Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 1 (31)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

UDC 004.855.5
DOI: https://doi.org/10.30837/2522-9818.2025.1.143

M. Koziuberda, D. Chernyshov

# A COMPARATIVE FRAMEWORK FOR ANALYZING DISTANCE METRICS IN HIGH-DIMENSIONAL SPACES

**Subject of the research** – developing a comprehensive framework to measure and analyze the relationships between different distance metrics in high-dimensional spaces. **Aim of the research** – to create a comparative framework that quantifies the "distance" between various distance metrics in high-dimensional settings. This framework aims to provide deeper insights into the interrelationships of these metrics and to guide practitioners in selecting the most appropriate metric for specific data analysis tasks. The research **tasks** include a theoretical formulation of methods to measure the "distance between distances", enabling a systematic comparison of different metrics. We conduct a thorough analysis of how these relationships evolve with increasing dimensionality. This involves developing mathematical models and employing visualization techniques to illustrate and interpret the relationships between metrics like the Manhattan distance, Euclidean distance, and others in high-dimensional spaces. A series of experiments are conducted on synthetic datasets to validate the theoretical findings and demonstrate the practical utility of the proposed framework. These datasets are carefully selected to cover a wide range of dimensionalities and data characteristics, ensuring a comprehensive evaluation of the framework's effectiveness. The **methodology** includes statistical analyses and visualization methods such as multidimensional scaling and heatmaps to represent the relationships between distance metrics clearly. The findings of the research are significant, revealing that the relationships between different distance metrics change notably as dimensionality increases. The results show patterns of convergence or divergence among certain metrics, providing valuable insights into their behavior in high-dimensional spaces. These insights are crucial for improving the accuracy and efficiency of data analysis techniques that rely on distance computations. **The conclusions** indicate that the proposed framework successfully quantifies the relationships between various distance metrics in high-dimensional spaces. By enhancing the understanding of how these metrics relate to one another, the research offers a valuable tool for selecting appropriate distance measures in high-dimensional data analysis. This contributes to more accurate and efficient analytical processes across various fields, including machine learning, data mining, and pattern recognition.

**Keywords:** distance metrics; high-dimensional spaces; metric comparison; $L_1$ norm; $L_2$ norm; data analysis; machine learning.

## Introduction

Distance metrics are fundamental tools in data analysis, machine learning, and pattern recognition, serving as the basis for various algorithms in clustering, classification, and information retrieval [1]. The choice of an appropriate distance metric is crucial, as it directly influences the performance and accuracy of these algorithms. In the era of big data, high-dimensional datasets have become increasingly common across fields such as image processing, genomics, finance, and social network analysis. These datasets present unique challenges due to the "curse of dimensionality", where the volume of the space increases so rapidly that the available data become sparse, making traditional analysis techniques less effective [2].

In high-dimensional spaces, the behavior of distance metrics can change dramatically. Phenomena such as the concentration of measure imply that in very high dimensions, distances between data points tend to become uniform regardless of the metric used. This effect can diminish the discriminative power of distance-based algorithms, leading to degraded performance [2]. Moreover, different distance metrics may exhibit varying sensitivities to specific data characteristics, such as scale, correlation, and noise, further complicating the analysis.

A significant challenge in this context is the lack of a systematic understanding of how different distance metrics relate to each other in high-dimensional spaces. While metrics like the Manhattan $(L_1)$ and Euclidean $(L_2)$ distances are well-understood in low-dimensional settings, their relationships and relative effectiveness in high-dimensional contexts are not fully explored [3, 4]. This gap in knowledge can hinder the ability of practitioners to select the most appropriate metric for their specific applications, potentially leading to suboptimal results.

Previous studies have investigated the asymptotic behavior of the Manhattan distance as dimensionality increases, highlighting its statistical properties and implications for empirical experiments [3]. For instance, Silva [3] analyzed the theoretical properties and statistical behavior of the Manhattan distance through mathematical derivations and computational simulations, revealing how its mean and variance exhibit predictable trends as dimensionality increases. Other research has focused on approximations of the Euclidean distance in multi-

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 1 (31)

dimensional spaces to reduce computational complexity while maintaining acceptable accuracy [5].

Additionally, efforts have been made to establish approximate relations between Manhattan and Euclidean distances, particularly in the context of experimental design using Latin hypercube sampling [4]. Das et al. [4] explored the relations and bounds between these distances, providing insights into which measure offers better space-filling properties in design of experiments (DoEs).

