

Review

Not peer-reviewed version

Clustering Methods for Analysis and Optimization of Time Series Data

[Nikolay Rumenov Kakanakov](#) and [Teodora Atanasova Mecheva](#)*

Posted Date: 9 February 2026

doi: 10.20944/preprints202602.0646.v1

Keywords: time series analysis; clustering; unsupervised learning; pattern discovery; data optimization; temporal data mining



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Clustering Methods for Analysis and Optimization of Time Series Data

Nikolay Kakanakov ^{1,2} and Teodora Mecheva ^{2,3,*}

¹ Technical University Sofia, Plovdiv branch

² Center of competence "Smart mechatronic, eco-and energy-saving systems and technologies"

³ Technical University Sofia, Plovdiv branch

* Correspondence: teodora.mecheva@tu-plovdiv.bg; Tel.: +359897010359

Abstract

Time series data, characterized by temporal dependencies, seasonality, and noise, are prevalent in domains such as healthcare, finance, energy, and transportation. Effective clustering of time series enables the discovery of patterns, supports forecasting, and facilitates data-driven decision-making. This paper provides a comprehensive review of time series clustering techniques, including conventional methods (e.g., k-means, hierarchical, and fuzzy clustering), similarity-based approaches (e.g., Dynamic Time Warping), feature-based methods, and deep learning models (e.g., autoencoders, convolutional and recurrent neural networks). The review analyzes tasks, application domains, performance outcomes, and key limitations, highlighting common challenges such as computational complexity, sensitivity to noise, and scalability issues. A particular focus is given to transport-related time series, including traffic flow, travel time, and congestion patterns, demonstrating how clustering can support traffic state classification, anomaly detection, and infrastructure planning. The analysis reveals a trade-off between accuracy, interpretability, and computational efficiency, emphasizing the need for scalable, robust, and domain-aware clustering frameworks. Finally, practical directions for future research are discussed, including lightweight hybrid approaches and transport-specific feature engineering to enhance clustering performance in real-world applications.

Keywords: time series analysis; clustering; unsupervised learning; pattern discovery; data optimization; temporal data mining

1. Introduction

Time series data consists of observations recorded sequentially over time and are characterized by temporal dependencies, trends, seasonality, and noise. In the context of time series data, similarity can be defined in terms of raw values, temporal dynamics, shape, or underlying generative processes. Traditional analytical techniques often focus on forecasting or statistical modeling, assuming a predefined structure or distribution. However, in many real-world scenarios, the underlying patterns are unknown or too complex to be modeled explicitly. This motivates the use of data mining techniques, among which clustering plays a central role [1–4].

Clustering methods aim to group similar objects such that time series within the same cluster exhibit comparable behavior, while series from different clusters are dissimilar. Unlike static data, time series exhibit strong temporal ordering and autocorrelation. Observations are often non-independent and may vary in length, scale, or sampling frequency. Additional challenges include missing values, noise, temporal misalignment, and concept drift. These properties complicate the direct application of classical clustering algorithms designed for independent and identically distributed data. Another important challenge is the definition of similarity. Two time series may represent the same underlying phenomenon while being shifted in time, scaled in amplitude, or evolving at different speeds. Consequently, the choice of similarity measures and feature extraction has a decisive impact on clustering performance and interpretability. Standard approach for time

series clustering includes preprocessing step—similarity measure and feature extraction, while deep approach includes deep learning methods [5–11].

There is no generally accepted classification for time series clustering methods. [12] considers time series clustering as a conventional technique and a subtype of Multivariate Time Series Anomaly Detection. [13] examined time series clustering methods in three main phases: data representation, similarity measure, and clustering algorithm. [14] explores contemporary clustering algorithms within the machine learning paradigm, focusing on five primary methodologies: centroid-based, hierarchical, density-based, distribution-based, and graph-based clustering. [15] classifies clustering techniques into hard clustering and soft clustering. [16] point out three major groups: depending upon whether they work directly with raw data, indirectly with features extracted from the raw data, or indirectly with models built from the raw data. [17] focuses on deep learning methods and considers 3 main components: network architecture, pretext loss, and clustering loss. All authors point out partitioning methods (k-means, k-medoids, fuzzy C-means, fuzzy C-medoids), hierarchical clustering (bottom-up and top-down), model-based (self-organizing map), and density-based Clustering (DBSCAN—Density-Based Spatial Clustering of Applications with Noise, OPTICS—Ordering Points To Identify the Clustering Structure). Most of clustering methods face significant challenges when dealing with high dimensional, noisy, and large-scale data.

2. Time Series Data Clustering

Clustering approaches provide effective tools for organizing and extracting insights from big and complicated datasets. The most appropriate clustering method is determined by the data properties and the application's unique needs. It is critical to assess the quality of clustering findings in order to assure the validity and use of the produced clusters [12–14].

In this relation Table 1 shows the overviewed research over clustering time series methods, techniques, tasks, applications, results, and limitations.

Table 1. Reviewed articles of clustering time series methods.

