Journal of Sustainable Development of Energy, Water and Environment Systems

http://www.sdewes.org/jsdewes

Year 2027, Volume 15, Issue 1, 1140766

Original Research Article

Characterization of latent micro-patterns in multivariate electrical consumption time series using a dark value score and unsupervised learning techniques

Kpatchaa Tombana Baba[*], *Eyouléki T. G. Palanga, and Kodjo Apeke*
Polytechnic School of Lome, University of Lome, Lome-TOGO
e-mail: vpalanga@univ-lome.tg

Cite as: Baba, K. T., Palanga, E. T. G., Apeke, K., Characterization of latent micro-patterns in multivariate electrical consumption time series using a dark value score and unsupervised learning techniques, J.sustain. dev. energy water environ. syst., 15(1), 1140766, 2027, DOI: https://doi.org/10.13044/j.sdewes.d14.0766

ABSTRACT

The increasing availability of high-frequency electricity consumption data has created new opportunities for understanding complex energy behaviors beyond dominant global patterns. However, conventional statistical and deep learning models primarily focus on average trends and often overlook weakly expressed local structures embedded within time series. This study proposes an exploratory framework for identifying latent micro-patterns in multivariate electrical consumption data using a combination of entropy-based feature scoring, temporal residual analysis, anomaly detection, and density-based clustering. The methodology is applied to the individual household electric power consumption dataset from the University of California, Irvine repository. A dark value score combining entropy and rarity is introduced to identify variables potentially containing hidden informational structures. Residuals extracted from Long Short-Term Memory predictions are analyzed using Isolation Forest and density-based clustering techniques to reveal statistically coherent micro-signals. Results show that these residual micro-structures are not random noise but exhibit identifiable temporal and statistical organization. Although clusters remain weakly separated, their distributions differ significantly from normal observations, supporting the existence of localized operational regimes invisible to dominant predictive models. The proposed framework provides an interpretable complementary perspective to traditional forecasting approaches and offers potential applications in anomaly detection, behavioral analysis, and intelligent energy monitoring.

KEYWORDS

Electrical consumption, Latent micro-pattern, Dark value score, Unsupervised learning, Residual analysis, Temporal anomaly analysis, Density-based clustering.

INTRODUCTION

The increasing availability of data from sensors [1], embedded systems, and digital platforms has profoundly transformed analytical and modelling approaches across many fields, particularly in energy [2] . In the context of electrical networks and household consumption, such data are now used for load forecasting, anomaly detection [3], energy optimisation, and understanding user behaviour in [2] and in [4]. Nevertheless, despite the apparent abundance of data [5] and the significant advances in machine learning methods [6], a substantial portion of the information contained in time series remains difficult to interpret and insufficiently exploited as in [7] and in [8].

Classical statistical models, such as linear regression methods, primarily rely on the assumption of global, linear, or weakly non-linear relationships between variables as state in

[*] Corresponding author

[9] and in [10]. While these methods offer appreciable interpretability, they quickly reach their limits in [11] when confronted with the intrinsic complexity of real systems, particularly when observed phenomena result from the superposition of heterogeneous, intermittent, or context-dependent behaviours. In contrast, deep learning models, notably recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and gated recurrent units (GRUs), are capable of capturing complex temporal dependencies and high-level non-linear relationships [12]. However, these models are often considered "black boxes," in [13] providing accurate predictions but offering limited insight into the underlying mechanisms [14].

This tension between predictive performance and interpretability has become a central challenge in data science research [15] applied to energy systems. Recent studies have attempted to combine hybrid approaches, integrating physical constraints, to improve model understanding as in[16]. Nevertheless, these approaches remain predominantly focused on optimising overall predictive accuracy rather than explicitly identifying local, rare, or weakly correlated informational structures that may nonetheless play a critical role in system dynamics [17].

This study is based on the observation that energy data contain micro-signals, that is, low-amplitude patterns localised in time or specific to certain subsets of observations, which are neither statistically dominant nor directly captured by classical methods [18]. Such micro-signals may correspond, for instance, to changes in user behaviour, non-trivial interactions between electrical appliances, or particular operational regimes that only appear under specific conditions. Although their contribution to overall variance is limited, their informational value can be high, particularly for fine-grained system understanding, early anomaly detection, or behavioural segmentation.

A methodological approach is introduced to highlight latent structures, drawing an analogy with physics, where dark matter represents an invisible component crucial for explaining observed universal dynamics. In this context, these latent structures are weakly visible in global statistics but become apparent through local inconsistencies, persistent residuals, or atypical behaviors within the data [19].

Unlike purely predictive approaches, the objective is not solely to improve model accuracy but to provide an exploratory and interpretable analytical framework that allows: (i) the isolation of informational subspaces overlooked by dominant models, (ii) the analysis of their distribution across variables and data clusters, and (iii) the formulation of plausible explanatory hypotheses without resorting to unverifiable assumptions. Emphasis is placed on methodological transparency and reproducibility of results.

The proposed methodology relies on a structured combination of preprocessing techniques, dimensionality reduction, unsupervised clustering, and residual analysis from temporal models. The originality of this approach lies in the systematic analysis of discrepancies between dominant representations (derived from global models such as LSTMs) and local structures revealed by complementary methods. These discrepancies are interpreted not as noise but as potentially informative signals, warranting further analysis.

To validate this approach, the methodology is applied to a widely used real-world dataset: the University of California Irvine (UCI) Machine Learning repository's "Individual Household Electric Power Consumption" dataset. This dataset offers several advantages: a long observation period, fine temporal granularity, and a variety of variables measuring different aspects of household electrical consumption. It thus provides a relevant testbed for studying the emergence of latent structures and heterogeneous behaviours [20].

An explicit comparison is conducted between the approach and a standard LSTM model used as a reference. This comparison extends beyond predictive performance to include residual analysis, the ability to discriminate distinct operational regimes, and the interpretability of results for a human analyst. Each identified cluster is analysed individually, and explanatory hypotheses are formulated based on observed statistical characteristics without unfounded extrapolation.

UNCORRECTED PROOF

Finally, beyond the presented case study, this work is situated within a broader perspective of operational applicability. The methodology is designed to be applicable to any type of tabular or temporal data, regardless of the domain, with minimal a priori assumptions. This generality opens the way for transforming the approach into an analytical support tool, suitable for both researchers and practitioners, capable of automatically generating interpretable reports to facilitate decision-making.

Thus, this study aims to contribute both methodologically, by proposing an original framework for exploring latent structures, and practically, by demonstrating the tangible benefits of analysing dark data in a real energy context.

MATERIALS AND METHODS

The objective of this study is not to maximize short-term predictive accuracy but to uncover latent, weakly dominant structures embedded in high-frequency electricity consumption data. These structures are assumed to carry localized and low-variance information that remains largely invisible to global models optimized for average performance. The proposed methodology combines global structure modeling, residual analysis, entropy-based feature screening, and unsupervised learning to identify and interpret such latent phenomena.