Despite these contributions, there remains a need for a comprehensive framework that systematically compares different distance metrics and quantifies their relationships in high-dimensional settings. Such a framework would provide deeper insights into how metrics like $L_1$, $L_2$, and others compare and contrast as dimensionality increases, offering a more nuanced understanding than traditional pairwise comparisons.

This research aims to address this challenge by developing a comprehensive framework for measuring and analyzing the relationships between various distance metrics in high-dimensional spaces. By introducing a method to quantify the "distance between distances", we seek to provide deeper insights into metric behavior in high dimensions. The framework is designed to capture the nuances of metric behavior, offering practical guidance for data analysts and machine learning practitioners.

The successful development of this comparative framework involves a detailed theoretical investigation coupled with extensive empirical validation. The theoretical component focuses on deriving mathematical formulations that define the relationships between different metrics, considering the effects of high dimensionality. The empirical component includes experiments on both synthetic and real-world datasets to demonstrate the practical implications of the theoretical findings. Visualization techniques such as multidimensional scaling and heatmaps are employed to illustrate the complex interrelationships between metrics clearly.

Integrating theoretical analysis with empirical experimentation is expected to yield a robust framework that not only enhances our understanding of distance metrics in high-dimensional spaces but also provides practical guidance for selecting appropriate metrics in various applications. This integrated approach is particularly relevant in fields where the choice of distance metric can significantly impact the performance of algorithms, such as clustering, classification, and nearest neighbor search [1, 2].

In the realm of distance metric analysis, various approaches have been proposed to compare and select appropriate metrics, ranging from heuristic methods based on domain knowledge to more systematic techniques involving metric learning [1]. However, these methods often focus on low-dimensional settings or specific applications, lacking generality in high-dimensional contexts. Our research fills this gap by providing a generalizable framework applicable across different domains and types of data.

By systematically analyzing and quantifying how distance metrics relate to each other in high-dimensional spaces, we aim to enhance the effectiveness of data analysis techniques that rely on these metrics. This contribution has the potential to improve the accuracy and efficiency of various machine learning algorithms, ultimately advancing the fields of data mining and pattern recognition.

The **objective** of this research is not only to develop a theoretical understanding of metric relationships but also to translate this understanding into practical tools and guidelines. By systematically analyzing and quantifying how distance metrics relate to each other in high-dimensional spaces, we aim to enhance the effectiveness of data analysis techniques that rely on these metrics. This contribution has the potential to improve the accuracy and efficiency of various machine learning algorithms, ultimately advancing the fields of data mining and pattern recognition. Through this research, we aspire to make a significant contribution to the understanding of distance metrics in high-dimensional data analysis. By providing both theoretical insights and practical tools, we hope to enable more informed decisions regarding metric selection, leading to improved outcomes in a wide range of applications where high-dimensional data are prevalent.

## Analysis of the problem and Existing Methods

Comparing and selecting appropriate distance metrics in high-dimensional spaces is a complex challenge that has significant implications for the effectiveness of machine learning algorithms [1, 2]. High-dimensional data often leads to phenomena such as the concentration of distances, where the distinction between the nearest and farthest points diminishes, making it difficult for algorithms like $k$-nearest neighbors $(k\text{-}NN)$ to perform effectively [1].

Aggarwal et al. [2] conducted a seminal study on the behavior of distance metrics in high-dimensional spaces, revealing that the choice of metric significantly impacts the meaningfulness of proximity measures. They found

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 1 (31) ISSN 2524-2296 (online)

that metrics such as the Manhattan distance ( $L_1$ norm) can be more preferable than the Euclidean distance ( $L_2$ norm) in high-dimensional applications due to their sensitivity to the dimensionality of the data.

Silva [3] investigated the asymptotic behavior of the Manhattan distance as the number of dimensions increases, providing insights into its statistical properties through mathematical derivations and computational simulations. The study highlighted predictable trends in the mean and variance of the Manhattan distance, emphasizing the need to understand how distance metrics behave as dimensionality grows.

Das et al. [4] explored approximate relations between the Manhattan and Euclidean distances in the context of Latin hypercube experimental design. Their work established bounds and relations between these distances, offering insights into which measure may provide better space-filling properties in design of experiments.

