ORIGINAL RESEARCH **Open access**

Graph Representation Learning and Boosting for User-Level Multi-Energy Load Anomaly Detection

Nora Schneider[1*], Tobias Frank[1]

Abstract

For data-driven operations in integrated energy systems, spotting anomalies is vital. Standard detection techniques tend to examine only a single energy load, often missing the spatial connections that exist among multivariate energy time series. A further major obstacle lies in handling the inherently skewed distribution of user-level multi-energy load records. This study puts forward EGBAD—an Ensemble Graph-Boosted Anomaly Detection system—that exploits the complementary strengths of graph relational reasoning and ensemble learning for user-level multi-energy loads. First, a dynamic method for constructing graphs based on multidimensional scaling (MDS) is introduced to turn multi-energy load measurements into graph-based representations. A graph convolutional network (GCN) then operates on these graphs to learn the spatiotemporal couplings present in the multi-energy load time series. To bolster detection reliability in the face of class imbalance, the complete training pipeline is situated within a Boosting ensemble scheme, where the importance of the minority class is systematically amplified with every boosting round. Tests on publicly obtainable real-world datasets confirm that the proposed model achieves heightened anomaly detection accuracy relative to most baseline approaches. Its performance is particularly striking under conditions of extreme label imbalance, where it attains top scores in both recall and F1-score for anomaly identification.

Keywords Anomaly detection, Multi-energy, Graph representation learning, Boosting

*Correspondence:
Nora Schneider
nora.schneider@rwth-aachen.de
[1] Department of Artificial Intelligence for Power Systems, Faculty of Electrical Engineering, RWTH Aachen University, Aachen, Germany

Introduction

User-level integrated energy systems are indispensable for the secure, stable, and low-carbon running of today's power infrastructures [1]. Their success depends heavily on high-integrity data to perform essential functions such as load forecasting [2] and resource allocation [3]. Yet, these user-level streams are susceptible to anomalies stemming from broken sensors, equipment faults, or cyber intrusions, eroding data quality and emphasizing the pressing need for resilient anomaly detection mechanisms [4].

Notwithstanding recent headway, spotting anomalies in user-level multi-energy load data is still profoundly difficult for several reasons:

1) Pronounced uncertainty and fluctuation. A multitude of external influences—weather, traffic conditions, occupant behavior—shape multi-energy loads. The resulting multi-source unpredictability produces erratic, volatile patterns, making anomaly detection far more challenging than for typical power system data [5].

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

2) Label shortage and severe imbalance. Datasets with exhaustive user-level labels are rare and often exhibit a pronounced long-tail distribution [6]. This acute class skew cripples the performance of supervised anomaly detection techniques.

3) Elaborate spatiotemporal linkages. Reliable anomaly identification requires capturing both temporal structure within each series and spatial interdependencies among variables such as meteorological conditions, human activities, and renewable generation. As seen in **Figure 1**, a highlighted segment around the 600–650 mark in an electricity load trace might appear out of the ordinary, but when viewed against coincident shifts in "HAC" and "temp" signals, a normal status becomes more plausible—validating the necessity of cross-series correlation analysis.

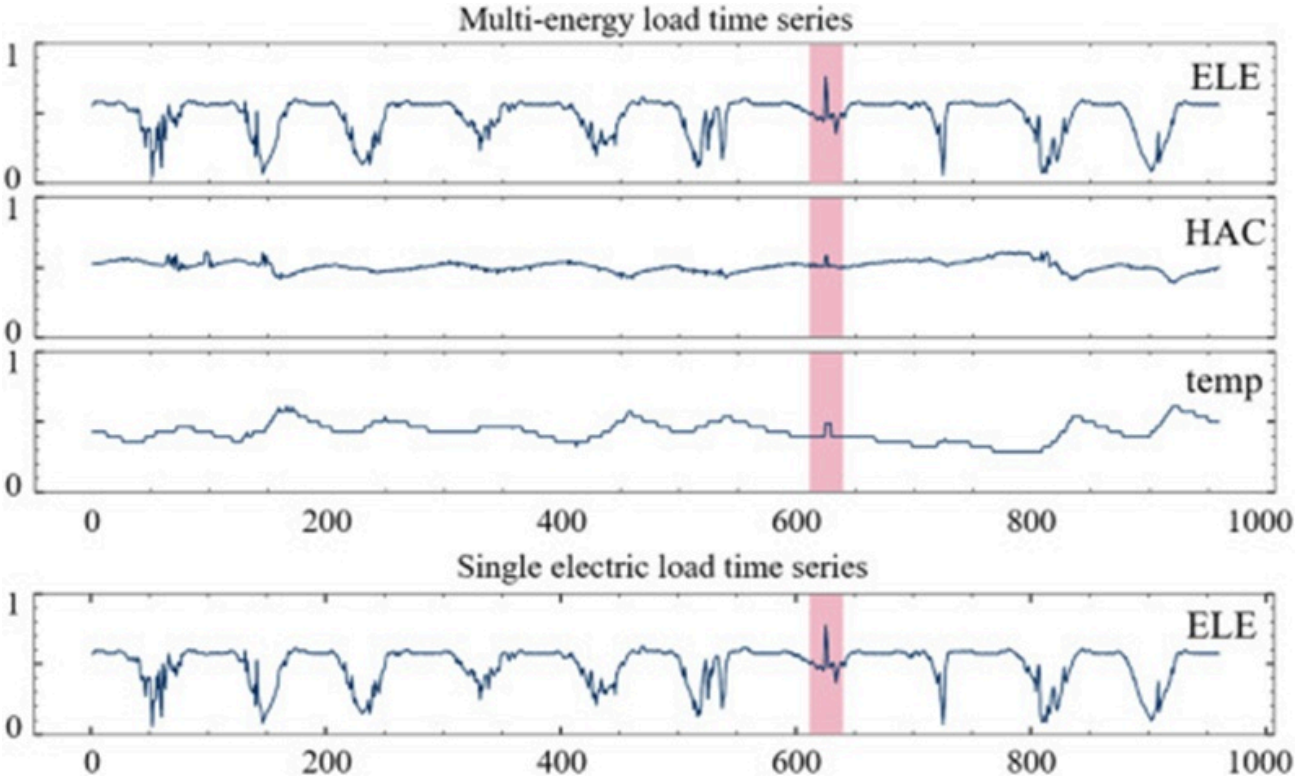


Figure 1. A potential anomaly (pink block) contrasted between multi-energy load and single electric load series.

Approaches to anomaly detection fall into two broad classes: statistical and artificial intelligence-driven. Statistical procedures, like the 3σ rule and interquartile range (IQR) rule [7], are conceptually uncomplicated and interpretable. However, because they rely heavily on static thresholds to declare outliers, their utility fades when faced with the high-dimensional, sparse, spatiotemporally intertwined data typical of user-level multi-energy systems.

Within the field of artificial intelligence, a further division exists between unsupervised and supervised methods. Unsupervised pathways typically isolate anomalies by grouping data and flagging points that stray far from dominant clusters as outliers [8]. Still, such pathways often hinge on prior distributional assumptions; they underperform in user-level multi-energy contexts where data are nonlinear and tightly coupled across time and space, eroding cluster boundary reliability and detection performance.

Concerning supervised routes, deep learning models have become the leading choice, with recurrent neural networks (RNNs) or Transformers routinely employed for learning long-range temporal patterns [9-12]. Their capacity to encode spatial cross-variable relationships, however, stays circumscribed, curtailing their suitability for the deeply interconnected nature of multi-energy load data. Graph Neural Networks (GNNs) have recently emerged as a promising alternative by treating multivariate time series as graphs and jointly modeling spatial and temporal dynamics [13-19]. Despite these

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

innovations, the majority of current models either presuppose fully annotated datasets or rely on unsupervised methods, hampering real-world deployment. In user-level multi-energy environments, acquisition constraints and privacy considerations mean only a handful of anomalies can normally be labeled, producing a weakly supervised setting plagued by severe imbalance that present methods are ill-suited to address.

To surmount these hurdles, this work presents EGBAD, an original architecture that weds graph-based representation learning with a Boosting ensemble paradigm. Our design faithfully captures the intricate correlations among multi-energy time series while maintaining strong robustness to skewed data distributions. EGBAD's foundation rests on three pillars: 1) dynamic graph construction, which employs multidimensional scaling to embed multivariate series into a graph format, encoding their spatiotemporal interplay; 2) a graph convolutional network serving as the core graph processing unit; 3) a Boosting Ensemble Framework that iteratively adjusts sample weights to spotlight elusive anomalies, thereby delivering stable results on imbalanced data.

The key contributions of this work are outlined below:

1) We devise EGBAD, an innovative framework combining graph representation learning with a Boosting ensemble. This synthesis enables effective characterization of intricate spatiotemporal dependencies while directly addressing severe data imbalance in user-level multi-energy settings.

2) In particular, we present an MDS-based dynamic graph construction technique that successfully extracts temporal dependencies and spatial coupling features from multi-energy load data. Additionally, a new imbalance-handling mechanism is introduced. By gradually heightening the weight of difficult instances inside a Boosting procedure, the tendency to overfit the normal class is reduced, and rare anomaly detection is sharpened.

3) Comprehensive testing on publicly accessible real-world datasets demonstrates that EGBAD outperforms cutting-edge baseline methods, especially in regimes of extreme class imbalance.

Related Work

This section surveys prior contributions to anomaly detection. We group the body of work into three strands: unsupervised learning-driven, supervised learning-driven, and graph-driven approaches. To enable a side-by-side appraisal, we compile a structured overview that distills the central properties, merits, and constraints of each strand, as shown in **Table 1.**

Table 1. Structured comparison of anomaly detection methods categorized by approach.

Reference	Method	Core Concept	Limitations
Unsupervised Learning-based Approaches			
[7]	3σ Rule	Detects anomalies using a standard deviation–based thresholding strategy.	Depend on predefined thresholds and statistical distribution assumptions, making them unsuitable for spatiotemporally coupled datasets.
[8]	IQR Rule	Identifies abnormal observations according to interquartile range boundaries.	
[20]	DBSCAN	Applies density-based clustering, treating	

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

		points outside dense regions as anomalies.	
[21]	k-means	Groups samples into clusters, where instances located far from cluster centroids are considered anomalous.	
	Isolation Forest	Utilizes random feature partitioning; anomalous samples tend to be isolated with fewer splits.	
Supervised Learning-based Approaches			
[10]	LSTM-DNT	Prediction-based method that combines LSTM forecasting with threshold-based detection, effectively capturing temporal dependencies.	Performance is highly influenced by prediction accuracy and threshold selection, limiting its effectiveness on spatiotemporally coupled multi-energy load datasets.
[12]	MHA	Prediction-based framework employing multi-head attention to capture long-range temporal dependencies across sequences.	
[22]	DLLD	Prediction-based approach built upon a CNN backbone to learn spatial correlations.	
[23]	MAD-GAN	Reconstruction-based method using adversarially trained autoencoders, primarily focusing on spatial dependency modeling.	Most methods emphasize either temporal or spatial patterns, while rarely learning both simultaneously.
[9]	LSTM-VAE	Reconstruction-based model integrating LSTM and variational autoencoders to capture temporal dynamics with probabilistic robustness.	
[11]	OmniAnomaly	Reconstruction-based framework utilizing stochastic RNNs to model temporal relationships.	
Graph-based Approaches			
[13]	Grid2Vec	Employs GNNs to learn grid-topology representations for fault detection, demonstrating strong spatial modeling capability.	Strongly reliant on predefined topology structures and sufficient labeled historical data.
[24]	GraphPMU	Combines GNNs with an encoder–decoder architecture to capture both spatial and temporal dependencies.	
[15]	DyGraphAD	Detects anomalies by jointly modeling inter-series correlations and intra-series temporal	Typically assumes the availability of reliable labels or balanced

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

		dynamics, thereby capturing spatiotemporal relationships.	datasets, which is impractical in user-level applications where anomalies are scarce and severely imbalanced.
[16]	STND	Integrates GNN and GRU architectures to learn spatiotemporal dependencies.	
[19]	GReLeN	Uses a variational autoencoder backbone together with a stochastic graph-learning mechanism to model spatiotemporal interactions.	
[17]	TLAN	Incorporates a temporal logical attention mechanism and an adaptive thresholding strategy to better adapt to spatiotemporal correlations.	
[14]	GDN	Learns dependency graphs among multivariate time series through sensor embeddings and deviation-based scoring.	
[18]	GAT	Utilizes graph attention networks to capture spatial and temporal relationships.	
Ours	**EGBAD**	Employs dynamic graph construction together with GCNs to comprehensively capture spatiotemporal coupling patterns in multi-energy load data. In addition, a Boosting-based ensemble learning strategy improves robustness under highly imbalanced data conditions.	—

Unsupervised learning-based approaches

Methods that operate without labels are gaining considerable traction in anomaly detection. Widely adopted clustering-based representatives encompass DBSCAN [25] and k-means [20]. A further well-known technique, Isolation Forest [21], distinguishes abnormal records by requiring fewer recursive partitions [26], thereby improving detection fidelity. Nevertheless, these schemes tend to rely on rigid thresholds and prior distributional assumptions, rendering them inadequate for pinpointing anomalies hidden in multi-energy load sequences characterized by deep spatiotemporal entanglement.