Methodological Framework and Research Hypotheses

Let $X \in R^{T \times p}$ denote a multivariate electricity consumption time series, where T is the number of time steps and p the number of observed variables. The proposed framework follows four successive stages:

1. Modeling the dominant global structure of the data using interpretable dimensionality reduction and temporal models.
2. Extracting residual signals not explained by the dominant structure.
3. Identifying localized and statistically rare temporal patterns within these residuals.
4. Grouping recurrent patterns using density-based clustering and interpreting their operational meaning.

The central assumption underlying this approach is that electricity consumption data contain latent micro-structures, defined here as localized temporal patterns [21] characterized by:

- low contribution to global variance,
- statistical rarity,
- internal temporal or statistical coherence.

Three research hypotheses are formulated:

<u>H1 – Existence of latent local structures</u>

Electricity consumption data contain local informational structures with low global variance, which are not fully captured by dominant global models.

<u>H2 – Non-random nature of micro-signals</u>

The detected micro-signals correspond to structured phenomena and cannot be considered as random noise.

<u>H3 – Internal organisation of micro-signals</u>

Micro-signals exhibit an internal organisation that can be analysed, even if this organisation is weak and does not lead to clearly separated groups.

UNCORRECTED PROOF

Data Preprocessing

The experiments are conducted on the UCI Individual Household Electric Power Consumption dataset, which provides one-minute resolution measurements of active power, reactive power, voltage, and sub-metered consumption .

Temporal Aggregation

To reduce high-frequency noise while preserving medium-term dynamics, the data are aggregated to an hourly resolution.

Missing Value Handling

Short missing intervals are interpolated using local temporal continuity. For longer gaps, column-wise mean imputation is applied [22]:

$$X_{i,j} = \begin{cases} x_{i,j}\ , & x_{i,j}\ available \\ \bar{x}_j\ , & x_{i,j}\ missing \end{cases} \quad (1)$$

Where:

- $X_{i,j}$: the value used for row i and column j after handling missing data.
- $x_{i,j}$: the original observed value in the dataset for row i, column j
- $\bar{x}_j$: the mean of column j, used to replace missing values.

Normalization

All variables are standardized to ensure comparability across features [23]:

$$X'_{i,j} = \frac{X_{i,j} - \mu_j}{\sigma_j} \quad (2)$$

Ensuring comparability across features for dimensionality reduction and clustering.
Where:

- $X'_{i,j}$: normalized value of row i, column j.
- $X_{i,j}$: value after filling missing data.
- μ_j : mean of column j.
- σ_j : standard deviation of column j.

Identification of Hidden Features: Entropy and Rarity

To prioritize variables potentially containing micro-signals, a dark value score combining entropy and rarity was computed [24]:

1. Entropy measures unpredictability [25]:

$$H(X_j) = -\sum_{k=1}^{n} p_k log(p_k) \quad (3)$$

UNCORRECTED PROOF

Where p_k is propability of values falling into histogram bin k. Higher entropy indicates greater diversity in the column suggesting latent micro-patterns.

2. Rarity emphasizes infrequent observations:

$$R(X_j) = \frac{1}{N}\sum_{i=1}^{N}(1 - f(x_{i,j})) \tag{4}$$

Where $f(x_{i,j})$ is the relative frequency of value $x_{i,j}$. Rare events receive higher weights.

3. Dark value for column j:

$$D(X_j) = H(X_j).R(X_j) \tag{5}$$

Columns with the highest dark values were selected for micro-signal analysis, as they are most likely to contain hidden, informative events.

Modeling Dominant Global Structure

To represent the dominant global structure of the data, both statistical and temporal modeling methods were applied to the normalized observations.

Principal Component Analysis (PCA)

PCA was applied to normalized data to identify directions of maximal variance [26]:

$$X' = XW \tag{6}$$

Where W contains the eigenvectors of the covariance matrix of X . PCA serves as an interpretable reference, quantifying how much of the variane is captured globally and how much remains unexplained [27].

LSTM Reference model

A univariate LSTM was trained separately on the columns with the highest dark value to capture temporal dependencies of each columns [28]:

$$\hat{y}_{t+1} = LSTM(X_t) \tag{7}$$

With prediction error:

$$E_t = y_t - \hat{y}_t \tag{8}$$

Extraction of Micro-Signals

Residuals from the LSTM and columns with high dark value were treated as new time series for micro-signal detection. Two complementary methods were applied:

- Statistical feature extraction: local mean, variance, skewness, and kurtosis per time window.

UNCORRECTED PROOF

- Isolation Forest: is based on the idea that rare and unusual observations are isolated more quickly than normal ones [29]. A low contamination rate (approximately 1%) is used to focus on the most atypical events. [30]

Observations classified as anomalies (y = -1) are interpreted as micro-signals, that is, local events with low global impact but potential informational value.

Unsupervised Clustering of Latent Structures

The detected micro-signals are projected into a normalised feature space and grouped using the DBSCAN algorithm [31]. DBSCAN is chosen because it [32]:

- identifies clusters with varying densities,
- isolates outlier points (cluster −1),
- does not require the number of clusters to be specified in advance.

Cluster quality is evaluated using the silhouette score, which is equal to −0.147 in this study. A negative silhouette score indicates that the clusters are neither compact nor well separated. This result suggests that micro-signals form weak, diffuse, and partially overlapping structures, rather than clearly defined regimes. This observation does not contradict Hypothesis H1. On the contrary, it confirms that the structures of interest are not dominant in a geometric sense, which is consistent with their low contribution to global variance.

The objective of the clustering stage is not to identify strongly separated macroscopic classes, but rather to explore weakly expressed local organizations embedded within residual space. Consequently, moderate or negative silhouette values are expected, since micro-signals naturally exhibit overlap, sparsity, and transitional behavior.

To test whether micro-signals differ statistically from normal observations, a Kolmogorov–Smirnov test [33] is applied to compare their distributions.