Research on approximating the Euclidean distance in multi-dimensional spaces has been conducted to reduce computational complexity while maintaining acceptable accuracy [5].

Comparative studies have also been carried out in specific applications. For instance, in clustering analysis for determining the spread of COVID-19 in Bekasi City, researchers compared the Euclidean and Manhattan distance calculations within the K-means clustering algorithm [6]. They found that the choice of distance metric affected the number of iterations and processing time, with the Manhattan distance leading to faster convergence [6]. This underscores the practical importance of selecting an appropriate distance metric based on the specific characteristics of the data and the analysis objectives.

Despite these studies, there remains a lack of a comprehensive framework that systematically compares different distance metrics and quantifies their relationships in high-dimensional settings. Existing methods often focus on specific metrics or applications, without providing a generalizable approach to measure the "distance between distances" across various metrics and dimensions.

Some studies have proposed methods for secure computation of distance metrics [7], and have analyzed the minimum Manhattan distance in permutations [8], but these works do not directly address the comparative analysis of distance metrics in high-dimensional spaces.

In the field of $k$-$NN$ algorithms, recent reviews have highlighted the need for enhanced methods to improve performance in high-dimensional data [1]. Modifications to exact $k$-$NN$ techniques have been proposed, but they often rely on heuristic approaches or are tailored to specific applications, rather than providing a systematic framework for metric comparison.

Overall, the challenge lies in developing a method that can quantify and analyze the relationships between different distance metrics as dimensionality increases, providing practical guidance for selecting appropriate metrics in high-dimensional data analysis.

## Problem solution

У цьому дослідженні коли ми розглядаємо граничну область $B(X)$ для множини $X$, ми маємо на увазі різницю між верхньою та нижньою апроксимацією $X$. Математично це може бути виражено так:

To address the challenge of systematically comparing distance metrics in high-dimensional spaces, we propose a framework that utilizes statistical sampling and correlation analysis to quantify the relationships between different distance metrics. Our method involves generating synthetic datasets with varying dimensionalities and analyzing how different distance metrics correlate in these high-dimensional settings.

We generate synthetic data by sampling points from a standard multivariate normal distribution. Specifically, we sample $n$ data points $x_i \in \mathbb{R}^d$, where component $x_i$ is drawn independently from a standard normal distribution $N(0,1)$. This approach ensures that the data are centered at the origin with identity covariance, providing an isotropic distribution equally spread in all directions. By varying the dimensionality $d$ from 2 to 1000, we investigate how the relationships between distance metrics evolve as the number of dimensions increases.

We focus on the family of Minkowski distance metrics, defined by the $L_p$ norm (1):

$$\|x\|_p = \left( \sum_{i=1}^{d} \left| x^p \right| \right)^{1/p}, \tag{1}$$

where $p \in \mathbb{R}$. We extend our analysis to include all computable values of $p$, including negative values and $p = \infty$, which corresponds to the Chebyshev distance. This comprehensive consideration of $p$ values allows us to examine a wide spectrum of distance metrics within the Minkowski family.

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 1 (31)

For each sampled point $x_i$, we compute its distance to the origin using various $L_p$ norms. This results in a set of distance measurements $\{\|x_i\|_p\}$ for each value of $p$. By focusing on distances from the origin, we eliminate pairwise dependencies and simplify the analysis while still capturing the essential characteristics of the distance metrics.

To quantify the relationships between different distance metrics, we perform a correlation analysis of the distance measurements obtained for different $pp$ values. Specifically, we calculate the Spearman rank correlation coefficient between pairs of distance metrics $L_{p1}$ and $L_{p2}$.

The Spearman rank correlation coefficient measures the monotonic relationship between two variables without assuming a linear relationship or normality, making it suitable for our analysis of non-linear dependencies between distance metrics.

By computing this correlation for various combinations of $p_1$ and $p_2$, we obtain a metric-pair correlation matrix that reflects how similar or dissimilar the distance metrics are in high-dimensional spaces. A higher correlation coefficient indicates that two metrics rank the distances of points similarly, implying they are more alike in their behavior.

We examine how the correlations between different distance metrics change as a function of dimensionality $dd$. By varying $dd$ from 2 to 1000, we observe the impact of increasing dimensions on the relationships between metrics. Our preliminary results indicate that after a certain dimensionality threshold, the correlations stabilize and reach a plateau. This suggests that in very high dimensions, the relative behavior of different distance metrics becomes consistent.