Source	Techniques	Tasks	Applications	Results	Limitations
[13]	Deep auto-encoders (DAE), deep convolutional auto-encoders (DCAE), sliding window, k-means	Grouping accelerometer data on cormorant movement	Biology	Indicates that DCAE shows the best behavior	High computational complexity
[18]	FLAG (Fused Lasso Generalized eigenvector method)	Shapelet discovery task	Finance, medical diagnosis, and weather forecasting	The proposed method is orders of magnitudes faster than existing shapelet-based methods, while achieving comparable or even better classification accuracies	High computational complexity; Difficult parameter selection; difficult implementation
[19]	Spectral clustering, sliding window	Distributing energy resources into the power grid	Energetics	Over a 94% reduction in error	Large computing resource

[20]	K-Means, K-Medoids, Spectral clustering, Self organized maps	Grouping recurrent plot of stock indexes	Finance	Introduced a feature-based clustering frame for grouping risk patterns	It does not reflect indicators characterizing long-term periods
[21]	Temporal Distortion Index (TDI), Dynamic Time Warping (DTW), Rank-difference-Rank score (RdR score)	Review over clustering techniques	-	Presents a comprehensive survey of data clustering	Only theoretical review
[22]	Peaks-Over-Threshold (POT), sliding window	Predicting sea levels and storm surges	Modeling of rare and extreme events	Near 500-year modelled dataset of sea level and non-tidal residual under pre-industrial conditions	Difficulty in characterizing the level of temporal clustering when focusing on relatively short timeframes
[23]	Spectral clustering	Finding similarities in timeseries EEG brain waves	Healthcare	The created models are accurate and can be used for timeseries classification	The choice of similarity measures and the choice of cluster prototype strongly influence the clustering results
[24]	30 unsupervised clustering algorithms	Finding similarities in electronic health records data	Healthcare	DTW and its lower-bound variants (i.e., LB-Improved and DTW-LB) are highly robust clustering algorithms	Cannot accommodate trajectories with varied lengths
[25]	Fuzzy Clustering Method (FCM), Iterative Self-Organizing Data Analysis Technique Algorithm (ISODATA), and Ordering Points To Identify the Clustering Structure (OPTICS)	Finding similarities in global sea level anomaly time series	Sea level change prediction	The ISODATA demonstrates superior clustering performance	Determining the number of clusters
[26]	Convolution Neural Network (CNN) and Long	Clustering multivariate	Climate prediction	The proposed approach outperforms	Need robust imputation,

	Short-Term Memory (LSTM)	spatiotemporal climate dataset		conventional and deep learning clustering algorithms	resource consuming
[27]	Clust3D	Disease prediction	Healthcare	Proposed method produces well separated clusters compared to existing heuristic methods	Sensible to high dimensional data
[28]	Medoid-Shape, K-Shape	In database IoT commodity time series data clustering	Trade	Extensive experiments show the high efficiency of proposed methods	Performs inefficiently with long subsequences
[29]	K-PCD	Building energy consumption predictions	Energetics	Clustering performance of the K-PCD algorithm is superior to traditional K-means algorithm	Sensitive to noise and outliers, high computational complexity
[30]	Tsfresh, Random Forest, Laplacian Score, and unsupervised Spectral Feature Selection, evidential c-means (ECM)	Analyzing barometers attributes related to pain intensity	Healthcare	Results show excellent separability and compactness	Poor scalability, difficult parameter tuning
[31]	A framework unifying pattern extraction and data prediction	Short-term load forecasting in smart grids	Energetics	Reduces the mean absolute percentage error by 2% to 5%	The performance on a real-world data set showed a decline
[32]	DTW, agglomerative hierarchical clustering, and principal component analysis	Predicting strata deformation	Geological hazard monitoring	Proposed method is an effective analysis method for strata deformation	High computational complexity, lack of scalability
[33]	DUET—a framework combining temporal clustering and channel clustering module	Multivariate time series forecasting	Financial investment, energy management, weather forecasting, and traffic optimization	Extensive experiments on 25 real-world datasets, demonstrate the state-of-the-art	Sensitive to noise, need parameter tuning

				performance of DUET	
[34]	A novel model-based approach based on Markov chain Monte Carlo	Clustering epidemiological data	Healthcare	The clusters are well separated, moreover	High computational cost
[35]	Dirichlet Process Mixture of Sparse Heteroskedastic Multi-output Gaussian Processes (DPM-SHMGP)	Clustering electricity load data	Energetics	The model's superior clustering performance compared to established clustering algorithms	computationally expensive, requires extensive hyperparameter tuning, and suffers from scalability, identifiability
[36]	Fuzzy C-means (FCM) clustering, the principle of justifiable granularity (PJG), and particle swarm optimization (PSO)	Prognosing Taiwan Weighted Stock Index (TAIEX) datasets and the Shanghai Composite Index (SHCI) datasets	Finance	the proposed model in this paper achieves higher forecasting accuracy than other models	Complex to implement, high computational cost
[37]	Autoencoder-based Iterative Modeling and Subsequence Clustering Algorithm	Multivariate Time-Series Sub-Sequence Clustering	Mechatronics	comparison with seven other state-of-the-art algorithms and eight datasets shows the capability and the increased performance of the algorithm	The method is sensitive to outliers
[38]	sliced Wasserstein k-means clustering	Identifying distinct market regimes	Finance	Using a grid of fixed projections throughout the algorithm simplifies the implementation and reduces the computational cost	Initialization sensitivity
[39]	Hierarchical Consensus Clustering (HCC) and the Statistics of Split Timeseries (SST)	Estimating cognitive states	Healthcare	Achieves 99% accuracy with lower computational cost.	Limited scalability, may lead to information loss
[40]	K-medoid	Reducing labeling cost	Healthcare	The proposed method increases accuracy in real world dataset	Sensitive to noises

