations. This implies that the reduced-dimensional representation $\mathbf{Y}$ retains approximately the geometric structure of the original dataset $\mathbf{X}$.

3.3.1 Dimensionality and Error Tolerance

The JL Lemma guides the selection of the reduced dimensionality k to achieve a balance between computational efficiency and the preservation of the original data structure. The dimension k is determined as a function of the number of data points n and the desired error tolerance ϵ , ensuring that the geometric properties of the dataset are maintained. The dimension k of the embedding space is determined as follows:

$$k \geq \frac{24}{3\epsilon^2 - 2\epsilon^3} \log(d) \quad (3.2)$$

The parameter ϵ denotes the permissible level of distortion in pairwise distances following the projection [111]. A lower value of ϵ implies more accurate preservation of these distances but necessitates a larger target dimension k , thereby increasing the computational cost. In contrast, a higher ϵ can reduce the required dimensionality k , allowing for faster computations at the expense of some accuracy [112].

In practice, the application of RP techniques typically involves empirical evaluation, where the value of k is adjusted to assess its influence on performance metrics in tasks such as classification or clustering. This empirical strategy, informed by the theoretical framework of the JL lemma, facilitates the selection of a suitable k that balances computational efficiency with fidelity in data representation.

After identifying an appropriate k , the RP process proceeds by constructing a random projection matrix R of dimensions $d \times k$, where each entry is sampled from a chosen probability distribution commonly Gaussian or a binary ± 1 scheme. The original high-dimensional data matrix X , with d features, is then transformed into a reduced-dimensional form Y using the projection:

$$X \in \mathbb{R}^{n \times d}, \quad R \in \mathbb{R}^{d \times k} \implies Y_{n \times k} = X \cdot R \in \mathbb{R}^{n \times k}$$

where R is the RP matrix and Y represents the projected data in the lower-dimensional space. This formulation highlights the role of RPs in reducing dimensionality while preserving the geometric structure of the data. It provides an efficient and theoretically supported method for processing high-dimensional datasets [113].

3.3.2 Plus-Minus One Random Projection (PMRP)

PMRPs are based on the idea of using a random matrix $\mathbf{R}$ where each entry r_{ij} is independently drawn from a Bernoulli distribution. Typically, each entry is +1 or -1 with equal probability [114]. A data point $\mathbf{X}$ in a high-dimensional space can be projected onto a lower-dimensional space given by $\mathbf{Y} = \mathbf{R}\mathbf{X}$. This special quality distinguishes the plus-minus one projection from other methods and adds to its effectiveness and simplicity [7]. The element selection procedure is binary, which is important for maintaining some features while lowering dimensionality. We define a kind of ordered randomness by assigning +1 or -1 to each entry in the projection matrix [114]. Plus/Minus One discrete element selection, which projection uses instead of continuous probability distributions, simplifies the procedure and can have favorable effects for particular kinds of data. With its members being either +1 or -1, the resulting projection matrix offers a special way to project high-dimensional data onto a lower-dimensional environment [84], [114]. This binary selection process often offers a lower computing overhead while maintaining critical information and lowering the dimensionality of the data compared to approaches that use continuous distributions [114].

Let R be the RP matrix of size $k \times d$ for reducing a d -dimensional data matrix $\mathbf{X}$ to k -dimensional space. Each element R_{ij} is drawn from a probability distribution with two values: +1 or -1, where:

$$R_{ij} = \begin{cases} +1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$$

3.3.3 Normal Random Projection

A common technique to reduce dimensionality is normal or Gaussian RPs, which is very helpful when handling highly dimensional data [40]. The fundamental concept is to use a random matrix to project the data onto a lower-dimensional subspace while preserving as many pairwise distances between points as possible [115]. In practice, a Gaussian random matrix can be used to create NRPs. Two fundamental parameters, the variance and the mean, define this distribution. Each component of the RP matrix is taken from a normal distribution, where the mean is zero and the variance is one when using a normal projection. Let us examine the implications of these parameters: Mean($\mu = 0$): The average or central value around which the random values are clustered is shown by the mean of the normal distribution [116], [117]. The numbers obtained from the normal distribution are symmetrically distributed about zero in this instance, where the mean is 0. Since the distribution is concentrated at the origin, this mean selection does not add any systematic bias. Variance ($\sigma^2 = 1$): This quantity expresses how widely distributed the random values are. This method ensures that the resulting lower-dimensional representation retains significant structures while lowering the dimensionality of the data [117].

To reduce a data matrix of p -dimension to a space of k -dimension, define R as a $k \times p$ projection matrix. Each R_{ij} element is derived separately from a normal distribution.

$$R_{ij} \sim \mathcal{N}(0, 1)$$

3.4 Formulation of the NHRP Method

This study introduces the Nonlinear Hybrid RP (NHRP) method. The method aims to capture the complex, nonlinear relationships in high-dimensional data by integrating nonlinear RPs with the strengths of HRP. This approach not only aims to preserve computational efficiency, but also establishes a balance between the preservation of distances and interpretability, making it adaptable to different data properties and analytical tasks. The proposed NHRPs are expressed as follows:

Polynomial Approach

$$H_1 = \alpha^2(R_1 \circ R_1) + 2\alpha(1 - \alpha)(R_1 \circ R_2) + (1 - \alpha)^2(R_2 \circ R_2) \quad (3.3)$$

where $\circ$ denotes the Hadamard (element-wise) product, and $0 \leq \alpha \leq 1$.

Trigonometric Approach

$$H_2 = \alpha \sin(R_1) + (1 - \alpha) \sin(R_2) \quad (3.4)$$

Note:

$$[\sin(R_1)]_{ij} = \sin((R_1)_{ij}), [\sin(R_2)]_{ij} = \sin((R_2)_{ij})$$

In the above equations, the R_1 and R_2 represent the matrices derived from two distinct random projection methods: NRP and PMRP and their coefficients sum up to one. The coefficients of R_1 and R_2 and the parameter α , which varies between 0 and 1, are essential to decide how to combine various approaches in the hybrid model. The optimal value of α can be determined through an optimization process that seeks to minimize the distortion of distances caused by projection.

In equation 3.4, each element of the resulting matrix H_2 is computed by taking the sine values corresponding to elements from R_1 and R_2 , multiplying them by α and $(1 - \alpha)$ respectively, and then adding the results together. This is a common technique in matrix operations and is useful when dealing with component-wise transformations.

We implemented a hybrid random projection, each integrating two individual random projection methods. Specifically, these hybrid projections consist of the combinations of NRP and PMRP.

3.4.1 Determining the Optimal Value of α

The ideal value of α is obtained by minimizing the disparity between the distances in the original high-dimensional space and the projected low-dimensional space. More formally, two data points $x, y \in X$ are provided; we define a distance preservation ratio as follows:

$$s = \frac{\|xH - yH\|_2}{\|x - y\|_2}$$

Our objective is to find α that minimizes the mean and variance of the distance preservation ratio, denoted $E[s]$. This can be formulated as:

$$\alpha^* = \arg \min_{\alpha} E[s]$$

Additionally, we can also consider minimizing the variance; i.e.

$$\alpha^* = \arg \min_{\alpha} Var[s]$$

considering that H is a nonlinear combination of R_1 and R_2 with the coefficient α , $E[s]$

is determined by minimizing α using standard optimization techniques. It is possible to address this optimization issue computationally by looking at the whole range of α (i.e., $[0, 1]$) and choosing the value that provides the lowest $E[s]$.

Algorithm 1 outlines the proposed framework for the NHRP and the optimization of the parameter α . In this method, $\mathbf{H}$ is formulated as a nonlinear combination of $\mathbf{R}_1$ and $\mathbf{R}_2$, weighted by the coefficient α . The function $E[s]$, which quantifies distortion, is dependent on α . The objective is to identify the optimal value of α that minimizes the distortion of dimensionality reduction while ensuring that the NHRP effectively preserves the geometric structure of the original data set. To address this optimization problem, a numerical strategy is employed, whereby α is systematically varied within the range $[0, 1]$. For each candidate value of α , the NHRP is applied to project the dataset into a lower-dimensional space. The performance of the projection is evaluated by calculating the average distortion, defined as the mean ratio of pairwise distances in the original high-dimensional space to those in the projected space. This metric serves as a quantitative measure of the adherence of the NHRP to the JL lemma [2]. The optimisation process is implemented using the `optim` function in the R programming environment, which facilitates the application of various optimization techniques. Specifically, the Limited-Memory Broyden–Fletcher–Goldfarb–Shanno algorithm with bounds (L-BFGS-B) is utilised due to its suitability for problems with bounded variables, such as the optimisation of α within $[0, 1]$ [118]. The L-BFGS-B algorithm efficiently identifies the optimal value of α that minimises $E[s]$, ensuring both computational efficiency and compliance with the specified constraints [118]. This framework provides a robust approach to optimise the NHRP method, balancing dimensionality reduction and distortion minimisation. By leveraging the theoretical foundation of random projections and employing advanced numerical optimisation techniques, this study

contributes a novel methodology to the field of dimensionality reduction in high-dimensional data analysis.

Algorithm 1 The proposed framework of the NHRP for dimensionality reduction with optimized α

- 1: **Input:** $\mathbf{X}$: Input data matrix of size $(n \times d)$; k : Target dimensionality after reduction
 - 2: **Output:** $\mathbf{Y}$: Reduced data matrix of size $(n \times k)$
 - 3: Generate two random projection matrices $\mathbf{R}_1$ and $\mathbf{R}_2$ of size $(d \times k)$ with each element drawn independently from any desired distribution;
 - 4: Set a range for the parameter α (e.g., $[0, 1]$);
 - 5: **for** each value of α **do**
 - Generate the hybrid random projection matrix $\mathbf{H}$ using any equation from equation (3-6);
 - Project the input data matrix $\mathbf{X}$ onto the random projection matrix $\mathbf{H}$ to obtain the reduced data matrix $\mathbf{Y}$: $\mathbf{Y} = \mathbf{X} \cdot \mathbf{H}$;
 - Evaluate the average distortion in pairwise distances, denoted as $E[s]$, using the distance preservation ratio equation;
 - 6: **end for**
 - 7: Select the optimal α^* that minimizes $E[s]$;
 - 8: Obtain $\mathbf{H}$ using α^* , $\mathbf{R}_1$, and $\mathbf{R}_2$;
 - 9: Project $\mathbf{X}$ onto $\mathbf{H}$ to obtain $\mathbf{Y}$: $\mathbf{Y} = \mathbf{X} \cdot \mathbf{H}$;
 - 10: **Return:** $\mathbf{Y}$
-

3.4.2 Supporting Lemmas

The following lemmas play a critical role in the development and analysis of the proposed NHRP method. They establish the mathematical framework that supports the methodology, ensuring the effective preservation of pairwise distances during dimensionality reduction. By including detailed proofs for these lemmas, we aim to provide a comprehensive understanding of the theoretical foundation of the NHRP approaches, offering insights into its mechanics and reinforcing its reliability in high-dimensional data analysis. The polynomial NHRP lemmas are shown in Appendix A.

The following are proofs of the lemmas of the trigonometric NHRP approach to provide a complete understanding of the theoretical foundation.

Lemma 1: For any pair of points $x, y \in X$, and for any $\alpha \in [0, 1]$, the Euclidean distance between the projections of x and y using the hybrid projection H_2 is less than or equal to the sum of the distances of the projections of x and y using the matrices R_1 and R_2 , respectively.

$$H_2 = \alpha \sin(R_1) + (1 - \alpha) \sin(R_2)$$

as defined in Equation 3.4.

Proof: Let H_2 (Equation 3.4) be the trigonometric NHRP, and let x, y be any two points in X . The Euclidean distance between their projections onto H_2 is given by:

$$\|x_{H_2} - y_{H_2}\|_2$$

Expanding this using the definition of H_2 :

$$\|x_{H_2} - y_{H_2}\|_2 = \|x(\alpha \sin(R_1) + (1 - \alpha) \sin(R_2)) - y(\alpha \sin(R_1) + (1 - \alpha) \sin(R_2))\|_2$$

$$= \|\alpha x \sin(R_1) + (1 - \alpha)x \sin(R_2) - \alpha y \sin(R_1) - (1 - \alpha)y \sin(R_2)\|_2$$

$$= \|\alpha(x \sin(R_1) - y \sin(R_1)) + (1 - \alpha)(x \sin(R_2) - y \sin(R_2))\|_2$$

By the **triangle inequality**, we obtain:

$$\|x_{H_2} - y_{H_2}\|_2 \leq \alpha \|x \sin(R_1) - y \sin(R_1)\|_2 + (1 - \alpha) \|x \sin(R_2) - y \sin(R_2)\|_2$$

Thus, we conclude that:

$$\|x_{H_2} - y_{H_2}\|_2 \leq \|x_{R_1} - y_{R_1}\|_2 + \|x_{R_2} - y_{R_2}\|_2$$

Lemma 2: Distance Preservation of Hybrid Projections

Let X be a high-dimensional dataset in $\mathbb{R}^d$, and let R_1 and R_2 be two different random projection (RP) matrices constructed using two distinct RP methods. Let $\alpha \in [0, 1]$. If we define H_2 as a hybrid projection matrix that is a nonlinear combination of R_1 and R_2 , given by:

$$H_2 = \alpha \sin(R_1) + (1 - \alpha) \sin(R_2),$$

then for any two data points $x, y \in X$, the following holds:

$$\frac{\|x_{H_2} - y_{H_2}\|_2}{\|x - y\|_2} \approx 1.$$

Proof:

We start with the following expression:

$$\|xH_2 - yH_2\|_2 = \|x(\alpha \sin R_1 + (1 - \alpha) \sin R_2) - y(\alpha \sin R_1 + (1 - \alpha) \sin R_2)\|_2.$$

Expanding this, we get:

$$\|xH_2 - yH_2\|_2 = \|\alpha x \sin R_1 + (1 - \alpha)x \sin R_2 - \alpha y \sin R_1 - (1 - \alpha)y \sin R_2\|_2.$$

Regrouping the terms:

$$\|xH_2 - yH_2\|_2 = \|\alpha(x \sin R_1 - y \sin R_1) + (1 - \alpha)(x \sin R_2 - y \sin R_2)\|_2.$$

By the triangle inequality:

$$\|xH_2 - yH_2\|_2 \leq \|\alpha(x \sin R_1 - y \sin R_1)\|_2 + \|(1 - \alpha)(x \sin R_2 - y \sin R_2)\|_2.$$

Applying the properties of norms:

$$\|xH_2 - yH_2\|_2 \leq \alpha\|x \sin R_1 - y \sin R_1\|_2 + (1 - \alpha)\|x \sin R_2 - y \sin R_2\|_2.$$

Since R_1 and R_2 are distance-preserving RPs, we can approximate:

$$\|x \sin R_1 - y \sin R_1\|_2 \approx \|x - y\|_2, \quad \|x \sin R_2 - y \sin R_2\|_2 \approx \|x - y\|_2.$$

Thus, we have:

$$\|xH_2 - yH_2\|_2 \leq \alpha\|x - y\|_2 + (1 - \alpha)\|x - y\|_2.$$

$$\|xH_2 - yH_2\|_2 \leq (\alpha + 1 - \alpha)\|x - y\|_2.$$

$$\|xH_2 - yH_2\|_2 \leq \|x - y\|_2.$$

Therefore, the ratio of the distances is approximately 1:

$$\frac{\|xH_2 - yH_2\|_2}{\|x - y\|_2} \approx 1.$$

This lemma highlights the ability of the hybrid projection matrix to preserve the relative distances between data points in a high-dimensional space when projected to a lower-dimensional space. It is essential to maintain the geometric structure of the dataset after dimensionality reduction. The assertion that

$$\frac{\|xH_2 - yH_2\|_2}{\|x - y\|_2} \approx 1$$

highlights the NHRP method's capability to preserve the geometric structure of data during dimensionality reduction. This ratio, ideally close to 1, indicates that the distances between data points remain consistent from the high-dimensional space to the projected space, aligning with the core principle of the JL lemma. The hybrid projection H_2 enables us to evaluate how the combination of two different projection matrices affects this distance preservation:

To further analyze, we expand:

$$\|xH_2 - yH_2\|_2 = \left[\begin{aligned} &\alpha^2 \|x \sin R_1 - y \sin R_1\|_2^2 \\ &+ (1 - \alpha)^2 \|x \sin R_2 - y \sin R_2\|_2^2 \\ &+ 2\alpha(1 - \alpha) \langle x \sin R_1 - y \sin R_1, x \sin R_2 - y \sin R_2 \rangle \end{aligned} \right]^{1/2}$$

Assuming that the projections R_1 and R_2 are independently scaled and their effects are approximately orthogonal due to their randomness, the cross-term is negligible. This simplifies to:

$$\|xH_2 - yH_2\|_2 \approx \sqrt{\alpha^2 \|x \sin R_1 - y \sin R_1\|_2^2 + (1 - \alpha)^2 \|x \sin R_2 - y \sin R_2\|_2^2}.$$

If both R_1 and R_2 approximately preserve distances according to the JL lemma, then:

$$\|xH_2 - yH_2\|_2 \approx \|x - y\|_2.$$

This validates the claim that the hybrid projection method retains the essential distance structure of the original dataset.

Lemma 3: Error Bound of Hybrid Projections

Given a high-dimensional dataset $X \subset \mathbb{R}^d$, two random projection (RP) matrices R_1 and R_2 , and a hybrid projection matrix $H_2 = \alpha \sin(R_1) + (1 - \alpha) \sin(R_2)$ for $\alpha \in [0, 1]$, the expected distortion of distances introduced by H_2 is bounded within a certain range dependent on α , the properties of R_1 and R_2 , and the dimensionality of the reduced space.

Proof: Let $x, y \in X$ be any two points. The distortion introduced by the hybrid

projection on the distance between x and y can be given by the difference:

$$\Delta = \left| \|x - y\|_2^2 - \|xH_2 - yH_2\|_2^2 \right|$$

Expanding H_2 , we have:

$$\Delta = \left| \|x - y\|_2^2 - \|\alpha \sin(xR_1 - yR_1) + (1 - \alpha) \sin(xR_2 - yR_2)\|_2^2 \right|$$

By the triangle inequality:

$$\Delta \leq \alpha \left| \|x - y\|_2^2 - \|x \sin R_1 - y \sin R_1\|_2^2 \right| + (1 - \alpha) \left| \|x - y\|_2^2 - \|x \sin R_2 - y \sin R_2\|_2^2 \right|$$

If we denote ϵ_1 and ϵ_2 as the maximum distortions introduced by R_1 and R_2 , respectively, we can bound Δ as follows:

$$\Delta \leq \alpha \epsilon_1 + (1 - \alpha) \epsilon_2$$

This shows that the distortion introduced by the hybrid projection is a weighted combination of the distortions of the individual projections, bounded by the maximum distortions of each projection.

Lemma 4: Sparsity Preservation

If R_1 and R_2 are RP matrices that preserve sparsity, then the hybrid projection matrix H_2 also approximately preserves the sparsity of the projected data.