The test yields a p-value close to 0.0000, leading to a clear rejection of the null hypothesis of identical distributions. This result shows that micro-signals follow a distribution that is statistically different from that of normal observations. Therefore, they cannot be considered random noise, which supports Hypothesis H2. As illustrated in the Figure 1 Autocorrelation comparison (H2),

UNCORRECTED PROOF

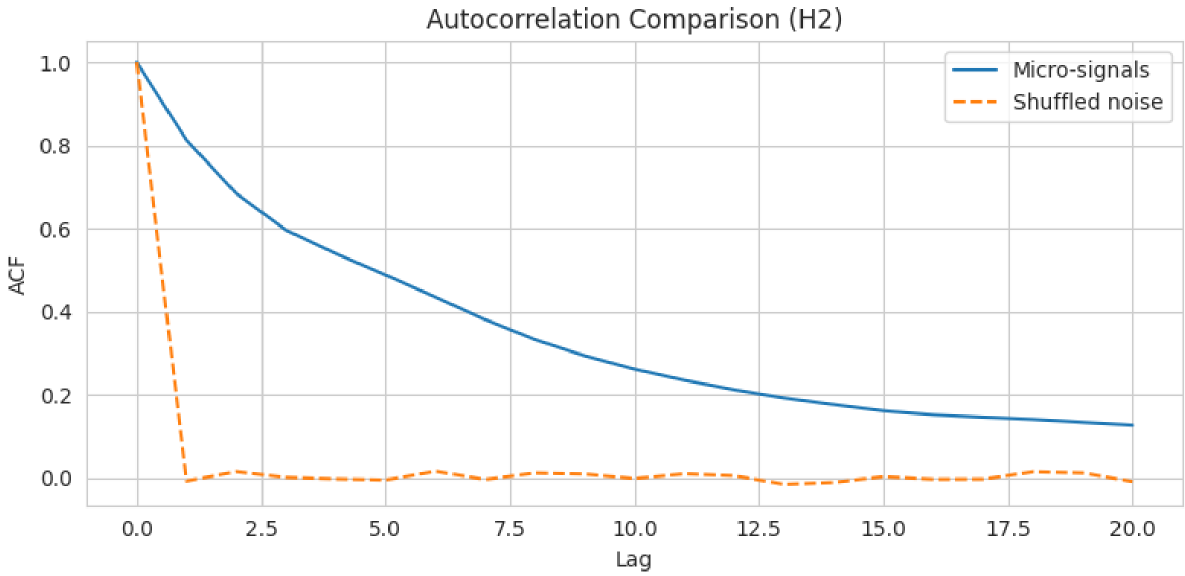

Figure 1 Autocorrelation comparison (H2)

UNCORRECTED PROOF

the autocorrelation profiles of the micro-signals exhibit persistent temporal dependencies, which differ significantly from those of the normal observations. This confirms that the detected structures are not random fluctuations but organized temporal phenomena.

Analysis and Interpretation of Clusters

Each cluster is analyzed in terms of:

- Distribution of original variables
- Contribution of LSTM residuals
- Temporal stability

Isolated and low-density clusters receive particular attention, as they may correspond to sporadic appliance usage, minor anomalies, or regime transitions. This interpretive stage directly addresses hypothesis H3.

RESULTS AND DISCUSSION

This section reports the main results of the proposed framework, focusing on the identification of high dark-value features, the detection of micro-signals, and the analysis of their latent structural organization.

Detection of High Dark-Value Features

The initial analysis focused on the identification of columns with high dark values, combining entropy and rarity measures to highlight variables likely to contain latent micro-signals. The Figure 2. Top dark value variables

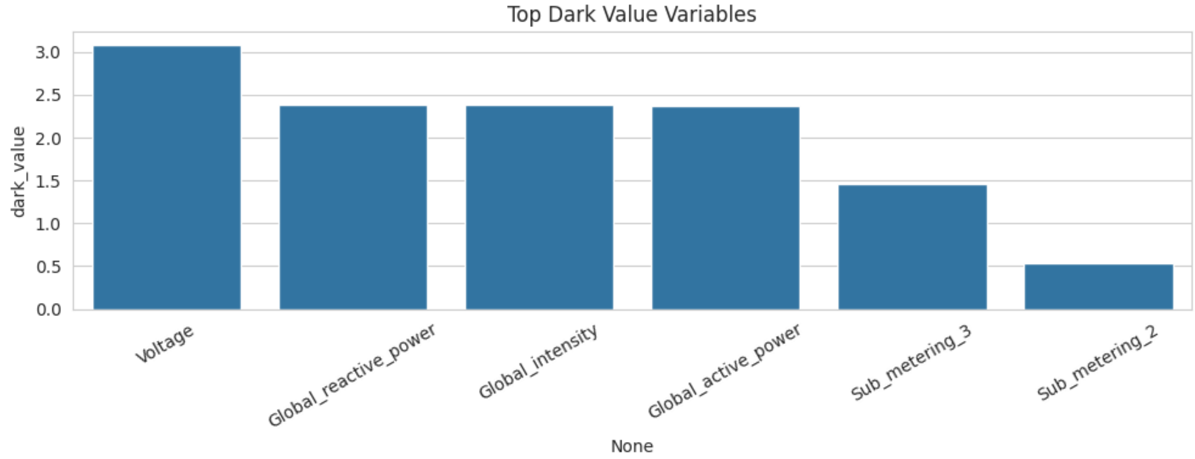

Figure 2. Top dark value variables

presents the top six (6) variables according to their dark value scores.

The analysis shows that the voltage exhibits the highest dark value, followed by the Global_reactive_power, the Global_intensity , the Global_active_power and the sub_metering variables measurements. High dark value indicates both diverse values (high entropy) and rare occurrences (high rarity), suggesting the presence of atypical events or micro-patterns within these variables. Variables with elevated dark values are prioritized for subsequent micro-signal detection. This aligns with Hypothesis H1, indicating that certain columns may contain structured low-frequency information not captured by global models.

Micro-Signal Detection Using Isolation Forest

To detect rare events or micro-signals, Isolation Forest was applied to the normalized numerical dataset. Outliers identified by the model represent observations that are isolated in feature space, suggesting anomalous but potentially informative behavior [34].

A total of 4703 micro-signals were detected.Figure 3. Micro-signals detected in voltage over time, Figure 4. Micro-signals detected in global_reactive_power over time and Figure 5. Micro-signals detected in global_active_power over time

