

ISSN 1840-4855
e-ISSN 2233-0046

Original scientific article
<http://dx.doi.org/10.70102/afts.2025.1833.453>

OPTIMISING FINANCIAL FORECASTING: THE POWER OF REDUCTION APPROACH

M. Sravan Kumar Reddy¹, C. Supraja²

¹Associate Professor, Department of Computer Science and Engineering, Rajeev Gandhi Memorial College of Engineering and Technology, Nandyal, Andhra Pradesh, India. e-mail: sravankumarreddy.m@rgmcet.edu.in, orcid: <https://orcid.org/0009-0001-7454-4030>

²Computer Science Department, Rajeev Gandhi Memorial College of Engineering and Technology, Nandyal, Andhra Pradesh, India. e-mail: cumbumsupraja@gmail.com, orcid: <https://orcid.org/0009-0009-0445-2374>

Received: June 12, 2025; Revised: August 29, 2025; Accepted: September 24, 2025; Published: October 30, 2025

SUMMARY

Given the complexity and multifaceted nature of the financial markets, effective aggregation of financial time series data underscores the optimization of predictive modeling in the finance industry. This research presents an innovative approach to clustering using auto encoders designed to distill informative representations out of S&P 500 financial time series data. Our particular methodology is horizontal (stock averages) and vertical (1-hour intraday frequency) dual-dimensional, which enables us to capture temporal patterns along with contextual richness. Comparative research demonstrates auto encoder-driven clustering enhances data quality and granularity, providing actionable understanding of market behavior. These implications applied to predictive modeling would also be considered under risk management, incipient investment strategies, or just pure advanced financial analyses.

Key words: *financial time series, clustering, auto encoders, s&p 500, dimensionality reduction, temporal patterns, predictive modeling.*

INTRODUCTION

In the gleaming world of finance, prediction models are of a very significant nature in sophisticated forecasting of market trends. Portfolio optimization, algorithmic trading, and risk analysis all run on prediction models [1], [2]. Noise, high-dimension, non-stationary, and nonlinear relations are the major challenges facing the traditional approaches used with machine learning in financial time series data [20]. Due to such complexity, exact modeling of market behavior tends to become rigid, making it necessary for sophisticated approaches to come forward for sustaining temporal and contextual interdependencies [3], [6]. Clustering, an all-purpose unsupervised machine learning paradigm, has been massive in contributing toward preprocessing, exploratory data analyses, and feature extraction in financial time series [21]. It lumps together assets exhibiting similar behavior for usage in anomaly detection, market segmentation, and portfolio diversification.

Classical clustering algorithms-k-means and hierarchical clustering-are inadequate to produce very fine clusters when complex financial data is concerned [5]. Such generic limitations are further compounded by the curse of dimensionality among other things that wilts their efficiency upon basing meaningful clusters. [22].

We strictly applied dimensionality reduction methods such as Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) to circumvent these challenges. These are methods to extract pertinent information through the transformation of high-dimensional data to lower-dimensional data [7]. Most of the time though, they may miss the nonlinear relationships inherent in financial time series. Deep learning approaches that have been seen as potential solutions to dimensionality reduction include autoencoders. Autoencoders are especially useful for finance because they can learn compressed nonlinear representations of high-dimensional data. [23], [22].

Table 1. Historical overview of key references in financial time series analysis and machine learning

Ref.	Year	Title	Key Contribution	Methodology/Focus
[1]	1970	Efficient capital markets: A review of theory and empirical work	Introduced the Efficient Market Hypothesis (EMH)	Market behavior analysis
[21]	2022	A comprehensive review of clustering techniques in financial time series analysis	Reviewed clustering applications for financial data	Clustering in finance
[3]	2016	Deep Learning	Comprehensive deep learning overview	Neural networks and AI
[22]	2008	Visualizing data using t-SNE	Proposed t-SNE for dimensionality reduction	Data visualization
[5]	2013	Auto-Encoding Variational Bayes	Introduced Variational Autoencoders (VAE)	Generative models
[23]	2013	Outlier Analysis	Discussed techniques for detecting data outliers	Outlier detection
[7]	1995	The profitability of technical trading rules in the Asian stock markets	Evaluated technical trading in Asian markets	Technical trading rules
[24]	2006	Reducing the dimensionality of data with neural networks	Proposed autoencoders for dimensionality reduction	Neural network applications
[9]	2010	Consumer credit-risk models via machine-learning algorithms	Applied ML algorithms to credit risk prediction	Financial risk modeling
[24]	2009	Learning deep architectures for AI	Explored deep learning architectures	Deep learning techniques

Table 1 offers a chronological panorama of the key references related to the analysis of financial time series and machine learning. The table affirms the significance of important studies, such as the Efficient Market Hypothesis proposed by Fama [1] and new clustering methods for financial data [2]. Useful aids such as autoencoders for dimensionality reduction [8] and t-SNE for visualization [4] provide a mark in the development from classical statistics toward complex deep learning algorithms. Other applications of machine learning in risk management [9] and technical trading [7] are also indicated, showing how these studies have more general implications in the finance domain [10].