Proof: Let $x \in X$ be a sparse vector. The sparsity of xH_2 can be analyzed by counting

the number of non-zero elements. Since H_2 is a nonlinear combination of R_1 and R_2 , defined as:

$$H_2 = \alpha \sin R_1 + (1 - \alpha) \sin R_2,$$

each element in xH_2 is derived from elements in xR_1 and xR_2 . If both R_1 and R_2 preserve sparsity, then the number of non-zero elements in xR_1 and xR_2 remains low. The nonlinear transformations $\sin R_1$ and $\sin R_2$ do not introduce a significant number of additional non-zero elements, and the weighted sum structure of H_2 ensures that the projection remains influenced by both matrices while maintaining sparsity.

Thus, the number of non-zero elements in xH_2 remains approximately close to that in xR_1 and xR_2 , implying that H_2 preserves the sparsity of x .

Furthermore, while Lemma 6 assumes that both R_1 and R_2 inherently preserve sparsity, the effect of H_2 is influenced by the interplay between $\sin R_1$ and $\sin R_2$. Even if R_1 and R_2 do not individually maintain sparsity, H_2 can still preserve sparsity under certain conditions. Specifically, the interaction between non-zero elements in xR_1 and xR_2 may lead to reductions in density due to structured transformations in H_2 . If R_1 is denser than R_2 , the term $(1 - \alpha) \sin R_2$ mitigates excessive density, thereby balancing sparsity. The key is to manage the contributions of both projection matrices such that H_2 does not significantly increase the density of xH_2 beyond what is observed in the original vector x . Thus, we conclude that H_2 approximately preserves the sparsity of the projected data.

Lemma 5: Computational Complexity Considering α

The computational complexity of applying the hybrid projection $H_2 = \alpha \sin R_1 + (1 - \alpha) \sin R_2$ to a dataset X depends not only on the dimensions of X and the properties of

R_1 and R_2 , but also on the computational cost of the nonlinear transformations and the optimization of α .

Proof: Consider a dataset $X \in \mathbb{R}^{n \times d}$, where n is the number of data points and d is the dimensionality. The computational complexity of applying H_2 involves three primary operations: matrix multiplications of X with R_1 and R_2 , element-wise sine transformations, and the linear combination of these results weighted by α and $(1 - \alpha)$, respectively.

1. **Matrix Multiplications:** Assuming R_1 and R_2 have dimensions $d \times k$, where k is the reduced dimensionality, the complexity of multiplying X with each of R_1 and R_2 is $O(ndk)$.

2. **Nonlinear Transformations:** The application of the sine function to the elements of XR_1 and XR_2 requires element-wise computations, contributing an additional $O(nk)$ complexity.

3. **Optimization of α :** The selection or optimization of α may introduce further computational overhead. If α is determined through a data-driven approach such as cross-validation or grid search, this adds to the total complexity. The complexity of optimizing α depends on the method used and can range from a constant-time operation (if α is predefined) to more complex iterative procedures.

Therefore, the total computational complexity is the sum of these three components. The dominant term remains the matrix multiplications $O(ndk)$, with an additional $O(nk)$ contribution from the nonlinear transformations. The optimization of α further influences the overall complexity depending on the approach used. Thus, the overall complexity of applying H_2 is a function of n , d , k , and the computational cost of optimizing α .

Lemma 6: Equivalence of Hybrid Projection Matrices with Optimal α

Let R_1 and R_2 be two RP matrices that are statistically similar in terms of distance

preservation. Define two hybrid projection matrices:

$$H_2 = \alpha \sin R_1 + (1 - \alpha) \sin R_2$$

$$H'_2 = \alpha' \sin R_2 + (1 - \alpha') \sin R_1.$$

If α^* is the optimal value minimizing the distance distortion for H_2 and α'^* is the optimal value for H'_2 , then $\alpha^* = 1 - \alpha'^*$, and the matrices H_2 and H'_2 are equivalent in their effectiveness for dimensionality reduction.

Proof:

1. **Distance Preservation Functions:** The distance preservation ratio for two data points $x, y \in X$ is given as

$$s = \frac{\|xH_2 - yH_2\|_2}{\|x - y\|_2}$$

and similarly for s' using H'_2 . The expectation of these ratios, $E[s]$ and $E[s']$, depend on α and α' , respectively. Since H_2 and H'_2 are nonlinear combinations of R_1 and R_2 , these expectations are functions of the corresponding alpha values.

2. **Symmetry in Distance Preservation:** Assuming R_1 and R_2 preserve distances similarly, the forms of $E[s]$ and $E[s']$ would be symmetric. This is because reversing the roles of R_1 and R_2 in the hybrid projection matrices does not alter the underlying statistical properties of the distance preservation. Therefore, the minimization problem for $E[s]$ with respect to α mirrors the minimization problem for $E[s']$ with respect to α' .

3. **Optimization:** Due to the symmetric nature of $E[s]$ and $E[s']$, the optimal values will satisfy the relationship

$$\alpha^* = 1 - \alpha'^*.$$

This relationship arises from the complementary roles of R_1 and R_2 in the two hybrid matrices.

4. **Equivalence Demonstration:** Under the optimal values α^* and α'^* , both H_2 and H'_2 achieve similar performance in terms of distance preservation. This implies that, for the optimal alpha values, switching the roles of R_1 and R_2 in the hybrid projection matrix does not affect its effectiveness in reducing dimensions while preserving the structure of the data. Hence, H_2 and H'_2 are equivalent in terms of their dimensionality reduction capabilities.

3.5 Simulation Study and Results

This section presents the simulation study and its findings, which compare the proposed NHRP technique with the HRP method and other traditional approaches.

3.5.1 Simulation Study

We thoroughly evaluate four methods for handling high-dimensional data: the classical RP methods (NRP and PMRP), the HRP method, and the proposed NHRP method. Our goal is to assess their performance, particularly focusing on the NHRP method. The simulation investigates how different dimensionality reduction strategies affect the preservation of pairwise distances in the projected data. We primarily analyze sparse datasets with varying proportions of non-zero values.

The evaluation metrics are based on the average distortion between pairwise distances in the original high-dimensional space and those in the projected lower-dimensional space.

Lower distortion values indicate better adherence to the JL lemma, signifying superior preservation of the original data distances after projection.

To comprehensively understand each method’s performance, we systematically vary several parameters in the simulation:

- **Sample Size (n):** We consider datasets with 50, 200, and 400 samples.
- **Data Dimensionality (d):** We test true dimensionalities of 500, 1000, 5000, and 10000.
- **Subspace Dimensionality (ϵ):** The projected dimension (k) is varied at three levels corresponding to ϵ values of 0.6, 0.8, and 0.9. For example, with an ϵ value of 0.6, an initial feature dimensionality of 500, and a sample size of 50, the projected data has 230 features. Similarly, for ϵ values of 0.8 and 0.9, the projected data have 166 and 153 features, respectively.
- **Sparse Data Density (δ):** We assess the impact of different proportions of non-zero values, considering densities of 20%, 60%, 80%, and 100%.

The simulation involves generating random matrices X with n samples and d features. Each method is applied to X to create a projected dataset Y , and the distortion between the distances in the original and projected spaces is computed and recorded across 1000 simulations. This systematic variation in sample size, data dimensionality, subspace dimensionality, and projection methods enables a thorough assessment of dimension-reduction techniques in diverse scenarios, providing valuable insights into their relative strengths, with a particular focus on the NHRP method.

3.5.2 Simulation Results

The simulation results in Tables 3.1 and 3.2 and Figure 3.1 and Figure 3.2 highlight the effectiveness of the NHRP methods compared to traditional RP techniques, such as NRP and PMRP, as well as the HRP method, in different scenarios. This section explores the significant findings and their implications.

- **Impact of Sample Size (n):** An analysis of **Table 3.1** reveals that increasing the sample size from $n = 50$ to $n = 400$, along with higher values of ϵ , generally leads to a reduction in mean distortion. However, the extent of this reduction varies across different projection methods.

For standard RP techniques, such as NRP, the distortion values remain constant across all sample sizes and δ values when $\epsilon = 0.6$, with NRP consistently reporting a mean distortion of 15.15 while PMRP, the distortion values remain relatively stable, they exhibit slight variations as the sample size increases. Specifically, at $\epsilon = 0.6$, the distortion is **10.71** for $n = 50$, **10.71** for $n = 200$, and **10.73** for $n = 400$. Although these differences are small, they suggest that HRP exhibits minimal sensitivity to sample size changes compared to more dynamic projection methods. This stability suggests that these methods do not exhibit sensitivity to sample size variations under the given conditions.

In contrast, the NHRP methods [polynomial-based (P_NHRP) and trigonometric-based (T_NHRP)] demonstrate notable improvements as the sample size increases. Specifically, the polynomial function achieves a reduction in distortion from 8.65 (SE = 0.057) at $n = 50$ to 8.63 (SE = 0.021) at $n = 400$ for $\epsilon = 0.8$ and $\delta = 0.2$. A similar trend is observed for the trigonometric function, which achieves lower distortion values

compared to the polynomial function. For example, at $\epsilon = 0.6$ and $\delta = 0.2$, the trigonometric function reduces distortion from 7.85 (SE = 0.005) at $n = 50$ to 7.82 (SE = 0.002) at $n = 400$. This consistency reinforces its effectiveness in preserving the structural properties of high-dimensional data.

These findings highlight the advantage of NHRP-based approaches, particularly under larger sample sizes, where they outperform conventional RP techniques in minimizing distortion. The significant reduction in distortion observed for NHRP suggests that the structured transformations embedded in these hybrid approaches enhance projection efficiency, especially in scenarios requiring robust distance preservation. The stability of standard RP methods and the sensitivity of NHRP to increasing sample sizes provide important insights into their respective applicability, with NHRP being particularly beneficial for large-scale datasets where distortion minimization is crucial.

- NHRP's effectiveness:** The consistency of maintaining pairwise distances during dimensionality reduction is demonstrated by the consistently reduced average distortions it exhibits compared to PMRP and NRP over a range of sample sizes, data dimensionality, and sparse data densities. Table 3.1, for example, shows that when $d = 500$, $n = 50$, and $\delta = 0.2$ are used, the mean distortions (standard errors) for NRP, PMRP, HRP, P_NHRP and T_NHRP are 15.15 (0.01), 15.15 (0.01), 10.71 (0.007), 10.17 (0.069) and 7.85 (0.005) respectively. As ϵ changes, these values adjust: for $\epsilon = 0.8$, they are 12.86 (0.011), 12.87 (0.01), 9.09 (0.007), 8.65 (0.057) and 6.66 (0.005); for $\epsilon = 0.9$, they are 12.35 (0.01), 12.35 (0.011), 8.73 (0.007), 8.33 (0.055) and 6.4 (0.005).
- Impact of Subspace Dimensionality (ϵ):** The simulation results demonstrate that

increasing ϵ generally leads to a reduction in average distortion, particularly for the NHRP methods. For instance, at $n = 50$ and $\delta = 0.2$, the mean distortion of the polynomial-based NHRP decreases from 10.17 (SE = 0.069) at $\epsilon = 0.6$ to 8.33 (SE = 0.055) at $\epsilon = 0.9$. A similar trend is observed in the T_NHRP, where the distortion reduces from 7.85 (SE = 0.005) to 6.4 (SE = 0.005) under the same conditions. These findings highlight that the structured transformations in NHRP provide enhanced adaptability to varying subspace dimensionalities, making them particularly effective in minimizing distortion while preserving the original data structure.

Compared to NRP and PMRP, HRP and NHRP exhibits a more pronounced decrease in distortion as ϵ increases, highlighting its efficiency in dimensionality reduction while preserving structural properties. For example, at $n = 200$ and $\delta = 0.6$, HRP maintains a lower distortion of 10.71 (SE = 0.004), whereas NRP and PMRP remain at 15.15 (SE = 0.005), showing minimal variation.

Furthermore, NHRP-based methods, particularly the trigonometric function, demonstrate even greater adaptability to changes in ϵ . The T_NHRP achieves a distortion of 6.39 (SE = 0.006) at $\epsilon = 0.9$, outperforming other techniques. These results underscore the advantage of HRP and NHRP in efficiently preserving pairwise distances across varying subspace dimensionalities.

- **Performance Consistency:** The results demonstrate that NHRP consistently outperforms PMRP, NRP, and HRP in reducing average distortion across different sample sizes (n), feature dimensions (d), and subspace dimensionalities (ϵ). For instance, at $n = 50$ and $\delta = 0.2$, the T_NHRP achieves the lowest mean distortion of 6.66 (SE = 0.005) at $\epsilon = 0.6$, significantly lower than HRP (9.09, SE = 0.007) and PMRP (12.87, SE = 0.012). This trend remains consistent across larger sample sizes, with

the T_NHRP maintaining its advantage at $n = 200$ and $n = 400$. The P_NHRP also exhibits steady improvements, further reinforcing the method’s robustness.

Overall, NHRP demonstrates remarkable stability and adaptability in preserving pairwise distances while minimizing distortion, making it a superior alternative for high-dimensional data applications.

Practical Implications: The simulation findings highlight the practical advantages of NHRP in real-world applications requiring dimensionality reduction. NHRP consistently preserves pairwise distances and maintains the structural integrity of high-dimensional data, making it highly suitable for tasks such as pattern recognition, clustering, and classification. Its ability to achieve lower distortion across varying sample sizes and subspace dimensionalities suggests that NHRP can effectively enhance the performance of machine learning models in domains where maintaining data geometry is crucial.

Beyond the primary findings discussed earlier, additional insights into the computational efficiency and scalability of the RP methods are provided in Supplementary Tables 1 and 2, which detail the computational time associated with each approach. Furthermore, Supplementary Tables 3–6 present the optimal values of α for the NHRP methods, offering a deeper understanding of their parameter optimization. Lastly, Supplementary Tables 7 and 8 report the 95% confidence intervals (CI) corresponding to the results in Tables 3.1 and 3.2, respectively, ensuring the statistical reliability and robustness of the findings. Additionally, Supplementary Figures 1 and 2 illustrate the effectiveness of the NHRP method across varying values of epsilon, providing a comprehensive evaluation of its performance under different parameter settings.

Table 3.1: Standard Error (SE) and Mean Distortions for each of the four random projection (RP) techniques utilised with dimension, $d=500$, across various sample sizes (n), across different densities (δ) and epsilon (ϵ).

		δ	0.2	0.6	0.8	1
n	epsilon	Function_Type	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)
50	0.6	NRP	15.15 (0.01)	15.15 (0.01)	15.15 (0.009)	15.15 (0.011)
		PMRP	15.15 (0.01)	15.16 (0.01)	15.17 (0.009)	15.16 (0.009)
		HRP	10.71 (0.007)	10.71 (0.007)	10.72 (0.007)	10.71 (0.007)
		P_NHRP	10.17 (0.069)	10.16 (0.066)	10.11 (0.074)	10.11 (0.063)
		T_NHRP	7.85 (0.005)	7.84 (0.005)	7.85 (0.005)	7.85 (0.005)
	0.8	NRP	12.86 (0.011)	12.89 (0.012)	12.87 (0.01)	12.87 (0.01)
		PMRP	12.87 (0.01)	12.87 (0.01)	12.88 (0.01)	12.85 (0.01)
		HRP	9.09 (0.007)	9.11 (0.008)	9.11 (0.007)	9.1 (0.007)
		P_NHRP	8.65 (0.057)	8.64 (0.058)	8.62 (0.064)	8.56 (0.06)
		T_NHRP	6.66 (0.005)	6.67 (0.006)	6.67 (0.005)	6.66 (0.005)
	0.9	NRP	12.35 (0.01)	12.35 (0.011)	12.34 (0.01)	12.34 (0.01)
		PMRP	12.35 (0.011)	12.36 (0.01)	12.36 (0.011)	12.35 (0.011)
		HRP	8.73 (0.007)	8.72 (0.009)	8.73 (0.008)	8.73 (0.008)
		P_NHRP	8.33 (0.055)	8.26 (0.054)	8.27 (0.061)	8.23 (0.059)
		T_NHRP	6.4 (0.005)	6.39 (0.006)	6.4 (0.005)	6.4 (0.005)
200	0.6	NRP	15.15 (0.006)	15.15 (0.006)	15.14 (0.006)	15.15 (0.006)
		PMRP	15.15 (0.005)	15.15 (0.005)	15.15 (0.005)	15.16 (0.006)
		HRP	10.71 (0.004)	10.71 (0.004)	10.71 (0.004)	10.71 (0.004)
		P_NHRP	10.15 (0.038)	10.17 (0.032)	10.06 (0.034)	10.19 (0.038)
		T_NHRP	7.85 (0.003)	7.85 (0.003)	7.85 (0.003)	7.85 (0.003)
	0.8	NRP	12.85 (0.006)	12.87 (0.006)	12.86 (0.006)	12.86 (0.006)
		PMRP	12.87 (0.005)	12.87 (0.004)	12.87 (0.004)	12.86 (0.005)
		HRP	9.09 (0.004)	9.1 (0.004)	9.1 (0.004)	9.1 (0.004)
		P_NHRP	8.63 (0.033)	8.57 (0.029)	8.58 (0.03)	8.55 (0.032)
		T_NHRP	6.66 (0.003)	6.67 (0.003)	6.66 (0.003)	6.66 (0.003)
	0.9	NRP	12.35 (0.006)	12.36 (0.007)	12.36 (0.006)	12.36 (0.006)
		PMRP	12.35 (0.005)	12.35 (0.005)	12.35 (0.005)	12.35 (0.005)
		HRP	8.73 (0.004)	8.73 (0.004)	8.73 (0.004)	8.74 (0.004)
		P_NHRP	8.28 (0.032)	8.21 (0.029)	8.26 (0.028)	8.23 (0.031)
		T_NHRP	6.4 (0.003)	6.4 (0.003)	6.4 (0.003)	6.4 (0.003)
400	0.6	NRP	15.15 (0.004)	15.16 (0.005)	15.16 (0.005)	15.15 (0.004)
		PMRP	15.15 (0.004)	15.15 (0.004)	15.15 (0.004)	15.15 (0.003)
		HRP	10.71 (0.003)	10.72 (0.003)	10.71 (0.003)	10.71 (0.003)
		P_NHRP	10.14 (0.025)	10.11 (0.023)	10.12 (0.026)	10.13 (0.024)
		T_NHRP	7.82 (0.002)	7.85 (0.002)	7.85 (0.002)	7.85 (0.002)
	0.8	NRP	12.88 (0.005)	12.86 (0.005)	12.87 (0.005)	12.86 (0.004)
		PMRP	12.87 (0.003)	12.86 (0.003)	12.86 (0.003)	12.86 (0.004)
		HRP	9.1 (0.003)	9.1 (0.003)	9.1 (0.003)	9.09 (0.003)
		P_NHRP	8.63 (0.021)	8.6 (0.018)	8.6 (0.019)	8.61 (0.022)
		T_NHRP	6.67 (0.002)	6.67 (0.002)	6.67 (0.002)	6.66 (0.002)
	0.9	NRP	12.35 (0.005)	12.35 (0.004)	12.36 (0.004)	12.35 (0.005)
		PMRP	12.35 (0.004)	12.35 (0.004)	12.35 (0.004)	12.35 (0.004)
		HRP	8.74 (0.003)	8.73 (0.003)	8.73 (0.004)	8.74 (0.003)
		P_NHRP	8.27 (0.02)	8.27 (0.017)	8.27 (0.02)	8.27 (0.021)
		T_NHRP	6.4 (0.002)	6.4 (0.002)	6.4 (0.002)	6.4 (0.002)

NRP = Normal Random Projection; PMRP = Plus-Minus One Random Projection; HRP = Hybrid Random Projection; P_NHRP = Polynomial-based Nonlinear Hybrid Random Projection; T_NHRP = Trigonometric-based Nonlinear Hybrid Random Projection.