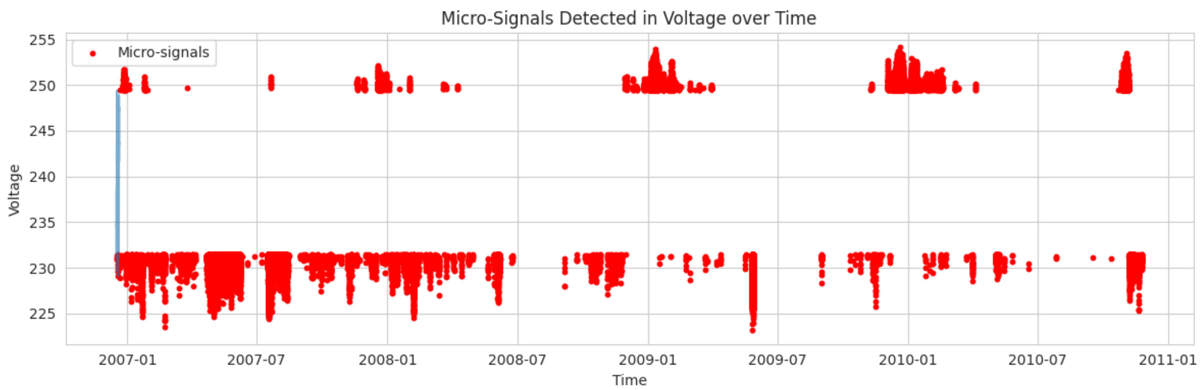

Figure 3. Micro-signals detected in voltage over time

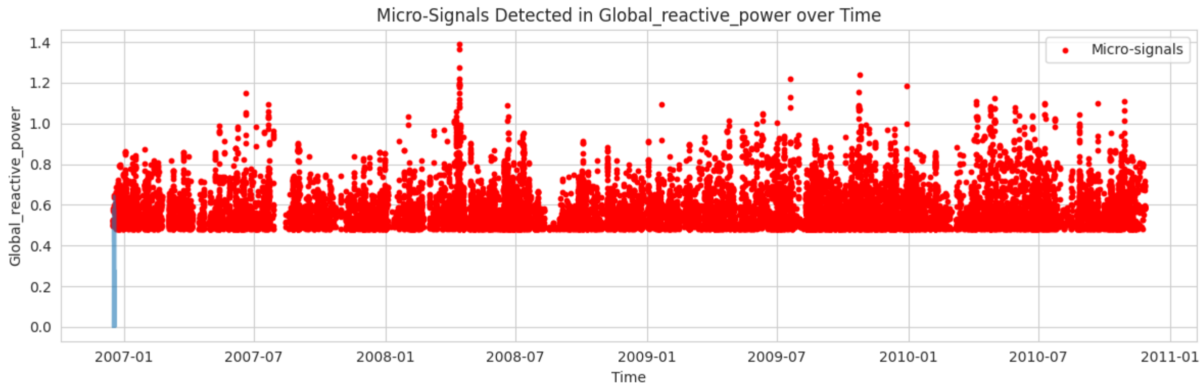

Figure 4. Micro-signals detected in global_reactive_power over time

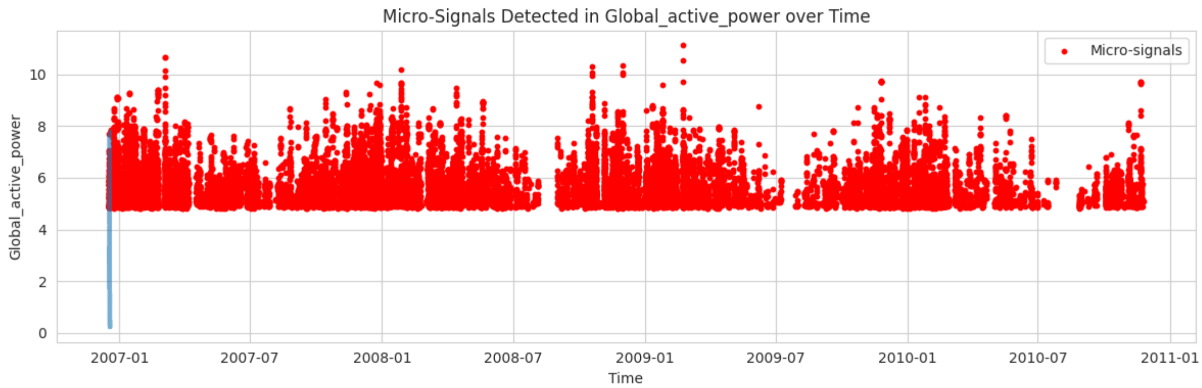

Figure 5. Micro-signals detected in global_active_power over time

illustrates the temporal occurrence of micro-signals for respectively Voltage, Global_reactive_power, and Global_active_power variables, with red dots indicating detected anomalies.

The identified micro-signals likely correspond to unusual consumption patterns, which may result from:

- Short-duration spikes in appliance usage,
- Errors or fluctuations in metering,
- External events affecting consumption (e.g., holidays, weather anomalies).

This confirms Hypothesis H2, showing that residual information contains coherent structures not captured by standard predictive models.

Clustering of Micro-Signals

To reveal coherent patterns among micro-signals, DBSCAN clustering was employed on standardized micro-signal data. Clusters were analyzed according to:

- Number of points per cluster,
- Temporal concentration,

- Feature distributions.