Supervised learning-based approaches

The scope of this subsection is narrowed to deep learning-oriented methods alone. Regarding how abnormalities are recognized, the pertinent literature can be divided into two branches: prediction-centric and reconstruction-centric strategies.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Frameworks built on RNNs or Transformers can encode temporal regularities and have been extensively leveraged for anomaly flagging. LSTM-DNT [10] discovers multivariate time series anomalies by exploiting prediction-error margins. Similarly, MHA [12] employs multi-head attention to extract temporal interdependencies across multiple sequences from charging-station operations. DLLD [22], designed on top of CNNs, jointly detects the presence and exact location of anomalies within power infrastructure. These strategies expose anomalies by comparing actual measurements with forecasted estimates and highlighting samples where the divergence is pronounced. Yet a fundamental vulnerability lies in their sensitivity to forecasting inaccuracies and cutoff choices. Their reliability often falters when applied to high-dimensional, spatiotemporally interwoven multi-energy load data, as predictive uncertainty may alternately mask genuine anomalies or lead to false positives.

MAD-GAN [23] achieves rapid anomaly extraction from multivariate time series via adversarially trained autoencoders. LSTM-VAE [9] couples an LSTM with a variational autoencoder to reconstruct the latent data manifold. OmniAnomaly [11] models the data distribution with a stochastic RNN and leverages inferred latent-variable representations to discriminate between normal behavior and anomalous events. These designs compress prototypical samples into a low-dimensional latent space and trigger anomaly alerts once reconstruction residuals surpass a learned bound. A recurring shortcoming, however, is that such reconstruction-based pipelines are predominantly erected on RNN or CNN foundations; they tend to emphasize either temporal or spatial structure while rarely addressing both concurrently. This narrow focus curtails their potency in multi-energy load environments governed by rich spatiotemporal interplay.

Graph-based approaches

Graphs are widely regarded as a powerful vehicle for expressing cross-series linkages, and a growing wave of research has engaged GNNs for anomaly detection. Grid2Vec [13] investigates representations of grid topology and leverages them for fault detection, yet its success depends heavily on specific topological configurations and ample historical labeled data. Likewise, STND [16] and GraphPMU [24] mobilize graph representation learning to encode spatiotemporal correlations and surface anomalous incidents. Complementary studies refine graph formation by learning sensor-specific embedding vectors [14] or by incorporating supplementary convolutional operations [15]. Building on Deng and Hooi [14], architectures such as GAT [18] and GReLeN [19] fuse graph attention mechanisms with variational autoencoders for feature extraction, achieving state-of-the-art performance on benchmark corpora. Despite such progress, these solutions still prove difficult to transplant into user-level multi-energy settings. They customarily assume the existence of dependable annotations or class-balanced collections, whereas practical data are weakly supervised and display a severe long-tail distribution.

Problem Statement

In this investigation, we consider a user-level unit hosting N sensors. The multi-energy load measurements, along with other explanatory variables, form a multivariate time series ensemble represented as $X \in R^{N \times T}$, where T denotes the sequence duration. Moreover, we denote the time series indexed by i over the full span T as Xi, and the vector of all series at a given timestamp t as Xt. Operating under a weakly supervised regime, the base learner is trained on unlabeled samples and produces detection outcomes on a labeled validation set, in keeping with the Boosting paradigm. The proposed construct functions as an end-to-end framework tasked with emitting a binary label $\hat{y}$ for every timestamp in the test partition. Aligned with common practice, label 0 corresponds to a normal sample and label 1 denotes an anomaly.

Materials and Methods

The complete schematic of our proposed model, EGBAD, is shown in **Figure 2**. EGBAD is designed to infer spatiotemporal relationships embedded in users' multi-energy loads using an ensemble graph-boosted methodology. The

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

architecture rests on three primary building blocks:

1) Dynamic graph construction: The multi-energy time series are sliced via a sliding window. Subsequently, each window is projected into a low-dimensional feature space via MDS; pairwise Euclidean distances then determine edge weights, yielding an ordered set of dynamic graphs.

2) Graph relational analysis founded on GCN: Base detection components anchored in GCNs are formulated to model inter-series interactions and temporal evolution across windows, enabling the apprehension of spatiotemporal dependencies.

3) Boosting ensemble learning framework: Base detectors are trained in succession; in each subsequent round, the importance weights of difficult cases are adjusted based on the preceding round, thereby counteracting majority-class overfitting and strengthening resilience to imbalance. A fully connected terminal layer with softmax outputs anomaly likelihoods.

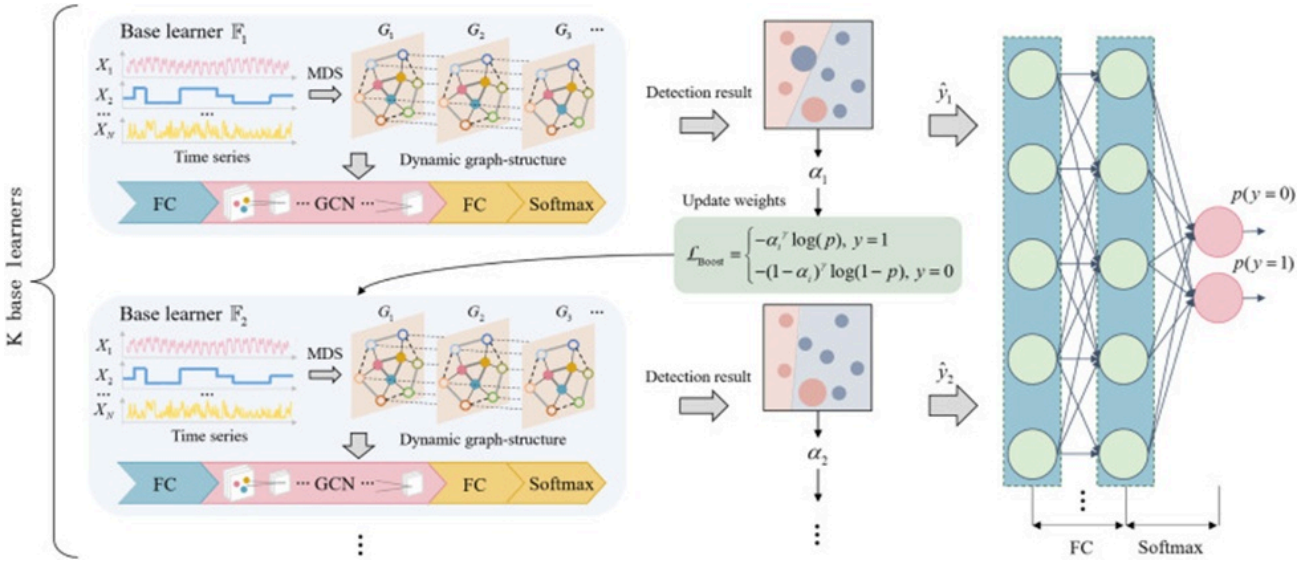


Figure 2. The overall framework of EGBAD, including 1) Dynamic graph construction via MDS; 2) GCN backbone to model spatiotemporal dependencies; and 3) Boosting ensemble strategy.

Dynamic graph construction

Crafting a meaningful graph representation is a prerequisite for applying GCNs to multi-energy load data. To embed the inherent spatiotemporal coupling, we turn to MDS for determining the graph's weighted edges. By safeguarding pairwise relationships, MDS supplies a disciplined mapping from inter-series dissimilarities directly into a graph topology. The complete procedure is illustrated in **Figure 3**.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

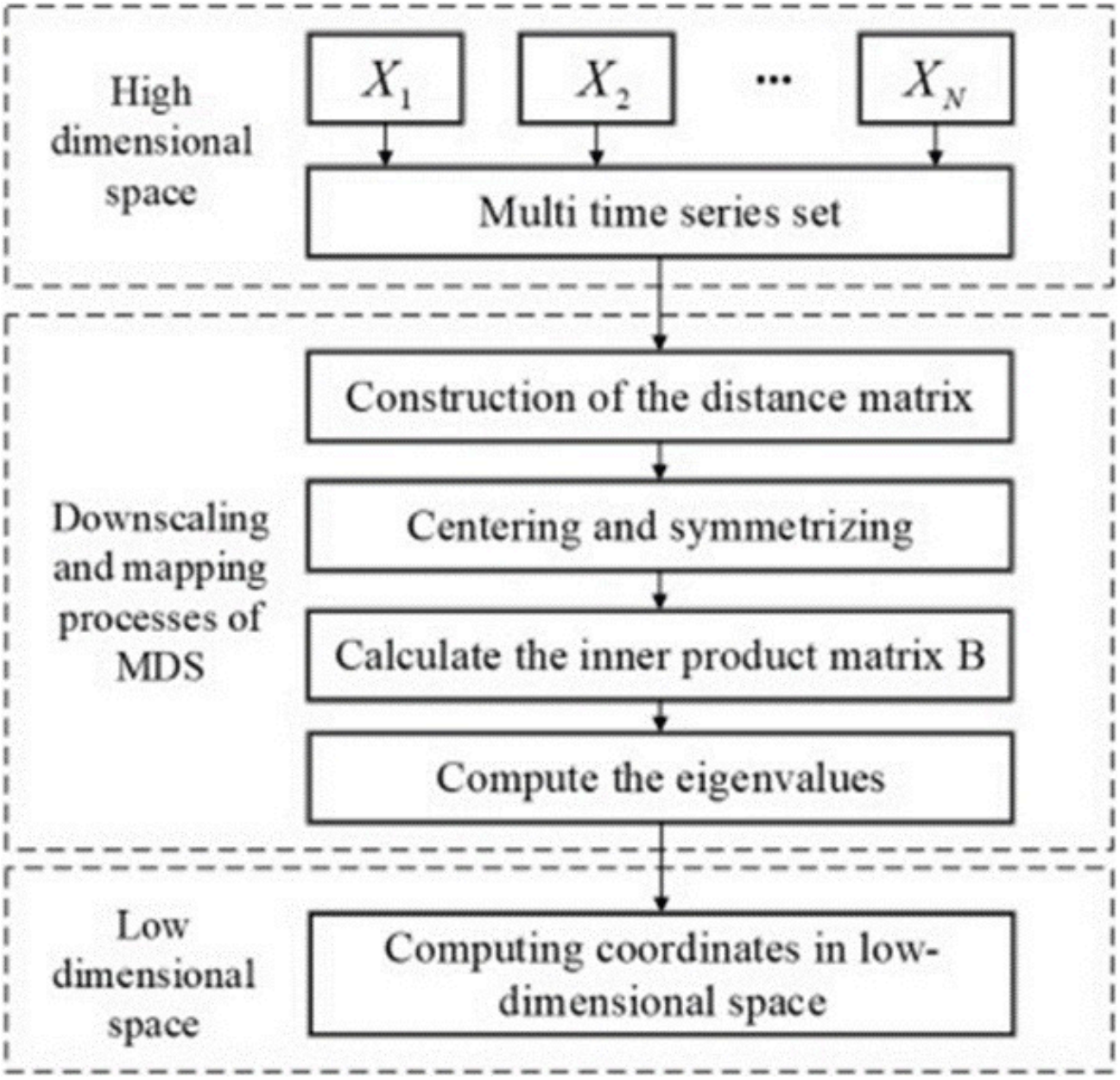


Figure 3. Workflow of multidimensional scaling.