While our approach is primarily empirical, it is grounded in the understanding that in high-dimensional spaces, the geometry of data changes significantly. For instance, it is known that distances tend to concentrate in high dimensions, a phenomenon related to the concentration of measure [2]. Although we do not derive new theoretical results, our empirical observations align with known theoretical behaviors, such as the stabilization of metric relationships at high dimensions.

## Research results

To investigate how the relationship between different distance metrics evolves with increasing dimensionality, we conducted experiments measuring the Spearman rank correlation coefficient between the $L_1$ (Manhattan) and $L_2$ (Euclidean) norms of data points sampled from a standard normal distribution. The dimensionality $d$ varied from 2 to 100, allowing us to observe trends over a wide range of dimensions (fig. 1).

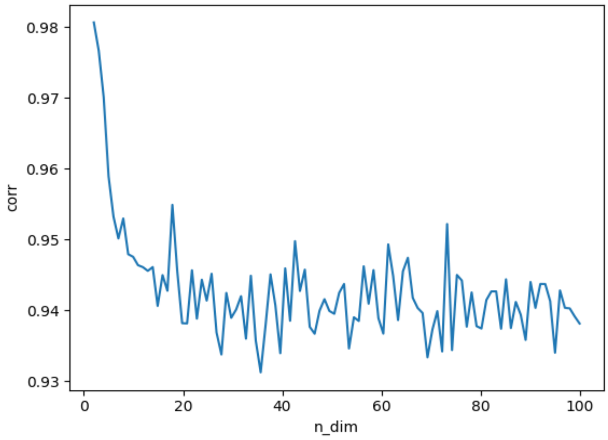


**Fig. 1.** Relatinship between dimensionality and correlation

The initial decrease in correlation can be attributed to the increasing complexity of the high-dimensional space. In low dimensions, the $L_1$ and $L_2$ norms are more closely related because the geometry of the space allows for less variation in the paths measured by these norms. Despite the decreasing trend, the correlation stabilizes around 0.94 in higher dimensions. This plateau indicates that while the norms are not perfectly correlated, they maintain a strong monotonic relationship even as dimensionality increases. This consistent correlation suggests that, in high-dimensional spaces, the relative rankings of distances from the origin according to the $L_1$ and $L_2$ norms remain similar.

In the next experiment, we investigate how the Spearman rank correlation between the $L_2$ norm (Euclidean distance) and various $L_p$ norms changes as we vary the value of $p$. By fixing $p_1 = 2$ (corresponding to the $L_2$ norm) and altering p over a specified range, we aim to understand how different $L_p$ norms relate to the $L_2$ norm in high-dimensional spaces (fig. 2).

This experiment demonstrates that the Spearman rank correlation between the $L_2$ norm and $L_p$ norms varies non-linearly with p. The correlation is highest at $p = 2$ and decreases for both lower and higher $p$ values. These findings highlight the importance of carefully

ISSN 2522-9818 (print)
Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 1 (31) ISSN 2524-2296 (online)

selecting the $p$ value in $L_p$ norms to match the specific needs of high-dimensional data analysis tasks.

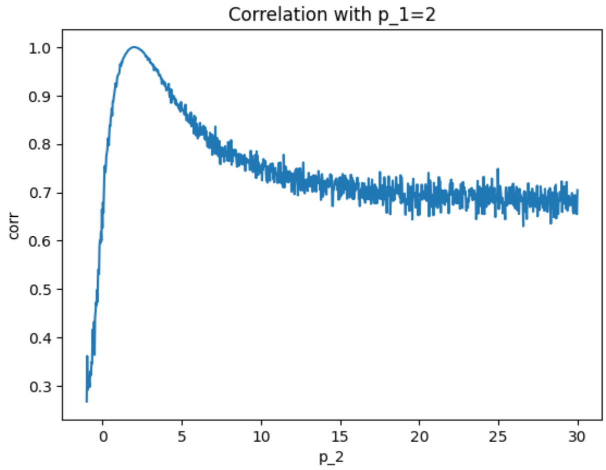


**Fig. 2.** Correlation between l2 norm and p_2 norms on $x$ axis

In this experiment, we explore the relationships between the $L_1$, $L_2$, $L_\infty$, and norms to understand how these distance metrics correlate with one another in high-dimensional spaces. Specifically, we aim to investigate whether the "distance" (as measured by dissimilarity or correlation) between $L_1$ and $L_2$ norms is approximately equal to that between $L_2$ and $L_\infty$ norms. Additionally, we examine how the correlation between $L_2$ and $L_p$ norms changes as p approaches infinity (table 1).