Figure 6. Number of micro-signals per cluster and per variable

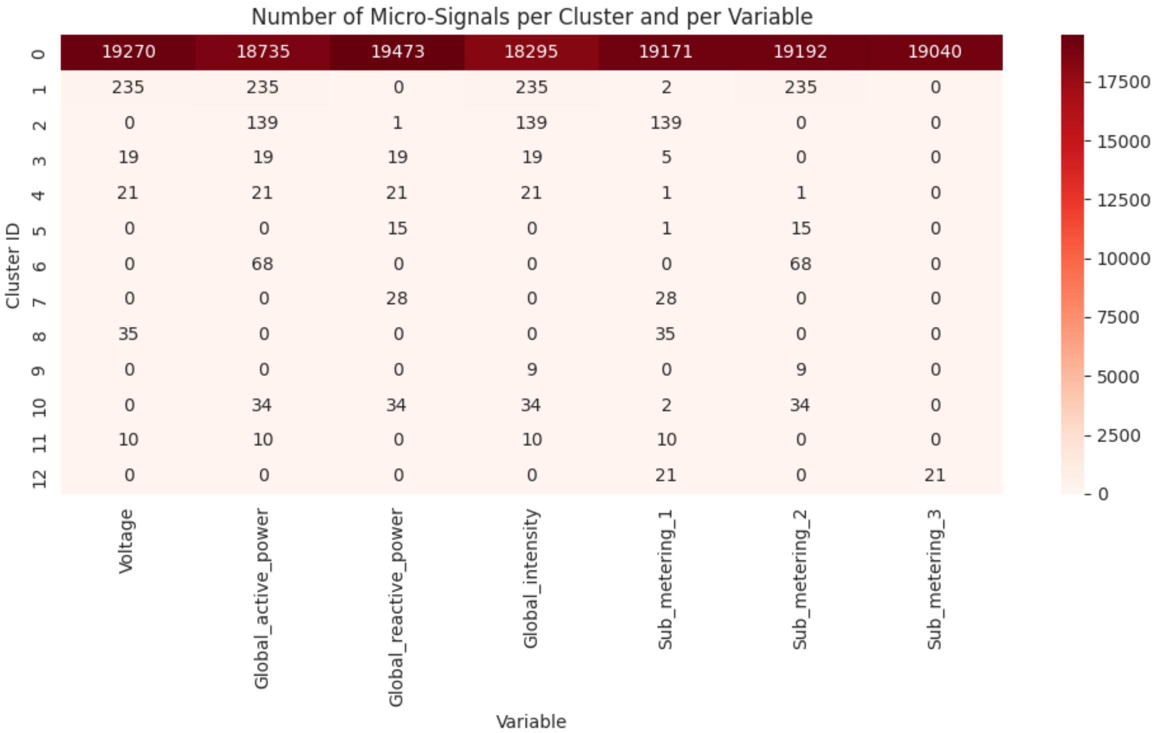

Figure 6. Number of micro-signals per cluster and per variable

presents the number of micro-signals per cluster and per variable, excluding cluster -1. The results highlight a non-uniform distribution of micro-events across both clusters and electrical variables, indicating that latent occurrences are unevenly distributed in the dataset. Clusters labeled **-1** correspond to highly isolated events, which may reflect measurement anomalies or unusual consumption episodes. These cases require further investigation using domain-specific knowledge. In contrast, the remaining clusters reveal recurring micro-patterns that likely represent low-frequency but structured behaviors, such as appliance usage during specific periods of the day or week.

Overall, these observations support Hypothesis H3 by emphasizing the relevance of rare but structurally consistent patterns within the electrical consumption dynamics.

Interpretation of weak cluster separability

The negative silhouette score (-0.147) does not necessarily indicate methodological failure. Instead, it reflects the intrinsic nature of latent micro-signals, which are expected to exhibit partial overlap and weak geometric separability due to their low variance contribution. Unlike conventional clustering tasks aiming to identify dominant regimes, the present study focuses on weakly expressed local structures embedded within global consumption dynamics. Therefore, diffuse cluster boundaries are consistent with the theoretical assumptions of the framework.

UNCORRECTED

Comparison with LSTM Predictions

A univariate LSTM model was trained to predict each variable, serving as a reference global model. Figure 7. LSTM prediction vs true signal for Voltage, Figure 8. LSTM prediction vs true signal for Global_reactive_power and Figure 9. LSTM prediction vs true signal for Global_active_power

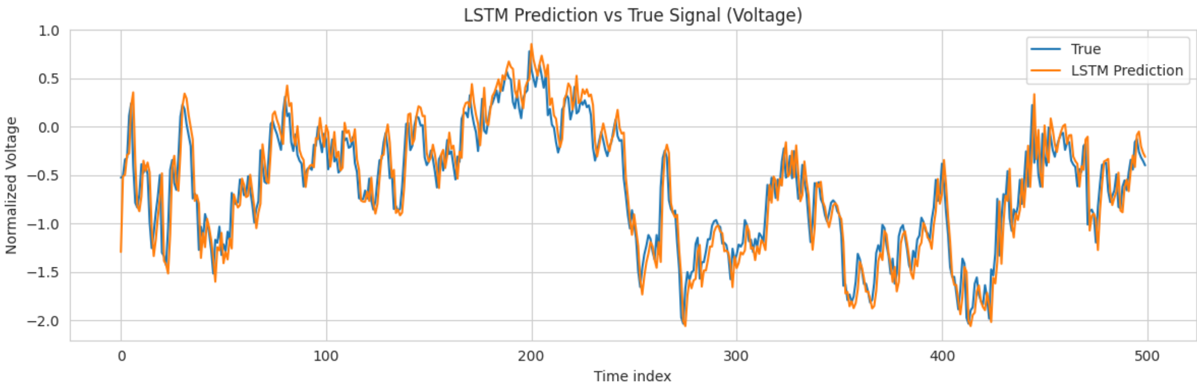

Figure 7. LSTM prediction vs true signal for Voltage

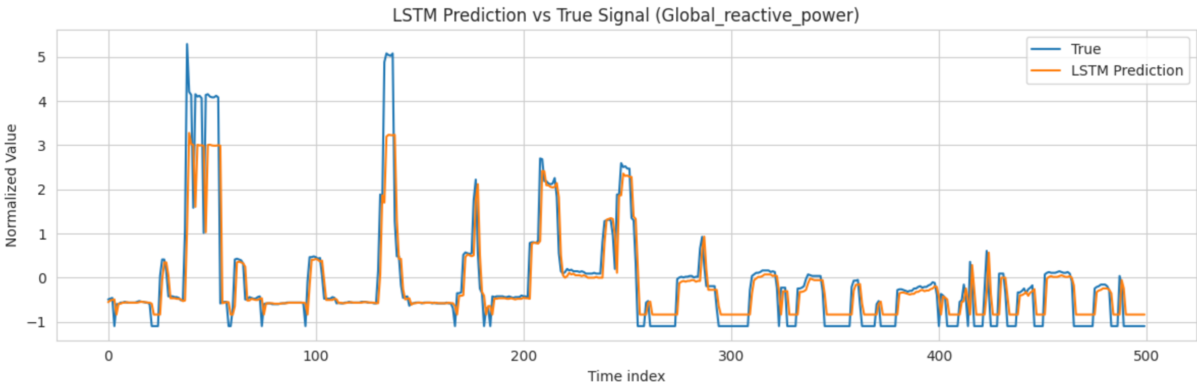

Figure 8. LSTM prediction vs true signal for Global_reactive_power

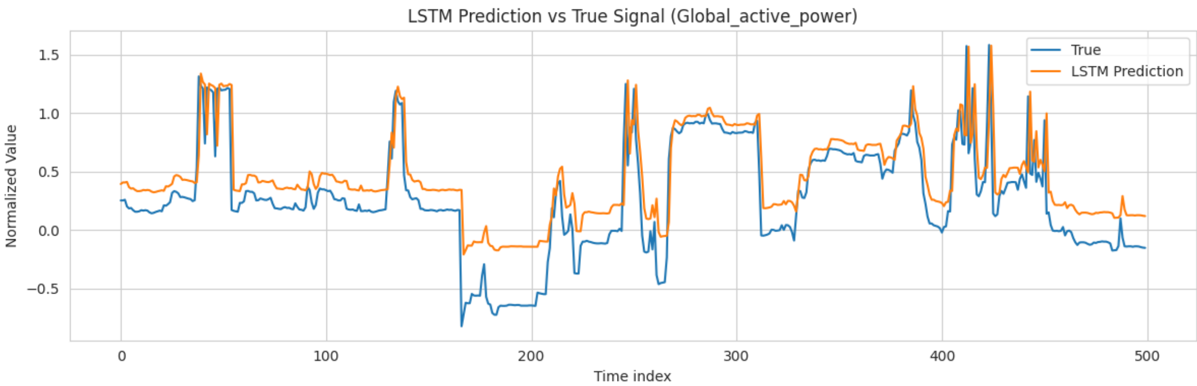

Figure 9. LSTM prediction vs true signal for Global_active_power

presents LSTM predictions for Voltage, Global_reactive_power and Global_active_power variables compared to actual values over a selected temporal window.