Consider datasets with p features and n samples in MDS:

$$\{X_1, X_2, ..., X_n | f_i, i = 1, 2, ...p\} \quad (1)$$

Euclidean distance is adopted as the metric within MDS. The gap between any pair of sample sequences is defined as $d_{ij} = \|X_i - X_j\|$, i, j ∈ 1,2,...,n, $\|\ .\|$ signifies the distance function. The distance matrix is:

$$\boldsymbol{D} = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{bmatrix} \quad (2)$$

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

The centering transformation applied to the distance matrix is given by:

$$E = I - \frac{1}{n} ii^{T} \quad (3)$$

where I denotes the identity matrix and i stands for an all-ones vector of length n.

The inner product matrix B is then computed as:

$$B = -\frac{1}{2} ED^{2} E \quad (4)$$

And the eigenvalues together with the eigenvectors are expressed as:

$$B = V \Lambda V^{T} = V \Lambda^{\frac{1}{2}} \Lambda^{\frac{1}{2}} V^{T} \quad (5)$$

Where V and Λ refer to the eigenvector and eigenvalue matrices of the inner product matrix B, respectively.

Within the context of this study, the high-dimensional feature space is reduced to a 2-dimensional plane; accordingly, the two largest eigenvalues (λ1, λ2) alongside their corresponding eigenvectors (v1, v2) are retained. The lower-dimensional coordinates become:

$$f = V \Lambda^{\frac{1}{2}} \quad (6)$$

Spatial interconnections among multi-energy variables are far from static—they shift over time. As a result, any fixed graph representation is bound to miss the vital spatiotemporal coupling inherent in the series. To remedy this shortfall, we propose a dynamic graph construction scheme, as shown in **Figure 4.**

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

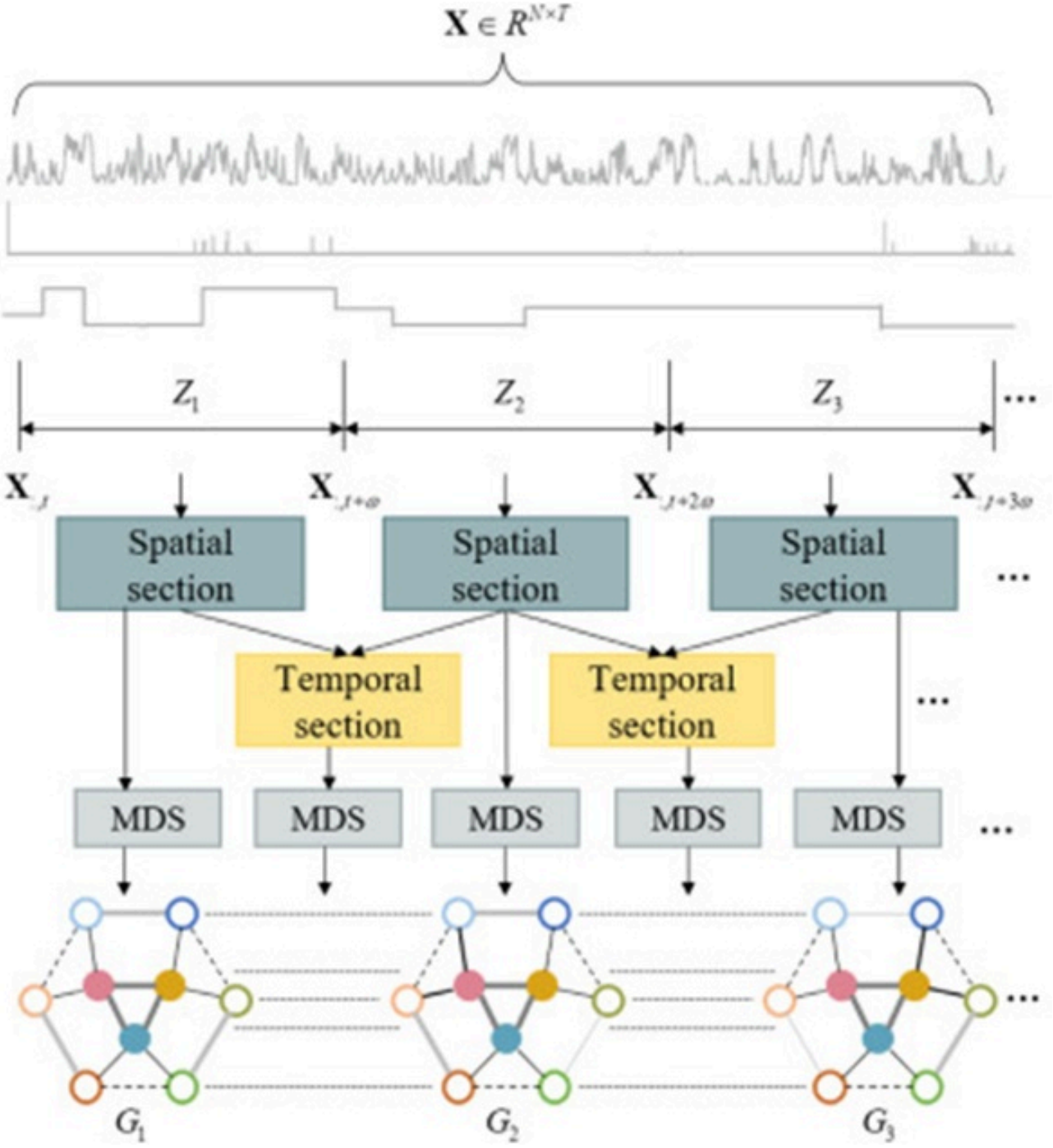


Figure 4. Dynamic graph construction based on MDS.

The time series are partitioned into distinct zones {Z} according to a sliding window ω:

$$Z_w = \left\{X^{(w-1)\omega}, X^{(w-1)\omega+1}, \ldots, X^{w\omega}\right\} \quad (7)$$

Within each temporal zone, static graphs are built. The distance optimization objective that MDS strives to minimize is:

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

$$\min \sum_{i<j} \left(\| X_i - X_j \| - d_{ij} \right) \quad (8)$$

In this paper, we aim to construct dynamic graphs that capture spatial and temporal correlations in multivariate time series. It therefore becomes necessary to model the interconnections between consecutive graphs. Given graph w and its successor graph w + 1, the MDS distance optimization objective is formulated as:

$$\min \sum_{\forall i,j} \left(\| X_i^w - X_j^{w+1} \| - d_{ij} \right) \quad (9)$$

Graph relational analysis based on GCN

After the multi-energy load data are converted into a sequence of graph structures, we employ a GCN to extract their deep spatiotemporal correlations. This GCN operates as the central analytical apparatus, digesting the graph representations to surface meaningful relational motifs.

For an undirected graph G = (V, E), the sets V and E stand for the nodes and edges, respectively. The adjacency matrix A encodes the nodal associations; its element Ai,j indicates the presence or absence of an edge linking node i and node j, being 1 if a link exists and 0 otherwise. The degree matrix D captures the graph's edge density. It is a diagonal matrix whose diagonal entries Dii equal the count of edges incident upon node i. The connection between the adjacency matrix and the degree matrix is depicted in **Figure 5.**

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

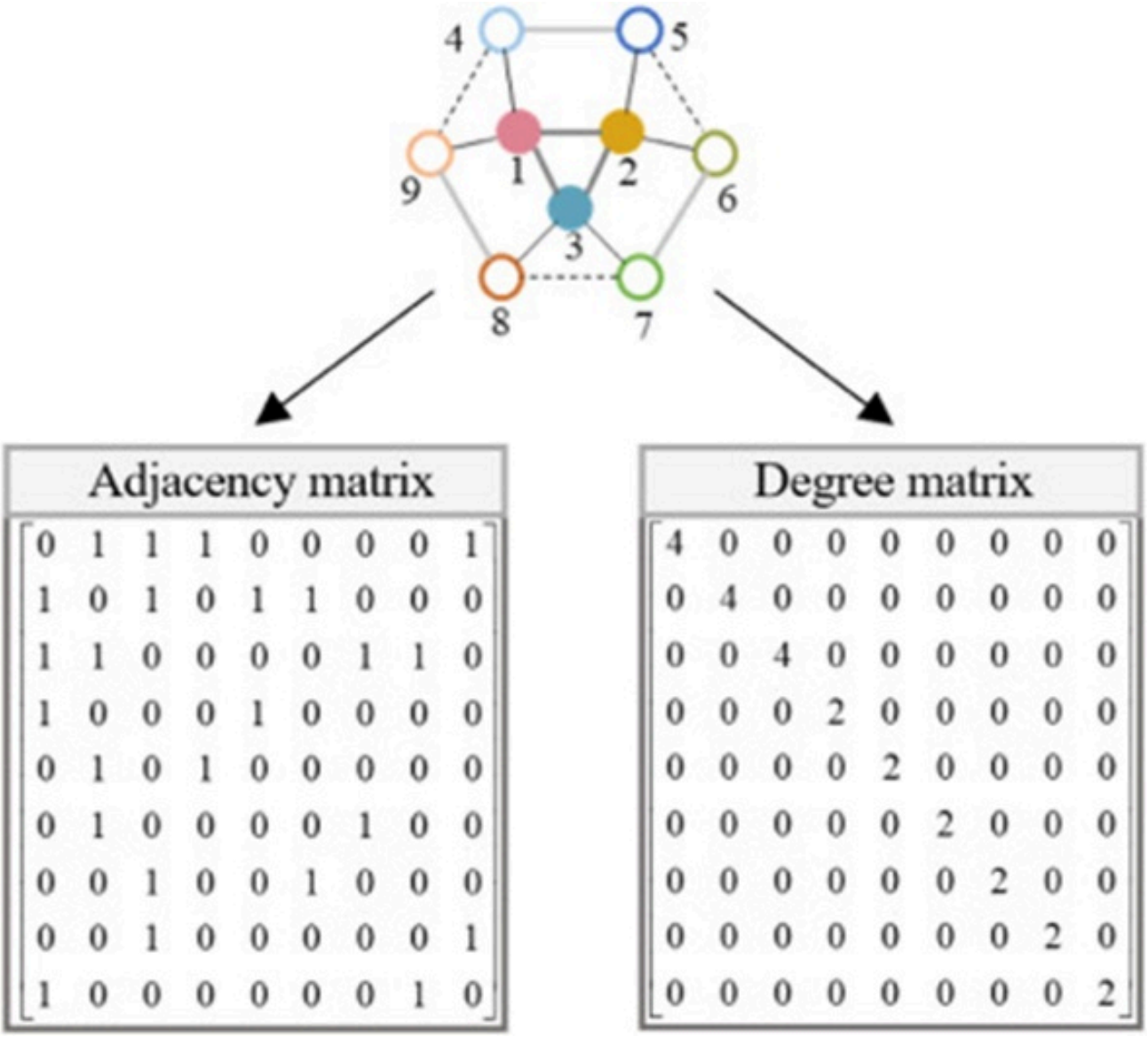


Figure 5. Adjacency matrix and degree matrix of graphs.

The Laplacian matrix L and the symmetrically normalized Laplacian matrix $\boldsymbol{L}^{sym}$ of graph G are expressed as follows:

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A} \quad (10)$$

$$\boldsymbol{L}^{sym} = \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{L}\boldsymbol{D}^{-\frac{1}{2}} \quad (11)$$

Adopting $\boldsymbol{L}^{sym}$ permits the simultaneous incorporation of information propagated from neighboring nodes and the node's intrinsic information. Moreover, the outputs of successive layers remain bounded because left-multiplication by D^{-1} and right-multiplication by D^{-1} perform row-wise and column-wise normalization, respectively. Given that $\boldsymbol{L}^{sym}$ is a positive semi-definite symmetric matrix, it assuredly possesses n linearly independent eigenvectors, thus enabling eigen decomposition.