This particular research presents a framework for 2D clustering using autoencoders for dimension reduction. This means that the financial time series data can be clustered logically. Hence, the present framework considers stock averages (horizontal) and contextual indicators such as volatility and trading volume (vertical). Placing the data with context and delineating it abundantly from a temporal perspective, this multi-modal approach further increases granularity and interpretability of the clustering results. By means of the S&P 500 index, our framework permits the exploration of market trends and behaviors in an integrated manner. In essence, the two-dimensional framework retains richer dynamics in financial markets, and the autoencoder framework preserves vital pieces of information.

RELATED WORK

Financial time series analysis widely uses clustering to cluster assets with similar behaviors which helps in portfolio management, forecasting, risk assessment and market predictions [11]. Clustering techniques help in creating investment strategies since they enable the detection of patterns and trends through the grouping of assets or time series with similar characteristics. Some of the traditional clustering methods such as k-means and hierarchical clustering find frequent application in analyzing financial data but are hampered by the higher dimensionality and complexity of the financial markets [25]. We have incorporated some of the methods of reducing dimensionality such as Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbour

Embedding (t-SNE). These methods decrease the data dimensionality by producing a lower-dimensional representation with important information preserved, hence making the clustering process more efficient [12],[13]. These techniques enable more effective comprehension of financial data through the capture of key trends and dynamics in the complex forms of financial data.

Autoencoders have become a trending deep learning method used for efficient dimensionality reduction in recent times. Autoencoders are architectures based on neural networks that can learn high-dimensional, complex data in nonlinear, compressed representations [14], [8]. They have shown great potential in the field of enhancing the accuracy of clustering and identifying the inherent patterns in financial time series. Autoencoders are more flexible and able to extract sophisticated, non-linear relationships from the data, making them applicable to contemporary financial datasets compared to the old techniques like PCA [15]. Autoencoders have proven to be exceptionally effective in areas like asset classification and fraud detection, where the data is often multidimensional and noisy [16].

In spite of all these developments, very little work has been done which can be seen to have merged multiple dimensions, especially temporal and contextual attributes, into clustering models. Contextual dimensions, like volatility and volume of trading, and temporal dimensions, like stock means over time, are important in terms of offering useful information about market behavior [17], [18]. The dimensions provide additional viewpoints that can contribute to the accuracy and level of detail in clustering results, thus allowing for a deeper analysis of market behavior. Contextual features give extra layers of information that can enhance clustering results by taking into account external market factors, and temporal features capture trends and patterns over time [19].

Table 2. Key Contributions and methodologies in financial time series clustering

Ref.	Key Contribution	Focus Area	Methodology Used	Application Domain	Key Findings
[11]	Overview of clustering methods in finance	Financial time series clustering	Various clustering algorithms	Finance	Survey of clustering techniques in finance
[25]	General clustering algorithms	Clustering techniques	Algorithmic approaches to clustering	General	Comprehensive analysis of clustering methods
[12]	Introduced t-SNE for dimensionality reduction	Dimensionality reduction	t-SNE	Data visualization	Effective reduction of high-dimensional data
[13]	PCA for dimensionality reduction	Dimensionality reduction	Principal Component Analysis (PCA)	Data analysis	Introduction to PCA in data reduction
[14]	Introduced autoencoders for nonlinear data reduction	Dimensionality reduction	Autoencoders	Deep learning applications	Nonlinear reduction of complex data
[15]	Deep learning for clustering financial data	Clustering in financial markets	Deep learning-based clustering	Finance	Improved clustering with deep learning
[16]	Anomaly detection in financial transactions	Anomaly detection	Deep learning	Financial transactions	Detection of anomalies in transactions
[17]	Hybrid clustering for portfolio optimization	Portfolio optimization	Hybrid clustering approach	Finance	Enhancing portfolio optimization via clustering
[18]	Multi-dimensional clustering for stock prediction	Stock market prediction	Multi-dimensional clustering	Stock market prediction	Improved prediction through multi-dimensional clustering
[19]	Incorporating volatility and volume in clustering	Contextual clustering	Clustering with volatility and volume features	Financial data analysis	Impact of volatility and volume on clustering results