[41]	Gramian Angular Field (GAF), Variational AutoEncoders (VAEs), and Deep Embedded Clustering (DEC), K-means	Clustering Bitcoin Tick-bar price	Finance	The method leads to financially interpretable clusters	Difficult parameter selection
[42]	K-medoids, COMB distance	SCADA (Supervisory Control And Data Acquisition) wind farm data clustering	Energetics	The clustering results contributed to the diagnosis of the wind flow and its interaction with the terrain; the clustering results may be used to perform anomaly detection	It does not consider turbulence index and air temperature
[43]	Euclidean distances, DTW, K-means clustering	Forecasts electricity consumption in a smart grid	Energetics	Time-series clustering method performed better than that using the total amount of electricity demand	Does not reflect electricity consumption data, apartment characteristics, and household characteristics
[44]	K-means, p-median, agglomerative clustering, a-lex, spectralCS, spectralAMI	Recognition of spatio-temporal traffic patterns	Transport	K-means and agglomerative clustering may be the most scalable methods	Does not reflect seasonality effects
[45]	DTW, Euclidean distance and MINDIST, K-means, K-medoids, and Hierarchical Agglomerative Clustering (HAC)	Public transportation smart card data clustering	Transport	Every algorithm has strengths and weaknesses, but generally, all perform well.	While smoothing out the time-series got rid of the noise, it also made it more difficult to discern patterns
[46]	K-Means, Sobolev distance	Classify voltage profiles obtained as numerical solutions of the PDE model for the case of symmetric Li/Li cells	Energetics	Cluster analysis can play a key role in discovering hidden structures within the data	Sensitivity to noise; The method is not scalable; The method requires to set the number of clusters

	MDU-Net comprises of two modules: Multi- resolution	Multivariate electricity time series prediction	Energetics	MDU-Net significantly outperforms state-of-the-art baselines in multivariate electricity time series prediction	Overfitting risk, reduced scalability, increased design complexity
[47]	Hierarchical Union learning (MRHU) and Differential Channel Clustering Fusion (DCCF)				

2.1. Methods

Among the methods considered, the most common are conventional techniques such as k-means, k-medoids, hierarchical agglomerative clustering, spectral clustering, and fuzzy C-means, [13,19,20,23,30,32,36,38–46]. Similarity-based methods, such as DTW, its variants, Euclidean distance, Sobolev distance, and hybrid distance measures, improve clustering performance for time-shifted or temporally distorted sequences [21,32,43,45,46]. Feature extraction and subsequence-based approaches, including sliding window techniques, shapelet-based methods, Tsfresh, and spectral feature selection, aim to capture discriminative local patterns in time series data [13,18,22,30,31]. Deep learning-based approaches, such as DAE, DCAT, CNN, convo LSTM, VAE, and hybrid deep architectures are preferred when dealing with complex, multivariate, and nonlinear time series. These models enable automatic feature learning and joint clustering or prediction [13,26,33,37,41,47]. Probabilistic and Bayesian models, including Dirichlet Process Mixture models and Markov Chain Monte Carlo (MCMC)-based approaches, provide flexible frameworks capable of modeling uncertainty and automatically determining the number of clusters [34,35].

2.2. Tasks

The primary problems addressed include time series clustering, similarity detection, pattern discovery, and forecasting [19,21,23,26,33]. Several studies focus on handling noisy, non-stationary, and temporally distorted data, which is common in real-world scenarios such as EEG signals, energy load profiles, and financial time series [23,26,32,43]. Time series clustering is often used as a preprocessing or exploratory step for higher-level tasks such as anomaly detection, classification, and decision support [22,30,42].

2.3. Application

[13] assumes that the proposed method can be used in finance, medical diagnosis, and weather forecasting. [18,33] point out several application domains—financial investment, energy management, weather forecasting, and traffic optimization. [18,23,24,27,30,34,39,40] investigate healthcare data. [22,25,26,32] examine problems in the field of natural phenomena. [20,28,38,41] deal with problems in the field of economics. [13] examines biological data. [19,29,31,35,42,46,47] address energetic issues. [44,45] applied clustering methods over transport data. [37] examines mechatronic issues. [21] provides only a theoretical overview and does not indicate a specific area of application.

2.4. Results

Most studies report significant improvements over baseline methods, such as higher clustering accuracy, reduced forecasting error, or improved separability of clusters [18,19,23,26,29,33,35,39,47]. Deep learning and hybrid approaches frequently achieve state-of-the-art performance, particularly for multivariate and large-scale datasets [26,33,37,41,47]. Some works also highlight improvements in computational efficiency or scalability compared to traditional shapelet-based or distance-based methods [18,38,39].

2.5. Limitations

Despite their effectiveness, many approaches suffer from high computational complexity and resource consumption, especially deep learning, DTW-based, and probabilistic models [13,18,26,32,34,35,47]. Sensitivity to noise and outliers is another common limitation, affecting both classical and advanced clustering techniques [29,37,40,46]. Several studies report challenges related to parameter tuning, scalability, and implementation complexity [18,30,33,35,36]. These limitations indicate that there is still no universally optimal solution for time series clustering, particularly in real-world, large-scale environments [21,24,44].

3. Clustering Transport Time Series: Use Case

3.1. Overview

Transport systems generate large volumes of time series data through traffic sensors, GPS-equipped vehicles, smart ticketing systems, and intelligent transportation infrastructure. Typical transport time series include traffic flow, vehicle speed, travel time, congestion index, passenger demand, and vehicle counts, often recorded at regular intervals across multiple spatial locations. These datasets are inherently multivariate, noisy, seasonal, and non-stationary, making them well-suited for clustering-based analysis [48–62].

A key objective in transport time series analysis is to identify segments with similar temporal behavior without assuming predefined traffic models. Clustering enables the discovery of recurring traffic patterns such as peak-hour congestion [44,49,61], off-peak free-flow conditions [50–52,55,56], incident-induced disruptions [48], seasonal demand variations [54], anomaly detection [54,60], or cybersecurity [58].

Different clustering paradigms can be applied to transport time series data. Distance-based methods such as DTW combined with partitioning algorithms (e.g., k-means or k-medoids) are effective for handling temporal misalignment between traffic patterns. Feature-based approaches extract statistical, frequency-domain, or shape-based features to reduce dimensionality and improve scalability. Model-based and deep learning approaches, including auto-encoders and recurrent neural networks, are increasingly used for clustering large-scale multivariate transport datasets, capturing complex temporal dependencies and spatial correlations. For example, road segments or sensor locations can be clustered based on daily or weekly traffic flow profiles, allowing transportation authorities to distinguish between stable corridors, bottleneck-prone areas, and highly variable routes.