Drawing on the convolution theorem:

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

$$f * g = \sum_{\tau=-\infty}^{+\infty} f(\tau)g(n-\tau) = F^{-1}[\hat{f}(\omega)\hat{g}(\omega)] \quad (12)$$

Where the left-hand side represents the discrete formulation of convolution, and the right-hand side asserts that convolving a graph signal f with a kernel g equates to computing the inverse Fourier transform of the product of their respective Fourier transforms.

Extending this principle to graph convolutional neural networks, the kernel g is expressed in diagonal matrix form:

$$\begin{bmatrix} \hat{g}(\lambda_1) & & \\ & \ddots & \\ & & \hat{g}(\lambda_n) \end{bmatrix} \quad (13)$$

where $\hat{g}$ (λ_k) = $\sum_{i=1}^{n}$ g (i)u_k^* (i) is the Fourier transform of g, and u_k^* represents the complex conjugate of the k-th eigenvector.

Analogously, for the Fourier transform $\hat{f}$ (λ_k) = $\sum_{i=1}^{n}$ f (i) u_k^* (i) of the graph signal f, its matrix formulation is $\hat{f}$=$\mathbf{U}^T$f.

Multiplying by U and carrying out the inverse transformation leads to the graph convolution expression:

$$f * g_G = \boldsymbol{U}\begin{bmatrix} \hat{g}(\lambda_1) & & \\ & \ddots & \\ & & \hat{g}(\lambda_n) \end{bmatrix}\boldsymbol{U}^T f \quad (14)$$

Where U and λ denote the eigenvectors and eigenvalues, respectively.

Paralleling the graph structure, $\boldsymbol{L}^{sym}$ serves as the medium for characterizing inter-nodal relationships and other non-Euclidean attributes. Because $\boldsymbol{L}^{sym}$ is a positive semi-definite symmetric matrix, it can undergo eigenvalue decomposition to obtain U and λ.

Within the deep learning pipeline, it is essential to instantiate convolutional kernel parameters that can be learned via backpropagation. The convolutional diagonal matrix diag($\hat{g}$(λ_i)) is therefore cast as the trainable kernel parameters:

$$\begin{bmatrix} \sum_{i=0}^{K} \theta_i \lambda_1^i & & \\ & \ddots & \\ & & \sum_{i=0}^{K} \theta_i \lambda_n^i \end{bmatrix} = g_\theta(\lambda) \quad (15)$$

Where $\theta_i \in$ is the parameter subject to error backpropagation-based optimization and can be initialized accordingly.

Hence, the layer-wise operation of the graph convolutional network is formulated as:

$$\text{Layer}: y = \sigma\left(\boldsymbol{U}g_\theta(\lambda)\boldsymbol{U}^T x\right) \quad (16)$$

where σ denotes the nonlinear activation, x represents the input, and y designates the output.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

In this work, the second-order Chebyshev polynomial [27] is employed as a substitute for the convolutional kernel:

$$g_\theta(\lambda) = \sum_{k=0}^{K-1} \theta'_k C_k(\lambda) \quad (17)$$

where Ck (.) is the k-th order Chebyshev polynomial, and $\theta' \in \mathbb{R}^K$ is the Chebyshev coefficient vector.

At this stage, the GCN instantiated in this paper can be written compactly as:

$$\text{Layer}: y = \sigma\left(U \sum_{k=0}^{K-1} \theta'_k C_k(\lambda) U^T x\right) \quad (18)$$

Its matrix-form expression becomes:

$$Y = \sigma(L^{sym} X W) \quad (19)$$

Where W is the linear weight matrix.

Anomaly detection within multi-energy load time series is cast as a binary classification task. Accordingly, a fully connected (FC) layer followed by a SoftMax layer is appended to the GCN stack, forming the elementary detector inside our EGBAD framework, as visualized in **Figure 6.**

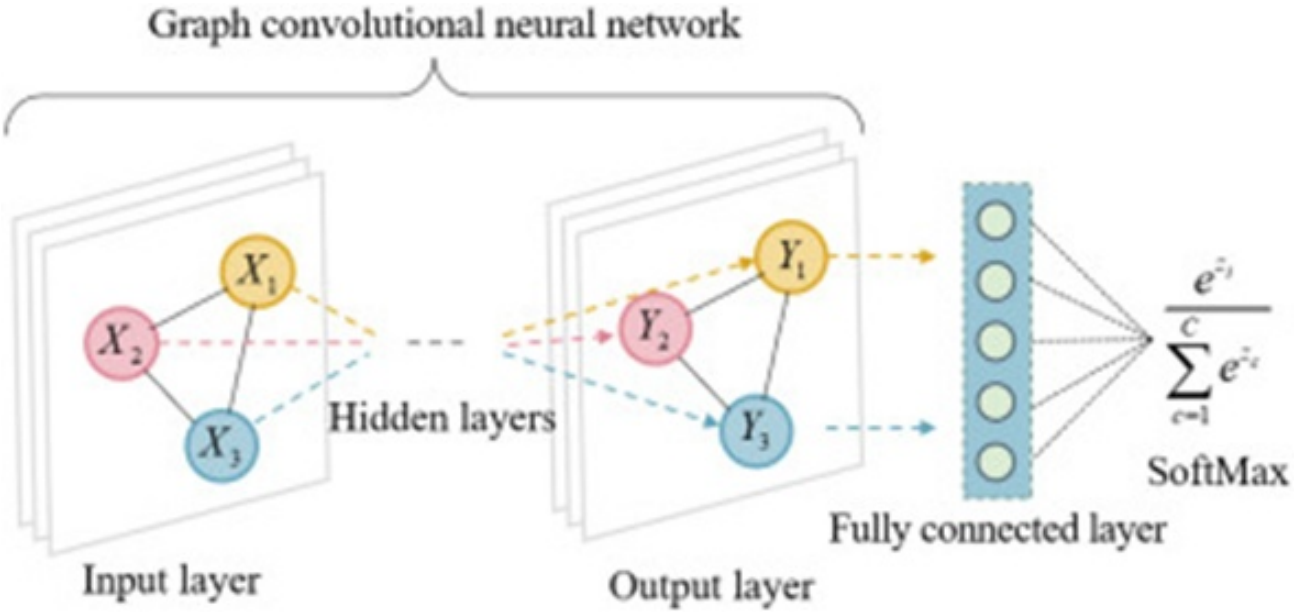


Figure 6. Model architecture of a single base learner in EGBAD.

In this sense, a single base detector in this paper can be concisely formulated as:

$$F = \text{Softmax}(\text{ReLU}(\sigma(L^{sym} X W^{(0)}) W^{(1)})) \quad (20)$$

Boosting ensemble learning framework

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Real-world user-level measurements pose two enduring obstacles: a shortage of labels and a severely imbalanced class ratio. Both issues stem from practical data-collection hurdles and stringent privacy regulations. Under such conditions, constructing vast, meticulously annotated databases is rarely feasible, thereby curbing the power of fully supervised models.

To navigate these hurdles, we tailor a Boosting-based ensemble training protocol specifically to user-level multi-energy loads. We discovered that the incremental, layer-by-layer training characteristic of boosting aligns naturally with the bias in data distribution introduced by class imbalance. In essence, boosting combines several mutually reinforcing base detectors into a single superior detector, thereby improving performance when labels are scarce and imbalance is severe [28].

Boosting strategies are broadly split into two camps: 1) amplifying the importance of wrongly classified instances, epitomized by Adaptive Boosting (AdaBoost); 2) approximating the negative gradient of a composite strong learner, as embodied by gradient boosting. For our purposes, we adopt a scheme reminiscent of AdaBoost. We engineer a bespoke loss function and incorporate a fully connected layer, together with a SoftMax layer, to take over the weighted voting role in the original formulation.

More formally, given an ensemble of base learners Fi, i = 1, 2, ..., K under the AdaBoost strategy, the ultimate decision emerges as a weighted sum across all base learners:

$$Y_{\text{Adaboost}}(x) = \sum_{i=1}^{K} \beta_i F_i(x) \quad (21)$$

AdaBoost relies on global optimization. While training the m-th layer, the objective function must incorporate the weightings inherited from all m − 1 prior learners:

$$\min \sum_{t=1}^{m-1} L\left(y_t, \sum_{i=1}^{K} \beta_i F_i(x)\right) \quad (22)$$

AdaBoost is computationally efficient and typically guards against overfitting, yet it retains sensitivity to outliers. When the class skew is extreme, the earliest learners regularly miss anomalies altogether. The ensuing over-amplification of misclassified cases destabilizes the sequential updates and propagates mistakes into downstream stages. As a consequence, the eventual detector may be shackled by the poor performance of its initial members. Our framework, by contrast, reformulates the sample-weighted loss to place explicit emphasis on samples that are inherently difficult to categorize correctly. This heightens anomaly recall while sidestepping the noted instability. **Figure 7** juxtaposes our Boosting framework against the conventional AdaBoost scheme.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

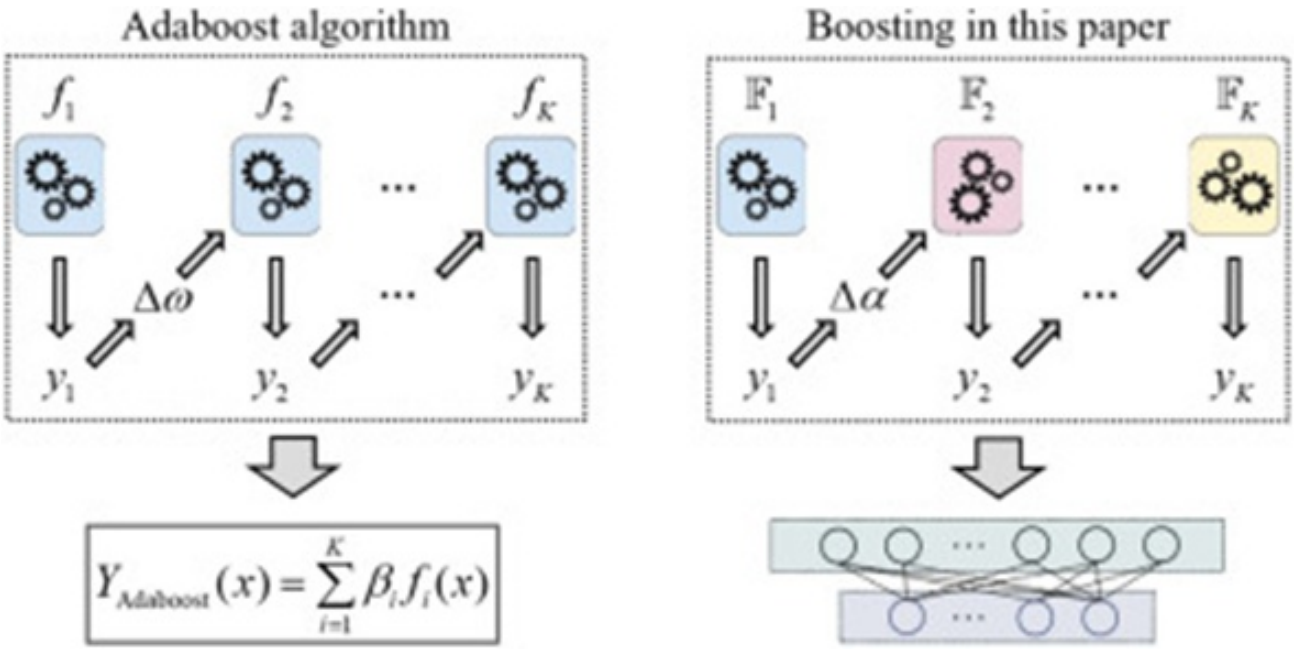


Figure 7. Adaboost against the proposed Boosting framework.