The list of major works in financial time series clustering and deep learning is summarized in Table 2. The approaches followed are highlighted in that the authors have used auto-encoders for nonlinear data reduction [15], different clustering techniques [11], as well as t-SNE for dimensionality reduction [13]. Table highlights various applications in the financial domain, such as portfolio optimization [18], anomaly detection on transactions [17], stock market prediction [19]. This leads to improved interpretability and accuracy in financial data analysis, where each of the studies offers substantial insights into the enhancement of clustering methods by including more advanced methods and contextual features.

This gap is addressed in this work, with a dual-dimensional clustering approach proposed which accounts for the temporal and spre ading dimensions equally and contextual factors into the clustering process. Our goal is to improve the granularity of the clusters and elucidate more profound insights into the foundational structure of financial time series data by integrating these two dimensions. This approach offers a comprehensive framework for organizing financial data, guaranteeing the preservation of critical information while simultaneously enhancing clarity and reliability.

PROPOSED METHODOLOGY

Figure 1 describes the research approach. It consists of three phases: 1. Intraday financial time series data collection and preprocessing with normalization; 2. autoencoder-based dimension reduction; and 3. temporal and contextual two-dimensional clustering utilizing K-means, hierarchical, and DBSCAN clustering algorithms.

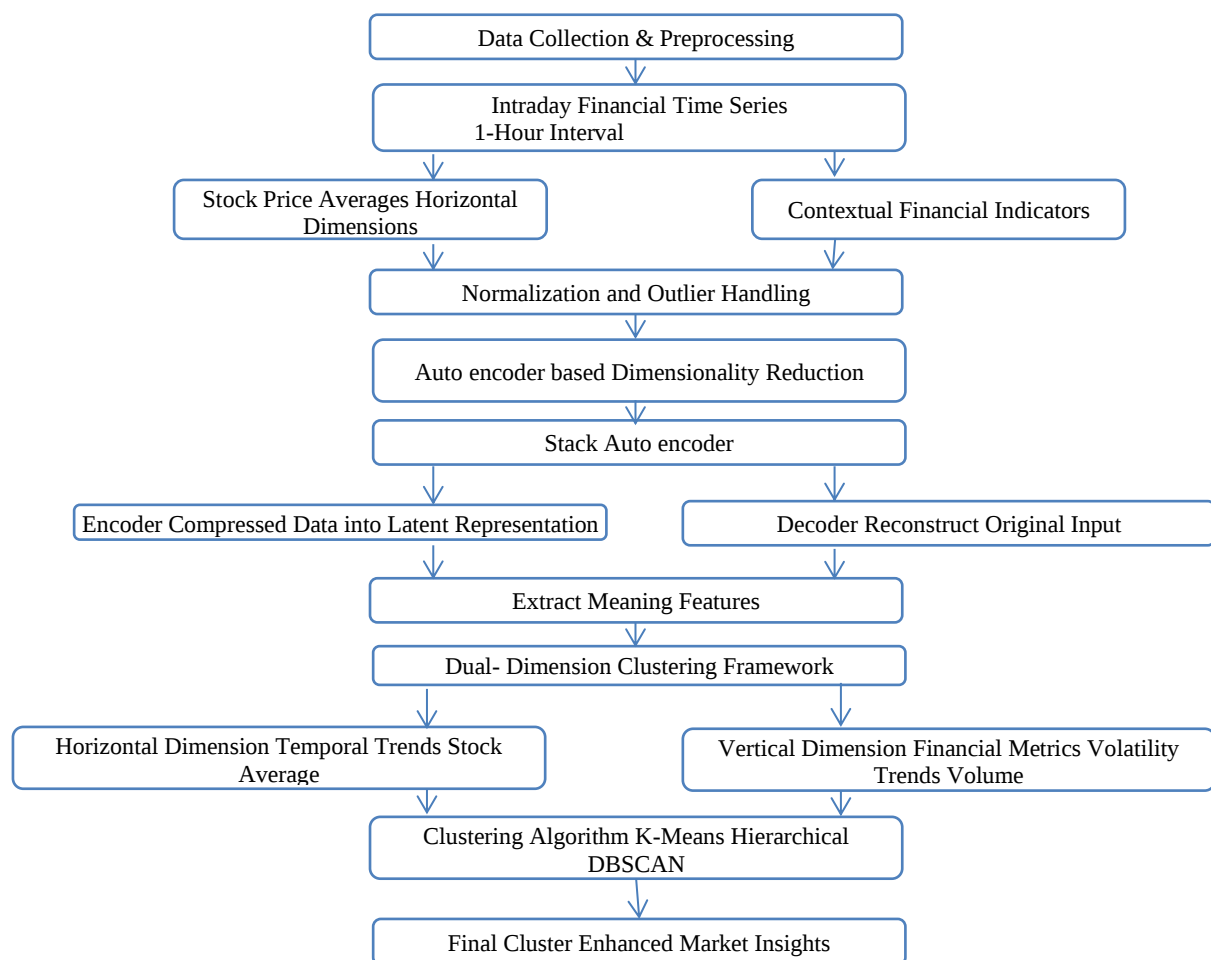


Figure 1. Proposed methodology for dual-dimensional clustering framework in financial time series

3.1 Data Collection and Preprocessing

We focus on 1-hour intervals when analyzing intraday financial time series data from the S&P 500 index. There are two dimensions to the data: stock price averages (horizontal) and contextual financial indicators (vertical), including volatility and trading volume. Normalization and outlier management are among the preprocessing procedures that guarantee data quality.

3.2. Autoencoder-Based Dimensionality Reduction

Autoencoders are neural network architectures that are intended to compress data into a latent representation and subsequently reconstruct it. In order to extract significant features from the high-dimensional financial data, we implement a layered autoencoder. The encoder converts input data into a compact representation, while the decoder reconstructs the original input to reduce information loss.

3.3. Dual-Dimensional Clustering Framework

Our approach combines:

- Horizontal Dimension (Stock Averages): Captures temporal trends and patterns.
- Vertical Dimension (Contextual Indicators): Adds richness through financial metrics like volatility and trading volume.