The outcomes of transport time series clustering include improved traffic state classification, congestion pattern recognition, incident detection, and infrastructure planning support. By grouping similar temporal behaviors, clustering results can guide adaptive traffic signal control, optimize public transport scheduling, and support data-driven decision-making in smart transportation systems.

3.2. Example

The goal of the analysis is to cluster temporal speed profiles (and optionally other variables such as counts or weight aggregates) by sensor/road segment/day to discover typical daily patterns, detect anomalies and incidents, and support planning and adaptive control. For the purposes of the analysis, is selected DTW + k-medoids pipeline, as this combination is practical and robust to temporal misalignment.

3.2.1. Data

The data is set in the following format:

Fields: Date;Direction;Plate number;Country;Speed;Class;Lane;Length;Weight

Sample rows (semicolon-delimited):

30.9.2023 r. 12:00:27;Center;159785927043;;49;1;L1;5500;1086
 30.9.2023 r. 12:00:33;Center;3658603332998141999;66;1;;L1;5200;826
 30.9.2023 r. 12:01:16;Ring road;4391747583219136573;66;0;;L1;5200;0
 30.9.2023 r. 12:01:17;Ring road;201064482549844772;;66;0;L1;5200;0
 30.9.2023 r. 12:02:03;Center;2661655543687701673;;67;1;L1;5600;429

Date: timestamp of the measurement.

Direction: spatial attribute (e.g., "Center").

Plate number: vehicle identifier (may be noisy or partially anonymized).

Country: extracted from previous attribute—Plate number

Speed: primary temporal variable for profiling.

Class;Lane;Length;Weight: numeric attributes usable for stratification or filtering.

3.2.2. Preprocessing

Preprocessing consists of three steps: construction of time series; normalization and Dynamic Time Warping.

Step 1: Construct time series:

Build time series per entity (e.g., daily speed profile per sensor or number vehicles per time interval).

Parse the CSV, standardize datetime formats.

Filter/segment data by direction, sensor, or road segment; aggregate (e.g., 1, 5, or 15 minutes).

Handle missing speed values (short gaps: linear interpolation; large gaps: exclude or flag).

Extract summary features (mean, std, peak time, autocorrelation).

Separate weekdays vs. weekends or cluster per day-of-week to handle weekly seasonality.

Step 2: Normalization.

To focus the clustering on the shape of daily traffic profiles rather than their absolute magnitude, each time series was normalized independently using z-score normalization. For a given series $\{x_t\}_{t=1}^T$ the normalized values were computed as:

$$x'_t = \frac{x_t - \mu_x}{\sigma_x}, \quad (1)$$

where μ_x and σ_x denote the mean value and standard deviation of the series. This transformation ensures zero mean value and unit variance for each profile, allowing DTW to capture relative temporal dynamics while reducing the influence of systematic speed level differences across days or road segments. Raw (unnormalized) series may additionally be retained for downstream analyses where absolute speed levels are of interest.

Step 3: Dynamic Time Warping.

DTW computes a similarity score between two time series by non-linearly aligning them along the temporal axis. Given two sequences

$$X = (x_1, \dots, x_n), Y = (y_1, \dots, y_m),$$

DTW defines a local cost function

$$d_{(i,j)} = |x_i - y_j| \quad (2)$$

and constructs a cumulative cost matrix $D \in R^{n \times m}$ according to the recurrence:

$$D(i,j) = d(i,j) + \min \begin{cases} D(i-1,j) \\ D(i,j-1) \\ D(i-1,j-1) \end{cases} \quad (3)$$

with boundary condition $D(1,1) = d(1,1)$. The DTW distance between X and Y is defined as the minimum cumulative cost $D(n,m)$ along a monotonic warping path from $(1,1)$ to (n,m) . To prevent unrealistic alignments and reduce computational complexity, a Sakoe–Chiba band constraint was applied.

3.2.3. Clustering

Clustering is performed using the k-medoids algorithm (Partitioning Around Medoids, PAM) applied to the precomputed pairwise DTW distance matrix. In contrast to k-means, k-medoids operates directly on arbitrary distance measures and selects actual observations as cluster representatives (medoids).

Given a set of time series X_1, \dots, X_N and their DTW distance matrix D , PAM iteratively minimizes the sum of within-cluster dissimilarities by optimizing the choice of medoids. Each time series is assigned to the cluster of the nearest medoid under the DTW distance. The number of clusters k is selected using a combination of internal validity criteria, including the silhouette coefficient computed on DTW distances, the elbow method, and the Davies–Bouldin index.

3.2.4. Evaluation and Interpretation

Cluster quality was assessed using internal validation metrics, including the silhouette score based on DTW distances and the intra-cluster DTW variance.

To support interpretability, cluster medoids are visualized as representative daily speed profiles, with individual cluster members overlaid to reveal characteristic temporal patterns such as peak timing, congestion duration, and recovery dynamics.

Spatial analysis is conducted by mapping cluster assignments to corresponding sensors or road segments, enabling the identification of recurrent traffic regimes, bottlenecks, and structurally similar corridors.

Time series exhibiting unusually large DTW distances to their assigned medoids are flagged as potential anomalies, indicative of incidents, sensor malfunctions, or atypical traffic conditions.

4. Discussion

The analysis of the reviewed literature reveals several important trends and open challenges in time series clustering research. A clear progression can be observed from classical clustering techniques toward more advanced similarity-based, feature-driven, and deep learning approaches. While conventional methods such as k-means, k-medoids, hierarchical clustering, and fuzzy clustering remain widely used due to their simplicity and interpretability, their effectiveness is often limited when applied to noisy, high-dimensional, or large-scale time series data.