Against this backdrop, we overhaul the traditional cross-entropy loss to adaptively zero in on cases with a high risk of misclassification. The enriched loss function reads:

$$L_{\text{Boost}} = \begin{cases} -\alpha_i{}^{\gamma}\log(p), y = 1 \\ -(1-\alpha_i)^{\gamma}\log(1-p), y = 0 \end{cases} \quad (23)$$

Where p corresponds to the model's predicted probability for the given class, and α is the hyperparameter that modulates the sample weight.

If a sample is an anomaly (y = 1) and the model correctly classifies it (p > 0.5), the associated loss term is assigned a lower weight. Should the opposite hold, a larger weight is applied instead.

Experiments

Datasets

The openly available real-world dataset used throughout this paper is sourced from the NEST [29] research platform established by the Swiss Federal Laboratories for Materials Science and Technology. This facility has disseminated multi-energy load records for three user-level smart building units, logged at 15-minute intervals. **Table 2** catalogues the detailed attributes of these datasets.

Table 2. Detailed characteristics of the real-world datasets.

Datasets	SolAce	UMAR	DFAB
Size	(35040,20)	(35040,28)	(35040,42)
Missing values	19359	50003	10457

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Training size	(27318,18)	(24943,26)	(27319,42)
Testing size	(6829,18)	(6236,26)	(6830,42)
Anomalies	1707	1559	1707
Anomaly rate	5%	5%	5%
Number of sensors	20	28	42

For the present study, every dataset was trimmed to a consistent two-year span covering July 2021 to July 2023. Upon removal of missing entries, normalization was carried out according to the expression below:

$$\hat{x} = \frac{x - \min(X)}{\max(X) - \min(X)} \quad (24)$$

To probe more rigorously how the degree of class imbalance affects model behavior, we employ a black-box attack methodology [30] to seed additional synthetic anomalies, enabling us to tune the imbalance ratio. The false data injection mechanism is specified as:

$$\begin{cases} x_{\mathrm{FDI}} \sim N(0, 0.1) \\ |x_{\mathrm{FDI}}| > 0.1 \end{cases} \quad (25)$$

Injected values adhere to a Gaussian distribution centered at zero with a variance of 0.1. Only values with absolute magnitudes greater than 0.1 are retained for insertion.

Implementation details

The GCN backbone opens with an input embedding layer of size 128, proceeds through two graph convolutional hidden layers (64 units apiece), and terminates in an output layer that maps onto a two-class prediction space (normal versus anomaly). The Boosting ensemble employs three base learners; each successive learner systematically heightens the weights of previously misclassified samples to refine the detection of scarce anomalies. Training is performed with a batch size of 64 and a learning rate of 0.001, relying on the Adam optimizer. **Table 3** collates the full model architecture and the associated training hyperparameters.

Table 3. model architecture and training parameters for EGBAD.

Component	Configuration / Value
Input embedding dimension	128
Number of GCN hidden layers	2
Hidden layer dimensionality	64
Output size	2 classes (normal vs. anomaly)
Number of base learners (K)	3 boosting stages
Weight coefficient α	0.25
Focusing factor γ	2.0
Batch size	64

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Learning rate	0.001
Number of training epochs	10
Optimization algorithm	Adam
Cross-validation strategy	5-fold cross-validation

Our experimental arrangement follows a weakly supervised blueprint with sparsely distributed labels: GCN fitting is performed without access to anomaly labels, whereas a compact labeled subset is reserved as the Boosting validation partition to drive sample reweighting. The definitive evaluation is performed on the test set using fully known ground-truth labels. Data Analysis. Every experiment was run on a server housing 4 Intel Xeon Gold 6254 processors and 2 NVIDIA A800 GPUs. To guarantee robustness, we cite scores averaged across 5-fold cross-validation.

Evaluation metrics

This investigation reports detection performance using a confusion matrix, in which anomalous points are treated as the positive class and normal points as the negative class **(Table 4).**

Table 4. Confusion matrix of anomaly detection.

True label \ Predicted label	**Anomaly**	**Normal**
Anomaly	True Positive (TP)	False Negative (FN)
Normal	False Positive (FP)	True Negative (TN)

The anomaly detection efficacy of the proposed model is quantified through precision (P), recall (R), and the F1-score, whose definitions follow:

$$P = \frac{TP}{TP + FP} \quad (26)$$

$$R = \frac{TP}{TP + FN} \quad (27)$$

$$F_{1-\text{Score}} = \frac{2 \times P \times R}{P + R} \quad (28)$$

A high precision value translates to a scarcity of false alarms; a high recall value indicates few overlooked anomalies; and the F1-score furnishes a harmonized, holistic indicator of overall performance.

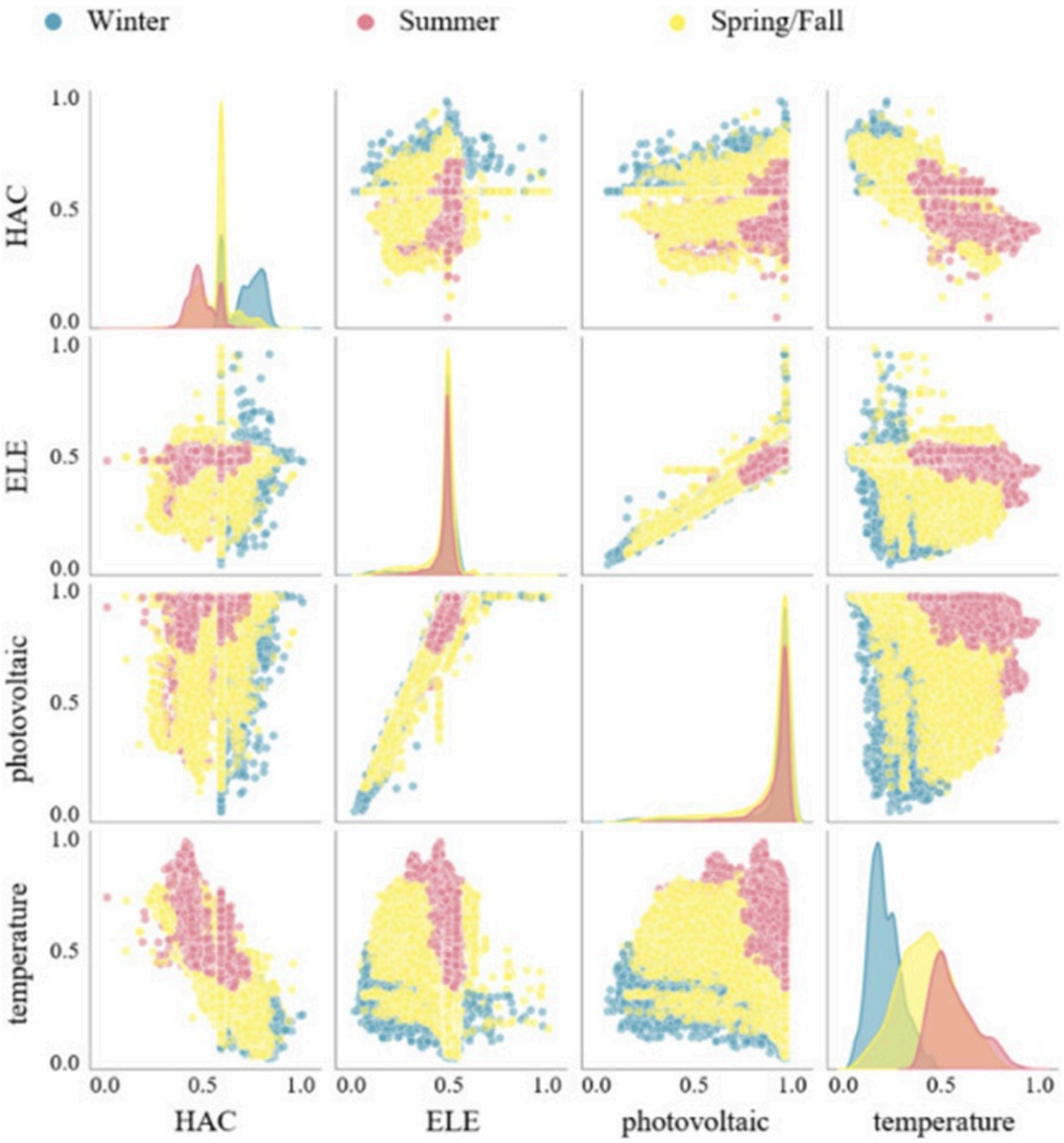
Data analysis

The datasets used in this work faithfully capture the distinctive real-world scenario of user-level multi-energy load data, characterized by strong spatiotemporal coupling and limited, weakly supervised labeling.

Figure 8 illustrates the spatial interconnections among electrical load, cooling/heating load, and auxiliary-feature variables across distinct seasons. Inspection of the off-diagonal tiles reveals substantial spatial coupling. For example,

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

between electrical load and photovoltaic output (row 3, column 2), as well as between cooling/heating loads and ambient temperature (row 4, column 1). The figure further documents how these spatial interrelations morph from winter to summer and through the shoulder seasons of spring and autumn, vividly capturing the authentic nature of spatiotemporal coupling. Although the overall linear correlation between electrical load and cooling/heating load appears weak (row 2, column 1), disaggregating the records by winter and summer lays bare clear, modelable patterns. A parallel trend surfaces for electrical load and temperature (row 4, column 2). Beyond this, the kernel density estimation plots arranged along the diagonal substantiate that the marginal distributions of each variable shift with the seasonal cycle, reinforcing the imperative of accounting for temporal dependencies in both univariate and multivariate analyses.



© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Figure 8. Visualization of spatiotemporal correlation for multi-energy load data. Diagonals show each variable's marginal distribution; off-diagonals plot pairwise scatter relationships, revealing season-dependent inter-series coupling.

Figure 9 presents the autocorrelation heatmaps and ACF (Autocorrelation Function) diagrams for electrical and cooling/heating loads. The left-side panels make clear that both series exhibit well-defined periodic temporal structure in spring and summer, whereas winter shows a notably different autocorrelation signature. The ACF traces further document that electrical load autocorrelation dissipates to near zero after roughly 4–5 hours, accompanied by a marked daily rhythm. Conversely, the cooling/heating load autocorrelation decays at a substantially more languid pace. These findings affirm the vital need to embed temporal dependencies, spatial correlations, and comprehensive spatiotemporal coupling into our model design and training workflow.

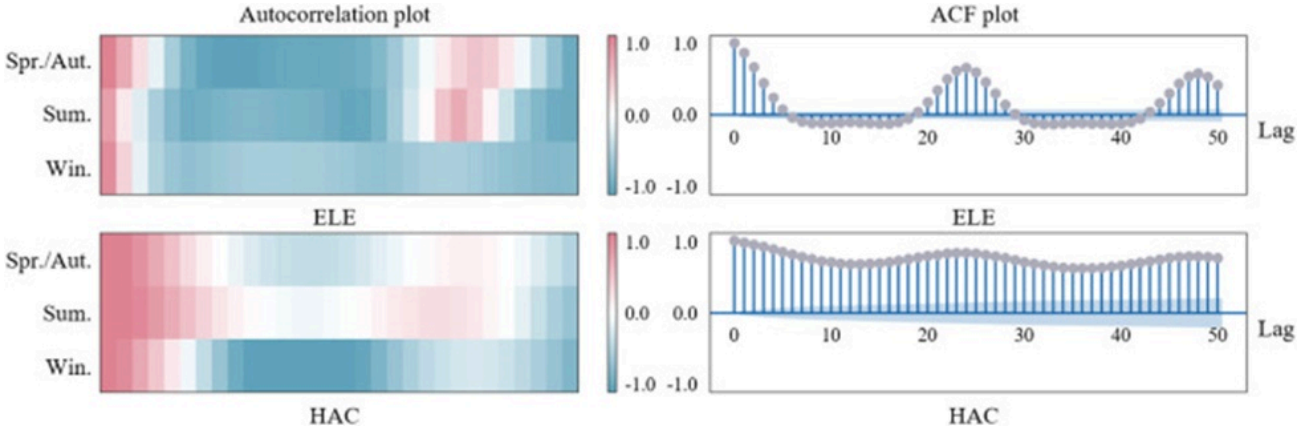


Figure 9. Temporal autocorrelation and ACF plots for multi-energy load data. Left: heatmap of lagged temporal autocorrelation. Right: ACF curves for each type of energy load, highlighting daily periodicity and decay patterns.