Table 3.2: **Standard Error (SE) and Mean Distortions for each of the four random projection (RP) techniques utilised with dimension, $d=1000$, across various sample sizes (n), across different densities (δ) and epsilon (ϵ).**

		δ	0.2	0.6	0.8	1
n	epsilon	Function_Type	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)
50	0.6	NRP	15.95 (0.009)	15.94 (0.012)	15.94 (0.01)	15.94 (0.01)
		PMRP	15.95 (0.01)	15.96 (0.01)	15.95 (0.01)	15.94 (0.01)
		HRP	11.28 (0.007)	11.27 (0.008)	11.28 (0.006)	11.27 (0.007)
		P_NHRP	10.71 (0.071)	10.65 (0.064)	10.78 (0.069)	10.59 (0.076)
		T_NHRP	8.26 (0.005)	8.25 (0.006)	8.26 (0.005)	8.26 (0.006)
	0.8	NRP	13.59 (0.009)	13.58 (0.009)	13.6 (0.009)	13.58 (0.01)
		PMRP	13.58 (0.012)	13.59 (0.011)	13.59 (0.011)	13.59 (0.011)
		HRP	9.61 (0.009)	9.62 (0.007)	9.61 (0.007)	9.61 (0.007)
		P_NHRP	9.09 (0.061)	9.08 (0.06)	9.05 (0.07)	8.98 (0.065)
		T_NHRP	7.04 (0.006)	7.04 (0.005)	7.04 (0.005)	7.04 (0.005)
	0.9	NRP	13.03 (0.01)	13.02 (0.01)	13.02 (0.011)	13 (0.011)
		PMRP	13.02 (0.01)	13.02 (0.009)	13 (0.01)	13.01 (0.008)
		HRP	9.2 (0.007)	9.22 (0.008)	9.2 (0.006)	9.19 (0.007)
		P_NHRP	8.7 (0.058)	8.69 (0.056)	8.68 (0.068)	8.58 (0.058)
		T_NHRP	6.74 (0.006)	6.75 (0.006)	6.74 (0.005)	6.73 (0.005)
200	0.6	NRP	15.96 (0.006)	15.95 (0.005)	15.95 (0.005)	15.94 (0.006)
		PMRP	15.96 (0.005)	15.96 (0.005)	15.95 (0.005)	15.94 (0.005)
		HRP	11.28 (0.004)	11.28 (0.003)	11.28 (0.004)	11.27 (0.004)
		P_NHRP	10.7 (0.039)	10.64 (0.044)	10.66 (0.035)	10.68 (0.036)
		T_NHRP	8.27 (0.003)	8.26 (0.003)	8.27 (0.003)	8.26 (0.003)
	0.8	NRP	13.58 (0.005)	13.6 (0.005)	13.59 (0.005)	13.59 (0.005)
		PMRP	13.59 (0.005)	13.59 (0.004)	13.58 (0.004)	13.58 (0.005)
		HRP	9.6 (0.004)	9.61 (0.004)	9.6 (0.004)	9.61 (0.004)
		P_NHRP	9.09 (0.033)	9.11 (0.034)	9.02 (0.028)	9.08 (0.03)
		T_NHRP	7.04 (0.003)	7.04 (0.003)	7.04 (0.003)	7.04 (0.003)
	0.9	NRP	13.01 (0.005)	13.02 (0.006)	13.01 (0.005)	13.02 (0.005)
		PMRP	13.02 (0.005)	13.02 (0.005)	13.02 (0.005)	13.03 (0.006)
		HRP	9.2 (0.004)	9.2 (0.004)	9.2 (0.004)	9.21 (0.004)
		P_NHRP	8.7 (0.03)	8.72 (0.033)	8.66 (0.028)	8.7 (0.028)
		T_NHRP	6.74 (0.003)	6.74 (0.003)	6.74 (0.003)	6.75 (0.003)
400	0.6	NRP	15.95 (0.005)	15.96 (0.004)	15.95 (0.004)	15.95 (0.005)
		PMRP	15.95 (0.003)	15.96 (0.004)	15.96 (0.004)	15.96 (0.003)
		HRP	11.28 (0.003)	11.29 (0.003)	11.28 (0.003)	11.28 (0.003)
		P_NHRP	10.64 (0.024)	10.69 (0.024)	10.65 (0.024)	10.65 (0.025)
		T_NHRP	8.26 (0.002)	8.27 (0.002)	8.27 (0.002)	8.26 (0.002)
	0.8	NRP	13.58 (0.005)	13.58 (0.004)	13.58 (0.004)	13.59 (0.004)
		PMRP	13.58 (0.004)	13.58 (0.004)	13.59 (0.004)	13.58 (0.004)
		HRP	9.6 (0.003)	9.6 (0.003)	9.61 (0.003)	9.61 (0.003)
		P_NHRP	9.05 (0.024)	9.09 (0.021)	9.09 (0.02)	9.04 (0.022)
		T_NHRP	7.04 (0.002)	7.04 (0.002)	7.04 (0.002)	7.04 (0.002)
	0.9	NRP	13.02 (0.004)	13.02 (0.004)	13.02 (0.004)	13.02 (0.004)
		PMRP	13.02 (0.003)	13.01 (0.003)	13.02 (0.003)	13.02 (0.004)
		HRP	9.2 (0.003)	9.2 (0.003)	9.2 (0.003)	9.2 (0.003)
		P_NHRP	8.69 (0.023)	8.73 (0.021)	8.71 (0.019)	8.65 (0.021)
		T_NHRP	6.74 (0.002)	6.74 (0.002)	6.74 (0.002)	6.74 (0.003)

NRP = Normal Random Projection; PMRP = Plus-Minus One Random Projection; HRP = Hybrid Random Projection; P_NHRP = Polynomial-based Nonlinear Hybrid Random Projection; T_NHRP = Trigonometric-based Nonlinear Hybrid Random Projection.

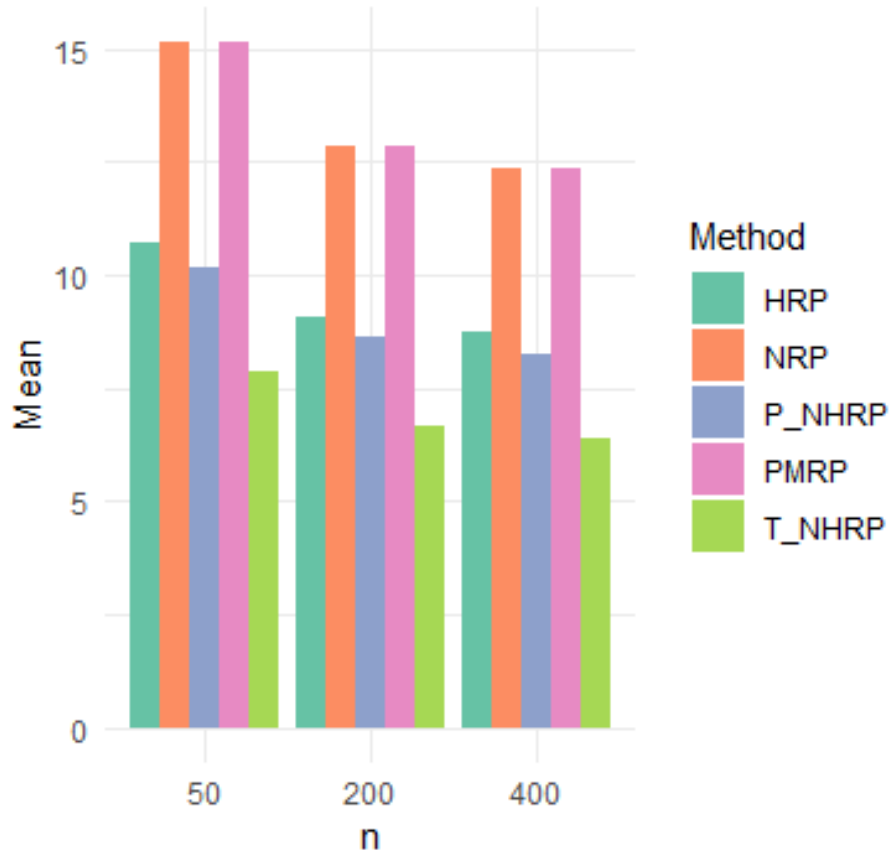


Figure 3.1: Comparison of Methods across Different n Values for $d = 500$.

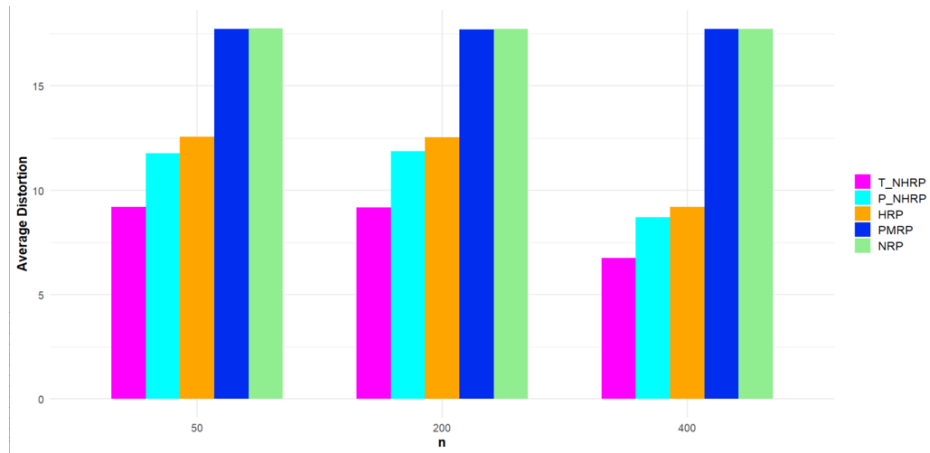


Figure 3.2: Comparison of Methods across Different n Values for $d = 5000$.

CHAPTER 4

APPLICATIONS OF THE DIMENSIONALITY REDUCTION METHODS ON MACHINE LEARNING DATASETS

4.1 Description of the Sample Datasets

We evaluated the performance of the proposed NHRP method through a comparative analysis with conventional NRP, PMRP, and HRP techniques. This evaluation is carried out using five well-established datasets sourced from the Machine Learning Repository and the Modified National Institute of Standards and Technology (MNIST) database. The datasets include toxicity data, period changer data, human activity recognition (HAR) data, Fashion MNIST, and Digit MNIST. These datasets are chosen because of their significance in the machine learning community and the diverse challenges they pose for dimensionality

reduction and data analysis.

The NHRP method, along with NRP, PMRP, and HRP, is applied to these datasets for dimensionality reduction. The effectiveness of each approach is then assessed based on the distortion ratio, measuring the discrepancy between the original and projected data in both the training and testing sets.

- I. **Period Changer Dataset:** This dataset, introduced in [119], focuses on the interaction of non-toxic molecules with the core clock protein CRY1. It consists of 90 molecules, each represented by 1177 unique molecular descriptors. Given its high-dimensional feature space, this dataset serves as a challenging benchmark for assessing dimensionality reduction techniques.
- II. **Toxicity Dataset:** This dataset, as described in [120], contains 171 molecular structures relevant to the CRY1 protein, which plays a critical role in circadian rhythm regulation. Each molecule is characterized by 1203 descriptors, and the dataset presents a binary classification problem, distinguishing 56 toxic molecules from 115 non-toxic ones. The complexity of its molecular structure makes it a suitable candidate for evaluating dimensionality reduction methods.
- III. **Human Activity Recognition (HAR) Dataset:** The HAR dataset [121] comprises sensor data collected from smartphones worn by 30 participants performing six distinct activities: walking, walking upstairs, walking downstairs, sitting, standing, and lying down. The dataset is widely used in activity recognition tasks and provides a challenging real-world application for dimensionality reduction techniques.
- IV. **Fashion MNIST Dataset:** This dataset, introduced by [122], consists of 70,000 grayscale images, each with a resolution of 28×28 pixels, resulting in 784 features

per image. The dataset is divided into a training set of 60,000 images and a test set of 10,000 images, with images categorized into ten classes: T-shirt/Top, Trouser, Pullover, Dress, Coat, Sandal, Shirt, Sneaker, Bag, and Ankle Boot.

V. Digit MNIST Dataset: A subset of the MNIST database, this dataset consists of grayscale images of handwritten digits (0 to 9), each represented as a 28×28 pixel matrix. In this study, a reduced version of the dataset is used, containing 3000 samples per digit for class labels 0, 1, 2, and 3. Each image is represented as a flattened vector, making it suitable for dimensionality reduction experiments.

By leveraging these datasets, the study provides an empirical comparison of NHRP against conventional dimensionality reduction techniques, demonstrating its effectiveness in preserving structural integrity while reducing computational complexity.

4.2 Performance Assessment of NHRP

To evaluate the performance of the dimensionality reduction methods, the datasets are partitioned into training and testing subsets, with 80% of the data allocated to the training set and the remaining 20% reserved for the test set. For the Human Activity Recognition (HAR) and Fashion MNIST datasets, a random selection of 2,000 samples is made. The effectiveness of the NHRP, HRP, NRP and PMRP methods is assessed by analyzing the preservation of distance between data points in both the training and testing sets.

This evaluation allows for the examination of how well each method maintains the underlying relationships within the data as it is projected to a lower-dimensional space. The distance preservation ratio is computed for both the training and test sets, which offers valuable information on the robustness and generalizability of each method in different contexts.

By comparing these ratios, we can better understand the performance of each technique in both the learning phase and validation phase.

This comparative analysis is crucial for identifying the most effective dimensionality reduction approach for various high-dimensional datasets, thereby informing data preprocessing strategies that are suitable for a wide range of applications.

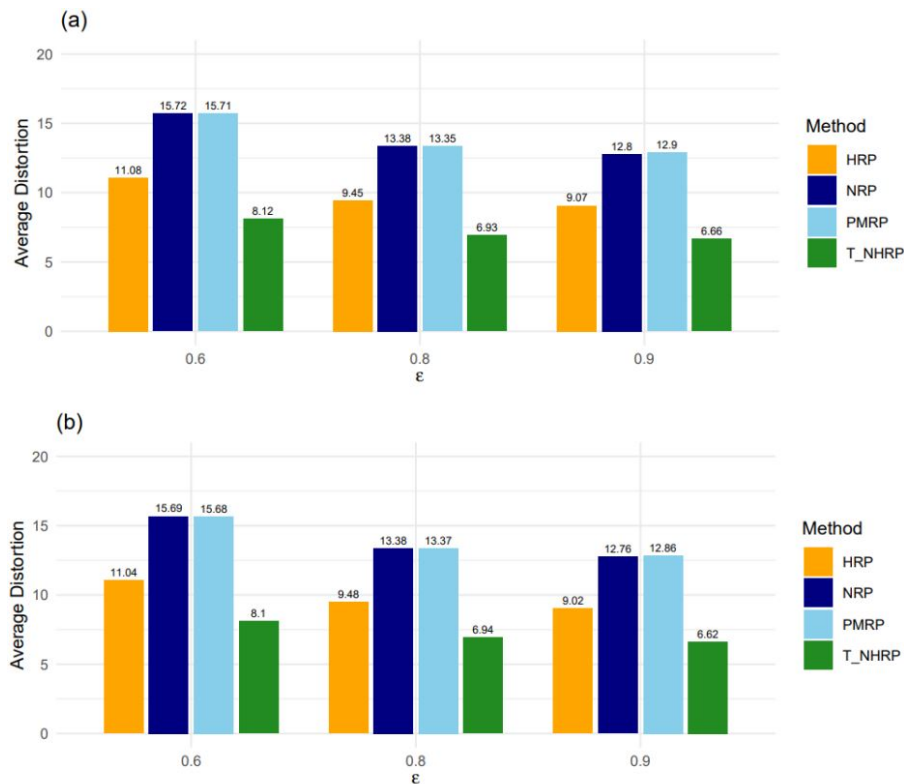


Figure 4.1: Comparison of (a) test and (b) train average distortion for NRP, PMRP, HRP and T_NHRP methods for digit mnist data.

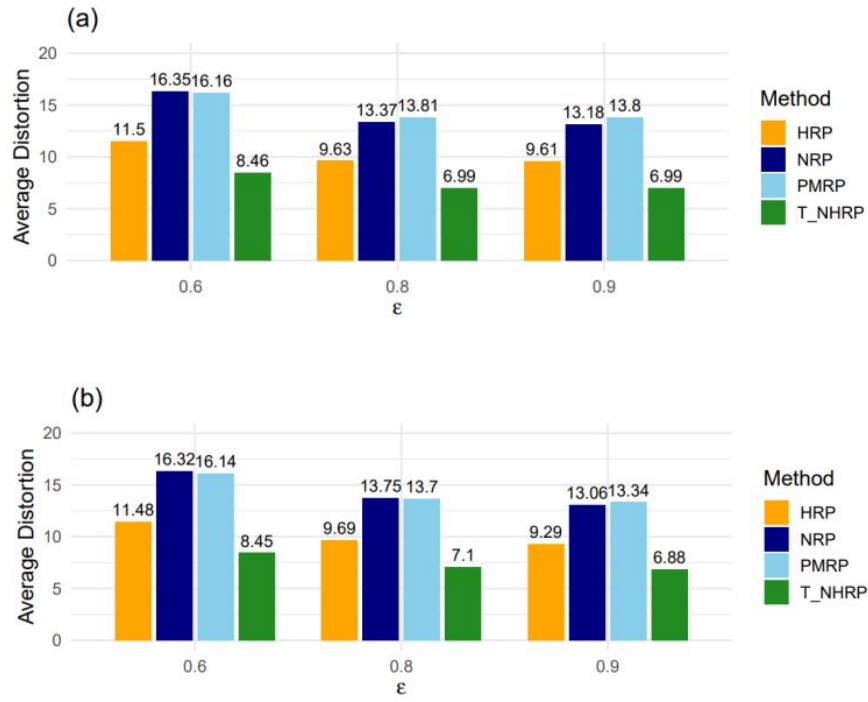


Figure 4.2: Comparison of (a) test and (b) train average distortion for NRP, PMRP, HRP and T_NHRP methods for toxicity data.