Centroid-based methods, for instance, perform well on small- to medium-scale datasets with low noise, whereas DTW-based or deep learning approaches are better suited for multivariate, non-linear, and temporally misaligned series. In the context of transport datasets, temporal misalignments due to varying congestion patterns make similarity-based clustering, particularly DTW or hybrid deep learning models, more effective than classical approaches.

Nevertheless, these advanced methods introduce a trade-off: improved accuracy and pattern capture often come at the cost of reduced interpretability and increased computational requirements. Feature-based and subsequence-oriented techniques partially alleviate these limitations by reducing dimensionality and capturing local discriminative patterns, yet they require careful feature selection, window sizing, and parameter tuning. Deep learning frameworks further enhance performance on complex multivariate time series but demand substantial computational resources and large datasets, which can hinder practical adoption.

Across all categories, a recurring issue is the balance between clustering accuracy, scalability, and interpretability. No single method consistently outperforms others across all datasets and application domains, highlighting the importance of application-driven method selection. This underscores the need for hybrid frameworks that combine simplicity, scalability, and robust handling of temporal dynamics, particularly for large-scale, noisy, and spatially dependent transport datasets.

In the context of transport time series data, the challenges are further amplified by strong seasonality, spatial dependencies, and non-stationary behavior. The reviewed transport-related studies demonstrate that clustering can successfully uncover recurring traffic patterns, congestion

regimes, and anomalous events. However, existing approaches often fail to jointly capture temporal dynamics, spatial correlations, and scalability requirements, indicating a clear research gap for integrated and efficient clustering frameworks tailored to intelligent transportation systems.

5. Conclusions

This paper presented a review of time series clustering techniques, analyzing a wide range of methods, application domains, achieved results, and inherent limitations. The reviewed studies confirm that time series clustering is a powerful tool for exploratory data analysis, pattern discovery, and decision support in complex temporal datasets.

Classical clustering techniques remain relevant due to their simplicity and interpretability, but they are insufficient for modern large-scale and multivariate time series. Similarity-based and feature-driven methods improve clustering quality but often suffer from high computational complexity and parameter sensitivity. Deep learning and probabilistic models achieve state-of-the-art performance in many scenarios, yet their practical adoption is constrained by resource demands, scalability issues, and reduced interpretability.

In transport applications, effective time series clustering has the potential to significantly enhance traffic monitoring, congestion management, anomaly detection, and infrastructure planning. As transportation systems continue to evolve toward data-driven and intelligent solutions, robust and scalable clustering methods will play a crucial role in supporting smart mobility and decision-making processes.

Future work will focus on practical extensions of existing time series clustering methods with an emphasis on transport-related data. Rather than developing highly complex models, the goal is to evaluate and adapt established clustering techniques for traffic flow, speed, and travel time data. A key direction is the comparison of distance-based and feature-based approaches to identify recurrent traffic patterns such as peak-hour congestion and off-peak conditions.

Additionally, lightweight hybrid solutions combining classical clustering algorithms with basic dimensionality reduction techniques can be explored to ensure scalability and applicability to large transport datasets. The clustering results will be assessed in practical cases, such as traffic state classification and anomaly detection, to support data-driven decision-making in intelligent transportation systems.

Author Contributions: Conceptualization, T.M.; methodology, T.M. and N.K.; formal analysis, T.M. and N.K.; investigation, T.M.; resources, N.K. and T.M.; data curation, T.M.; writing—original draft preparation, T.M.; writing—review and editing, T.M. and N.K.; visualization, T.M.; supervision, N.K.; project administration, N.K.; funding acquisition, N.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Regional Development Fund within the OP “Research, Innovation and Digitalization Programme for Intelligent Transformation 2021-2027”, Project CoC “Smart Mechatronics, Eco- and Energy Saving Systems and Technologies”, No. BG16RFPR002-1.014-0005.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DBSCAN	Density-Based Spatial Clustering of Applications with Noise
OPTICS	Ordering Points To Identify the Clustering Structure
SCADA	Supervisory Control And Data Acquisition
DAE	Deep Auto-encoders
DCAE	Deep Convolutional Auto-encoders
FLAG	Fused LAsso Generalized eigenvector method
TDI	Temporal Distortion Index
DTW	Dynamic Time Warping

RdR	Rank-difference-Rank
FCM	Fuzzy Clustering Method
SODATA	Iterative Self-Organizing Data Analysis Technique Algorithm
OPTICS	Ordering Points To Identify the Clustering Structure
CNN	Convolution Neural Network
LSTM	Long Short Term Memory
ECM	Evidential c-means
FCM	Fuzzy C-means
PJG	Principle of Justifiable Granularity
PSO	Particle Swarm Optimization
HCC	Hierarchical Consensus Clustering
SST	Statistics of Split Timeseries
GAF	Gramian Angular Field
VAEs	Variational AutoEncoders
DEC	Deep Embedded Clustering
SCADA	Supervisory Control And Data Acquisition
HAC	Hierarchical Agglomerative Clustering
MRHU	Multi-resolution Hierarchical Union learning
DCCF	Differential Channel Clustering Fusion
PAM	Partitioning Around Medoids