- The LSTM captures global trends effectively but fails to reproduce localized micro-spikes.
- Residuals between LSTM predictions and actual measurements in Figure 10. LSTM residual for Voltage, Figure 11. LSTM residual for Global_reactive_power and Figure 12. LSTM residual for Global_active_power,

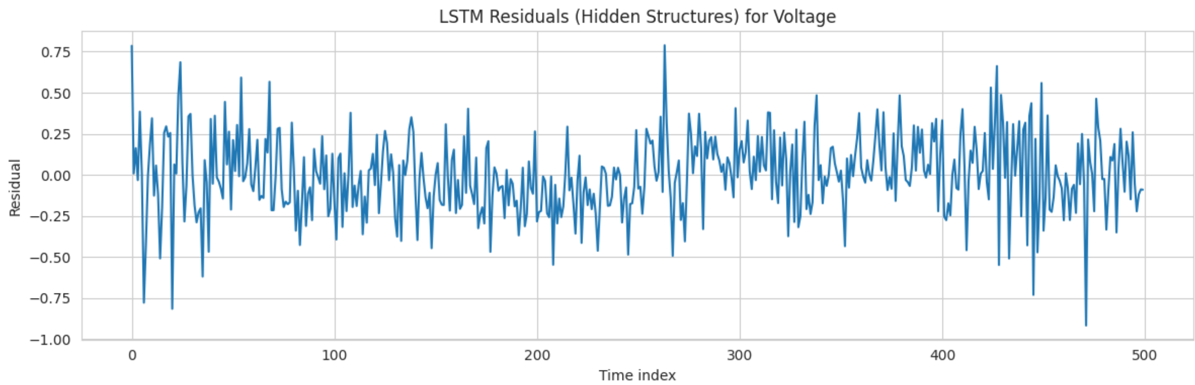

Figure 10. LSTM residual for Voltage

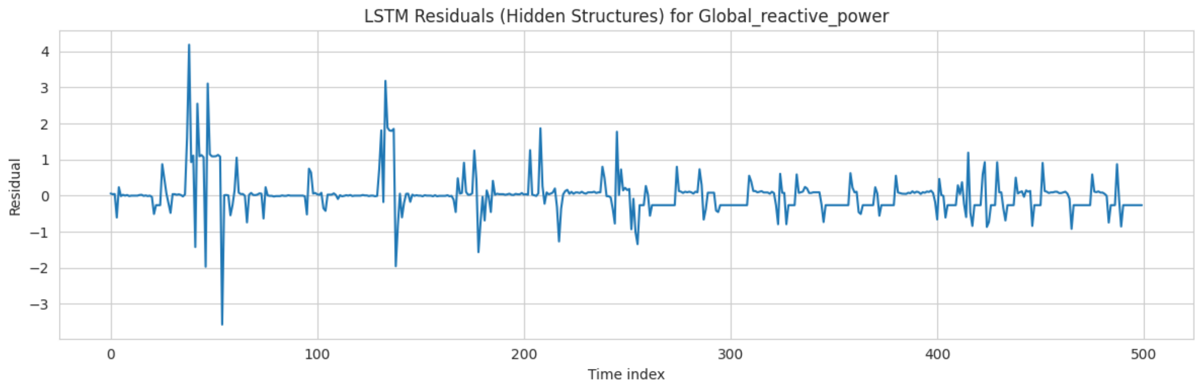

Figure 11. LSTM residual for Global_reactive_power

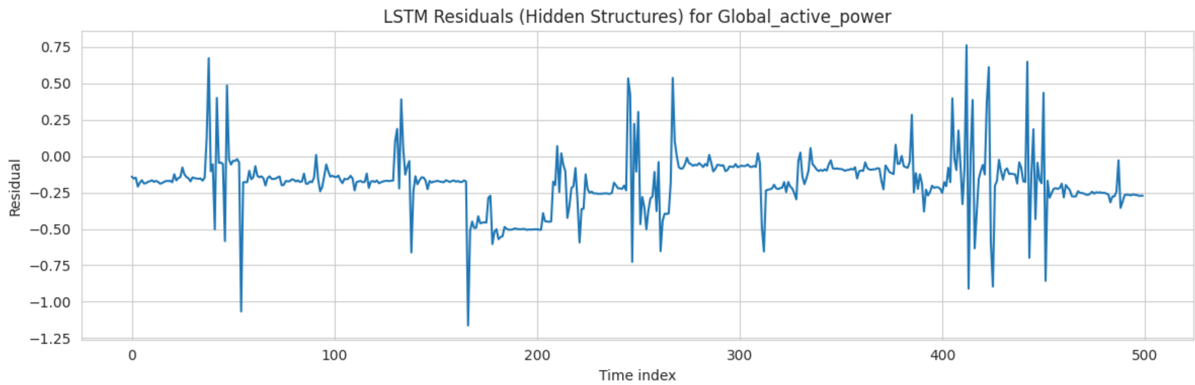

Figure 12. LSTM residual for Global_active_power

correspond closely to micro-signals detected by Isolation Forest.

This complementarity demonstrates that the recurrent temporal prediction model captures the dominant and smooth temporal dynamics of electricity consumption, whereas the isolation-based anomaly detector focuses on irregular residual deviations [35].These residual deviations correspond to localized behaviors that are too weak or too infrequent to significantly influence the global learning process, yet they remain statistically meaningful [36]. In this sense, the anomaly detector reveals hidden local structures overlooked by the global model.

Limitations and Considerations

While the methodology uncovers informative micro-structures:

- Isolated micro-signals (cluster -1) require domain expertise for interpretation. Blind statistical analysis cannot distinguish between rare but meaningful events and sensor errors. [37]
- The choice of DBSCAN parameters influences cluster identification. Sensitivity analysis may be required for optimal detection [38].
- The LSTM model is univariate and relatively simple; more complex models may reduce residual variance but risk masking low-frequency structures [39].
- The framework remains exploratory and should not be interpreted as a causal model of electricity consumption behavior.
- In practical energy systems, such micro-signals may correspond to: transient appliance activation, irregular user behavior, local voltage instability, intermittent reactive power fluctuations, or early signatures of abnormal operational conditions.

CONCLUSION

This study introduced an exploratory and interpretable framework for detecting latent micro-patterns in multivariate household electricity consumption time series. In contrast to conventional forecasting approaches that mainly focus on global predictive accuracy, the objective was to identify weak, low-amplitude, and statistically rare structures embedded within the data. The main contribution of this work is the definition of a dark value score, which combines entropy and rarity to highlight variables that are likely to contain hidden informational content.

The results show that the detected micro-signals cannot be regarded as random noise. Statistical tests indicate that their distributions differ from those of normal observations, and autocorrelation analysis reveals temporal organization. Although cluster separability remains limited, this behavior is consistent with the underlying assumption that such micro-patterns contribute marginally to global variance while retaining internal structure.