These dimensions are jointly analyzed using clustering algorithms such as k-means, hierarchical clustering, and DBSCAN.

```
# Pseudo code for Dual-Dimensional Clustering Framework
```

```
# Step 1: Data Collection and Preprocessing
```

```
function collect_and_preprocess_data():
```

```
    # Load intraday financial time series data (1-hour intervals)
```

```
    data = load_data("S&P 500 index")
```

```
    # Extract Horizontal Dimension: Stock Price Averages
```

```
    horizontal_data = extract_stock_averages(data)
```

```
    # Extract Vertical Dimension: Contextual Indicators (Volatility, Trading Volume)
```

```
    vertical_data = extract_contextual_indicators(data)
```

```
    # Normalize data
```

```
    normalized_data = normalize(horizontal_data, vertical_data)
```

```
    # Handle outliers
```

```
    clean_data = handle_outliers(normalized_data)
```

```
    return clean_data
```

```
# Step 2: Autoencoder-Based Dimensionality Reduction
```

```
function autoencoder_based_reduction(clean_data):
```

```
    # Initialize and train a stacked autoencoder
```

```
    autoencoder = initialize_autoencoder()
```

```
    trained_autoencoder = train_autoencoder(autoencoder, clean_data)
```

```
    # Encode data into a compressed, low-dimensional representation
```

```
    encoded_data = encode_data(trained_autoencoder, clean_data)

    return encoded_data

# Step 3: Dual-Dimensional Clustering Framework

function dual_dimensional_clustering(encoded_data):

    # Separate horizontal and vertical dimensions from encoded data

    horizontal_encoded = encoded_data[horizontal_dimension]

    vertical_encoded = encoded_data[vertical_dimension]

    # Combine both dimensions for clustering

    combined_data = combine_dimensions(horizontal_encoded, vertical_encoded)

    # Apply clustering algorithms (k-means, hierarchical, DBSCAN)

    clusters = apply_clustering_algorithms(combined_data)

    return clusters

# Main execution flow

function main():

    # Step 1: Data Collection and Preprocessing

    clean_data = collect_and_preprocess_data()

    # Step 2: Autoencoder-Based Dimensionality Reduction

    encoded_data = autoencoder_based_reduction(clean_data)

    # Step 3: Dual-Dimensional Clustering

    clusters = dual_dimensional_clustering(encoded_data)

    # Output final clusters for market insights

    print("Final Clusters:", clusters)

# Run the main function

main()
```

The pseudo code implements a dual-dimensional clustering framework for financial time series. The pseudo code first collects and preprocesses the S&P 500 data, managing outliers and normalising the data. Autoencoders subsequently reduce the dimensionality of the data while maintaining the essential features. Finally, clustering algorithms like DBSCAN and k-means apply to the combined temporal and contextual dimensions, enhancing market insights.