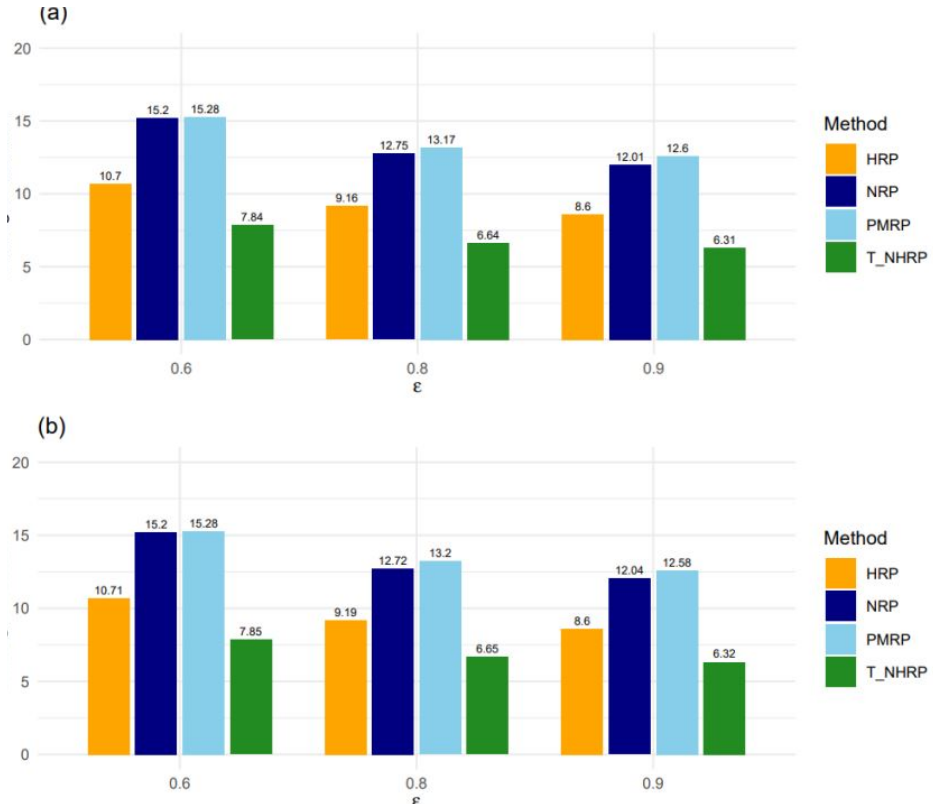


Figure 4.3: Comparison of (a) test and (b) train average distortion for NRP, PMRP, HRP and T_NHRP methods for HAR data.

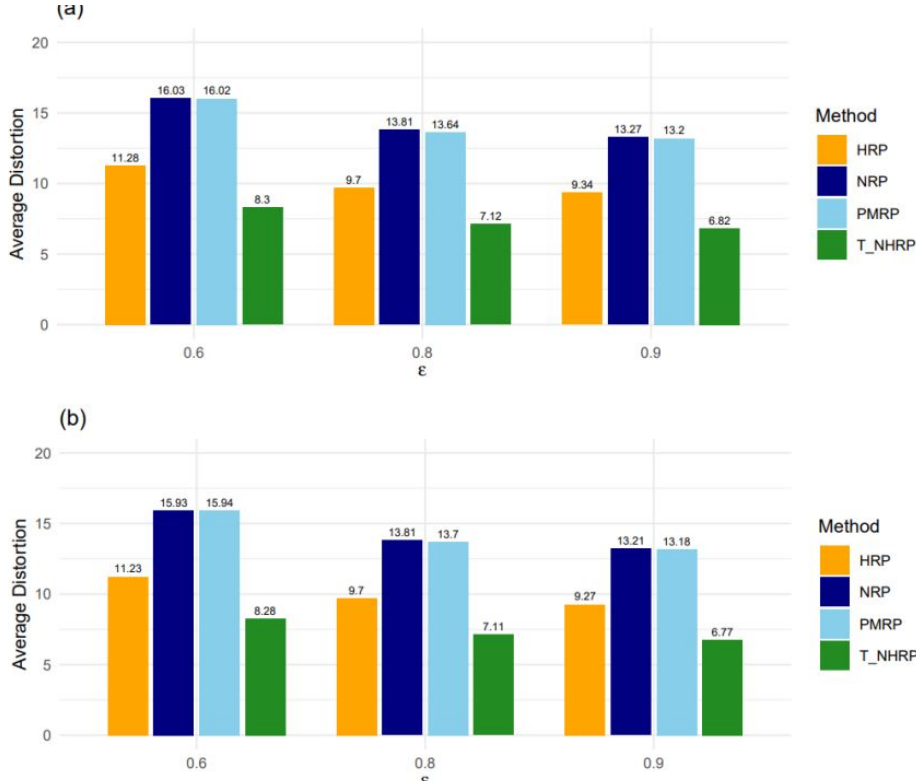


Figure 4.4: **Comparison of (a) test and (b) train average distortion for NRP, PMRP, HRP and T_NHRP methods for period changer data.**

Figures 4.1 through 4.4 present a comprehensive comparative analysis of four random projection (RP) techniques: NRP, PMRP, HRP, and the proposed T_NHRP across various machine learning datasets. These figures illustrate how dimensionality reduction impacts distortion levels at different values of ϵ , facilitating clear insights into each method's performance trends. Each figure is divided into two vertically arranged subplots: (a) depicts the average distortion of the test data, and (b) shows the corresponding training data distortion.

In the case of the Digit MNIST dataset (Figure 4.1), the training distortions for NRP at $\epsilon = 0.6, 0.8$, and 0.9 are 15.69, 13.38, and 12.76, respectively. The corresponding test distortions are 15.72, 13.38, and 12.80. PMRP follows a similar pattern, with training distortions of 15.68, 13.37, and 12.86, and test distortions of 15.71, 13.35, and 12.90. HRP

achieves lower distortions, recording training values of 11.04, 9.48, and 9.02, and test values of 11.08, 9.45, and 9.07. T_NHRP consistently produces the lowest distortion across all ϵ levels, with training distortions of 8.10, 6.94, and 6.62, and corresponding test distortions of 8.12, 6.93, and 6.66.

A similar trend is observed in the Toxicity dataset (Figure 4.2), where T_NHRP again records the lowest distortion values 8.45, 7.10, and 6.88 at $\epsilon = 0.6, 0.8$, and 0.9 . In comparison, NRP results in distortions of 16.32, 13.75, and 13.06; PMRP in 16.14, 13.70, and 13.34; and HRP in 11.48, 9.69, and 9.29. These findings are consistent with other datasets, including HAR and Period Changer (Figures 4.3 and 4.4), which confirm the robustness of the T_NHRP technique.

Another notable observation across all data sets is the nearly identical distortion levels between training and test data, indicating that all RP methods, particularly T_NHRP, generalize well to unseen data. Overall, these results underscore the superior performance of T_NHRP in minimizing information loss during dimensionality reduction, consistently outperforming NRP, PMRP, and HRP.

CHAPTER 5

DISCUSSION, CONCLUSION AND FUTURE PERSPECTIVES

5.1 Discussion

In this thesis, we have explored the growing integration of RP with various dimensionality reduction techniques to address the challenges posed by high-dimensional data. Through a brief literature review and the development of a NHRP method, this research contributes to the understanding and improvement of hybrid dimensionality reduction techniques.

The literature review categorizes hybrid approaches into four main types, each offering distinct advantages and limitations depending on the specific application requirements. By focusing on the integration of Random Projections with machine learning, graph embeddings, non-linear techniques, and the novel approach using the integration of two random projections, this thesis provides a comprehensive framework for understanding the interplay between different dimensionality reduction strategies.

One of the most prominent benefits of RP-based hybrid techniques is their scalability

and efficiency. By leveraging the Johnson-Lindenstrauss (JL) lemma, RP ensures approximate preservation of pairwise distances in the reduced dimensional space. This characteristic has been widely exploited in preprocessing high-dimensional datasets for machine learning models, enhancing both computational speed and accuracy. However, while RP excels in linear transformations, its effectiveness diminishes when handling complex, nonlinear relationships. To address this, RP has been successfully integrated with non-linear methods like t-SNE and manifold learning, striking a balance between computational efficiency and the preservation of intricate data structures.

Incorporating RP into machine learning pipelines has proven effective for tasks such as high-dimensional classification, clustering, and anomaly detection. Notably, its combination with ensemble learning methods has reduced feature redundancy and overfitting, thus improving model performance. However, challenges remain, particularly in scenarios involving feature-dependent algorithms, where aggressive dimensionality reduction can result in significant information loss.

The application of RP in graph embeddings has also garnered attention. By reducing the dimensionality of graph adjacency matrices or embeddings, RP aids in tasks like link prediction and node classification while ensuring scalability for large-scale networks. However, this integration often comes with a trade-off: the potential loss of graph structure and relational information during projection. Efforts such as sparse RP and hybrid approaches with Laplacian Eigenmaps have mitigated these issues, providing a balance between computational efficiency and accuracy.

Hybrid RP techniques combining linear and nonlinear methods, such as the integration of kernel methods or autoencoders, have demonstrated potential in managing data with nonlinear structures. These methods enable the extraction of meaningful patterns while

maintaining efficiency. Similarly, the novel integration of RP1 with RP2 offers a promising avenue for balancing computational efficiency and precision. By blending dense and sparse projection techniques, this hybrid approach minimizes distance distortion and enhances representation, making it a robust choice for resource-constrained environments.

Furthermore, dimensionality reduction remains a fundamental challenge in high-dimensional data analysis, where the preservation of structural integrity and computational efficiency are often competing objectives. In this study, we introduce and evaluate the NHRP method, a novel approach that extends conventional Hybrid Random Projection (HRP) techniques by incorporating nonlinear transformations to enhance distance preservation while maintaining computational efficiency.

NHRP builds upon Normal Random Projection (NRP) and Plus-Minus One Random Projection (PMRP), leveraging the strengths of both techniques while introducing a nonlinear transformation function. By integrating the dense, theoretically sound nature of NRP with the sparse and computationally efficient properties of PMRP, NHRP optimally balances projection density and distortion minimization. The nonlinear formulation further refines this balance, ensuring improved adaptability across various high-dimensional datasets. The efficiency of NHRP in maintaining the fundamental geometrical and statistical characteristics of the data via a dimensionality reduction process is demonstrated by a number of rigorous mathematical arguments. These proofs validate the NHRP’s capacity to preserve data integrity after projection by showing that it complies with strict high dimensional data processing criteria. By carefully using matrix features like sparsity and density, the NHRP fundamental design optimises the error boundaries set by the Johnson-Lindenstrauss (JL) lemma for both the NRP and PMRP.

The mathematical foundations of NHRP are grounded in the JL lemma, which provides

theoretical guarantees for distance preservation under random projections. Unlike standard HRP, which linearly combines projection matrices, NHRP applies a nonlinear transformation to the hybridization process, reducing distortion and improving structural preservation. This modification strategically manages sparsity and density trade-offs, ensuring that the projected data retains essential geometric and statistical properties. Furthermore, the incorporation of a nonlinear component mitigates the limitations imposed by purely linear hybrid approaches, addressing potential distortions arising from suboptimal weighting of projection matrices.

To empirically validate the effectiveness of NHRP, we conduct an extensive comparative analysis against NRP, PMRP, and HRP across multiple real-world and synthetic datasets. The datasets include applications in medical diagnostics (e.g., lung cancer dataset) and complex pattern recognition (e.g., Darwin dataset), ensuring a diverse range of test scenarios. Performance was evaluated based on average distortion, pairwise distance preservation, and computational efficiency. Results indicate that while NRP and PMRP exhibit nearly identical average distortions due to their shared theoretical foundations, NHRP consistently outperforms both standalone methods and HRP. NHRP achieves superior pairwise distance retention compared to HRP, demonstrating lower variance in distortion across different projection scales. This improvement is attributed to the nonlinear weighting mechanism, which prevents excessive deviation from the original data structure. Despite introducing a nonlinear transformation, NHRP maintains a computational complexity comparable to HRP, ensuring feasibility for large-scale applications. The optimized hybridization process reduces unnecessary computations, thereby minimizing the trade-off between accuracy and efficiency. HRP, which exhibits marginal instability in certain high-dimensional scenarios, NHRP consistently delivers robust performance across different datasets and error tolerance

levels (ϵ). This suggests that the nonlinear formulation enhances stability, making it more adaptable to varying data distributions.

5.2 Conclusion and Future Perspectives

This review of the literature highlights the transformative potential of hybrid DR techniques based on RP in the field of dimensionality reduction. By integrating RP with machine learning models, graph embeddings, nonlinear methods, and novel combinations like NRP and PMRP, researchers have developed scalable and efficient solutions for high-dimensional data analysis. These hybrid approaches leverage RP’s computational advantages while addressing its limitations, such as the inability to capture complex data structures or preserve task-specific information.

The effectiveness of the Nonlinear Hybrid Random Projection (NHRP) extends beyond theoretical advantages, providing practical benefits for dimensionality reduction in high-dimensional data analysis. By integrating structured and adaptive projection mechanisms, NHRP achieves superior preservation of structural and statistical properties while ensuring computational efficiency. This characteristic is particularly crucial in machine learning and deep learning applications, where the quality of feature representations significantly impacts model performance. Unlike conventional random projection techniques, which either employ dense transformations (as in NRP) or sparse mappings (as in PMRP), NHRP strategically balances these properties to optimize data representation with minimal information loss.

A critical observation from the results is that NHRP consistently outperforms HRP on various metrics, particularly in distortion minimization. For example, in Table 3.1, NHRP methods (polynomial and trigonometric transformations) demonstrate lower average distortions compared to HRP on multiple values ϵ . Specifically, for $\epsilon = 0.6$ and $n = 400$,

the polynomial-based NHRP reduces distortion to 6.66 (SE = 0.002), while HRP exhibits a higher distortion of 8.73 (SE = 0.007). This pattern remains consistent across different sample sizes and δ values, indicating that NHRP not only preserves pairwise distances more effectively but also adapts better to varying data complexities.

Despite its improved performance, NHRP introduces additional complexity in parameter tuning, particularly concerning the blending coefficient α . The optimal selection of α is highly dependent on dataset characteristics such as sparsity, density, and intrinsic dimensionality. Unlike HRP, which relies on a predefined combination of NRP and PMRP, NHRP requires fine-tuning to balance the influence of nonlinear transformations. This adaptability makes NHRP more robust but also necessitates additional computational overhead during parameter optimization. Furthermore, the interpretability of NHRP transformations remains a challenge, as the nonlinear mappings can obscure the contribution of individual projection techniques, making analysis more complex compared to HRP.

From a computational efficiency perspective, HRP maintains a relatively stable runtime, whereas NHRP, particularly with polynomial and trigonometric transformations, incurs a slight increase in processing time. However, the trade-off between computational cost and distortion minimization suggests that NHRP remains a preferable choice for high-dimensional applications where accuracy is a priority. The results further indicate that as ϵ increases, the gap in performance between HRP and NHRP widens, reinforcing the advantage of incorporating nonlinear transformations in high-dimensional projections.

Future research in this domain should focus on the development of adaptive hybrid projection methods that dynamically adjust to the characteristics of the input data. Enhancing the robustness of RP-based techniques against noise and incomplete datasets is another critical area. Additionally, integrating domain-specific knowledge into RP techniques could

provide tailored solutions for challenges in fields like neuroscience, security, and physics. To enhance the practical usability of NHRP, future research should focus on developing adaptive algorithms for dynamically optimizing α based on data characteristics. Machine learning-driven approaches for parameter tuning could automate this process, improving NHRP’s applicability across various datasets. Additionally, integrating NHRP with deep learning architectures could offer novel solutions for dimensionality reduction in large-scale artificial intelligence applications, where efficient data compression without significant information loss is essential.

In conclusion, hybrid methods offer great potential for advancing dimensionality reduction techniques, addressing the challenges of high-dimensional data, and enabling more efficient, accurate, and scalable applications across diverse domains. Overall, NHRP presents a significant advancement over HRP by achieving lower distortion rates while maintaining adaptability across different dataset structures. Although it introduces additional complexity in parameter selection and computational cost, its superior ability to preserve geometric and statistical properties makes it a powerful tool for high-dimensional data analysis. Addressing the challenges associated with parameter optimization and interpretability will be crucial for maximizing its impact in scientific and industrial applications. The NHRP method has emerged as a significant contribution to the field of data science, particularly in dimensionality reduction. Its ability to effectively balance reducing data complexity and preserving crucial information paves the way for more accurate and efficient data analysis in a myriad of high-dimensional data.

APPENDIX A

This appendix presents the detailed mathematical proofs for the polynomial NHRP approach that support the key theoretical results discussed in the main chapters of this thesis. By placing these derivations here, we maintain the clarity and readability of the core discussion while still providing rigorous justifications for all stated lemmas and theorems.

Lemma 1. For any pair of points $x, y \in X$, and for any $\alpha \in [0, 1]$, the Euclidean distance between the projections of x and y using the hybrid projection H_1 is less than or equal to the sum of the distances of the projections of x and y using the matrices R_1 and R_2 , respectively.

Proof. Let H_1 (equation 3.3) be the polynomial NHRP defined by:

$$H_1 = \alpha^2(R_1 \circ R_1) + 2\alpha(1 - \alpha)(R_1 \circ R_2) + (1 - \alpha)^2(R_2 \circ R_2)$$

where $\circ$ denotes the Hadamard (element-wise) product, and $0 \leq \alpha \leq 1$.

The projections of x and y using H_1 are:

$$xH_1 = x [\alpha^2(R_1 \circ R_1) + 2\alpha(1 - \alpha)(R_1 \circ R_2) + (1 - \alpha)^2(R_2 \circ R_2)]$$

$$yH_1 = y [\alpha^2(R_1 \circ R_1) + 2\alpha(1 - \alpha)(R_1 \circ R_2) + (1 - \alpha)^2(R_2 \circ R_2)]$$

Subtracting and factoring out $(x - y)$:

$$xH_1 - yH_1 = (x - y) [\alpha^2(R_1 \circ R_1) + 2\alpha(1 - \alpha)(R_1 \circ R_2) + (1 - \alpha)^2(R_2 \circ R_2)]$$

Taking the Euclidean norm and applying the triangle inequality:

$$\begin{aligned} \|xH_1 - yH_1\|_2 &\leq \alpha^2\|(x - y)(R_1 \circ R_1)\|_2 \\ &\quad + 2\alpha(1 - \alpha)\|(x - y)(R_1 \circ R_2)\|_2 \\ &\quad + (1 - \alpha)^2\|(x - y)(R_2 \circ R_2)\|_2 \end{aligned}$$

If the entries of R_1 and R_2 are appropriately normalized, the Hadamard product preserves element-wise structure and ensures each term remains bounded, contributing to stable distance preservation in the projected space. ■

$$\|xH_1 - yH_1\|_2 \leq \|xR_1 - yR_1\|_2 + \|xR_2 - yR_2\|_2$$

Lemma 2: Distance Preservation of Hybrid Projections

Let X be a high-dimensional dataset in $\mathbb{R}^d$, and let R_1 and R_2 be two different random projection (RP) matrices constructed using two distinct RP methods. Let $\alpha \in [0, 1]$. If we define H_1 as a hybrid projection matrix that is a nonlinear combination of R_1 and R_2 , then for any two data points $x, y \in X$, the following holds:

$$\frac{\|xH_1 - yH_1\|_2}{\|x - y\|_2} \approx 1.$$

Suppose R_1 and R_2 are two RP matrices constructed via two different RP methods.