Overall, the proposed framework provides an additional analytical layer to traditional predictive models by transforming residual information into interpretable insights. Beyond the case of household electricity consumption, it can be extended to other multivariate time series systems where subtle local dynamics play a significant role.

Future work could integrate exogenous variables, extend the method to other time series, and explore hybrid models combining deep learning with explicit detection of rare structures. Such approaches may improve understanding of complex, weakly observable phenomena in energy systems and beyond.

REFERENCES

[1] S. Syafii, K. Krismadinata, F. Fahmi, F. Azizah, and I. N. Izrillah, ‘Enhancing Electric Vehicle Charging Stations through Internet of Things Technology for Optimizing Photovoltaic and Battery Storage Integration’, *J. Sustain. Dev. Energy Water Environ.*

UNCORRECTED PROOF

Syst., vol. [13], no. [4], p. [1]-[15], Dec. 2025. doi : https://doi.org/10.13044/j.sdewes.d13.0609

[2] 'Smart Factories and Energy Efficiency in Industry 4.0 - Distributed Time-Sensitive Systems - Wiley Online Library'. Accessed: Feb. 08, 2026. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781394197798.ch4

[3] 'Early Electrical Fault Detection in Power Systems Using Data Analytics - ProQuest'. Accessed: Feb. 08, 2026. [Online]. Available: https://www.proquest.com/openview/db4cacbeaa5cb63c2d09a5996b0559f2/1?pq-origsite=gscholar&cbl=18750&diss=y

[4] Y. Bokovi, K. Moyème, S. Kwami Séname, T. Pidéname, and L. Yendoubé, 'Machine Learning Electrical Load Forecasting: an application in microgrid energy consumption with adaboost regressor approach and a comparative study with hybrid method based on LSTM and MLP approaches', *J. Sustain. Dev. Energy Water Environ. Syst.*, vol. 13, no. 4, pp. 1–23, Dec. 2025, doi: 10.13044/j.sdewes.d13.0606.

[5] Z. Garip, E. Ekinci, and A. Alan, 'Day-ahead solar photovoltaic energy forecasting based on weather data using LSTM networks: a comparative study for photovoltaic (PV) panels in Turkey', *Electr. Eng.*, vol. 105, no. 5, pp. 3329–3345, Oct. 2023, doi: 10.1007/s00202-023-01883-7.

[6] 'Unleashing Big Data's Potential: The Mediating Role of Predictive Analytics in Decision-making for Technology and Innovation Sector | Integrating 4IR and 5IR Technologies for Digital Business Innovation | Books Gateway | Emerald Publishing'. Accessed: Feb. 08, 2026. [Online]. Available: https://www.emerald.com/books/edited-volume/20966/chapter-abstract/107781795/Unleashing-Big-Data-s-Potential-The-Mediating-Role?redirectedFrom=fulltext

[7] A. Unni and H. K. Channi, 'Data Analytics for Performance Optimization in Renewable Energy', in *Optimization in Sustainable Energy*, John Wiley & Sons, Ltd, 2026, pp. 291–327. doi: 10.1002/9781394242139.ch9.

[8] Y. Chen, C. Lin, J. Liu, and D. Yu, 'One-hour-ahead solar irradiance forecast based on real-time K-means++ clustering on the input side and CNN-LSTM', *J. Atmospheric Sol.-Terr. Phys.*, vol. 266, p. 106405, Jan. 2025, doi: 10.1016/j.jastp.2024.106405.

[9] M. Goodarzi, N. Haghtalab, A. Shayesteh, and T. Delavar, 'Climatic Determinants of Energy Optimization in LEED-Certified Buildings: A Multi-Method Machine-Learning Approach', Feb. 02, 2026, *Social Science Research Network, Rochester, NY*: 6168263. doi: 10.2139/ssrn.6168263.

[10] O. Timur and H. Y. Üstünel, 'Short-Term Electric Load Forecasting for an Industrial Plant Using Machine Learning-Based Algorithms', *Energies*, vol. 18, no. 5, Art. no. 5, Jan. 2025, doi: 10.3390/en18051144.

[11] D. Maulud and A. M. Abdulazeez, 'A Review on Linear Regression Comprehensive in Machine Learning', *J. Appl. Sci. Technol. Trends*, vol. 1, no. 2, pp. 140–147, Dec. 2020, doi: 10.38094/jastt1457.

[12] S. P. Yadav, S. Zaidi, A. Mishra, and V. Yadav, 'Survey on Machine Learning in Speech Emotion Recognition and Vision Systems Using a Recurrent Neural Network (RNN)', *Arch. Comput. Methods Eng.*, vol. 29, no. 3, pp. 1753–1770, May 2022, doi: 10.1007/s11831-021-09647-x.

[13] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, 'A Survey of Methods for Explaining Black Box Models', *ACM Comput Surv*, vol. 51, no. 5, p. 93:1-93:42, Aug. 2018, doi: 10.1145/3236009.

[14] K. Oublal, Q. Bouniot, Q. Gan, S. Clémençon, and Z. Akata, 'TimeSAE: Sparse Decoding for Faithful Explanations of Black-Box Time Series Models', Jan. 14, 2026, *arXiv*: arXiv:2601.09776. doi: 10.48550/arXiv.2601.09776.

[15] J. Hussain, *Deep Learning Black Box Problem*. 2019. Accessed: Feb. 08, 2026. [Online]. Available: https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-393479