Figure 10 presents the label distribution for three smart-building datasets: DFAB (20% labeled), UMAR (25% labeled), and SolAce (12% labeled). Factoring in the 5% anomaly prevalence, labeled anomalies constitute just 1% of DFAB's records (20% × 5%), further exacerbating the imbalance. This combination of restricted labels and exceedingly rare anomalies prompts traditional supervised models to fixate on the majority class, habitually missing the minority (anomaly) instances.

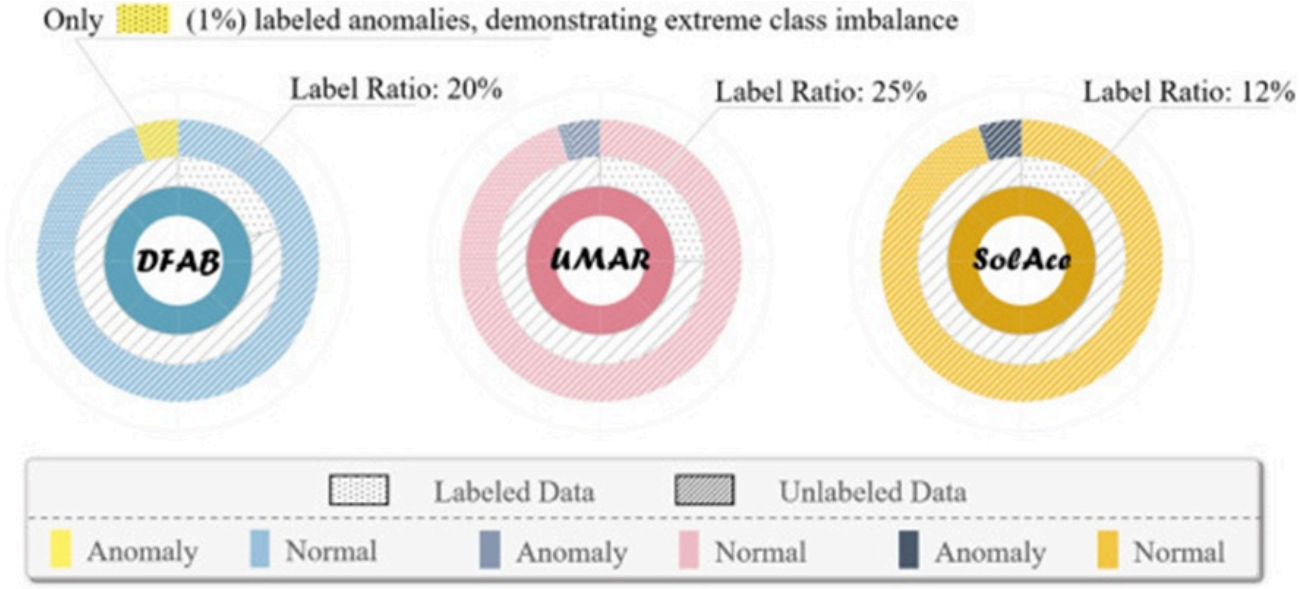


Figure 10. Label distribution and anomaly proportion of three multi-energy load datasets.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Impact of different graph construction methods

The judicious crafting of graphs forms the bedrock for attaining high-precision modeling. We juxtapose the resultant anomaly detection performance yielded by different graph construction strategies, as tabulated in **Table 5.**

Table 5. Impact of different graph construction methods.

	F1-score(%)	Recall(%)	Precision(%)
k-means_6	86.3	85.1	87.5
k-means_7	94.4	93.6	95.2
k-means_8	94.7	93.4	96.0
k-means_9	90.7	88.7	92.8
k-means_10	84.3	84.4	84.3
DBSCAN	94.1	92.3	95.9
FCM	95.6	93.5	97.8
Dy-MDS	**96.8**	**95.3**	**98.3**

The graph construction variants under evaluation include k-means clustering, DBSCAN clustering, and Fuzzy C-means clustering (FCM). The recipe for building graphs proceeds in two stages: first, the time series are clustered using the above techniques to produce associated groupings, with each time series mapped to a “node.” Subsequently, “edge” connections are forged by evaluating Pearson correlation coefficients within each cluster.

As shown in **Table 5**, the chosen graph construction method directly shapes the model’s final detection performance. In k-means clustering, the resulting detection quality depends on the number of clusters k. The model peaks when k is set to 7 or 8, plausibly because the UMAR dataset used in this case study naturally organizes into 7–8 categories—for instance, room temperature, temperature control valves, rooftop photovoltaics, and the like. DBSCAN clustering, which dispenses with the need to pre-specify the number of clusters, delivers detection results marginally inferior to those of k-means with k = 8. FCM reaches its zenith when the cluster count is fixed at 8, nudging slightly past k-means in detection accuracy. This may stem from the fuzzy membership inherent in certain feature variables, with FCM supplying richer modeling capacity to handle the ambiguity induced by indistinct feature boundaries.

Our dynamic MDS-based graph construction method clinches the top anomaly detection accuracy. The reason is that the MDS procedure can precisely gauge inter-feature relationships, culminating in a more faithful graph structure. Moreover, the production of dynamic graphs ensures that the downstream GCN module can continually follow the temporal evolution of the feature variables.

Comparison study

The anomaly detection figures for the three user-level units are displayed in **Tables 6 and 7**, which pertain to heating-cooling load and electrical load, respectively. The collection of baseline competitor models spans: the 3σ rule and boxplot derived from the IQR rule, Isolation Forest [21], Random Forest, LSTM-VAE [9], TLAN [17], GAT [18], and GDN [14]. These models are all strong contenders and collectively represent a wide spectrum of methodological traditions. Specifically:

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

• The 3σ rule and boxplot embody traditional statistical models.

• Isolation forest is a pervasively used unsupervised anomaly detection technique.

• Random forest, a classical supervised machine learning model, was likewise included in our baselines.

• LSTM-VAE, a sequential autoencoder grounded in reconstruction error, is a customary baseline in the anomaly detection literature.

• TLAN, which integrates a temporal logical attention mechanism with an adaptive thresholding strategy, can model spatiotemporal correlations.

• GAT, which merges spatial and temporal graph attention networks, is routinely adopted for modeling spatiotemporal dependencies in anomaly detection.

• GDN (Graph Deviation Network) fuses sensor embeddings, a graph attention-powered module, and deviation scoring to infer the dependency graph among multivariate time series.

Table 6. Anomaly detection results on heating and cooling load. Results are reported as mean ± std. The best score in each column is boldfaced.

	Average			**SolAce**			**UMAR**			**DFAB**	
	F1(%)	**R(%)**	**P(%)**	**F1(%)**	**R(%)**	**P(%)**	**F1(%)**	**R(%)**	**P(%)**	**F1(%)**	**R(%)**
3σ	45.4±1.5	29.5±1.3	**99.1±0.0**	56.7±1.4	39.7±1.4	**99.2±0.0**	49.8±1.6	33.3±1.4	98.2±0.1	26.7±1.7	15.4±
boxplot	55.1±3.2	39.8±3.3	91.7±0.1	59.7±3.5	42.8±3.6	98.7±0.0	44.0±3.4	29.6±3.1	86.2±0.2	61.7±2.8	46.9±
IF	69.0±2.6	65.9±4.5	73.3±0.7	76.2±2.3	75.5±4.3	76.9±0.6	58.6±3.3	50.0±4.9	70.7±0.7	72.3±2.1	72.1±
RF	80.1±1.5	69.1±2.3	95.6±0.3	76.3±1.5	65.2±2.1	92.0±0.4	87.4±1.7	78.5±2.7	98.6±0.2	76.5±1.4	63.5±
LSTM-VAE	76.3±1.3	64.6±1.7	95.9±0.3	86.0±1.0	77.0±1.6	97.2±0.3	79.6±1.1	70.0±1.7	92.3±0.4	63.3±1.7	46.7±
TLAN	81.3±1.7	70.4±2.4	97.6±0.3	82.7±1.6	71.8±2.4	97.6±0.3	90.1±1.3	83.4±2.1	98.0±0.3	71.1±2.0	56.1±
GAT	89.0±0.6	82.1±0.9	97.5±0.1	87.5±0.6	81.5±1.1	94.3±0.2	94.5±0.4	90.2±0.7	99.3±0.1	85.0±0.7	74.6±
GDN	78.4±0.9	66.9±1.2	97.4±0.2	82.4±0.8	71.0±1.2	98.1±0.1	88.4±0.5	81.9±1.0	96.1±0.2	64.3±1.3	47.9±
EGBAD	**96.8±0.3**	**95.3±0.5**	98.3±0.1	**96.4±0.3**	**94.7±0.6**	98.2±0.1	**97.4±0.3**	**95.8±0.5**	**99.0±0.0**	**96.5±0.3**	**95.3±**

Table 7. Anomaly detection results on electrical load. Results are reported as mean ± std. The best score in each column is boldfaced.

	Average			**SolAce**			**UMAR**			**DFAB**	
	F1(%)	**R(%)**	**P(%)**	**F1(%)**	**R(%)**	**P(%)**	**F1(%)**	**R(%)**	**P(%)**	**F1(%)**	**R(%**
3σ	30.2±1.6	18.0±1.1	**99.4±0.1**	36.5±1.6	22.4±1.2	**99.2±0.2**	30.4±1.7	18.0±1.2	**99.4±0.1**	23.8±1.4	13.5±
boxplot	36.4±2.2	22.3±1.6	98.9±0.1	39.8±2.3	24.9±1.8	98.8±0.1	36.4±2.0	22.3±1.5	98.2±0.1	33.1±2.2	19.8±

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

IF	72.9±1.3	82.0±3.1	68.4±0.5	57.3±0.9	85.5±3.6	43.1±0.5	84.4±1.3	84.2±2.7	84.7±0.4	76.9±1.6	76.2±
RF	83.3±1.2	72.0±1.7	99.0±0.2	83.9±1.1	73.1±1.6	98.4±0.2	87.0±1.0	77.6±1.7	98.9±0.2	79.0±1.2	65.3±
LSTM-VAE	75.5±0.8	64.3±1.1	95.3±0.1	87.7±0.7	80.7±1.2	94.6±0.1	82.4±0.6	72.6±1.0	95.1±0.1	56.3±1.2	39.7±
TLAN	80.8±1.0	70.7±1.6	95.2±0.4	81.0±1.1	71.2±1.6	94.0±0.4	87.3±0.8	80.8±1.5	94.9±0.5	74.1±1.2	60.1±
GAT	89.6±0.4	83.0±0.7	97.5±0.1	89.0±0.3	84.0±0.6	94.7±0.1	93.0±0.4	88.1±0.7	98.5±0.1	86.8±0.4	77.0±
GDN	81.0±0.5	70.3±0.8	96.3±0.2	80.3±0.5	69.4±0.8	95.2±0.1	85.8±0.5	78.1±0.8	95.3±0.2	77.0±0.6	63.3±
EGBAD	**94.5±0.3**	**93.5±0.5**	95.6±0.1	**93.6±0.2**	**94.3±0.5**	92.9±0.1	**94.0±0.4**	**92.3±0.6**	95.7±0.2	**96.0±0.3**	**93.8±**

As shown in **Table 6**, the EGBAD model introduced in this paper achieves the highest F1-score, balancing precision and recall in the context of anomaly detection. The mean precision, recall, and F1-score come to 98.3%, 95.3%, and 96.8%, respectively. The advantage of EGBAD over the runner-up, GAT, is statistically significant ($P = 0.0020 < 0.05$).

Conventional statistical techniques, such as the 3σ rule and the box plot's interquartile range rule, achieve strong anomaly-detection precision of 99.1% and 91.7%, respectively. In contrast, their recall figures for anomalies are considerably weaker than those of competing models, reaching merely 29.5% and 39.8%. This stems from their reliance solely on distributional assumptions and cutoff thresholds, which fail to accommodate the high-dimensional, intricate nature and multi-source variability of the data environments of user-level units in distribution networks.