This property is significant because it ensures that the hybrid method does not introduce

a correlation between previously uncorrelated dimensions, thereby maintaining the integrity of the data structure in the reduced space.

We start by expanding the distance between xH_1 and yH_1 :

$$\|xH_1 - yH_1\|_2 = \|\alpha^2 xR_1 + 2\alpha(1-\alpha)xR_2 + (1-\alpha)^2 xR_2 - (\alpha^2 yR_1 + 2\alpha(1-\alpha)yR_2 + (1-\alpha)^2 yR_2)\|_2$$

We can regroup the terms as follows:

$$\|xH_1 - yH_1\|_2 = \|\alpha^2(xR_1 - yR_1) + (1-\alpha)^2(xR_2 - yR_2) + 2\alpha(1-\alpha)(xR_2 - yR_2)\|_2$$

Now, by applying the triangle inequality, we get:

$$\|xH_1 - yH_1\|_2 \leq \|\alpha^2(xR_1 - yR_1)\|_2 + \|(1-\alpha)^2(xR_2 - yR_2)\|_2 + \|2\alpha(1-\alpha)(xR_2 - yR_2)\|_2$$

Since the norms are linear, we simplify the expression:

$$= \alpha^2\|xR_1 - yR_1\|_2 + (1-\alpha)^2\|xR_2 - yR_2\|_2 + 2\alpha(1-\alpha)\|xR_2 - yR_2\|_2$$

The terms involving R_2 can be combined:

$$= \alpha^2\|xR_1 - yR_1\|_2 + ((1-\alpha)^2 + 2\alpha(1-\alpha))\|xR_2 - yR_2\|_2$$

We can simplify the second term:

$$(1 - \alpha)^2 + 2\alpha(1 - \alpha) = (1 - \alpha)(1 - \alpha + 2\alpha) = (1 - \alpha)(1 + \alpha) = 1 - \alpha^2$$

Thus, the expression becomes:

$$= \alpha^2 \|xR_1 - yR_1\|_2 + (1 - \alpha^2) \|xR_2 - yR_2\|_2$$

By the JL lemma, we know that the random projections R_1 and R_2 approximately preserve pairwise distances. Therefore, we can approximate:

$$\|xR_1 - yR_1\|_2 \approx \|x - y\|_2 \quad \text{and} \quad \|xR_2 - yR_2\|_2 \approx \|x - y\|_2$$

Substitute these approximations into the expression:

$$\|xH_1 - yH_1\|_2 \approx \alpha^2 \|x - y\|_2 + (1 - \alpha^2) \|x - y\|_2$$

Factor out $\|x - y\|_2$:

$$\|xH_1 - yH_1\|_2 \approx \|x - y\|_2 (\alpha^2 + (1 - \alpha^2))$$

Simplify the expression inside the parentheses:

$$\alpha^2 + (1 - \alpha^2) = 1$$

Thus, we have:

$$\|xH_1 - yH_1\|_2 \approx \|x - y\|_2$$

Therefore, we conclude:

$$\frac{\|xH_1 - yH_1\|_2}{\|x - y\|_2} \approx 1$$

This shows that the hybrid projection method H_1 preserves distances between data points, validating that it retains the essential geometric structure of the original dataset. This result aligns with the JL lemma and demonstrates the effectiveness of the NHRP method in maintaining pairwise distances during dimensionality reduction.

Lemma 3: Error Bound of Hybrid Projections

Given a high-dimensional dataset $X \subset \mathbb{R}^d$, two random projection matrices R_1 and R_2 , and a hybrid projection matrix

$$H_1 = \alpha^2(R_1 \circ R_1) + 2\alpha(1 - \alpha)(R_1 \circ R_2) + (1 - \alpha)^2(R_2 \circ R_2)$$

for $\alpha \in [0, 1]$, the expected distortion of distances introduced by H_1 is bounded within a certain range dependent on α , the properties of R_1 and R_2 , and the dimensionality of the reduced space.

Proof:

Let $x, y \in X$ be two points in the dataset. The distortion introduced by the hybrid projection on the distance between x and y is given by the difference:

$$\Delta = \left| \|x - y\|_2^2 - \|xH_1 - yH_1\|_2^2 \right|$$

Expanding $\|xH_1 - yH_1\|_2^2$:

We now expand $\|xH_1 - yH_1\|_2^2$ based on the given expression for H_1 :

$$\|xH_1 - yH_1\|_2^2 = \|\alpha^2(xR_1 - yR_1) + 2\alpha(1 - \alpha)(xR_1 - yR_1) + (1 - \alpha)^2(xR_2 - yR_2)\|_2^2$$

Now, applying the triangle inequality:

$$\Delta = \left| \|x - y\|_2^2 - (\alpha^2\|xR_1 - yR_1\|_2^2 + 2\alpha(1 - \alpha)\|xR_1 - yR_1\|_2^2 + (1 - \alpha)^2\|xR_2 - yR_2\|_2^2) \right|$$

We can now simplify and express the difference in terms of the distortions introduced by the individual random projections R_1 and R_2 . Let ϵ_1 and ϵ_2 represent the maximum distortions introduced by R_1 and R_2 , respectively:

$$\epsilon_1 = \max_{x,y} \left| \|x - y\|_2^2 - \|xR_1 - yR_1\|_2^2 \right|$$

$$\epsilon_2 = \max_{x,y} \left| \|x - y\|_2^2 - \|xR_2 - yR_2\|_2^2 \right|$$

Thus, we can bound the distortion Δ as follows:

$$\Delta \leq \alpha\epsilon_1 + (1 - \alpha)\epsilon_2$$

This shows that the distortion introduced by the hybrid projection matrix H_1 is a weighted combination of the distortions introduced by R_1 and R_2 , which are bounded by the maximum distortions of each individual projection matrix.

The error bound for the hybrid projection is given by:

$$\Delta \leq \alpha\epsilon_1 + (1 - \alpha)\epsilon_2$$

This demonstrates that the distortion introduced by the hybrid projection matrix H_1 is bounded by a weighted sum of the distortions of the individual random projections R_1 and R_2 , as required.

Lemma 4: Sparsity Preservation

If R_1 and R_2 are random projection matrices that maintain sparsity, the hybrid projection matrix H_1 will also approximately preserve the sparsity of the projected data.

Proof:

Let $x \in X$ represent a sparse vector. The sparsity of xH_1 can be analyzed by counting the number of non-zero components in the projected vector. Since H_1 is a linear combination of the matrices R_1 and R_2 , each element in xH_1 can be expressed as a weighted sum of the projections xR_1 and xR_2 . Specifically, we have:

$$xH_1 = \alpha^2(xR_1) + 2\alpha(1 - \alpha)(xR_2) + (1 - \alpha)^2(xR_1)$$

Since both xR_1 and xR_2 are sparse, due to the properties of the random projection matrices R_1 and R_2 , the number of non-zero components in these projections is relatively low. Consequently, the number of non-zero elements in their linear combination will also remain small, implying that the sparsity of x is largely preserved under the transformation by H_1 .

Although Lemma 6 assumes that both R_1 and R_2 individually preserve sparsity, it is important to note that even if these matrices do not strictly preserve sparsity, the hybrid matrix H_1 can still maintain a level of sparsity. The interaction between the non-zero

components of xR_1 and xR_2 might lead to cancellations or limited increases in the number of non-zero elements, especially when the parameter α is chosen in a way that balances the contributions of both projections.

If R_1 is denser than R_2 , a smaller value of α would be preferred, and conversely, if R_2 is denser, a larger α might be more suitable. Therefore, careful management of the trade-off between density and sparsity is essential to ensure that the sparsity of the projected vector xH_1 does not significantly increase compared to that of the original vector x .

In conclusion, if R_1 and R_2 preserve sparsity, then the hybrid projection matrix H_1 will also approximately preserve the sparsity of the data.

Lemma 5: Computational Complexity Considering α

The computational complexity of applying the hybrid projection $H_1 = \alpha^2(R_1 \circ R_1) + 2\alpha(1 - \alpha)(R_1 \circ R_2) + (1 - \alpha)^2(R_2 \circ R_2)$ to a dataset X depends not only on the dimensions of X and the properties of R_1 and R_2 , but also on the computational cost of the nonlinear transformations and the optimization of α .

Proof:

Let the dataset $X \in \mathbb{R}^{n \times d}$, where n is the number of data points and d is the dimensionality. The hybrid projection matrix H_1 is defined as:

$$H_1 = \alpha^2(R_1 \otimes R_1) + 2\alpha(1 - \alpha)(R_1 \otimes R_2) + (1 - \alpha)^2(R_2 \otimes R_2)$$

The computational complexity of applying H_1 involves the following key steps:

1. **Matrix Multiplications:** To apply the projection, we first need to compute XR_1 and XR_2 . The matrix multiplication of X (of size $n \times d$) with R_1 and R_2 (of size $d \times k$) is $O(ndk)$.

2. **Tensor Products:** Next, we compute the tensor products $R_1 \otimes R_1$, $R_1 \otimes R_2$, and $R_2 \otimes R_2$. The time complexity of computing the tensor product of two matrices of size $d \times k$ is $O(d^2k^2)$, which needs to be computed for the three terms in the expression for H_1 . This results in a complexity of $O(d^2k^2)$.
3. **Nonlinear Transformations:** The application of nonlinear transformations (e.g., sine) on the results from the tensor products involves element-wise computations, contributing an additional $O(nk)$ complexity for each transformation.
4. **Linear Combination of Projections:** After applying the nonlinear transformations, we compute the linear combination of these projections weighted by α and $1 - \alpha$, which contributes a complexity of $O(nk)$.
5. **Optimization of α :** The parameter α may be selected through an optimization procedure, such as cross-validation or grid search. The complexity of optimizing α depends on the method used. If α is predefined, this step contributes $O(1)$ to the overall complexity. If optimization is required, the complexity may increase depending on the number of iterations and the method used.

Thus, the total computational complexity is the sum of these components:

$$O(ndk) + O(d^2k^2) + O(nk) + O(nk)$$

The dominant term is the tensor product computation $O(d^2k^2)$, while the matrix multiplications and nonlinear transformations each contribute $O(ndk)$ and $O(nk)$, respectively. The optimization of α either adds a constant complexity $O(1)$ or an iterative complexity depending on the optimization strategy. Therefore, the overall computational complexity of applying H_1 is:

$$O(ndk) + O(d^2k^2) + O(nk) + \mathcal{O}_\alpha$$

where $\mathcal{O}_\alpha$ represents the complexity associated with the optimization of α , which may range from a constant to more complex iterative procedures.

Lemma 6: Equivalence of Hybrid Projection Matrices with Optimal α

Let R_1 and R_2 be two RP matrices that are statistically similar in distance preservation. Define two hybrid projection matrices $H = \alpha^2(R_1 \circ R_1) + 2\alpha(1-\alpha)(R_1 \circ R_2) + (1-\alpha)^2(R_2 \circ R_2)$ and $H' = \alpha'(R_2 \circ R_2) + 2\alpha'(1-\alpha')(R_1 \circ R_2) + (1-\alpha')^2(R_1 \circ R_1)$.

If α^* is the optimal value minimizing the distance distortion for H_1 and α'^* is the optimal value for H' , then $\alpha^* = 1 - \alpha'^*$ and the matrices H and H' are equivalent in their effectiveness of dimensionality reduction.

Proof:

1. Distance Preservation Functions:

The distance preservation ratio for two data points $x, y \in X$ is given by:

$$s = \frac{\|xH - yH\|_2^2}{\|x - y\|_2^2}$$

and similarly for s' using H' :

$$s' = \frac{\|xH' - yH'\|_2^2}{\|x - y\|_2^2}$$

The expectation of these ratios, $E[s]$ and $E[s']$, depends on α and α' respectively. Since H and H' are nonlinear combinations of R_1 and R_2 , these expectations are functions of the corresponding α and α' values. In both cases, the goal is to minimize the distortion in the distances after projection.

2. Symmetry in Distance Preservation:

Assuming that R_1 and R_2 preserve distances similarly, the forms of $E[s]$ and $E[s']$ will be symmetric. This symmetry arises because reversing the roles of R_1 and R_2 in the hybrid projection matrices does not alter the underlying statistical properties of the distance preservation. Specifically, $E[s]$ depends on α while $E[s']$ depends on α' , but the functional form of the expectation is symmetric in terms of R_1 and R_2 . Thus, the minimization problem for $E[s]$ with respect to α mirrors the minimization problem for $E[s']$ with respect to α' .

3. Optimization:

Given the symmetric nature of $E[s]$ and $E[s']$, the optimal values of α and α' must satisfy the relationship:

$$\alpha^* = 1 - \alpha'^*$$

This relationship arises from the complementary roles of R_1 and R_2 in the two hybrid matrices. Since the statistical distance preservation properties of the matrices are symmetric, the optimal values must be complementary to each other to achieve the minimal distortion.

4. Equivalence Demonstration:

Under the optimal values α^* and α'^* , both H and H' will achieve similar performance in terms of distance preservation. This implies that, for the optimal values, switching the roles of R_1 and R_2 in the hybrid projection matrix does not affect its effectiveness in reducing dimensions while preserving the structure of the data. Hence, H and H' are equivalent in terms of their dimensionality reduction capabilities.

Thus, we have shown that for optimal values of α^* and α'^* , the hybrid projection matrices H and H' are equivalent in their performance for dimensionality reduction.

APPENDIX B

Additional insights into the computational efficiency and scalability of the RP methods are provided in Appendix B Tables 1 and 2, which detail the computational time associated with each approach. Furthermore, Appendix B Tables 3–6 present the optimal values of α for the NHRP methods, offering a deeper understanding of their parameter optimization. Additionally, Appendix B Table 7 and 8 report the 95% confidence intervals (CI) corresponding to the results in Tables 3.1 and 3.2, respectively, ensuring the statistical reliability and robustness of the findings. Lastly, Appendix B Figures 1, 2 and 3 illustrate the effectiveness of the NHRP method across varying values of epsilon, providing a comprehensive evaluation of its performance under different parameter settings.

Table B.1: The time taken to calculate each average distortion and the standard error for each method employed for $d = 500$.

d	n	NP (mean T.)	NP (sd T.)	PM(mean T.)	PM (sd T.)	HRP(mean T.)	HRP(sd T.)	P_NHRP(mean T.)	P_NHRP(sd T.)	T_NHRP(mean T.)	T_NHRP(sd T.)
500	50	0.00027	3.0E-05	0.00027	2.0E-05	0.00606	2.0E-04	0.01525	1.1E-03	0.01969	6.0E-04
500	50	0.00021	1.0E-05	0.00021	1.0E-05	0.00518	3.5E-04	0.01043	3.9E-04	0.01467	2.4E-04
500	50	0.0002	1.0E-05	0.0002	2.0E-05	0.0045	1.4E-04	0.01004	9.3E-04	0.01431	2.3E-04
500	50	0.0002	1.0E-05	0.00024	2.0E-05	0.00517	3.8E-04	0.01035	6.0E-04	0.01413	2.6E-04
500	50	0.00019	2.0E-05	0.00023	2.0E-05	0.00535	1.6E-04	0.01338	7.9E-04	0.01482	2.3E-04
500	50	0.0002	2.0E-05	0.00018	1.0E-05	0.00511	2.8E-04	0.01088	6.5E-04	0.01276	3.2E-04
500	50	0.00022	2.0E-05	0.00024	2.0E-05	0.00584	4.8E-04	0.01332	8.9E-04	0.01536	2.4E-04
500	50	0.00012	2.0E-05	0.00012	1.0E-05	0.00277	1.4E-04	0.00722	5.5E-04	0.00895	2.6E-04
500	50	0.00015	1.0E-05	0.00015	2.0E-05	0.00335	1.9E-04	0.00822	6.2E-04	0.00939	2.7E-04
500	50	0.00019	2.0E-05	0.00021	2.0E-05	0.00524	4.3E-04	0.01188	9.8E-04	0.01381	3.9E-04
500	50	0.00023	2.0E-05	0.00021	2.0E-05	0.00503	2.1E-04	0.01093	4.6E-04	0.01383	1.7E-04
500	50	0.00022	1.0E-05	0.00023	2.0E-05	0.00502	1.6E-04	0.01226	7.7E-04	0.01591	3.3E-04
500	200	0.00256	9.0E-05	0.00252	8.0E-05	0.05762	2.4E-03	0.1271	7.5E-03	0.09116	2.5E-03
500	200	0.00214	4.0E-05	0.00216	4.0E-05	0.05047	3.7E-03	0.12204	8.8E-03	0.07794	8.3E-04
500	200	0.00161	5.0E-05	0.00162	3.0E-05	0.03694	1.1E-03	0.08918	5.0E-03	0.05992	9.2E-04
500	200	0.00225	3.0E-05	0.00231	4.0E-05	0.05172	1.6E-03	0.1161	5.8E-03	0.08278	7.0E-04
500	200	0.00188	7.0E-05	0.00188	6.0E-05	0.04459	1.9E-03	0.10277	7.4E-03	0.06916	2.1E-03
500	200	0.00154	3.0E-05	0.00159	5.0E-05	0.03582	1.9E-03	0.08082	5.5E-03	0.05796	4.9E-04
500	200	0.00154	3.0E-05	0.00155	3.0E-05	0.03446	5.1E-04	0.08958	6.2E-03	0.05789	4.4E-04
500	200	0.0013	3.0E-05	0.00139	4.0E-05	0.02894	5.7E-04	0.07285	4.8E-03	0.04869	6.0E-04
500	200	0.00175	7.0E-05	0.00177	8.0E-05	0.03962	1.4E-03	0.08748	4.5E-03	0.06543	2.0E-03
500	200	0.00143	3.0E-05	0.00147	3.0E-05	0.03385	5.7E-04	0.08887	5.7E-03	0.0542	5.0E-04
500	200	0.00149	3.0E-05	0.00148	3.0E-05	0.03236	5.0E-04	0.07891	4.8E-03	0.05412	4.7E-04
500	200	0.00139	3.0E-05	0.00146	3.0E-05	0.03235	5.0E-04	0.07571	4.7E-03	0.05377	4.4E-04
500	400	0.00521	6.0E-05	0.00529	9.0E-05	0.11296	1.7E-03	0.30409	2.4E-02	0.1714	1.5E-03
500	400	0.00726	1.9E-04	0.00716	1.6E-04	0.16803	1.1E-02	0.44743	3.1E-02	0.23814	5.0E-03
500	400	0.00987	2.9E-04	0.01013	3.6E-04	0.21973	8.2E-03	0.50764	3.8E-02	0.32804	1.1E-02
500	400	0.00912	6.2E-04	0.00925	6.0E-04	0.19094	9.0E-03	0.50695	4.0E-02	0.28504	8.1E-03
500	400	0.00554	1.6E-04	0.0057	1.9E-04	0.11989	3.1E-03	0.27361	1.4E-02	0.18036	3.9E-03
500	400	0.00561	1.0E-04	0.00561	1.1E-04	0.1247	4.7E-03	0.33373	2.1E-02	0.18263	2.4E-03
500	400	0.00843	2.4E-04	0.00841	2.4E-04	0.18848	7.1E-03	0.49306	3.8E-02	0.2835	7.8E-03
500	400	0.00552	1.2E-04	0.00562	1.4E-04	0.1216	4.9E-03	0.28062	1.4E-02	0.17897	2.3E-03
500	400	0.0047	1.1E-04	0.00459	1.0E-04	0.10061	2.0E-03	0.25402	1.8E-02	0.15423	3.0E-03
500	400	0.0055	1.5E-04	0.00561	1.5E-04	0.11963	2.9E-03	0.31833	1.9E-02	0.18056	3.9E-03
500	400	0.00811	3.6E-04	0.00785	2.3E-04	0.17103	5.9E-03	0.43549	3.3E-02	0.255	5.9E-03
500	400	0.00498	8.0E-05	0.00506	8.0E-05	0.11068	2.0E-03	0.25789	1.4E-02	0.16834	2.5E-03