UNCORRECTED PROOF

[16] K. T. Baba and E. Palanga, 'A Multi-Resolution Approach Based on the Integration of a Nonlinear Physical Model and Long Short-Term Memory Network for Photovoltaic Power Modeling and Forecasting', *J. Sustain. Dev. Energy Water Environ. Syst.*, vol. [14], no. [2], p. [1]-[10], Jun. 2026. doi: https://doi.org/10.13044/j.sdewes.d14.0677
[17] R. Aruna, R. Madavan, A. Karthick, and V. Mohanavel, 'Black-box modeling and smart control strategies for proton exchange membrane fuel cell stability', *Electr. Power Syst. Res.*, vol. 254, p. 112574, May 2026, doi: 10.1016/j.epsr.2025.112574.
[18] M. Barate, E. T. G. Palanga, A. S. A. Ajavon, and K. Agbossou, 'Stochastic Characterization of Faults in Electrical Transmission Networks: Case Study of the Electrical Community of Benin', *J. Sustain. Dev. Energy Water Environ. Syst.*, vol. 13, no. 1, pp. 1–17, Mar. 2025, doi: 10.13044/j.sdewes.d12.0531.
[19] M. Belay, S. Blakseth, A. Rasheed, and P. Rossi, 'Unsupervised Anomaly Detection for IoT-Based Multivariate Time Series: Existing Solutions, Performance Analysis and Future Directions', *Sensors*, vol. 23, p. 2844, Mar. 2023, doi: 10.3390/s23052844.
[20] M. Ahmad, S. Ullah Bazai, S. Hussain, A. I. Ashirova, Y. J. Erkaboy Ugli, and U. Aslam Bhatti, 'Predicting Household Electricity Consumption Using Machine Learning and Big Data Analytics', in *2025 IEEE 2nd International Conference on Deep Learning and Computer Vision (DLCV)*, Jun. 2025, pp. 1–7. doi: 10.1109/DLCV65218.2025.11088841.
[21] Z. Hu, Y. Sun, Y. Zhao, W. Wu, L. Kong, and K. Chen, 'MSIF-SSTR: A "Quick Smuggler" Smuggling Speedboat Trajectory Recognition Method Based on Multi-source Information Fusion', *Expert Syst. Appl.*, p. 131525, Feb. 2026, doi: 10.1016/j.eswa.2026.131525.
[22] M. D. A. Rusbandi, S. Fauziati, and G. D. Putra, 'Techniques for Handling Missing Values in Customers Electricity Data: A Systematic Literature Review', in *2024 2nd International Conference on Technology Innovation and Its Applications (ICTIIA)*, Sep. 2024, pp. 1–6. doi: 10.1109/ICTIIA61827.2024.10761291.
[23] L. Ding, D. Tang, W. Wei, F. Li, and W. Cai, 'Data Normalization Analysis for Electirc Energy Meter', in *2020 Chinese Control And Decision Conference (CCDC)*, Aug. 2020, pp. 4258–4263. doi: 10.1109/CCDC49329.2020.9164076.
[24] D. Hock, M. Kappes, and B. Ghita, 'Using multiple data sources to detect manipulated electricity meter by an entropy-inspired metric', *Sustain. Energy Grids Netw.*, vol. 21, p. 100290, Mar. 2020, doi: 10.1016/j.segan.2019.100290.
[25] L. G. Gómez Acosta and M. Chacón Pacheco, 'Entropy and Statistical Complexity in Bioelectrical Signals: A Literature Review', *Signals*, vol. 6, no. 1, p. 7, Mar. 2025, doi: 10.3390/signals6010007.
[26] A. Zseni, A. Horváth, G. Z. Macher, D. Sipos, and J. Pécsinger, 'Using Multivariate Statistical Analysis for Examining the Relationship between Food Waste Generation and Socio-economic Factors', *J. Sustain. Dev. Energy Water Environ. Syst.*, vol. 13, no. 3, pp. 1–16, Sep. 2025, doi: 10.13044/j.sdewes.d13.0579.
[27] M. L. Jallow and H. Jiang, 'Financial Inclusion, Household Consumption and Carbon Emission: Evidence from Sub-Saharan Africa', *J. Sustain. Dev. Energy Water Environ. Syst.*, vol. 13, no. 1, pp. 1–21, Mar. 2025, doi: 10.13044/j.sdewes.d13.0538.
[28] Y. Bokovi, M. Kabe, K. S. Sedzro, T. Pidéname, and Y. Lare, 'Electrical Load Forecasting Using Machine Learning Approach: Simulation of a Microgrid Energy Consumption with Adaboost Regressor.', *J. Sustain. Dev. Energy Water Environ. Syst. JSDEWES*, vol. 13, no. 4, p. 1, Dec. 2025, doi: 10.13044/j.sdewes.d13.0606.
[29] S. Ahmed, Y. Lee, S.-H. Hyun, and I. Koo, 'Unsupervised Machine Learning-Based Detection of Covert Data Integrity Assault in Smart Grid Networks Utilizing Isolation Forest', *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 10, pp. 2765–2777, Oct. 2019, doi: 10.1109/TIFS.2019.2902822.
[30] W. S. Al Farizi, I. Hidayah, and M. N. Rizal, 'Isolation Forest Based Anomaly Detection: A Systematic Literature Review', in *2021 8th International Conference on Information*
UNCORRECTED PROOF

Technology, Computer and Electrical Engineering (ICITACEE), Sep. 2021, pp. 118–122. doi: 10.1109/ICITACEE53184.2021.9617498.

[31] L. Fan *et al.*, 'A Clustering Method of Electric Power Customers Based on DBSCAN Algorithm', in *2024 IEEE 3rd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, Feb. 2024, pp. 1236–1244. doi: 10.1109/EEBDA60612.2024.10485751.

[32] D. Deng, 'DBSCAN Clustering Algorithm Based on Density', in *2020 7th International Forum on Electrical Engineering and Automation (IFEEA)*, Sep. 2020, pp. 949–953. doi: 10.1109/IFEEA51475.2020.00199.

[33] 'Kolmogorov–Smirnov Test: Overview - Berger - Major Reference Works - Wiley Online Library'. Accessed: Feb. 08, 2026. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat06558

[34] M. Wu, 'Exploiting Micro-Signals for Physiological Forensics', in *Proceedings of the 2020 ACM Workshop on Information Hiding and Multimedia Security*, in IH&MMSec '20. New York, NY, USA: Association for Computing Machinery, Jun. 2020, p. 1. doi: 10.1145/3369412.3396882.

[35] S. M. Al-Selwi *et al.*, 'RNN-LSTM: From applications to modeling techniques and beyond—Systematic review', *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 5, p. 102068, Jun. 2024, doi: 10.1016/j.jksuci.2024.102068.

[36] R. Alsini, A. Almakrab, A. Ibrahim, and X. Ma, 'Improving the outlier detection method in concrete mix design by combining the isolation forest and local outlier factor', *Constr. Build. Mater.*, vol. 270, p. 121396, Feb. 2021, doi: 10.1016/j.conbuildmat.2020.121396.

[37] T. Barbariol and G. A. Susto, 'TiWS-iForest: Isolation forest in weakly supervised and tiny ML scenarios', *Inf. Sci.*, vol. 610, pp. 126–143, Sep. 2022, doi: 10.1016/j.ins.2022.07.129.

[38] W.-C. Kuo, C.-H. Chen, and C.-C. Wang, 'Enabling strategic bidding in ancillary services through a robust probabilistic load forecasting framework: The DREAM approach', *Electr. Power Syst. Res.*, vol. 254, p. 112687, May 2026, doi: 10.1016/j.epsr.2025.112687.

[39] 'Stochastic Reliability Enhancement of Renewable-Rich Power System via LSTM Forecasting and Degradation-Aware Resource Optimization | Iranian Journal of Science and Technology, Transactions of Electrical Engineering | Springer Nature Link'. Accessed: Feb. 08, 2026. [Online]. Available: https://link.springer.com/article/10.1007/s40998-025-01010-1

UNCORRECTED PROOF