**Table 1**. *Correlation table between* $L_1$, $L_2$, *and* $L_\infty$ *metrics*

| | $L_1$ | $L_2$ | $L_\infty$ |
|---|---|---|---|
| $L_1$ | 1 | 0.95 | 0.89 |
| $L_2$ | 0.95 | 1 | 0.95 |
| $L_\infty$ | 0.89 | 0.95 | 1 |

The correlations $\rho_{1,2}$ and $\rho_{2,\infty}$ are very close in value, suggesting that the "distance" or dissimilarity between $L_1$ and $L_2$ norms is approximately equal to that between $L_2$ and $L_\infty$ norms. This suggests a kind of symmetry in the relationships among these norms in high-dimensional spaces, which may not be as apparent in lower dimensions.

When comparing $L_p$ $L_2$, as $p$ approaches infinity, the $L_p$ norm increasingly emphasizes the largest components of the vectors. This results in the $L_p$ norm converging to the $L_\infty$ norm.

## Practical Implications

The comparative framework presented in this study has significant potential for real-world implementation, particularly in fields where high-dimensional data analysis is a core challenge. In e-commerce recommendation systems, where vast amounts of user and product attributes are generated daily, practitioners often rely on distance-based methods to measure similarities between items or users [9, 10]. Implementing the proposed framework can help them determine which distance metric is best suited for capturing meaningful relationships in datasets containing thousands of features, such as product descriptions, user demographics, and behavioral logs. By systematically quantifying the relationships among metrics like Manhattan, Euclidean, and other Minkowski norms, companies can more confidently select a metric that preserves essential characteristics of user-item interactions, thus improving recommendation accuracy and user satisfaction.

In the financial sector, high-dimensional data arise in areas such as portfolio optimization [11], fraud detection, and algorithmic trading [12], where feature spaces range from historical price series to textual sentiment indicators. A robust, data-driven evaluation of distance metrics helps identify the most appropriate norm to capture subtle patterns in time series and complex relationships among assets. When dealing with high-frequency trading scenarios or large-scale risk modeling, the correct choice of distance metric can also reduce computational overhead and minimize latency.

In image recognition and computer vision applications, large volumes of pixel information and extracted feature vectors are often managed in domains ranging from autonomous driving to medical imaging [13]. When selecting the most effective distance metric for tasks like image retrieval or similarity-based classification, analyzing correlations among norms under different dimensional settings allows teams to anticipate performance trade-offs. This becomes especially critical in medical contexts where even minor improvements in classification accuracy can translate into earlier detection of diseases. Moreover, integrating this framework into workflow pipelines can streamline the process of feature extraction by aligning metric choice with specific datasets, reducing the trial-and-error phase that often burdens data science projects.

Sensor networks and Internet of Things deployments also stand to benefit [14, 15]. Industrial

ISSN 2522-9818 (print)
ISSN 2524-2296 (online) Innovative technologies and scientific solutions for industries. 2025. No. 1 (31)

monitoring systems produce high-dimensional data encompassing sensor readings, event logs, and contextual information. Employing this comparative framework enables the identification of norms that robustly differentiate normal operational conditions from anomalies, helping to prevent downtime or safety incidents in factories, power plants, or transportation systems. By exploiting the insights into how metric relationships evolve with dimensionality, it becomes easier to design anomaly detection pipelines that remain effective at scale. In terms of practical implementation, the development of user-friendly libraries that encapsulate the proposed framework's functionalities can lower the barrier to entry for data scientists who may not have a strong mathematical background in norm analysis.

Such tools can integrate with popular machine learning libraries, automatically generating correlation heatmaps or multidimensional scaling visualizations to guide metric selection. An organization interested in adopting the framework could embed these utilities in its data engineering platform, thereby incorporating distance metric analysis into routine data processing steps. Adapting the framework to specialized hardware accelerators like GPUs can further reduce computation time, particularly when sampling from large datasets or exploring many pairs of distance metrics. Finally, the advances in interpretability provided by the proposed framework have value in regulated industries like healthcare or finance, where practitioners must justify their methods to stakeholders. Demonstrating that a chosen distance metric aligns with the intrinsic geometry of the data helps address concerns regarding fairness, bias, and other ethical considerations. Through these avenues of application – from recommendation systems and algorithmic trading to medical image classification and industrial monitoring – the proposed comparative framework offers a practical means of improving the accuracy, efficiency, and reliability of high-dimensional data analysis.