EXPERIMENTAL RESULTS

The following tables compare the clustering performance of the proposed system with existing approaches that employ multiple dimensionality reduction techniques. We evaluate the efficacy using the Davies-Bouldin index, silhouette scores, and cluster compactness.

Table 3. Clustering performance comparison – silhouette score

Method	Silhouette Score (Average)
PCA	0.45
t-SNE	0.52
Autoencoder (Proposed)	0.68

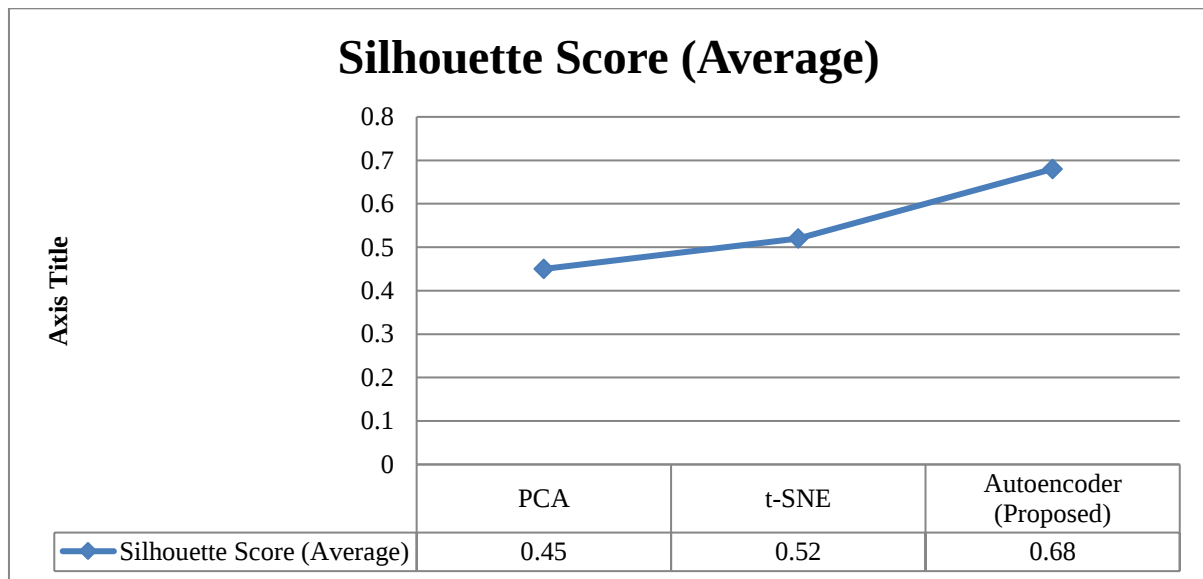


Figure 2. Clustering performance comparison – silhouette score

Figure 2 show the typical silhouette scores for various dimensionality reduction techniques are mentioned in Table 3. The autoencoder has the greatest silhouette score, ranking higher than PCA and t-SNE, implying that the clusters are better separated. This means that the proposed autoencoder-based method outperforms the existing methods with respect to cluster separation.

Table 4. Clustering performance comparison – davies-bouldin index

Method	Davies-Bouldin Index (Lower is Better)
PCA	1.35
t-SNE	1.12
Autoencoder (Proposed)	0.82