References

1. Liu, Y.; Tay, D.; Mellow, D. Time Series. **2025**, DOI: 10.1002/9781405198431.wbeal1216.pub2.
2. Zielinski, T.; Hay, J.; Millar, A.J. Period Estimation and Rhythm Detection in Timeseries Data Using BioDare2, the Free, Online, Community Resource. *Methods Mol Biol.* **2022**; 2398:15-32. DOI: 10.1007/978-1-0716-1912-4_2.
3. Paparrizos J.; Bogireddy s.P.T.R. Time-Series Clustering: A Comprehensive Study of Data Mining, Machine Learning, and Deep Learning Methods. *Proc. VLDB Endow.* **2025**, 18, 11, 4380–4395. DOI: 10.14778/3749646.3749700.
4. K. Zhang et al. Self-Supervised Learning for Time Series Analysis: Taxonomy, Progress, and Prospects, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2024**, 46, 10, 6775-6794, DOI: 10.1109/TPAMI.2024.3387317
5. Chaudhry, M.; Shafi, I.; Mahnoor, M.; Vargas, D.L.R.; Thompson, E.B.; Ashraf, I. A Systematic Literature Review on Identifying Patterns Using Unsupervised Clustering Algorithms: A Data Mining Perspective. *Symmetry* **2023**, 15, 1679. DOI: 10.3390/sym15091679.
6. Cong, M.; Li, Z.; Wu, Y.; Lu, Q.; Li, M., Zhou; Z. Evaluating and Optimizing Energy and Comfort Performance in Airport Cooling Systems Through Dynamic Occupancy Modeling and Time-series Clustering. *Building and Environment*, **2025**, 274, 112781 DOI: 10.1016/j.buildenv.2025.112781.
7. Rahkmawati, Y.; & Annisa, S. Clustering Time Series Using Dynamic Time Warping Distance in Provinces in Indonesia Based on Rice Prices. *TIERS Information Technology Journal* **2023**, 4, 2, 115-121, DOI: 10.38043/tiers.v4i2.5081.
8. Özkoç, E.E. Clustering of Time-Series Data. *Data Mining—Methods, Applications and Systems*, **2020**, 6, DOI: 10.5772/intechopen.84490.
9. Lu, Y.; Zhang Y.; Chen, S. A Review of Time Series Data Mining Methods Based on Cluster Analysis, *35th Chinese Control and Decision Conference (CCDC)*, Yichang, China, **2023**, pp. 4198-4202, DOI: 10.1109/CCDC58219.2023.10326549
10. 46.Musau, V. M.; Gaetan, C.; Girardi, P. Clustering of Bivariate Satellite Time Series: A Quantile Approach. *Environmetrics*, **2022**, 33, 7, e2755, DOI: 10.1002/env.2755
11. Oyewole, G. J.; Thopil, G. A. Data Clustering: Application and Trends. *Artificial intelligence review*, **2023**, 56, 7, 6439-6475, DOI: 10.1007/s10462-022-10325-y.
12. Belay, M.A.; Blakseth, S.S.; Rasheed, A.; Salvo Rossi, P. Unsupervised Anomaly Detection for IoT-Based Multivariate Time Series: Existing Solutions, Performance Analysis and Future Directions. *Sensors* **2023**, 23, 2844. DOI: 10.3390/s23052844.

13. Alqahtani, A.; Ali, M.; Xie, X.; Jones, M.W. Deep Time-Series Clustering: A Review. *Electronics* **2021**, *10*, 3001. DOI: 10.3390/electronics10233001
14. Wani, A.A. Comprehensive Analysis of Clustering Algorithms: Exploring Limitations and Innovative Solutions. *PeerJ Computer Science* **10**, **2024**, e2286, DOI: 10.7717/peerj-cs.2286.
15. Gatea, A.N.; Hashim, H.S.; AL-Asadi H.A.A.; Türeli, D.K.; Abduljabbar, Z.A.; Nyangaresi, V.O. Clustering in Data Analysis: Comprehensive Insights into Techniques and Challenges. In: Silhavy, R., Silhavy, P. (eds) *Software Engineering: Emerging Trends and Practices in System Development*. CSOC **2025**. Lecture Notes in Networks and Systems, vol 1562. Springer, Cham. DOI: 10.1007/978-3-032-04725-0_7
16. Liao, T.W. Clustering of Time Series Data—a Survey, *Pattern Recognition*, **2005**, *38*, 11, 1857-1874, DOI: 10.1016/j.patcog.2005.01.025.
17. Lafabregue, B.; Weber, J.; Gançarski, P. et al. End-to-end Deep Representation Learning for Time Series Clustering: a *Comparative Study*. *Data Min Knowl Disc* **36**, **2022**, 29–81, DOI: 10.1007/s10618-021-00796-y.
18. Hou, L.; Kwok, J.; & Zurada, J. Efficient Learning of Timeseries Shapelets. *Proceedings of the AAAI Conference on Artificial Intelligence*, **2016**, *30*, 1. DOI: 10.1609/aaai.v30i1.10178.
19. Blakely, L.; Reno M. J.; Feng M., Spectral Clustering for Customer Phase Identification Using AMI Voltage Timeseries, *IEEE Power and Energy Conference at Illinois (PECI)*, **2019** Champaign, IL, USA, 1-7, DOI: 10.1109/PECI.2019.8698780.
20. Wu J.; Zhang Z.; Tong R.; Zhou Y.; Hu Z.; Liu K., Imaging Feature-Based Clustering of Financial Time Series, *PLoS ONE*, **2023**, *18*, 7, e0288836. DOI: 10.1371/journal.pone.0288836.
21. Guerard, G.; Djebali, S., Mixed Data Clustering Survey and Challenges. *SN COMPUT. SCI.*, **2025**, *6*, 939. DOI: 10.1007/s42979-025-04439-7.
22. Jenkins, L.J.; Haigh, I.D.; Camus, P. et al. The Temporal Clustering of Storm Surge, Wave Height, and High Sea Level Exceedances Around the UK Coastline. *Nat Hazard*, **2023**, *115*, 1761–1797. DOI: 10.1007/s11069-022-05617-z.
23. Sukhorukova, N.; Willard-Turton, J.; Garwoli, G.; Morgan, C.; Rokey, A., Spectral Clustering and Long Timeseries Classification. *The ANZIAM Journal*. **2024**, *66*, 2, 121-131. DOI: 10.1017/S1446181124000105
24. Mariam, A.; Javidi, H.; Zabor, E.C.; Zhao, R.; Radivoyevitch, T.; Rotroff D.M., Unsupervised Clustering of Longitudinal Clinical Measurements in Electronic Health Records. *PLOS Digit Health.*, **2025**, *3*, 10, e0000628. DOI: 10.1371/journal.pdig.0000628.
25. Sun, Q.; Wan, J.; Liu, S. Feature-Based Clustering of Global Sea Level Anomaly Time Series. *Sci Rep* **15**, **2025** 35483. DOI: 10.1038/s41598-025-19269-z.
26. Faruque, O.; Nji, F.N.; Cham, M.; Salvi, R.M.; Zheng, X.; Wang, J. Deep Spatiotemporal Clustering: A Temporal Clustering Approach for Multi-dimensional Climate Data. **2025**. Available online: <https://arxiv.org/abs/2304.14541>, (22.01.2026)
27. Papagiannopoulos, O.D.; Pezoulas, V.C.; Papaloukas, C.; Fotiadis, D.I. 3D clustering of gene expression data from systemic autoinflammatory diseases using self-organizing maps (Clust3D). *Computational and Structural Biotechnology Journal*, **23**, **2024**, 2152-2162, 2001-0370, DOI: 10.1016/j.csbj.2024.05.003.
28. Su, Y.; Liang, K.Y.; Song S. In-Database Time Series Clustering. *Proc. ACM Manag.* **2025**, *3*, 1, 46. DOI: 10.1145/3709696.
29. Yang, H.; Ran, M.; Feng, H.; Hou, D. K-PCD: A new clustering algorithm for building energy consumption time series analysis and predicting model accuracy improvement. *Applied Energy*, **2025**, *377*, DOI: 10.1016/j.apenergy.2024.124584.
30. Soubeiga, A.; Antoine, V.; Corteval, A.; Kerckhove, N.; Moreno, S.; Falih, I.; Phalip, J. Clustering and Interpretation of time-series trajectories of chronic pain using evidential c-means. *Expert Systems with Applications*, **2025**, *260*, DOI: 10.1016/j.eswa.2024.125369.
31. Liu, Y.; Zheng, R.; Liu, M.; Zhu, J.; Zhao, X.; Zhang, M. Short-Term Load Forecasting Model Based on Time Series Clustering and Transformer in Smart Grid. *Electronics* **2025**, *14*, 230. DOI: 10.3390/electronics14020230.
32. Ge, Y.; Zhou, Q.; Ren, X.; Chen, J.; Mei, D.. A Combined Model for Evaluating Strata Deformation Characteristics of Intertidal Zone Utilizing Improved Time Series Clustering Method. *IEEE Transactions on Instrumentation and Measurement*, **2025**, *74*, 1-15, 2543215, DOI: 10.1109/TIM.2025.3597675.