## Conclusions

This research has presented a comprehensive framework for analyzing and quantifying the relationships between different distance metrics in high-dimensional spaces. By employing statistical sampling from a standard multivariate normal distribution and performing correlation analyses between various $L_p$ norms, we have gained valuable insights into how these metrics behave and relate to each other as dimensionality increases.

Our findings reveal that the Spearman rank correlation between the $L_1$ and $L_2$ norms decreases with increasing dimensionality but stabilizes at a strong positive correlation in high-dimensional spaces. Specifically, the correlation coefficient decreases from approximately 0.98 in two dimensions to around 0.94 in dimensions beyond ten, indicating that despite the initial decline, the $L_1$ and $L_2$ norms maintain a consistent relationship in higher dimensions. This suggests that, for high-dimensional data, the relative rankings of distances according to these norms remain similar, which has practical implications for algorithm performance and computational efficiency in data analysis tasks.

Further exploration into the correlation between the $L_2$ norm and various $L_p$ norms across a wide range of $p$ values demonstrated a non-linear relationship. The correlation peaks at $p = 2$ , as expected, since we are comparing the $L_2$ norm with itself, and decreases for both lower and higher $pp$ values. Notably, as $p$ approaches infinity, corresponding to the $L_\infty$ norm, the correlation with the $L_2$ norm decreases, reflecting the increasing emphasis on the largest components in the data vectors. This behavior highlights the importance of selecting appropriate p values in $L_p$ norms based on the specific characteristics of the data and the analysis objectives.

The comparison among the $L_1$, $L_2$, and $L_\infty$ norms revealed that the "distance" or dissimilarity between $L_1$ and $L_2$ norms is approximately equal to that between $L_2$ and $L_\infty$ norms. The strong correlations observed among these norms suggest a form of symmetry in their relationships within high-dimensional spaces. This symmetry indicates that transitioning from one norm to another involves similar changes in how the norms aggregate component information, which can inform decisions when selecting distance metrics for high-dimensional data analysis.

These insights contribute to a deeper understanding of the behavior of distance metrics in high-dimensional settings. They underscore the necessity for practitioners to consider the implications of their choice of distance metric, especially in applications where the dimensionality of data is high. Selecting a metric that aligns with the desired sensitivity to data features can enhance algorithm performance, improve computational efficiency, and ultimately lead to more accurate and reliable results in machine learning and data mining tasks.

Сучасний стан наукових досліджень та технологій в промисловості. 2025. № 1 (31)
ISSN 2522-9818 (print)
ISSN 2524-2296 (online)

However, it is important to acknowledge the limitations of this study. The experiments were conducted using synthetic data sampled from a standard normal distribution with independent and identically distributed components. While this approach provides a controlled environment for analysis, real-world data often exhibit correlations, non-normal distributions, and other complexities not captured in our synthetic datasets. Consequently, the findings may not directly generalize to all types of data distributions. Additionally, the focus on distances from the origin simplifies the analysis but may not capture all aspects relevant in practical applications where pairwise distances between data points are crucial.

Future research could extend this framework to include different data distributions, incorporate real-world datasets, and explore the effects of feature correlations and other data characteristics. The development of theoretical models to explain and predict the observed relationships between distance metrics in high-dimensional spaces would further enhance the utility of this framework. Such advancements would provide more comprehensive guidance for selecting appropriate distance measures, thereby improving the effectiveness of algorithms that rely on distance computations across various domains.

In conclusion, this study offers a valuable contribution to the understanding of distance metrics in high-dimensional spaces by providing both empirical evidence and practical insights. The proposed framework enables a systematic comparison of different distance metrics, highlighting the nuanced ways in which they relate to each other as dimensionality increases. By informing the selection of distance metrics, this work has the potential to enhance the accuracy and efficiency of high-dimensional data analysis techniques, ultimately advancing the fields of machine learning, data mining, and pattern recognition.