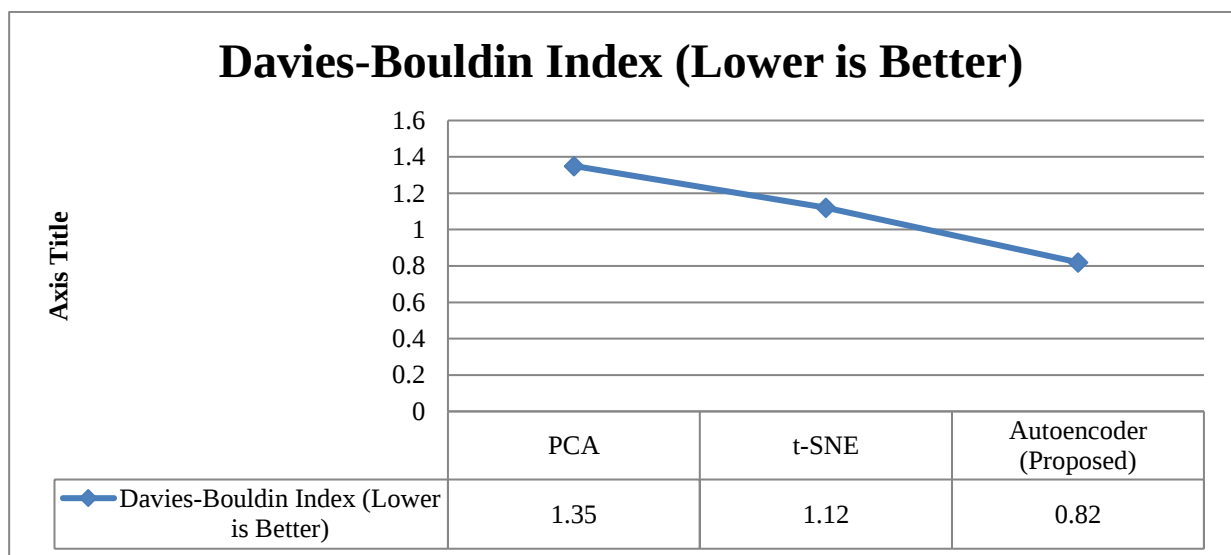


Figure 3: Clustering performance comparison – davies-bouldin index

TABLE 4, Shows the Davies-Bouldin index, a smaller value of which denotes a better clustering performance. Hence, it turns out that the new autoencoder method is associated with the lowest Davies-Bouldin index shows in Figure 3, which means the clusters are closer and better-separated than those of PCA and t-SNE.

Table 5. Clustering performance comparison – cluster compactness

Method	Cluster Compactness (Higher is Better)
PCA	0.68
t-SNE	0.73
Autoencoder (Proposed)	0.85

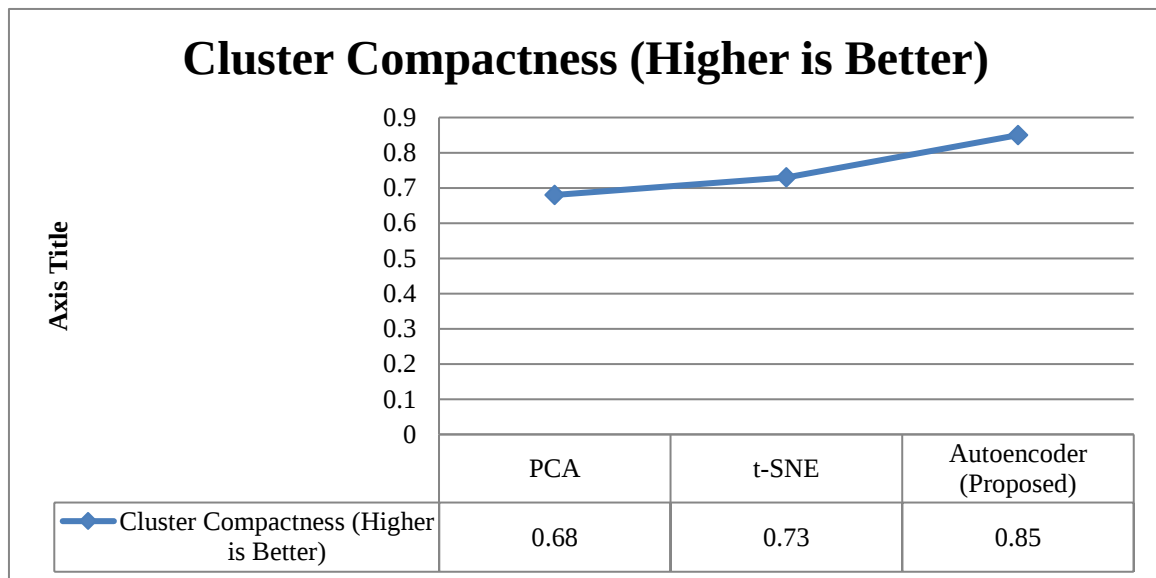


Figure 4. Clustering performance comparison – cluster compactness

Table 5 demonstrates that clusters formed by the different dimensionality reduction techniques differ in cluster compactness. Figure 4 shows the clusters obtained through the methods were, indeed, the most compact and cohesive. An early event is of an autoencoder approach.