The unsupervised Isolation Forest model achieves similar precision and recall across all datasets, suggesting that data imbalance has a limited effect on unsupervised techniques. That said, the typical unsupervised model lacks sophistication, which hampers its ability to represent the multifaceted spatial-temporal interdependencies among various feature variables. As a result, its precision, recall, and F1-score are generally lower.

Graph-based models uniformly register high anomaly detection precision, all surpassing 97.0%. Their anomaly recall also shows some advancement over other model families. This confirms that modelling the spatial-temporal interdependencies among multi-energy load sequences and other feature variables can meaningfully elevate model performance. Given the extreme skew in the anomaly dataset, most models achieve lower recall for anomalies than they do for precision. The model proposed here, built on a Boosting ensemble framework, places greater emphasis on samples that are frequently misclassified, thereby achieving an anomaly recall of 95.3%. GAT has the next-highest recall, potentially due to its node-level attention mechanism, which incorporates feature amplification and multi-head attention that operate as an ensemble, thereby strengthening its grasp of minority-sample characteristics. Notably, the EGBAD framework proposed in this work did not yield the highest precision, underscoring that managing the precision-recall trade-off remains a challenging problem in anomaly detection. Even so, EGBAD still attained the top overall F1-score on the entire dataset. **Figure 11** presents the anomaly detection outputs of the three highest-performing models on a portion of the data. Every selected model displays high precision in spotting anomalies. The proposed EGBAD model achieved 100% precision on this subset.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

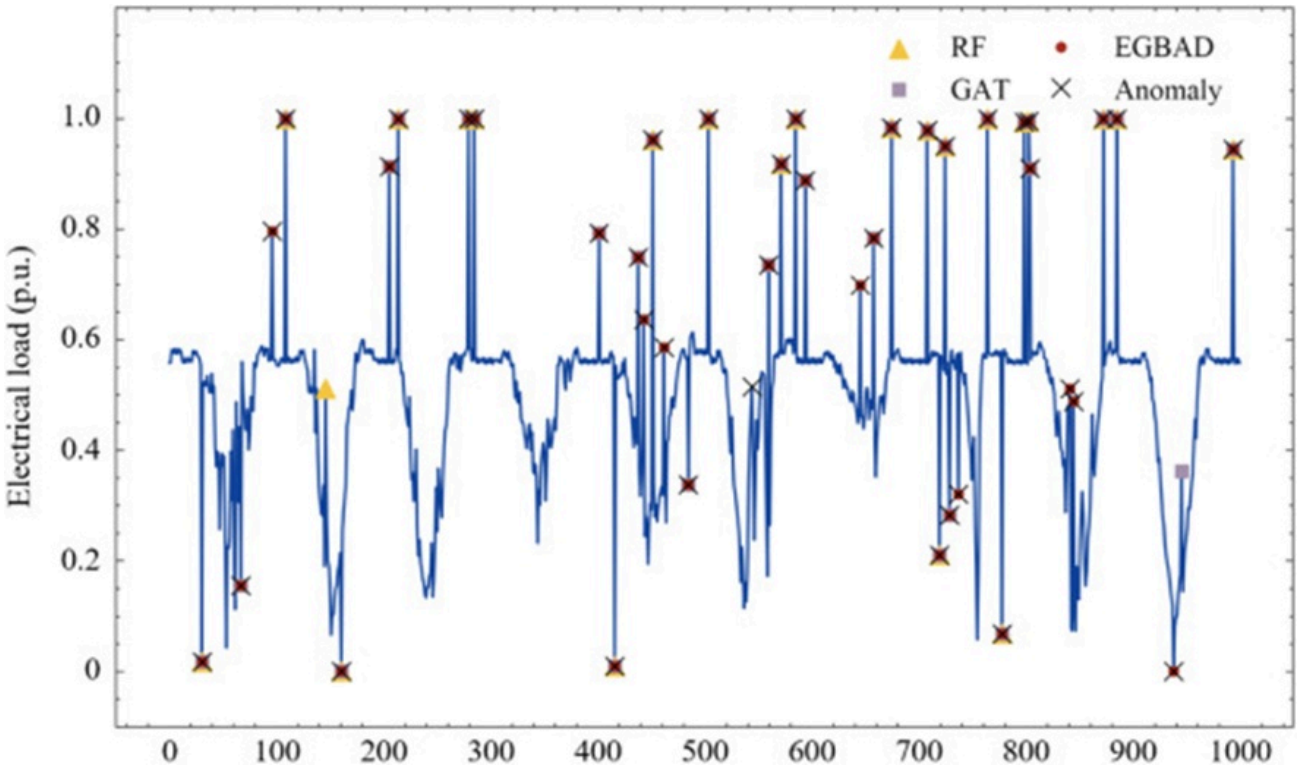


Figure 11. Subset-level anomaly detection results produced by the three most competitive models.

Table 8 reports the training and inference runtimes for each model. EGBAD completes training in 132 s on average and achieves an inference latency of 3 ms, both of which are lower than GAT (141 s and 11 ms). These figures highlight how compact yet potent EGBAD is: structurally, it only stacks two GCN layers on top of a fully connected layer. At the same time, the Boosting framework handles data imbalance head-on. Because of this lightweight profile, EGBAD is a strong candidate for practical industrial deployment.

Table 8. Average time consumption of EGBAD and other competitive baselines.

	LSTM-VAE	TLAN	GAT	EGBAD
Avg. training time	105s	186s	141s	132s
Avg. inference latency	4ms	15ms	11ms	3ms

EGBAD: Superior performance under data imbalance

To examine anomaly detection behaviour across a gradient of data imbalance and to test the stability of the proposed EGBAD model, the anomaly ratio is manipulated following the protocol. When the anomaly percentage rises, the dataset shifts closer to a balanced state. Table 9 collates model detection outcomes under these changing anomaly ratios.

Table 9. Detection performance under varying degrees of imbalance. Results are reported as mean ± std. The best results for each anomaly ratio are boldfaced.

Anomaly ratio	EGBAD			GAT			RF			IF	
	F1(%)	R(%)	P(%)	F1(%)	R(%)	P(%)	F1(%)	R(%)	P(%)	F1(%)	R(

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

1%	**92.9±0.3**	90.4±0.6	95.6±0.2	81.9±0.6	70.7±0.9	97.4±0.2	68.8±2.3	53.5±2.8	96.2±0.4	69.0±1.9	70.3
2%	**94.1±0.3**	93.0±0.6	95.2±0.2	87.2±0.5	78.5±0.8	98.1±0.1	76.9±1.9	64.3±2.5	95.8±0.4	70.7±1.8	71.8
5%	**95.6±0.2**	94.4±0.5	97.0±0.1	89.2±0.4	82.6±0.8	97.3±0.1	81.7±1.3	70.5±2.0	97.3±0.3	70.9±1.7	73.9
10%	**96.2±0.3**	94.3±0.5	98.1±0.1	91.2±0.4	85.2±0.7	98.0±0.1	85.5±0.9	76.8±1.4	96.5±0.3	75.0±1.8	74.4
15%	**96.4±0.2**	95.2±0.4	97.7±0.1	96.3±0.3	94.5±0.5	98.2±0.0	93.7±0.6	89.2±1.1	98.7±0.3	80.5±1.8	79.4
20%	95.5±0.2	94.5±0.5	96.6±0.2	**95.9±0.3**	94.4±0.5	97.5±0.1	93.8±0.5	92.1±1.1	95.6±0.3	80.2±1.9	79.9
25%	**96.3±0.2**	94.9±0.4	97.8±0.1	96.0±0.2	93.9±0.4	98.3±0.1	94.1±0.5	92.0±1.1	96.2±0.2	59.7±2.1	55.3
30%	95.8±0.2	94.6±0.4	97.1±0.1	**96.9±0.2**	95.7±0.5	98.1±0.1	93.9±0.6	92.4±1.2	95.5±0.2	60.1±1.7	59.6
Average	**95.4±0.3**	**93.9±0.5**	96.9±0.1	91.8±0.4	86.9±0.6	**97.9**±0.1	86.1±1.1	78.9±1.7	96.5±0.3	70.8±1.8	70.6

Eight experimental configurations are constructed with distinct anomaly ratios to probe how models behave as imbalance levels vary. The baselines picked for comparison cover IF, RF, and GAT. As **Table 9** makes clear, EGBAD records the best F1-score in six of the eight scenarios. In terms of averages, EGBAD leads in anomaly recall and F1-score at 93.9% and 95.4%, respectively, whereas GAT leads in anomaly precision at 97.9%. As shown in **Figure 12**, the anomaly ratio exerts a pronounced effect on the recall of supervised models. Isolation Forest, by contrast, keeps precision and recall within a narrow band, signalling that data imbalance scarcely disrupts its recall behaviour. At the same time, its overall detection accuracy is modest, with both precision and recall oscillating between 0.6 and 0.8. In imbalanced datasets, both RF and GAT exhibit a recall lagging behind their precision. This pattern reflects an optimisation bias: during training, these models allocate disproportionate attention to the head data (normal samples) to boost overall accuracy, thereby underestimating their sensitivity to the minority class (anomalies). Conversely, the EGBAD framework introduced here maintains steady precision and recall, both above 93%. Moreover, the right-hand panel of **Figure 12** confirms that even under a 1% anomaly rate—an extreme imbalance scenario—EGBAD jointly sustains high precision and high recall.

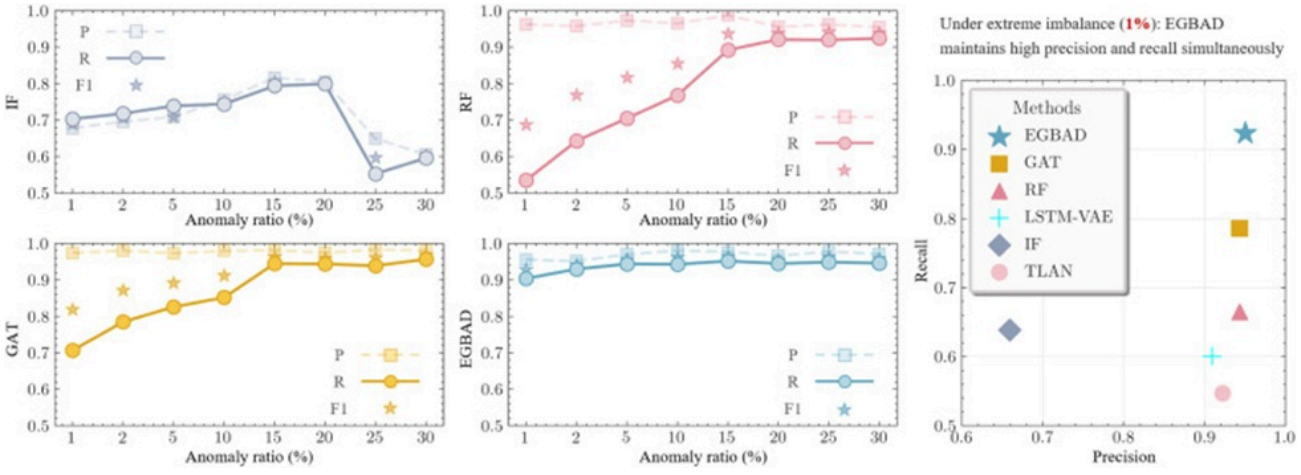


Figure 12. Performance comparison of anomaly detection methods: Evaluation metrics at various anomaly ratios and EGBAD superiority at extreme imbalance.