**References**

1. Halder, R. K., Uddin, M. N., Uddin, Md. A., Aryal, S., and Khraisat, A. (2024), "Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications". *Journal of Big Data, Springer Science and Business Media LLC.* Vol. 11, No. 1. Aug. 11. DOI: 10.1186/s40537-024-00973-y
2. Aggarwal, C. C., Hinneburg, A., and Keim, D. A. (2001), "On the Surprising Behavior of Distance Metrics in High Dimensional Space". *Lecture Notes in Computer Science, Springer Berlin Heidelberg.* Vol. 1973. P. 420–434. DOI: 10.1007/3-540-44503-x_27
3. Silva, E. C. de M. (2024), "Asymptotic behavior of the Manhattan distance in n-dimensions: Estimating multidimensional scenarios in empirical experiments". *arXiv*. DOI: 10.48550/ARXIV.2406.15441
4. Jamali, A. R. M. J. U. U., and Alam, Md. A. (2019), "Approximate relations between Manhattan and Euclidean distance regarding Latin hypercube experimental design". *Journal of Physics: Conference Series, IOP Publishing*. Vol. 1366, No. 1. Nov. 1, 012030 p. DOI: 10.1088/1742-6596/1366/1/012030
5. Cardarilli, G. C., Di Nunzio, L., Fazzolari, R., Nannarelli, A., Re, M., and Spano, S. (2020), "NN-Dimensional Approximation of Euclidean Distance". *IEEE Transactions on Circuits and Systems II: Express Briefs, IEEE*. Vol. 67, No. 3. Mar., P. 565–569. DOI: 10.1109/tcsii.2019.2919545
6. Cahya, F. N., Mahatma, Y., and Rohimah, S. R. (2023), "Perbandingan Metode Perhitungan Jarak Euclidean dengan Perhitungan Jarak Manhattan pada K-Means Clustering Dalam Menentukan Penyebaran Covid di Kota Bekasi". *JMT: Jurnal Matematika dan Terapan, Universitas Negeri Jakarta.* Vol. 5, No. 1. Feb. 28, P. 43–55. DOI: 10.21009/jmt.5.1.5
7. Liu, W., and Zhang, W. (2020), "A Quantum Protocol for Secure Manhattan Distance Computation". *IEEE Access, IEEE.* Vol. 8. P. 16456–16461. DOI: 10.1109/access.2020.2966800
8. Blackburn, S. R., Homberger, C., and Winkler, P. (2019), "The minimum Manhattan distance and minimum jump of permutations". *Journal of Combinatorial Theory, Series A, Elsevier BV.* Vol. 161. Jan., P. 364–386. DOI: 10.1016/j.jcta.2018.09.002
9. Fkih, F. (2022), "Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison", *Journal of King Saud University - Computer and Information Sciences,* Vol. 34, no. 9, Elsevier, P. 7645–7669, Oct. 2022, DOI: 10.1016/j.jksuci.2021.09.014
10. Vančura, V.; Kordík, P., Straka, M. (2024), "beeFormer: Bridging the Gap Between Semantic and Interaction Similarity in Recommender Systems", *arXiv,* DOI: 10.48550/ARXIV.2409.10309
11. Yang, J. (2025), "DFF: Decision-Focused Fine-tuning for Smarter Predict-then-Optimize with Limited Data," *arXiv*, DOI: 10.48550/ARXIV.2501.01874
12. Monteiro, T. (2024), "AI-Powered Energy Algorithmic Trading: Integrating Hidden Markov Models with Neural Networks", *arXiv*, DOI: 10.48550/ARXIV.2407.19858
13. Samantaray, A. K., Rahulkar, A. D. (2019), "Comparison of Similarity Measurement Metrics on Medical Image Data", *in 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE*, P. 1–5, DOI: 10.1109/icccnt45670.2019.8944781

14. Maudet, G., Batton-Hébert, M., Maillé, P., Toutain, L. (2022), "Emission Scheduling Strategies for Massive-IoT: Implementation and Performance Optimization", *in NOMS 2022 IEEE/IFIP Network Operations and Management Symposium, IEEE*, P. 1–4, Apr. 25, DOI: 10.1109/noms54207.2022.9789769
15. Shafique, A., Asad, M., Aslam, M., Shaukat, S., Cao, G. (2022), "Multi-hop similarity-based-clustering framework for IoT-oriented software-defined wireless sensor networks", *IET Wireless Sensor Systems*, Vol. 12, No. 2, Institution of Engineering and Technology (IET), P. 67–80, DOI: 10.1049/wss2.12037