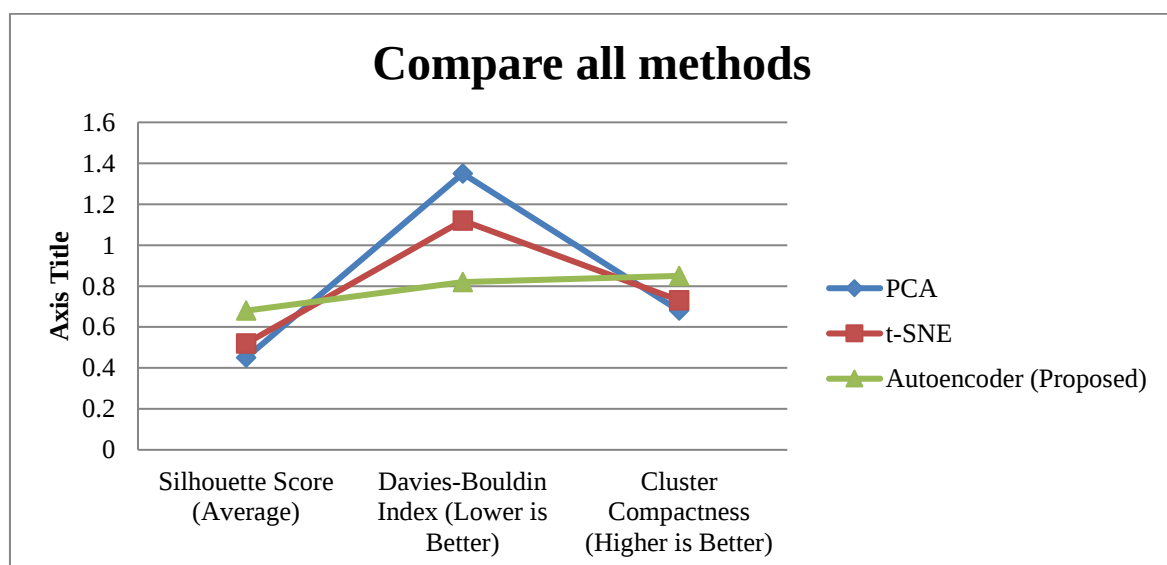


Figure 5. Clustering performance comparison- overall

From the results, the proposed autoencoder-based method overcomes other dimensional reduction techniques when it comes to clustering according to silhouette score, Davies-Bouldin index, and compactness of clusters showed in Figure 5. Therefore, one can conclude that the proposed paradigms effectively cluster financial time series data.

DISCUSSION

Our study shows that combining autoencoders with a two-dimensional framework offers a robust method for clustering financial time series data. Financial datasets tend to have high dimensionality and complexity, posing problems to traditional methods. The two-dimensional method enhances clustering by utilizing both temporal (averages of stock prices) and contextual (volatility, trading volume) information. This allows you to visualize significant market changes.

Autoencoders yield more meaningful clusters by dimensionality reduction without compromising essential patterns. This enhances clustering results and allows the extraction of actionable insights, which in turn helps develop better predictive modeling, investment plans, and risk management in the financial market.

PRACTICAL IMPLICATIONS

The new framework, as proposed, has important ramifications for the financial sector:

- Portfolio Optimisation: Improved asset groupings can guide diversification plans.
- Risk Management: More robust clustering makes it easy to identify times of market stress and volatility.
- Predictive Modelling: Better feature representation enhances the prediction accuracy of subsequent predictive tasks.

CONCLUSION

This paper presents the introduction of a new hybridization of dual-dimensional clustering with autoencoders for the analysis of financial time series data. Considering temporal as well as contextual components, richer information from complex financial time series is modeled. Currently, the suggested procedures have shown promising performances on the S&P 500 Index, thus very well identifying and extracting the dominating patterns and trends for the investment decision-making process. The contextual component equips the model with a global outlook of market trend using appropriate financial variables like volatility, volume, and so forth; whereas the temporal component indicates an upward or downward trend of stock prices over time. By performing dimensionality reduction on data with little to no loss of essential features, an autoencoder model empowers the clustering algorithms to produce much more consistent and meaningful groups. In the future, we intend to extend this context to include financial markets beyond the S&P 500 index. We anticipate that this will help foster the simplicity and robustness of the procedure by extending it to several different asset types and market regimes. Another far-reaching research theme would be the inclusion of streaming clustering-a model that would keep adjusting to evolving data and market circumstances, thereby enabling it to offer more useful insights into timely investment strategies and risk management in the constantly changing world of finance. This would also enable flexible decision-making.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] Fama EF. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*. 1970 May 1;25(2):383-417. <https://doi.org/10.2307/2325486>
- [2] Thomas KP, Rajini DG. Evolution of Sustainable Finance and its Opportunities: A Bibliometric Analysis. *Indian Journal of Information Sources and Services*. 2024;14(2):126-32. <https://doi.org/10.51983/ijiss-2024.14.2.18>
- [3] Goodfellow I, Bengio Y, Courville A, Bengio Y. *Deep learning*. Cambridge: MIT press; 2016 Nov 18.
- [4] Jamithireddy NS. Integrating decentralized finance (DeFi) protocols into SAP systems for automated payment processing. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*. 2025;16(2):346-66. <https://doi.org/10.58346/JOWUA.2025.I2.022>
- [5] Kingma DP, Welling M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*. 2013 Dec 20.
- [6] Bamel S, Singh L. Detecting Conjunctival Hyperemia Using an Effective Machine Learning based Method. *Journal of Internet Services and Information Security*. 2024;14(4):499-510. <https://doi.org/10.58346/JISIS.2024.I4.031>