Table B.2: The time taken to calculate each average distortion and the standard error for each method employed for $d = 10000$.

d	n	NP (mean T.)	NP (sd T.)	PM(mean T.)	PM (sd T.)	HRP(mean T.)	HRP(sd T.)	P.NHRP(mean T.)	P.NHRP(sd T.)	T.NHRP(mean T.)	T.NHRP(sd T.)
10000	50	0.00513	7.0E-05	0.00534	5.0E-05	0.13375	5.0E-03	0.31677	1.7E-02	0.49331	2.7E-03
10000	50	0.00496	5.0E-05	0.00523	3.0E-05	0.12886	2.6E-03	0.3238	1.5E-02	0.48455	8.2E-04
10000	50	0.00503	5.0E-05	0.0053	6.0E-05	0.13025	2.7E-03	0.31884	1.5E-02	0.49005	3.0E-03
10000	50	0.00458	4.0E-05	0.00488	5.0E-05	0.12809	5.9E-03	0.3032	1.8E-02	0.46903	2.6E-03
10000	50	0.00364	5.0E-05	0.00386	5.0E-05	0.10474	5.2E-03	0.24392	1.1E-02	0.35657	2.7E-03
10000	50	0.00353	3.0E-05	0.00376	3.0E-05	0.10503	8.9E-03	0.25451	1.5E-02	0.35	4.4E-04
10000	50	0.0035	3.0E-05	0.00374	3.0E-05	0.10196	5.8E-03	0.23955	1.3E-02	0.34737	4.2E-04
10000	50	0.00354	3.0E-05	0.0037	3.0E-05	0.09888	5.3E-03	0.22812	1.3E-02	0.34971	1.3E-03
10000	50	0.00333	4.0E-05	0.0035	3.0E-05	0.09103	3.0E-03	0.22998	1.2E-02	0.32807	2.2E-03
10000	50	0.00323	3.0E-05	0.00345	3.0E-05	0.08875	4.2E-03	0.23189	1.4E-02	0.32286	4.0E-04
10000	50	0.00325	3.0E-05	0.00341	4.0E-05	0.08787	3.3E-03	0.23611	1.6E-02	0.32266	9.3E-04
10000	50	0.00318	2.0E-05	0.0034	3.0E-05	0.0938	6.3E-03	0.20421	1.2E-02	0.32096	3.5E-04

Table B.3: **Optimal Alpha for The P_NHRP when $d = 500$.**

d	n	ϵ	Optimal α (P_NHRP)	
500	50	0.6	0.2	0.224218785
500	50	0.6	0.6	0.24037575
500	50	0.6	0.8	0.198755282
500	50	0.6	1	0.238055134
500	50	0.8	0.2	0.168050778
500	50	0.8	0.6	0.227239427
500	50	0.8	0.8	0.243936105
500	50	0.8	1	0.200548836
500	50	0.9	0.2	0.222581035
500	50	0.9	0.6	0.207301323
500	50	0.9	0.8	0.23341607
500	50	0.9	1	0.173882098
500	200	0.6	0.2	0.201268722
500	200	0.6	0.6	0.220761765
500	200	0.6	0.8	0.211277528
500	200	0.6	1	0.201243627
500	200	0.8	0.2	0.228706781
500	200	0.8	0.6	0.245705017
500	200	0.8	0.8	0.202050065
500	200	0.8	1	0.215147336
500	200	0.9	0.2	0.198300441
500	200	0.9	0.6	0.260736886
500	200	0.9	0.8	0.223822001
500	200	0.9	1	0.231423311
500	400	0.6	0.2	0.21619254
500	400	0.6	0.6	0.22333804
500	400	0.6	0.8	0.212895866
500	400	0.6	1	0.226036157
500	400	0.8	0.2	0.207852889
500	400	0.8	0.6	0.214669006
500	400	0.8	0.8	0.236644146
500	400	0.8	1	0.244886286
500	400	0.9	0.2	0.22538178
500	400	0.9	0.6	0.204247828
500	400	0.9	0.8	0.215427385
500	400	0.9	1	0.218918146

Table B.4: **Optimal Alpha for The T_NHRP when $d = 500$.**

d	n	ϵ	Optimal α (T_NHRP)	
500	50	0.6	0.2	0.620551296
500	50	0.6	0.6	0.618276528
500	50	0.6	0.8	0.622077297
500	50	0.6	1	0.613815914
500	50	0.8	0.2	0.622636627
500	50	0.8	0.6	0.614417512
500	50	0.8	0.8	0.625721121
500	50	0.8	1	0.609990898
500	50	0.9	0.2	0.622951735
500	50	0.9	0.6	0.623418502
500	50	0.9	0.8	0.616533188
500	50	0.9	1	0.622221459
500	200	0.6	0.2	0.624795088
500	200	0.6	0.6	0.62380197
500	200	0.6	0.8	0.622668415
500	200	0.6	1	0.624551406
500	200	0.8	0.2	0.625428712
500	200	0.8	0.6	0.620228739
500	200	0.8	0.8	0.621253138
500	200	0.8	1	0.615073809
500	200	0.9	0.2	0.618677085
500	200	0.9	0.6	0.620667275
500	200	0.9	0.8	0.618236583
500	200	0.9	1	0.617175148
500	400	0.6	0.2	0.619702
500	400	0.6	0.6	0.618470368
500	400	0.6	0.8	0.619599989
500	400	0.6	1	0.619706448
500	400	0.8	0.2	0.622451154
500	400	0.8	0.6	0.61964972
500	400	0.8	0.8	0.621744806
500	400	0.8	1	0.618119471
500	400	0.9	0.2	0.62120767
500	400	0.9	0.6	0.618699486
500	400	0.9	0.8	0.6205232
500	400	0.9	1	0.624804492

Table B.5: **Optimal Alha for the P_NHRP when d=5000.**

d	n	ϵ	Optimal α (P_NHRP)	
5000	50	0.6	0.2	0.229755028
5000	50	0.6	0.6	0.242506151
5000	50	0.6	0.8	0.179381736
5000	50	0.6	1	0.219634935
5000	50	0.8	0.2	0.267085392
5000	50	0.8	0.6	0.223889339
5000	50	0.8	0.8	0.290467596
5000	50	0.8	1	0.201455019
5000	50	0.9	0.2	0.207424601
5000	50	0.9	0.6	0.199444561
5000	50	0.9	0.8	0.241691715
5000	50	0.9	1	0.253957792
5000	200	0.6	0.2	0.216431018
5000	200	0.6	0.6	0.219918495
5000	200	0.6	0.8	0.214207638
5000	200	0.6	1	0.244997649
5000	200	0.8	0.2	0.230322215
5000	200	0.8	0.6	0.20449827
5000	200	0.8	0.8	0.226607344
5000	200	0.8	1	0.224214023
5000	200	0.9	0.2	0.221008314
5000	200	0.9	0.6	0.208590266
5000	200	0.9	0.8	0.265018929
5000	200	0.9	1	0.214044191
5000	400	0.6	0.2	0.211040071
5000	400	0.6	0.6	0.209251124
5000	400	0.6	0.8	0.198195526
5000	400	0.6	1	0.214081006
5000	400	0.8	0.2	0.231782575
5000	400	0.8	0.6	0.230275596
5000	400	0.8	0.8	0.213328892
5000	400	0.8	1	0.230292438
5000	400	0.9	0.2	0.217835168
5000	400	0.9	0.6	0.217765529
5000	400	0.9	0.8	0.236068465
5000	400	0.9	1	0.210987571

Table B.6: **Optimal Alpha for the T_NHRP when d=5000.**

d	n	ϵ	Optimal α (T_NHRP)	
5000	50	0.6	0.2	0.624420432
5000	50	0.6	0.6	0.621723186
5000	50	0.6	0.8	0.618904852
5000	50	0.6	1	0.619768937
5000	50	0.8	0.2	0.626762335
5000	50	0.8	0.6	0.614943561
5000	50	0.8	0.8	0.623616703
5000	50	0.8	1	0.616426369
5000	50	0.9	0.2	0.621623748
5000	50	0.9	0.6	0.625351579
5000	50	0.9	0.8	0.6172419
5000	50	0.9	1	0.617499638
5000	200	0.6	0.2	0.621955435
5000	200	0.6	0.6	0.620637328
5000	200	0.6	0.8	0.619698581
5000	200	0.6	1	0.620325929
5000	200	0.8	0.2	0.622251601
5000	200	0.8	0.6	0.616960181
5000	200	0.8	0.8	0.620868864
5000	200	0.8	1	0.618358172
5000	200	0.9	0.2	0.618796527
5000	200	0.9	0.6	0.615195989
5000	200	0.9	0.8	0.619437709
5000	200	0.9	1	0.622448127
5000	400	0.6	0.2	0.61924799
5000	400	0.6	0.6	0.622873362
5000	400	0.6	0.8	0.620929269
5000	400	0.6	1	0.619045339
5000	400	0.8	0.2	0.621369658
5000	400	0.8	0.6	0.622404061
5000	400	0.8	0.8	0.62090362
5000	400	0.8	1	0.619453185
5000	400	0.9	0.2	0.619984787
5000	400	0.9	0.6	0.6185863
5000	400	0.9	0.8	0.623258898
5000	400	0.9	1	0.621024525

Table B.7: **95% Confidence Intervals of Average Distortions for the five random projection (RP) methods employed with dimensions, $d = 500$ across different densities (δ) and epsilon (ϵ).**

δ			0.2	0.6	0.8	1
n	Method	epsilon	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)
50	NRP	0.6	15.15 $\pm$ 0.0196	15.15 $\pm$ 0.0196	15.15 $\pm$ 0.0176	15.15 $\pm$ 0.0216
	PMRP		15.15 $\pm$ 0.0196	15.16 $\pm$ 0.0196	15.17 $\pm$ 0.0176	15.16 $\pm$ 0.0176
	HRP		10.71 $\pm$ 0.0137	10.71 $\pm$ 0.0137	10.72 $\pm$ 0.0137	10.71 $\pm$ 0.0137
	P_NHRP		10.17 $\pm$ 0.1352	10.16 $\pm$ 0.1294	10.11 $\pm$ 0.1450	10.11 $\pm$ 0.1235
	T_NHRP		7.85 $\pm$ 0.0098	7.84 $\pm$ 0.0098	7.85 $\pm$ 0.0098	7.85 $\pm$ 0.0098
	NRP	0.8	12.86 $\pm$ 0.0216	12.89 $\pm$ 0.0235	12.87 $\pm$ 0.0196	12.87 $\pm$ 0.0196
	PMRP		12.87 $\pm$ 0.0196	12.87 $\pm$ 0.0196	12.88 $\pm$ 0.0196	12.85 $\pm$ 0.0196
	HRP		9.09 $\pm$ 0.0137	9.11 $\pm$ 0.0157	9.11 $\pm$ 0.0137	9.1 $\pm$ 0.0137
	P_NHRP		8.65 $\pm$ 0.1117	8.64 $\pm$ 0.1137	8.62 $\pm$ 0.1254	8.56 $\pm$ 0.1176
	T_NHRP		6.66 $\pm$ 0.0098	6.67 $\pm$ 0.0118	6.67 $\pm$ 0.0098	6.66 $\pm$ 0.0098
	NRP	0.9	12.35 $\pm$ 0.0196	12.35 $\pm$ 0.0216	12.34 $\pm$ 0.0196	12.34 $\pm$ 0.0196
	PMRP		12.35 $\pm$ 0.0216	12.36 $\pm$ 0.0196	12.36 $\pm$ 0.0216	12.35 $\pm$ 0.0216
	HRP		8.73 $\pm$ 0.0137	8.72 $\pm$ 0.0176	8.73 $\pm$ 0.0157	8.73 $\pm$ 0.0137
	P_NHRP		8.33 $\pm$ 0.1078	8.26 $\pm$ 0.1058	8.27 $\pm$ 0.1196	8.23 $\pm$ 0.1156
	T_NHRP		6.4 $\pm$ 0.0098	6.39 $\pm$ 0.0118	6.4 $\pm$ 0.0098	6.4 $\pm$ 0.0098
200	NRP	0.6	15.15 $\pm$ 0.0118	15.15 $\pm$ 0.0118	15.14 $\pm$ 0.0118	15.15 $\pm$ 0.0118
	PMRP		15.15 $\pm$ 0.0098	15.15 $\pm$ 0.0098	15.15 $\pm$ 0.0098	15.16 $\pm$ 0.0098
	HRP		10.71 $\pm$ 0.0078	10.71 $\pm$ 0.0078	10.71 $\pm$ 0.0078	10.71 $\pm$ 0.0078
	P_NHRP		10.15 $\pm$ 0.0745	10.17 $\pm$ 0.0627	10.06 $\pm$ 0.0666	10.19 $\pm$ 0.0745
	T_NHRP		7.85 $\pm$ 0.0059	7.85 $\pm$ 0.0059	7.85 $\pm$ 0.0059	7.85 $\pm$ 0.0059
	NRP	0.8	12.85 $\pm$ 0.0118	12.87 $\pm$ 0.0118	12.86 $\pm$ 0.0118	12.86 $\pm$ 0.0118
	PMRP		12.87 $\pm$ 0.0098	12.87 $\pm$ 0.0078	12.87 $\pm$ 0.0078	12.86 $\pm$ 0.0098
	HRP		9.09 $\pm$ 0.0078	9.1 $\pm$ 0.0078	9.1 $\pm$ 0.0078	9.1 $\pm$ 0.0078
	P_NHRP		8.63 $\pm$ 0.0647	8.57 $\pm$ 0.0568	8.58 $\pm$ 0.0588	8.55 $\pm$ 0.0627
	T_NHRP		6.66 $\pm$ 0.0059	6.67 $\pm$ 0.0059	6.66 $\pm$ 0.0059	6.66 $\pm$ 0.0059
	NRP	0.9	12.35 $\pm$ 0.0118	12.36 $\pm$ 0.0137	12.35 $\pm$ 0.0118	12.35 $\pm$ 0.0118
	PMRP		12.35 $\pm$ 0.0098	12.35 $\pm$ 0.0098	12.35 $\pm$ 0.0098	12.35 $\pm$ 0.0098
	HRP		8.73 $\pm$ 0.0078	8.73 $\pm$ 0.0078	8.73 $\pm$ 0.0078	8.74 $\pm$ 0.0078
	P_NHRP		8.28 $\pm$ 0.0627	8.21 $\pm$ 0.0568	8.26 $\pm$ 0.0549	8.23 $\pm$ 0.0608
	T_NHRP		6.4 $\pm$ 0.0059	6.4 $\pm$ 0.0059	6.4 $\pm$ 0.0059	6.4 $\pm$ 0.0059
400	NRP	0.6	15.15 $\pm$ 0.0078	15.16 $\pm$ 0.0098	15.16 $\pm$ 0.0098	15.15 $\pm$ 0.0078
	PMRP		15.15 $\pm$ 0.0078	15.15 $\pm$ 0.0078	15.15 $\pm$ 0.0078	15.15 $\pm$ 0.0059
	HRP		10.71 $\pm$ 0.0059	10.72 $\pm$ 0.0059	10.71 $\pm$ 0.0059	10.71 $\pm$ 0.0059
	P_NHRP		10.14 $\pm$ 0.049	10.11 $\pm$ 0.0451	10.12 $\pm$ 0.051	10.13 $\pm$ 0.0470
	T_NHRP		7.85 $\pm$ 0.0039	7.85 $\pm$ 0.0039	7.85 $\pm$ 0.0039	7.85 $\pm$ 0.0039
	NRP	0.8	12.88 $\pm$ 0.0098	12.86 $\pm$ 0.0098	12.87 $\pm$ 0.0098	12.86 $\pm$ 0.0078
	PMRP		12.87 $\pm$ 0.0059	12.86 $\pm$ 0.0059	12.86 $\pm$ 0.0059	12.86 $\pm$ 0.0078
	HRP		9.1 $\pm$ 0.059	9.1 $\pm$ 0.059	9.1 $\pm$ 0.059	9.09 $\pm$ 0.059
	P_NHRP		8.63 $\pm$ 0.0412	8.6 $\pm$ 0.0353	8.6 $\pm$ 0.0372	8.61 $\pm$ 0.0431
	T_NHRP		6.67 $\pm$ 0.0039	6.67 $\pm$ 0.0039	6.67 $\pm$ 0.0039	6.66 $\pm$ 0.0039
	NRP	0.9	12.35 $\pm$ 0.0098	12.35 $\pm$ 0.0078	12.36 $\pm$ 0.0078	12.35 $\pm$ 0.0098
	PMRP		12.35 $\pm$ 0.0078	12.35 $\pm$ 0.0078	12.35 $\pm$ 0.0078	12.35 $\pm$ 0.0078
	HRP		8.74 $\pm$ 0.059	8.73 $\pm$ 0.059	8.73 $\pm$ 0.078	8.74 $\pm$ 0.059
	P_NHRP		8.27 $\pm$ 0.0392	8.27 $\pm$ 0.0333	8.27 $\pm$ 0.0392	8.27 $\pm$ 0.0412
	T_NHRP		6.4 $\pm$ 0.0039	6.4 $\pm$ 0.0039	6.4 $\pm$ 0.0039	6.4 $\pm$ 0.0039

Table B.8: **95% Confidence Intervals of Average Distortions for the five RP methods employed with dimensions, $d = 1000$ across different densities (δ) and epsilon (ϵ)**