*Надійшла (Received) 05.02.2025*

*Відомості про авторів / About the Authors*

**Чернишов Дмитро Владиславович** – Харківський національний університет радіоелектроніки, аспірант комп'ютерних наук, кафедра системотехніки, Харків, Україна; e-mail: dmytro.chernyshov@nure.ua; ORCID ID: https://orcid.org/0009-0003-2773-7467

**Козюберда Михайло Валентинович –** аспірант кафедри програмної інженерії Харківський національний університет радіоелектроніки, Харків, Україна; e-mail: mykhailo.koziuberda@nure.ua; ORCID ID: https://orcid.org/0009-0004-9533-427X

**Chernyshov Dmytro** – Kharkiv National University of Radio Electronics, Department of System Engineering, PhD of Computer Science, Kharkiv, Ukraine.

**Koziuberda Mykhailo Valentynovych –** postgraduate student of the Department of Software Engineering Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

# ПОРІВНЯЛЬНА СИСТЕМА ДЛЯ АНАЛІЗУ МЕТРИК ВІДСТАНІ У ВИСОКОВИМІРНОМУ ПРОСТОРІ

**Предмет дослідження** – розроблення комплексної рамкової методології для вимірювання та аналізу взаємозв'язків між різними метриками відстані у багатовимірних просторах. **Мета дослідження** – створити порівняльну методологію, яка кількісно оцінює "відстань" між різними метриками відстані в умовах високої розмірності. Ця методологія має на меті надати глибше розуміння взаємозв'язків між зазначеними метриками та допомогти фахівцям у виборі найбільш відповідної метрики для конкретних завдань аналізу даних. **Завдання дослідження** охоплюють теоретичне формулювання методів вимірювання "відстані між відстанями", що дає змогу систематично порівнювати різні метрики. Необхідно ґрунтовно проаналізувати зміни цих взаємозв'язків залежно від зростання розмірності. Для цього передбачено розробити математичні моделі та застосувати методи візуалізації з метою ілюстрації та інтерпретації взаємозв'язків між такими метриками, як манхеттенська відстань, евклідова відстань та інші у високорозмірних просторах. Серію експериментів заплановано провести на синтетичних наборах даних для верифікації теоретичних висновків і демонстрації практичної корисності запропонованої **методології**. Ці набори даних ретельно обрано таким чином, щоб охопити широкий спектр розмірностей і характеристик даних, що забезпечує всебічну оцінку ефективності методології. У межах дослідження застосовано статистичні аналізи та методи візуалізації (зокрема багатовимірне шкалювання і теплові карти), які дають змогу чітко подати взаємозв'язки між метриками відстані. **Досягнуті результати** свідчать про те, що взаємозв'язки між різними метриками відстані суттєво змінюються зі зростанням розмірності. Спостерігаються як збіги, так і розбіжності між окремими метриками, що дає важливі знання про їх поведінку у високорозмірних просторах. Ці висновки мають вирішальне значення для підвищення точності й ефективності методів аналізу даних, заснованих на обчисленні відстані. **Висновки** демонструють, що запропонована рамкова методологія успішно дає кількісну оцінку взаємозв'язків між різними метриками відстані у високорозмірних просторах. Розширене розуміння того, як ці метрики взаємопов'язані, дає змогу обґрунтовано обрати потрібну метрику у високорозмірному аналізі даних, сприяючи більш точним і ефективним аналітичним процедурам у сферах машинного навчання, інтелектуального аналізу даних та розпізнавання образів.

**Ключові слова:** метрики відстані; багатовимірні простори; порівняння метрик; норма $L_1$; норма $L_2$; аналіз даних; машинне навчання.

*Бібліографічні описи / Bibliographic descriptions*

Чернишов Д. В., Козюберда М. В. Порівняльна система для аналізу метрик відстані у високовимірному просторі. *Сучасний стан наукових досліджень та технологій в промисловості*. 2025. № 1 (31). С. 143–150. DOI: https://doi.org/10.30837/2522-9818.2025.1.143

Chernyshov, D., Koziuberda, M. (2025), "Comparative framework for analyzing distance metrics in high-dimensional spaces", *Innovative Technologies and Scientific Solutions for Industries*, No. 1 (31), P. 143–150. DOI: https://doi.org/10.30837/2522-9818.2025.1.143