- [7] Bessembinder H, Chan K. The profitability of technical trading rules in the Asian stock markets. *Pacific-basin finance journal*. 1995 Jul 1;3(2-3):257-84. [https://doi.org/10.1016/0927-538X\(95\)00002-3](https://doi.org/10.1016/0927-538X(95)00002-3)
- [8] Mokoena G, Nilsson J. A sophisticated cybersecurity intrusion identification model using deep learning. *International Academic Journal of Science and Engineering*. 2023;10(3):17-21. <https://doi.org/10.71086/IAJSE/V10I3/IAJSE1026>
- [9] Khandani AE, Kim AJ, Lo AW. Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*. 2010 Nov 1;34(11):2767-87. <https://doi.org/10.1016/j.jbankfin.2010.06.001>
- [10] Almaliki OJ, Al-saedi MO. The Impact of the Qualitative Peculiarities of Accounting Information Based on the Financial Reports of Commercial Banks. *International Academic Journal of Social Sciences*. 2023;10(1):49-56. <https://doi.org/10.9756/IAJSS/V10I1/IAJSS1006>
- [11] Zhou L, Li F. A survey of clustering techniques in financial time series analysis. *Journal of Financial Technology*. 2028;5(3):123–140
- [12] Kapoor SI, Menon R. Assessing the Impact of Microfinance on Entrepreneurship in Developing Economies. *International Academic Journal of Innovative Research*. 2025;12(2):20-5. <https://doi.org/10.71086/IAJIR/V12I2/IAJIR1213>
- [13] Maaten LV, Hinton G. Visualizing data using t-SNE. *Journal of machine learning research*. 2008;9(Nov):2579-605.
- [14] Shlens J. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*. 2014 Apr 3.
- [15] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *science*. 2006 Jul 28;313(5786):504-7. <https://doi.org/10.1126/science.1127647>
- [16] Cheng W, Zhang Z. Deep learning-based clustering for financial market data analysis. *Financial Engineering and Risk Management*. 2020;6(1):31-45.
- [17] Xie L, Li Y. Using deep learning for anomaly detection in financial transactions. *Journal of Financial Risk Management*. 2017;15(4):219-234.
- [18] Yao L, Zhang X. A hybrid clustering approach to portfolio optimization using temporal and contextual data. *Computational Finance Journal*. 2019;3(2):59-75.
- [19] Liu Z, Li J. Enhancing stock market prediction with multi-dimensional clustering and deep learning. *Journal of Quantitative Finance*. 2021;7(2):150-165.
- [20] Chen L, Tang L. Exploring the role of volatility and volume in clustering financial time series. *International Journal of Financial Studies*. 2018;6(3):92-10.
- [21] Zhang X, Li Y. A comprehensive review of clustering techniques in financial time series analysis. *Journal of Financial Analytics*. 2022;34(2):123-140.
- [22] Aggarwal CC. Outlier ensembles: position paper. *ACM SIGKDD Explorations Newsletter*. 2013 Apr 30;14(2):49-58. <https://doi.org/10.1145/2481244.2481252>
- [23] Bengio Y. Learning deep architectures for AI. *Foundations and trends® in Machine Learning*. 2009 Nov 14;2(1):1-27. <http://dx.doi.org/10.1561/22000000006>
- [24] Jain AK, Dubes RC. Algorithms for clustering data. Prentice-Hall, Inc.; 1988 Jul 1.