33. Qiu, X.; Wu, X.; Lin, Y.; Guo, C.; Hu, J.; Yang, B. DUET: Dual Clustering Enhanced Multivariate Time Series Forecasting. Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1 (KDD '25). Association for Computing Machinery, 2025, New York, NY, USA, 1185–1196. DOI: 10.1145/3690624.3709325
34. Corradin, R.; Danese, L.; KhudaBukhsh, W. R.; Ongaro, A. Model-based clustering of time-dependent observations with common structural changes. *Statistics and Computing*, 2026, 36(1), 1-15, DOI: 10.1007/s11222-025-10756-x.
35. Han, L.; Cheng, S.; Yuan, W.; Zhang, X.; Wang, J. A novel nonparametric Bayesian model for time series clustering: Application to electricity load profile characterization. *Future Generation Computer Systems*, 2026, 174, 107929, DOI: 10.1016/j.future.2025.107929.
36. Chen, H.; Gao, X.; Wu, Q. An Enhanced Fuzzy Time Series Forecasting Model Integrating Fuzzy C-Means Clustering, the Principle of Justifiable Granularity, and Particle Swarm Optimization. *Symmetry* 2025, 17, 753. DOI: 10.3390/sym17050753.
37. Köhne, J.; Henning, L.; & Gühmann, C. Autoencoder-based iterative modeling and multivariate time-series subsequence clustering algorithm. *IEEE Access* 2023, 11, 18868-18886, DOI: 10.1109/ACCESS.2023.3247564.
38. Luan, Q.; Hamp, J. Automated regime classification in multidimensional time series data using sliced Wasserstein k-means clustering. *Data Science in Finance and Economics* 2025, 5, 3, 387-418, DOI: 10.3934/DSFE.2025016.
39. Ramakrishna, J. S.; Ramasangu, H. Classification of cognitive states using clustering-split time series framework. *Computer Assisted Methods in Engineering and Science* 2024, 3, 2, 241-260, DOI: 10.24423/comes.2024.448.
40. Koyuncu, F. S.; İnkaya, T. An Active Learning Approach Using Clustering-Based Initialization for Time Series Classification. *International Symposium on Intelligent Manufacturing and Service Systems* 2023, 224-235, DOI: 10.1007/978-981-99-6062-0_21.
41. Najafgholizadeh, A.; Nasirkhani, A.; Mazandarani, H. R.; Soltanalizadeh, H. R.; Sabokrou, M. Imaging time series for deep embedded clustering: a cryptocurrency regime detection use case. *27th International Computer Conference, Computer Society of Iran (CSICC)* 2022, 1-6. IEEE, DOI: 10.1109/csicc5295.2022.9780526.
42. Martins, A.A.; Vaz, D.C.; Silva, T.A.N.; Cardoso, M.; Carvalho, A. Clustering of Wind Speed Time Series as a Tool for Wind Farm Diagnosis. *Math. Comput. Appl.* 2024, 29, 35. DOI: 10.3390/mca29030035
43. Kim, H.; Park, S.; Kim, S. Time-series Clustering and Forecasting Household Electricity Demand Using Smart Meter Data. *Energy Reports*, 2023, 9, 4111-4121, DOI: 10.1016/j.egyr.2023.03.042
44. Cebeauer, M.; Jenelius, E.; Gundlegård, D.; Burghout, W. Revealing Representative Day-types in Transport Networks Using Traffic Data Clustering. *Journal of Intelligent Transportation Systems*, 2024, 28, 5, 695-718, DOI: 10.1080/15472450.2023.2205020
45. Monteiro, L. D.; Llavori, R. B.; & Canut, C. G. Time-Series Clustering for Public Transport Mobility Analysis. *University Master's Degree Thesis, Universitat Jaume I*, 2024, Available online: <http://hdl.handle.net/10234/698367> (02.02.2026).
46. Quarta, M. G.; Sgura, I.; Frittelli, M.; Barreira, R.; Bozzini, B. Shape Classification of Battery Cycling Profiles via K-Means Clustering based on a Sobolev distance. *Journal of Computational and Applied Mathematics*, 2026, 117365, DOI: 10.1016/j.cam.2026.117365.
47. Guan, Y.; Zheng, C.; Shi, Y.; Wang, G.; Wu, L.; Chen, Z.; Li, H. MDU-Net: Multi-resolution learning and differential clustering fusion for multivariate electricity time series forecasting. *Information Systems*, 2026, 102693, DOI: 10.1016/j.is.2026.102693.
48. Zhang, X.; Chen, H.; Chen, J. et al. A hybrid machine learning-enhanced MCDM model for transport safety engineering. *Sci Rep* 15, 2025, 36467, DOI: 10.1038/s41598-025-21297-8.
49. Zhou, F.; Yao, J.; Yin, H. Identifying the Passenger Transport Corridors in an Urban Rail Transit Network Based on OD Clustering. *Sustainability* 2025, 17, 9127. DOI: 10.3390/su17209127.
50. Castro, A.; Ludke, G.; Dias, V.; Longras, S.; Sule, E.; Vilhena, E.; Rocha, A. Modal Distribution Diversification and Intermodal Transport Analysis in Europe: A Comprehensive Investigation of Freight Transport Patterns. *Logistics* 2025, 9, 162. DOI: 10.3390/logistics9040162.