δ			0.2	0.6	0.8	1
n	Method	epsilon	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)
50	NRP	0.6	15.95 $\pm$ 0.0176	15.94 $\pm$ 0.0235	15.94 $\pm$ 0.0196	15.94 $\pm$ 0.0196
	PMRP		15.95 $\pm$ 0.0196	15.96 $\pm$ 0.0196	15.95 $\pm$ 0.0196	15.94 $\pm$ 0.0196
	HRP		11.28 $\pm$ 0.0137	11.27 $\pm$ 0.0157	11.28 $\pm$ 0.0118	11.27 $\pm$ 0.0137
	P_NHRP		10.71 $\pm$ 0.1392	10.65 $\pm$ 0.1254	10.78 $\pm$ 0.1352	10.59 $\pm$ 0.149
	T_NHRP		8.26 $\pm$ 0.0098	8.25 $\pm$ 0.0118	8.26 $\pm$ 0.0098	8.26 $\pm$ 0.0098
	NRP	0.8	13.59 $\pm$ 0.0176	13.58 $\pm$ 0.0176	13.6 $\pm$ 0.0176	13.58 $\pm$ 0.0196
	PMRP		13.58 $\pm$ 0.0235	13.59 $\pm$ 0.0216	13.59 $\pm$ 0.0216	13.59 $\pm$ 0.0216
	HRP		9.61 $\pm$ 0.0176	9.62 $\pm$ 0.0137	9.61 $\pm$ 0.0137	9.61 $\pm$ 0.0137
	P_NHRP		9.09 $\pm$ 0.1196	9.08 $\pm$ 0.1176	9.05 $\pm$ 0.1372	8.98 $\pm$ 0.1274
	T_NHRP		7.04 $\pm$ 0.1176	7.04 $\pm$ 0.0098	7.04 $\pm$ 0.0098	7.04 $\pm$ 0.0098
	NRP	0.9	13.03 $\pm$ 0.0196	13.02 $\pm$ 0.0196	13.02 $\pm$ 0.0216	13 $\pm$ 0.0216
	PMRP		13.02 $\pm$ 0.0196	13.02 $\pm$ 0.0176	13 $\pm$ 0.0196	13.01 $\pm$ 0.0157
	HRP		9.2 $\pm$ 0.0137	9.22 $\pm$ 0.0157	9.2 $\pm$ 0.0118	9.19 $\pm$ 0.0137
	P_NHRP		8.7 $\pm$ 0.1137	8.69 $\pm$ 0.1098	8.68 $\pm$ 0.1333	8.58 $\pm$ 0.1137
	T_NHRP		6.74 $\pm$ 0.1176	6.75 $\pm$ 0.1176	6.74 $\pm$ 0.0098	6.73 $\pm$ 0.0098
200	NRP	0.6	15.95 $\pm$ 0.0176	15.94 $\pm$ 0.0235	15.94 $\pm$ 0.0196	15.94 $\pm$ 0.0196
	PMRP		15.95 $\pm$ 0.0196	15.96 $\pm$ 0.0196	15.95 $\pm$ 0.0196	15.94 $\pm$ 0.0196
	HRP		11.28 $\pm$ 0.0137	11.27 $\pm$ 0.0157	11.28 $\pm$ 0.0118	11.27 $\pm$ 0.0137
	P_NHRP		10.71 $\pm$ 0.1392	10.65 $\pm$ 0.1254	10.78 $\pm$ 0.1352	10.59 $\pm$ 0.149
	T_NHRP		8.26 $\pm$ 0.0098	8.25 $\pm$ 0.0118	8.26 $\pm$ 0.0098	8.26 $\pm$ 0.0118
	NRP	0.8	13.59 $\pm$ 0.0176	13.58 $\pm$ 0.0176	13.6 $\pm$ 0.0176	13.58 $\pm$ 0.0196
	PMRP		13.58 $\pm$ 0.0235	13.59 $\pm$ 0.0216	13.59 $\pm$ 0.0216	13.59 $\pm$ 0.0216
	HRP		9.61 $\pm$ 0.0176	9.62 $\pm$ 0.0137	9.61 $\pm$ 0.0137	9.61 $\pm$ 0.0137
	P_NHRP		9.09 $\pm$ 0.1196	9.08 $\pm$ 0.1176	9.05 $\pm$ 0.1196	8.89 $\pm$ 0.1274
	T_NHRP		7.04 $\pm$ 0.0118	7.04 $\pm$ 0.0098	7.04 $\pm$ 0.0098	7.04 $\pm$ 0.0098
	NRP	0.9	13.03 $\pm$ 0.0196	13.02 $\pm$ 0.0196	13.02 $\pm$ 0.0216	13.02 $\pm$ 0.0216
	PMRP		13.02 $\pm$ 0.0196	13.02 $\pm$ 0.0176	13 $\pm$ 0.0196	13.01 $\pm$ 0.0157
	HRP		9.2 $\pm$ 0.0137	9.22 $\pm$ 0.0157	9.2 $\pm$ 0.0118	9.19 $\pm$ 0.0137
	P_NHRP		8.7 $\pm$ 0.1137	8.69 $\pm$ 0.1098	8.68 $\pm$ 0.1333	8.58 $\pm$ 0.1137
	T_NHRP		6.74 $\pm$ 0.0118	6.75 $\pm$ 0.0118	6.74 $\pm$ 0.0098	6.73 $\pm$ 0.0098
400	NRP	0.6	15.95 $\pm$ 0.0098	15.96 $\pm$ 0.0078	15.95 $\pm$ 0.0078	15.95 $\pm$ 0.0098
	PMRP		15.95 $\pm$ 0.0059	15.96 $\pm$ 0.0078	15.96 $\pm$ 0.0078	15.96 $\pm$ 0.0059
	HRP		11.28 $\pm$ 0.0059	11.29 $\pm$ 0.0059	11.28 $\pm$ 0.0059	11.28 $\pm$ 0.0059
	P_NHRP		10.64 $\pm$ 0.0470	10.69 $\pm$ 0.0470	10.65 $\pm$ 0.0470	10.65 $\pm$ 0.049
	T_NHRP		8.26 $\pm$ 0.0039	8.27 $\pm$ 0.0039	8.27 $\pm$ 0.0039	8.26 $\pm$ 0.0039
	NRP	0.8	13.58 $\pm$ 0.0098	13.58 $\pm$ 0.0078	13.58 $\pm$ 0.0078	13.59 $\pm$ 0.0078
	PMRP		13.58 $\pm$ 0.0078	13.58 $\pm$ 0.0078	13.59 $\pm$ 0.0078	13.58 $\pm$ 0.0078
	HRP		9.6 $\pm$ 0.0059	9.6 $\pm$ 0.0059	9.61 $\pm$ 0.0059	9.61 $\pm$ 0.0059
	P_NHRP		9.05 $\pm$ 0.0470	9.09 $\pm$ 0.0412	9.09 $\pm$ 0.0392	9.04 $\pm$ 0.0431
	T_NHRP		7.04 $\pm$ 0.0039	7.04 $\pm$ 0.0039	7.04 $\pm$ 0.0039	7.04 $\pm$ 0.0039
	NRP	0.9	13.02 $\pm$ 0.0078	13.02 $\pm$ 0.0078	13.02 $\pm$ 0.0078	13.02 $\pm$ 0.0078
	PMRP		13.02 $\pm$ 0.0059	13.01 $\pm$ 0.0059	13.02 $\pm$ 0.0059	13.02 $\pm$ 0.0078
	HRP		9.2 $\pm$ 0.0059	9.2 $\pm$ 0.0059	9.2 $\pm$ 0.0059	9.2 $\pm$ 0.0059
	P_NHRP		8.69 $\pm$ 0.0451	8.73 $\pm$ 0.0412	8.71 $\pm$ 0.0372	8.65 $\pm$ 0.0412
	T_NHRP		6.74 $\pm$ 0.0039	6.74 $\pm$ 0.0039	6.74 $\pm$ 0.0039	6.74 $\pm$ 0.0059

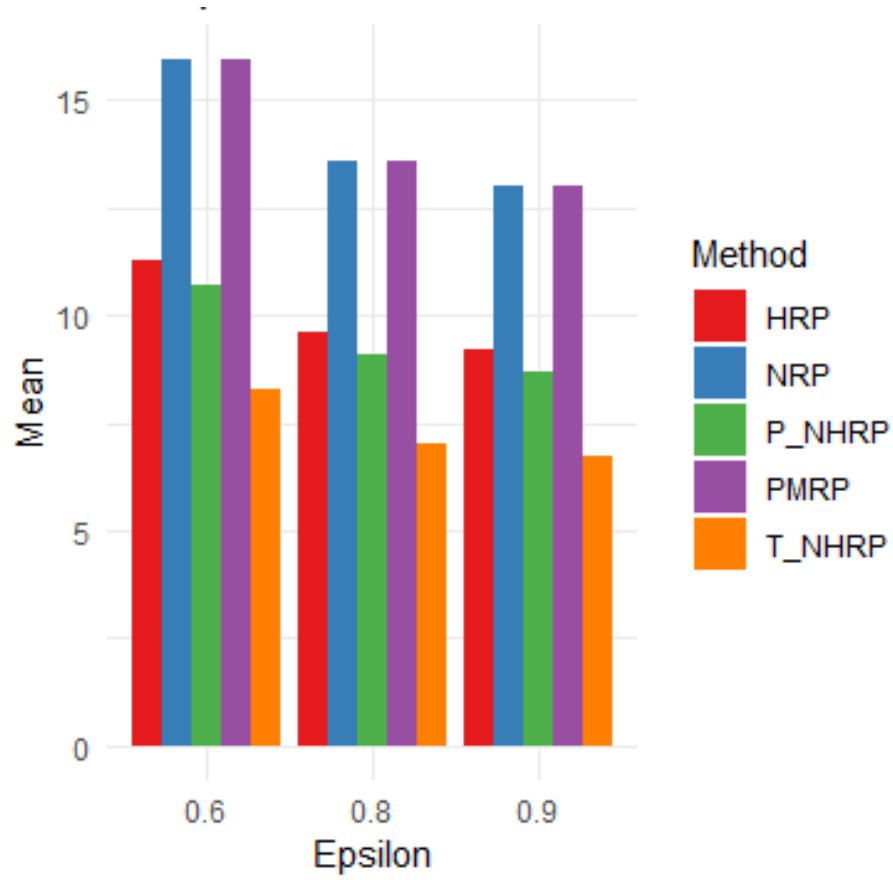


Figure B.1: Comparison of Methods across Different Epsilons for $d = 1000$

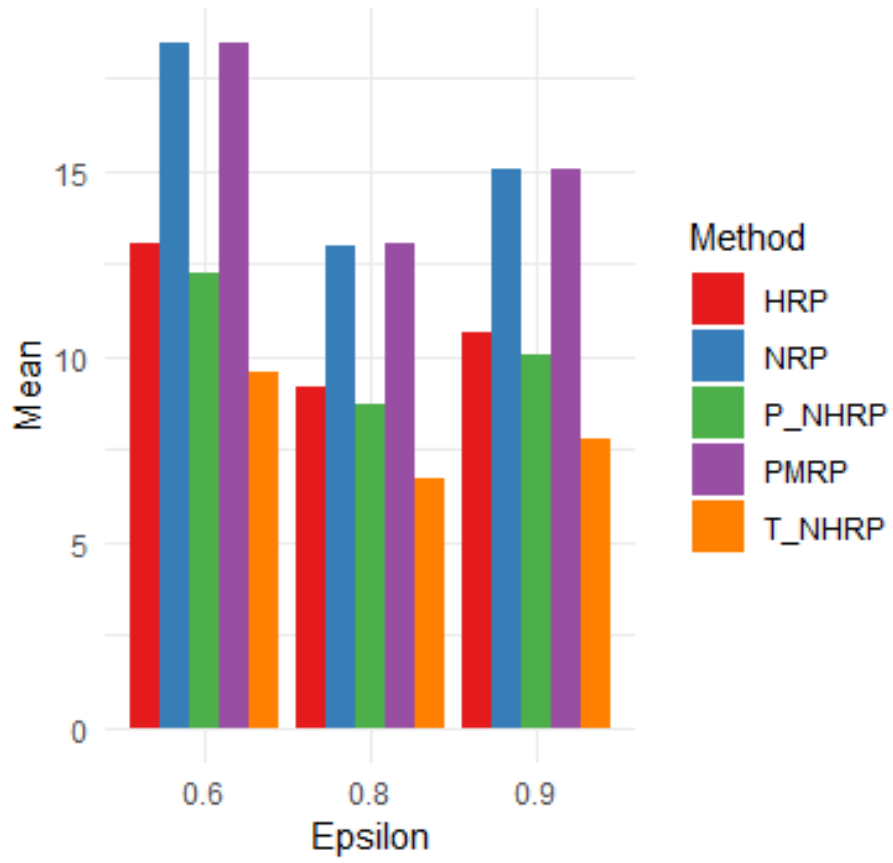


Figure B.2: Comparison of Methods across Different Epsilons for $d = 10000$

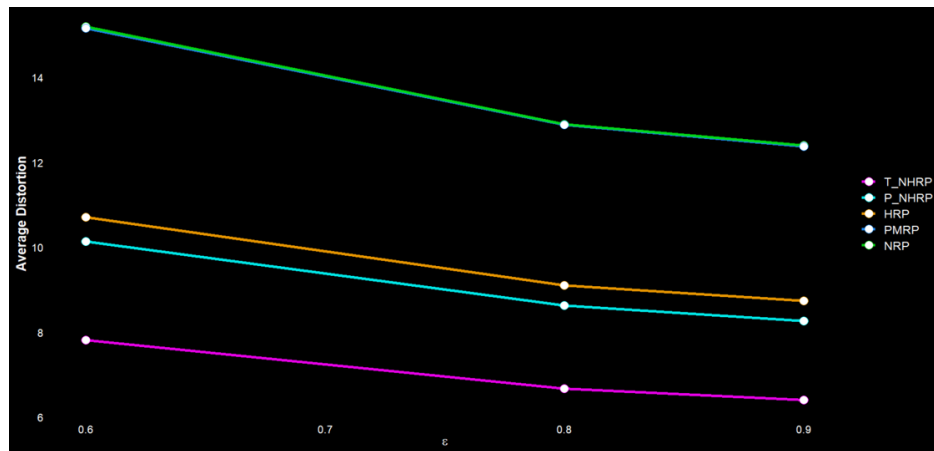


Figure B.3: Average Distortion for the Random projection (RP) methods employed with $d = 1000$ and $n=50$

REFERENCES

- [1] Y. Yahaya, J. O. Ajadi, R. A. Sanusi, Z. Sawlan, and N. A. Adegoke, “Hybrid random projection technique for enhanced representation in high-dimensional data,” *Expert Systems with Applications*, vol. 262, p. 125569, 2025.
- [2] W. B. Johnson, “Extensions of lipshitz mapping into hilbert space,” in *Conference modern analysis and probability, 1984*, 1984, pp. 189–206.
- [3] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, “On the surprising behavior of distance metrics in high-dimensional space,” in *Proceedings of the International Conference on Database Theory*, 2001, pp. 420–434.
- [4] E. Bingham and H. Mannila, “Random projection in dimensionality reduction: Applications to image and text data,” in *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2001, pp. 245–250.
- [5] P. Li, T. Hastie, and K. W. Church, “Very sparse random projections,” in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 287–296.
- [6] S. Vempala, *The random projection method*. American Mathematical Soc., 2005, vol. 65.

- [7] D. Achlioptas, “Database-friendly random projections: Johnson-lindenstrauss with binary coins,” *Journal of Computer and System Sciences*, vol. 66, no. 4, pp. 671–687, 2003.
- [8] S. Dasgupta and A. Gupta, “An elementary proof of the johnson-lindenstrauss lemma,” International Computer Science Institute, Tech. Rep. 99(006), 2003.
- [9] A. Hinneburg and D. A. Keim, “Optimal grid-clustering: Towards breaking the curse of dimensionality in high-dimensional clustering,” in *Proceedings of the 25th International Conference on Very Large Data Bases*, 2003, pp. 506–517.
- [10] S. Papadimitriou, P. Raghavan, H. Tamaki, and S. Vempala, “Latent semantic indexing: A probabilistic analysis,” in *Proceedings of the twenty-first annual international ACM SIGIR conference on Research and development in information retrieval*, 1998, pp. 50–57.
- [11] S. Kaski, “Dimensionality reduction by random mapping: Fast similarity computation for clustering,” in *Proceedings of the 1998 IEEE International Joint Conference on Neural Networks*, vol. 1, 1998, pp. 413–418.
- [12] S. Dasgupta, “Learning mixtures of gaussians,” in *Proceedings of the 40th Annual Symposium on Foundations of Computer Science*, 1999, pp. 634–644.
- [13] X. Z. Fern and C. E. Brodley, “Random projection for high dimensional data clustering: A cluster ensemble approach,” in *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, 2003, pp. 186–193.

- [14] Y. Yu, F. Huang, and Q. Gu, “Compressive sensing for robust multi-view subspace learning,” in *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, 2015, pp. 1930–1939.
- [15] Y. Yahaya, J. O. Ajadi, R. A. Sanusi, Z. Sawlan *et al.*, “Hybrid random projection: Integrating dense and sparse techniques for enhanced representation in high-dimensional data,” *Expert Systems with Applications*, 2024.
- [16] A. L. Samuel, “Some studies in machine learning using the game of checkers,” *IBM Journal of Research and Development*, 1959.
- [17] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [18] P. Schlexer *et al.*, “Machine learning for heterogeneous catalyst design and discovery,” *Nature Reviews Materials*, 2019.
- [19] B. R. Goldsmith *et al.*, “Machine learning for heterogeneous catalysis,” *Annual Review of Chemical and Biomolecular Engineering*, 2018.
- [20] A. Esteva *et al.*, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, 2017.
- [21] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” *Nature Medicine*, 2019.
- [22] P. Rajpurkar *et al.*, “Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning,” *arXiv preprint arXiv:1711.05225*, 2017.
- [23] V. I. Jurtz *et al.*, “An introduction to deep learning on biological sequence data: examples and solutions,” *Bioinformatics*, 2017.

- [24] C. Angermueller *et al.*, “Deep learning for computational biology,” *Molecular Systems Biology*, 2016.
- [25] A. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” *Computers and Electronics in Agriculture*, 2018.
- [26] B. Chlingaryan, A. Sukkarieh S. & Whelan, “Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review,” *Computers and Electronics in Agriculture*, 2018.
- [27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems* 25, 2012, pp. 1097–1105.
- [28] L. Y. D. e. a. Hinton, G. E. Deng, “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,” *IEEE Signal Processing Magazine*, 2012.
- [29] F. Ahmed, T. Mahmood, M. Riaz, and N. Abbas, “Comprehensive review of high-dimensional monitoring methods: trends, insights, and interconnections,” *Quality Technology & Quantitative Management*, pp. 1–25, 2024.
- [30] J. Fan, F. Han, and H. Liu, “Challenges of big data analysis,” *National science review*, vol. 1, no. 2, pp. 293–314, 2014.
- [31] V. Tercero-Gómez, V. Aguilar-Lleyda, A. Cordero-Franco, and W. Conover, “A distribution-free cusum chart for joint monitoring of location and scale based on the combination of wilcoxon and mood statistics,” *Quality and Reliability Engineering International*, vol. 36, no. 4, pp. 1422–1453, 2020.