51. Matseliukh, Y.; Lytvyn, V.; Bublyk, M. K-means Clustering Method in Organizing Passenger Transportation in a Smart City. *International Conference on Computational Linguistics and Intelligent Systems* **2025**, 15-16, Available online: <https://ceur-ws.org/Vol-3983/paper17.pdf> (05.02.2026)
52. Babapourdijojin, M.; Corazza, M.V.; Gentile, G. Systematic Analysis of Commuting Behavior in Italy Using K-Means Clustering and Spatial Analysis: Towards Inclusive and Sustainable Urban Transport Solutions. *Future Transp.* **2024**, *4*, 1430-1456. DOI: 10.3390/futuretransp4040069.
53. Cebecauer, M.; Jenelius, E.; Gundlegård, D.; Burghout, W. Revealing Representative Day-types in Transport Networks Using Traffic Data Clustering. *Journal of Intelligent Transportation Systems* **2024**, *28*(5), 695–718. DOI: 10.1080/15472450.2023.2205020
54. Cheng, Q.; Hong, K.; Huang, K.; Liu, Z. Evaluating Effectiveness and Identifying Appropriate Methods for Anomaly Detection in Intelligent Transportation Systems. *IEEE Transactions on Intelligent Transportation Systems*. **2025**, DOI: 10.1109/TITS.2025.3580960.
55. Cheng, C.-H.; Tsai, M.-C.; Cheng, Y.-C. An Intelligent Time-Series Model for Forecasting Bus Passengers Based on Smartcard Data. *Appl. Sci.* **2022**, *12*, 4763. DOI: 10.3390/app12094763
56. Chen, M. Y.; Chiang, H. S.; & Yang, K. J. Constructing cooperative intelligent transport systems for travel time prediction with deep learning approaches. *IEEE Transactions on Intelligent Transportation Systems*, **2022**, *23*(9), 16590-16599, DOI: 10.1109/TITS.2022.3148269.
57. Gal-Tzur, A. Transport-Related Synthetic Time Series: Developing and Applying a Quality Assessment Framework. *Sustainability* **2025**, *17*, 1212. DOI: 10.3390/su17031212.
58. Yu, Y.; Zeng, X.; Xue, X.; Ma, J. LSTM-Based Intrusion Detection System for VANETs: A Time Series Classification Approach to False Message Detection. *IEEE Transactions on Intelligent Transportation Systems*, **2022**, *23*(12), 23906-23918, 10.1109/TITS.2022.3190432.
59. Chowdhury, M.; Dey, K.; Apon, A. (Eds.). *Data Analytics for Intelligent Transportation Systems. Elsevier, 2nd Edition*, 2024, ISBN: 978-0-443-13878-2, DOI: 10.1016/C2022-0-02815-4.
60. Wang, Y.; Du, X.; Lu, Z.; Duan, Q.; Wu, J. Improved LSTM-based time-series anomaly detection in rail transit operation environments. *IEEE Transactions on Industrial Informatics*, **2022**, *18*(12), 9027-9036., DOI: 10.1109/TII.2022.3164087
61. Blom Västberg, Oskar & Karlström, Anders & Jonsson, R. Daniel & Sundberg, Marcus. (2019). A Dynamic Discrete Choice Activity-Based Travel Demand Model. *Transportation Science*. *54*. DOI: 10.1287/trsc.2019.0898.
62. Blom, A. Day-type clustering of travel speed data for intelligent transportation systems. **2024**, Available online: <https://www.diva-portal.org/smash/get/diva2:1909560/FULLTEXT02.pdf> (02.02.2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.