- [32] M. R. Ahmad and S. E. Ahmed, “On the distribution of the t^2 statistic, used in statistical process monitoring, for high-dimensional data,” *Statistics & Probability Letters*, vol. 168, p. 108919, 2021.
- [33] G. Liu, C. Kan, Y. Chen, and H. Yang, “Model-driven parametric monitoring of high-dimensional nonlinear functional profiles,” in *2014 IEEE international conference on automation science and engineering (CASE)*. IEEE, 2014, pp. 722–727.
- [34] Y. Zhang and Y. Zhang, “Process monitoring, fault diagnosis and quality prediction methods based on the multivariate statistical techniques,” *IETE Technical Review*, vol. 27, no. 5, pp. 406–420, 2010.
- [35] S. W. Choi, C. Lee, J.-M. Lee, J. H. Park, and I.-B. Lee, “Fault detection and identification of nonlinear processes based on kernel pca,” *Chemometrics and intelligent laboratory systems*, vol. 75, no. 1, pp. 55–67, 2005.
- [36] D. Sachin *et al.*, “Dimensionality reduction and classification through pca and lda,” *International journal of computer Applications*, vol. 122, no. 17, 2015.
- [37] D. Maulud and A. M. Abdulazeez, “A review on linear regression comprehensive in machine learning,” *Journal of Applied Science and Technology Trends*, vol. 1, no. 2, pp. 140–147, 2020.
- [38] X. Wu, Z. Chu, X. ZHENG, X. Wang, and P. Yang, “Sparse random projection compressive data gathering in lossy wireless sensor networks,” *Chinese Journal of Computers*, vol. 41, no. 53, pp. 1–16, 2018.
- [39] W. B. Johnson, J. Lindenstrauss, and G. Schechtman, “Extensions of lipschitz maps into banach spaces,” *Israel Journal of Mathematics*, vol. 54, no. 2, pp. 129–138, 1986.

- [40] E. Bingham and H. Mannila, “Random projection in dimensionality reduction: applications to image and text data,” in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 2001, pp. 245–250.
- [41] T. Mahmood, F. Ahmed, M. Riaz, and N. Abbas, “High-dimensional control charts with application to surveillance of grease damage in bearings of wind turbines,” *Production & Manufacturing Research*, vol. 12, no. 1, p. 2377739, 2024.
- [42] B. Iglewicz and D. C. Hoaglin, “How to detect and handle outliers (the asqc basic reference in quality control),” 1993.
- [43] I. T. Jolliffe and J. Cadima, “Principal component analysis: a review and recent developments,” *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, 2016.
- [44] S. Wold, K. Esbensen, and P. Geladi, *Principal Component Analysis*. Springer, 1987, no. 1-3.
- [45] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
- [46] R. P. McDonald, *Factor analysis and related methods*. Psychology Press, 2014.
- [47] I. Borg and P. Groenen, *Modern Multidimensional Scaling: Theory and Applications*. Springer, 2005.
- [48] A. Hyvärinen and E. Oja, “Independent component analysis: Algorithms and applications,” *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

- [49] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, “Canonical correlation analysis: An overview with application to learning methods,” *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, 2004.
- [50] E. Candès, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [51] L. Van der Maaten and G. Hinton, “Visualizing data using t-sne.” *Journal of machine learning research*, vol. 9, no. 11, 2008.
- [52] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [53] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [54] A. C. Belkina, C. O. Ciccolella, R. Anno, R. Halpert, J. Spidlen, and J. E. Snyder-Cappione, “Automated optimized parameters for t-distributed stochastic neighbor embedding improve visualization and analysis of large datasets,” *Nature communications*, vol. 10, no. 1, p. 5415, 2019.
- [55] M. Leland, H. John, and M. James, “Uniform manifold approximation and projection for dimension reduction,” *arXiv preprint arXiv:1802.03426*, 2018.
- [56] E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. Kwok, L. G. Ng, F. Ginhoux, and E. W. Newell, “Dimensionality reduction for visualizing single-cell data using umap,” *Nature biotechnology*, vol. 37, no. 1, pp. 38–44, 2019.

- [57] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” *arXiv preprint arXiv:1312.6114*, 2013.
- [58] A. Ali, M. Hamraz, D. M. Khan, W. Deebani, and Z. Khan, “A random projection k nearest neighbours ensemble for classification via extended neighbourhood rule,” *arXiv preprint arXiv:2303.12210*, 2023.
- [59] I. Atsonios, “Phoneme recognition using sparse random projections and ensemble classifiers,” in *Advances in Nonlinear Speech Processing: International Conference on Nonlinear Speech Processing, NOLISP 2009, Vic, Spain, June 25-27. Revised Selected Papers*. Springer, 2010, pp. 191–198.
- [60] J. Butcher, D. Verstraeten, B. Schrauwen, C. Day, and P. Haycock, “Extending reservoir computing with random static projections: a hybrid between extreme learning and rc,” in *18th European Symposium on Artificial Neural Networks (ESANN 2010)*. D-Side, 2010, pp. 303–308.
- [61] S. Mirniaharikandehei, M. Heidari, G. Danala, S. Lakshmivarahan, and B. Zheng, “Applying a random projection algorithm to optimize machine learning model for predicting peritoneal metastasis in gastric cancer patients using ct images,” *Computer methods and programs in biomedicine*, vol. 200, p. 105937, 2021.
- [62] V. Poornima *et al.*, “A hybrid model for prediction of diabetes using machine learning classification algorithms and random projection,” 2023.
- [63] A. M. Ibrahim, M. Farouk, and M. W. Fakhr, “Privacy preserving image retrieval using multi-key random projection encryption and machine learning decryption,” *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 42, no. 2, pp. 155–174, 2024.

- [64] S. M. Abd-Alhalem, H. S. Marie, W. El-Shafai, T. Altameem, R. S. Rathore, and T. M. Hassan, “Cervical cancer classification based on a bilinear convolutional neural network approach and random projection,” *Engineering Applications of Artificial Intelligence*, vol. 127, p. 107261, 2024.
- [65] M. Zhu and S. Ermon, “A hybrid approach for probabilistic inference using random projections,” in *International Conference on Machine Learning*. PMLR, 2015, pp. 2039–2047.
- [66] C. K. I. Williams, “Using the nyström method to speed up kernel machines,” *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 13, pp. 682–688, 2001.
- [67] A. Grover and J. Leskovec, “node2vec: Scalable feature learning for networks,” in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016, pp. 855–864.
- [68] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” *arXiv preprint arXiv:1609.02907*, 2016.
- [69] N. Goyal, A. Goel, T. Garg, N. Sachdeva, and P. Kumaraguru, “Efficient knowledge graph embeddings via kernelized random projections,” in *International Conference on Big Data Analytics*. Springer, 2023, pp. 198–209.
- [70] J. Hu, B. Hooi, and B. He, “Efficient heterogeneous graph learning via random projection,” *IEEE Transactions on Knowledge and Data Engineering*, 2024.
- [71] H. Chen, S. F. Sultan, Y. Tian, M. Chen, and S. Skiena, “Fast and accurate network embeddings via very sparse random projection,” in *Proceedings of the 28th ACM international conference on information and knowledge management*, 2019, pp. 399–408.

- [72] B. Raducanu and F. Dornaika, “A supervised non-linear dimensionality reduction approach for manifold learning,” *Pattern Recognition*, vol. 45, no. 6, pp. 2432–2444, 2012.
- [73] F. Luo, L. Zhang, B. Du, and L. Zhang, “Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5336–5353, 2020.
- [74] L. F. R. Ribeiro, P. H. P. Saverese, and D. R. Figueiredo, “struc2vec: Learning node representations from structural identity,” in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2017, pp. 385–394.
- [75] B. Perozzi, R. Al-Rfou, and S. Skiena, “Deepwalk: Online learning of social representations,” in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014, pp. 701–710.
- [76] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [77] L. Cheng and C. You, “Hybrid non-linear dimensionality reduction method framework based on random projections,” in *2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*. IEEE, 2016, pp. 43–48.
- [78] A. Mohan and M. Venkatesan, “Hybrid dimensionality reduction technique for hyperspectral images using random projection and manifold learning,” in *Trends and Applications in Knowledge Discovery and Data Mining: PAKDD 2020 Workshops, DSFN, GII, BDM, LDRC and LBD, Singapore, May 11–14, 2020, Revised Selected Papers 24*. Springer, 2020, pp. 116–127.

- [79] C. Ayaz, L. Scalfi, B. A. Dalton, and R. R. Netz, “Generalized langevin equation with a nonlinear potential of mean force and nonlinear memory friction from a hybrid projection scheme,” *Physical review E*, vol. 105, no. 5, p. 054138, 2022.
- [80] I. Lahmar, A. Zaier, M. Yahia, and R. Boaullegue, “Fuzzy kernel weighted random projection ensemble clustering for high dimensional data,” in *International Conference on Hybrid Intelligent Systems*. Springer, 2022, pp. 177–187.
- [81] B. Yang and Z. Peng, “A hybrid dimensionality reduction technique for hyperspectral images using random projection and manifold learning,” *Journal of Applied Remote Sensing*, vol. 11, no. 3, pp. 243–252, 2020.
- [82] A. Ashok and M. A. Neifeld, “Compressive imaging: Hybrid projection design,” in *Imaging Systems*. Optica Publishing Group, 2010, p. IWD3.
- [83] B. Schölkopf, A. Smola, and K.-R. Müller, “Nonlinear component analysis as a kernel eigenvalue problem,” *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [84] D. Achlioptas, “Database-friendly random projections: Johnson-lindenstrauss with binary coins,” *Journal of Computer and System Sciences*, vol. 66, no. 4, pp. 671–687, 2003.
- [85] P. Li, T. J. Hastie, and K. W. Church, “Very sparse random projections,” in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2007, pp. 287–296.
- [86] G. Yu, Y. Xia, J. Li, and H. Bai, “Discriminative random projections for face recognition,” *Pattern Recognition Letters*, vol. 79, pp. 67–72, 2016.

- [87] M. S. Charikar, “Similarity estimation techniques from rounding algorithms,” in *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC)*, 2002, pp. 380–388.
- [88] N. Goyal and E. Ferrara, “Graph embedding techniques, applications, and performance: A survey,” *Knowledge-Based Systems*, vol. 151, pp. 78–94, 2018.
- [89] W. L. Hamilton, R. Ying, and J. Leskovec, “Representation learning on graphs: Methods and applications,” *arXiv preprint arXiv:1709.05584*, 2017.
- [90] P. Dhal and C. Azad, “A comprehensive survey on feature selection in the various fields of machine learning,” *Applied Intelligence*, vol. 52, no. 4, pp. 4543–4581, 2022.
- [91] P. Li, T. J. Hastie, and K. W. Church, “Very fast random projections,” *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 249–258, 2006.
- [92] S. Dasgupta, “Experiments with random projection,” in *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2002, pp. 143–151.
- [93] N. Ailon and B. Chazelle, “The fast johnson-lindenstrauss transform and approximate nearest neighbors,” *SIAM Journal on Computing*, vol. 39, no. 1, pp. 302–322, 2009.
- [94] S. Dasgupta, “Experiments with random projection,” in *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI2000)*, 2000, pp. 143–151.
- [95] J. He, R. Lawrence, and L. Saul, “Random projections for manifold learning,” *Journal of Machine Learning Research*, vol. 13, pp. 1351–1371, 2012.

- [96] R. I. Arriaga and S. Vempala, “An algorithmic theory of learning: Robust concepts and random projection,” in *Proceedings of the 40th Annual Symposium on Foundations of Computer Science (FOCS)*, 1999, pp. 616–623.
- [97] R. Hamid, Y. Fu, A. Martinez, and J. Ho, “Compact random features for fast machine learning on large datasets,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1854–1866, 2011.
- [98] H. Shashidhara, P. Joseph, and K. Srinivasa, “Improving motif refinement using hybrid expectation maximization and random projection,” in *Proceedings of the International Symposium on Biocomputing*, 2010, pp. 1–7.
- [99] H. Xie, J. Li, and H. Xue, “A survey of dimensionality reduction techniques based on random projection,” *arXiv preprint arXiv:1706.04371*, 2017.
- [100] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2002.
- [101] A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, “Clustering-based hyperspectral band selection using information measures,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4158–4171, 2007.
- [102] Z. L. X. Zhao and H. Q., “A hybrid framework for dimensionality reduction in hyperspectral imaging,” *Journal of Remote Sensing*, vol. 14, no. 4, pp. 765–780, 2022.
- [103] L. Cheng, C. You, and Y. Guan, “Random projections for non-linear dimensionality reduction,” *Journal of Machine Learning Research*, vol. 17, no. 1, pp. 1–25, 2016.

- [104] R. Clarke, H. W. Ressom, A. Wang, J. Xuan, M. C. Liu, E. A. Gehan, and Y. Wang, “The properties of high-dimensional data spaces: implications for exploring gene and protein expression data,” *Nature reviews cancer*, vol. 8, no. 1, pp. 37–49, 2008.
- [105] X. Desquesnes, A. Elmoataz, O. Lézoray, and V.-T. Ta, “Efficient algorithms for image and high dimensional data processing using eikonal equation on graphs,” in *International Symposium on Visual Computing*. Springer, 2010, pp. 647–658.
- [106] L. Cheng, L. Wang, and F. Gao, “Power system fault classification method based on sparse representation and random dimensionality reduction projection,” in *2015 IEEE Power & Energy Society General Meeting*. IEEE, 2015, pp. 1–5.
- [107] L. Cheng and C. You, “Hybrid non-linear dimensionality reduction method framework based on random projections,” in *2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, 2016, pp. 43–48.
- [108] I. K. Fodor, “A survey of dimension reduction techniques,” Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States), Tech. Rep., 2002.
- [109] B. Denham, R. Pears, and M. A. Naeem, “Enhancing random projection with independent and cumulative additive noise for privacy-preserving data stream mining,” *Expert Systems with Applications*, vol. 152, p. 113380, 2020.
- [110] T. Mo, L. Wang, Y. Wu, J. Huang, W. Liu, R. Yang, and X. Zhen, “Classifier ensemble with evolutionary optimisation enforced random projections,” *Expert Systems with Applications*, vol. 222, p. 119845, 2023.

- [111] R. Siddharth and G. Aghila, “Randpro-a practical implementation of random projection-based feature extraction for high dimensional multivariate data analysis in r,” *SoftwareX*, vol. 12, p. 100629, 2020.
- [112] S. Kang and H.-S. Oh, “Spherical random projection,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, p. qkae035, 2024.
- [113] P. Indyk and R. Motwani, “Approximate nearest neighbors: Towards removing the curse of dimensionality, 1998,” *URL citeseer.ist.psu.edu/article/indyk98approximate.html*, pp. 604–613, 2023.
- [114] W.-Y. Li and S.-Z. Zhang, “Binary random projections with controllable sparsity patterns,” *Journal of the Operations Research Society of China*, vol. 10, no. 3, pp. 507–528, 2022.
- [115] P. I. Wójcik and M. Kurdziel, “Training neural networks on high-dimensional data using random projection,” *Pattern Analysis and Applications*, vol. 22, pp. 1221–1231, 2019.
- [116] C. Zhang, N. Chen, and J. Wu, “Spatial rank-based high-dimensional monitoring through random projection,” *Journal of Quality Technology*, vol. 52, no. 2, pp. 111–127, 2020.
- [117] K. Kang, “Correlations between random projections and the bivariate normal,” *Data Mining and Knowledge Discovery*, vol. 35, no. 4, pp. 1622–1653, 2021.
- [118] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, “A limited memory algorithm for bound constrained optimization,” *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995.

- [119] Gül and F. Rahim, “Period changer,” 2022.
- [120] F. R. Gül, “Toxicity,” <https://archive.ics.uci.edu/ml/datasets/Toxicity>, 2022, uCI Machine Learning Repository.
- [121] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, “A public domain dataset for human activity recognition using smartphones,” in *Esann*, vol. 3, 2013, p. 3.
- [122] H. Xiao, K. Rasul, and R. Vollgraf, “Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms,” *arXiv preprint*, vol. arXiv:1708.07747, 2017.

VITAE

Personal Information

Name: Adeniran Usman

Email: usmanadedeji97@gmail.com

Address: 9, Daddy Monsor Street, Lagos State, 100271, Nigeria

Phone Number: +2347067259634

Date of Birth: 13th April, 1997

LinkedIn: linkedin.com/in/usman-adeniran

Education

MSc. Applied Statistics (May, 2025)

King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia

BSc. Statistics (First Class Honours – 4.61/5.0) (August, 2021)

University of Ilorin, Ilorin, Nigeria

Work Experience

Graduate Teaching Assistant, King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia. (2024 – Present)

- Python lab instructor for 100 undergraduate students across four sections, teaching and guiding them on using Python programming for mathematical expressions in Calculus II

Graduate Teaching Assistant, King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia. (2023 – 2024)

- MATLAB assistant lab instructor for 25 undergraduate students across a section, teaching and guiding them on using MATLAB for mathematical expressions in Calculus I.

National Youth Service Corps, I.T Support Officer, Mewar International University, Kuchikau, Nigeria

- Provided IT support and data management services.

Tutor, Department of Statistics, University of Ilorin

- Conducted tutorial sessions for undergraduate students in core statistics courses.

Tutor, Faculty of Physical Science, University of Ilorin, Ilorin, Nigeria

- Provided mathematics and statistics tutorials to first- and second-year students.

Data Analyst, iSolution Data Consultant, Unilorin, Nigeria. (2020 – 2021)

- Managed data analysis projects, developed reports, and provided insights using statistical software.

Data Analyst Intern, KPMG Virtual Internship. (2020)

- Assisted in data processing and statistical modeling for business analytics.

Research Experience and Publications

Current Thesis Research

- *Nonlinear Hybrid Random Projections for High Dimensional Data (Under Review)*

Current Thesis Research

- *Hybrid Dimensionality Reduction Techniques Based On Random Projections: A Literature Review. (Under Review)*

Research

- *Effects of Weather Variation on Students' Academic Performance at King Fahd University of Petroleum and Minerals (Under Review)*

Research

- *Enhancing Economic Forecasting with Bayesian Neural Networks: A Focus on GDP Prediction and Uncertainty Quantification in Nigeria. (Under Review)*

Other Research Undertaken

- *Factorial Analysis of the Effect of NPK Fertilizer: An Application to the Yield of White Yam*
- *Statistical Analysis on Coronavirus (Covid-19) Outbreak in Nigeria. (Undergraduate Thesis)*
- *Time Series Analysis of Oil Export & Exchange Rates*
- *The Attitude and Proficiency of Mathematics Teachers in the Use of ICT in Teaching in Lagos, Nigeria*

Technical Skills

- **Programming & Statistical Software:** R, Python, SPSS, Minitab, Jamovi, Power BI.
- **Data Analysis & Visualization:** Statistical modeling, machine learning, data simulation.
- **Project & Research Management:** Experimental design, statistical reporting.

Awards & Recognitions

- Saudi Universities Sport Federation (Bronze Award) (2023).
- Michael Taiwo Scholarship Award (2022).
- First Class Honours Award in Statistics (2021).
- Association of Statistics Student Tutor of the Year Award .
- Merit Award: Association of Statistics Students of Nigeria (ASSON), University of Ilorin Chapter.
- Best 3rd Year Statistics Student.

Leadership & Engagement

- Captain, Nigeria Community KFUPM Football Club, Dhahran, Saudi Arabia.
(2023 – Present)
- Assistant Presiding Officer (APO I) – Chairmen and Councillor Area Council Election, FCT-Abuja. (2022)
- Tutor, Statistics Muslim Students Society of Nigeria (MSSN). (2016 – 2021)

Volunteer Work

- Volunteer, KFUPM Medical Center Administration. (Jan 2024)

References

Available upon request.