The polar chart in **Figure 13** contrasts model performance under a range of anomaly ratios. EGBAD plainly demonstrates resilience to skewed data distributions. When the anomaly ratio dips below 15%, making the data more imbalanced, EGBAD secures the top F1-score. As the anomaly ratio climbs and the dataset becomes more balanced, RF and GAT make substantial performance gains.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

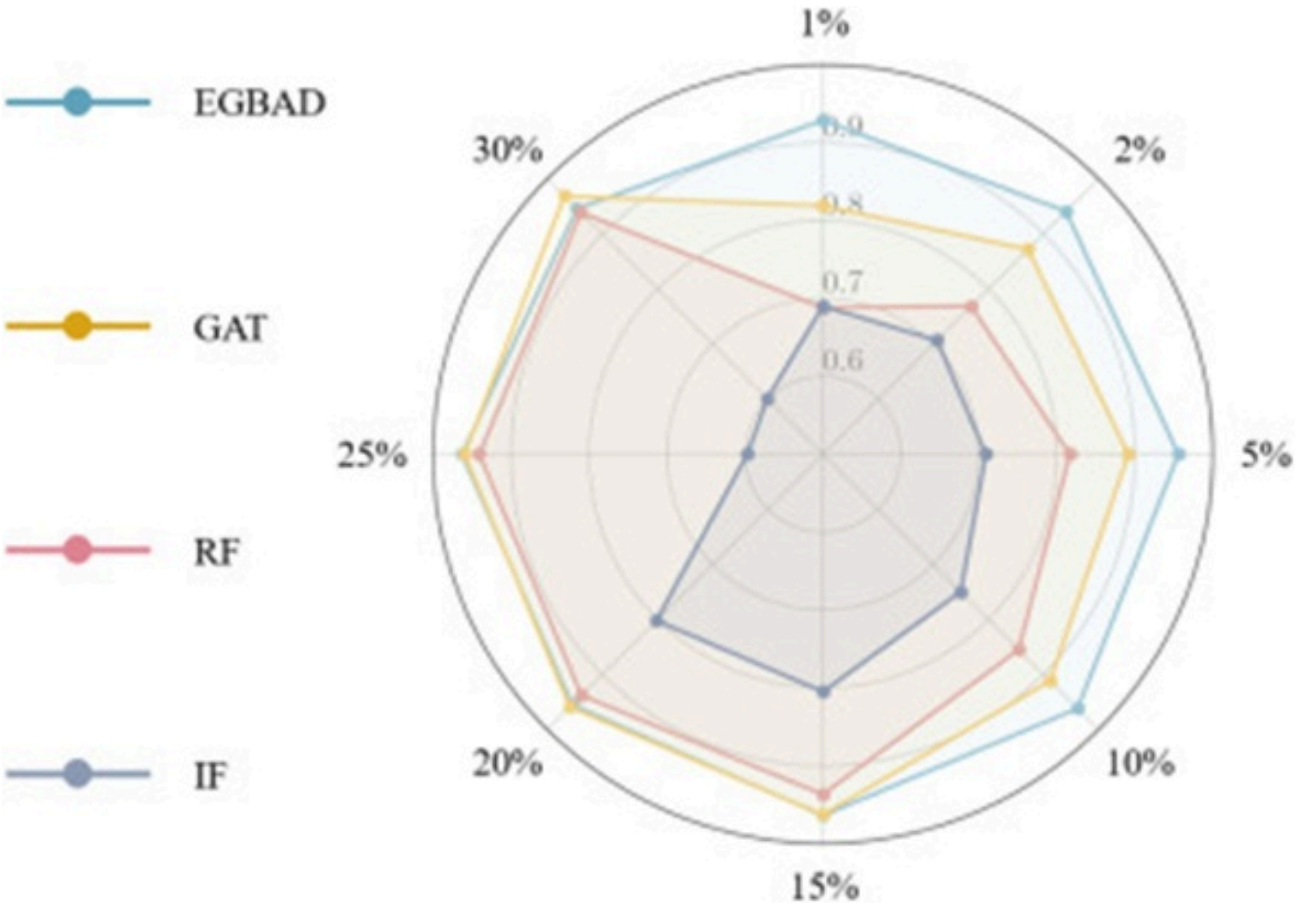


Figure 13. Radar Plot: F1-score of different models under various anomaly ratios. EGBAD maintains consistently strong performance across all ratios and exceeds 90% F1 even at the extreme 1% anomaly rate.

Ablation study

Model Architecture Ablation Study. This subsection disentangles the contributions of EGBAD's building blocks by testing several variants: removing the Boosting ensemble, switching to Adaboost, replacing the dynamic graph with a static one, and substituting GCN with GNN or GAT. The findings are compiled in **Table 10.** Dropping the Boosting ensemble inflicts a severe penalty: the F1-score sinks from 95.6% to 88.5%. This is because, without Boosting, EGBAD degrades to a plain graph-based detector and loses its ability to handle imbalanced data. Moving from a dynamic to a static graph and swapping GCN for alternative graph backbones both lower the F1-score, further corroborating the architectural decisions underpinning the model. Notably, the GAT-based configuration records an even lower F1-score than the GNN-based one, which may reflect the Boosting framework's preference for diverse weak learners as base learners. GAT, with a richer, more complex structure than GNNs, may not perform well when retrained repeatedly within the Boosting loop.

Table 10. Ablation study on model design.

	Avg	SolAce	UMAR	DFAB
EGBAD	**95.6**	**95.0**	**95.7**	**96.2**
w/o Boosting	88.5	88.1	91.6	85.7
Adaboost	91.7	91.0	92.8	91.4
Static graph	93.7	93.7	93.2	94.3

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

GCN→GNN	92.3	91.9	92.2	92.8
GCN→GAT	90.2	90.3	90.5	89.9

MDS Dimension Ablation Study. During dynamic graph construction, MDS compresses high-dimensional features into a lower-dimensional representation. Adopting a 2-dimensional mapping involves a trade-off between anomaly detection accuracy and computational efficiency. To verify this choice, an ablation study is performed by varying the MDS output dimension across d ∈ {2, 3, 4}, with the downstream F1-score (%) serving as the evaluation criterion, as reported in **Table 11.**

Table 11. Ablation study on MDS dimension.

	training time (s)	Avg	SolAce	UMAR	DFAB
d=2	132	94.5	93.6	94.0	96.0
d=3	149	94.5	93.6	94.0	96.0
d=4	176	94.5	93.5	93.9	96.1

The evidence shows that d = 2 strikes the most favourable balance: it delivers equivalent mean performance to d = 3 and d = 4 (94.5 vs. 94.5 vs. 94.5) while sidestepping the 33.3% rise in GCN training time observed at d = 4. This confirms that a 2D embedding conserves the essential pairwise structure while minimising computational overhead.

Boosting Stage Ablation Study. The Boosting ensemble framework functions as the cornerstone that allows EGBAD to tackle severe data imbalance. Within this framework, each base learner sequentially sharpens the focus on previously misclassified samples. Nevertheless, adding more stages does not automatically improve performance. In the experiments, the number of base learners is set to K = 3. **Figure 14** visualises the confusion matrices obtained with different choices of K.

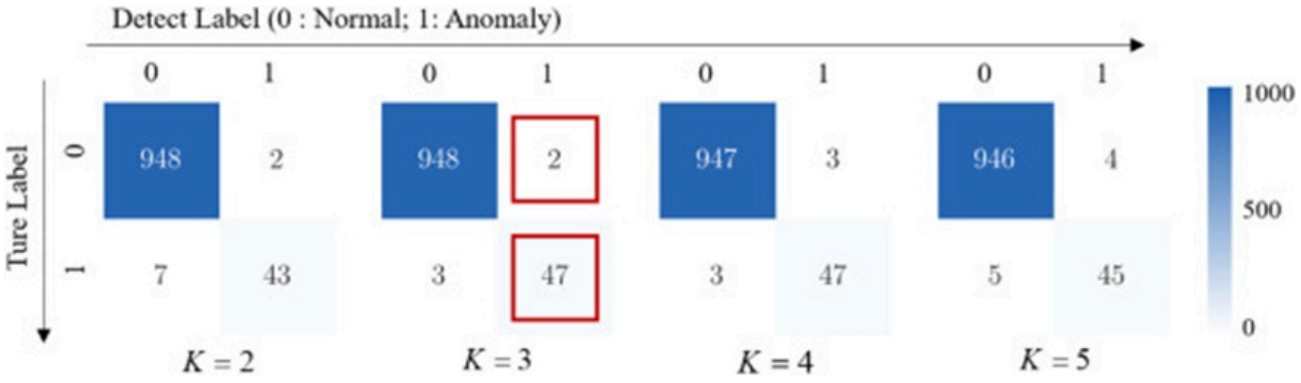


Figure 14. Anomaly detection confusion matrices under different Boosting stages. K = 3 yields the highest recall with the fewest false positives, and is therefore adopted in the main experiments as a practical trade-off.

As the figure illustrates, K = 3 detects the largest pool of anomalies (47) while generating the fewest false positives (2), thereby achieving the best precision–recall trade-off. Setting K = 2 yields comparatively weaker recall, suggesting that Boosting has not yet fully counteracted the effects of imbalance. On the other hand, larger K values lead to more false positives, indicating that repeatedly up-weighting hard samples invites overfitting. While this tendency could be curbed by adding additional cross-attention modules, we adopt K = 3 as a pragmatic middle ground between detection quality and model complexity.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Hyperparameter Analysis. **Figure 15** traces the impact of two key hyperparameters. Sliding window sizes from the set {24, 48, 72, 96, 120, 192} were evaluated, with 96 selected for the main experiments. The F1-score for anomaly detection rises steadily as the window expands, then levels off at 96; beyond this point, further enlargement yields negligible gains, a behaviour attributable to the 15-minute sampling period of the multi-energy load data. For GCN depth, increasing from 1 to 2 layers yields a meaningful boost, lifting the F1-score by 15.2%. However, deepening to 3 layers yields no additional improvement while inflating the training time by an average of 44 seconds per 10 epochs. Based on this trade-off, a depth of 2 is chosen as the most practical setting, balancing performance gains with computational cost.

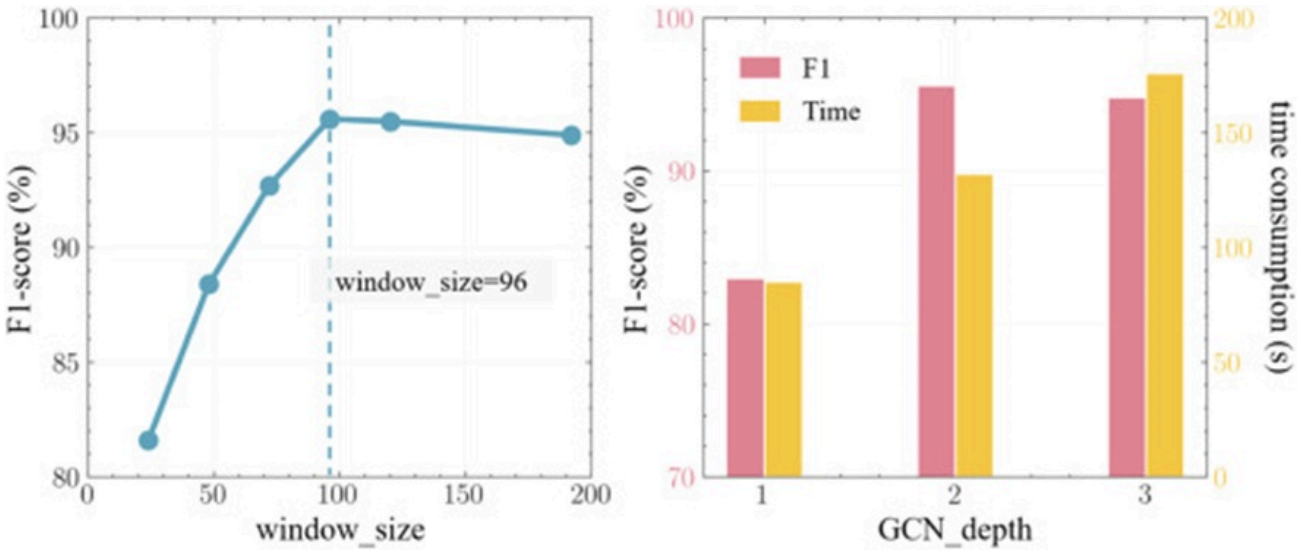
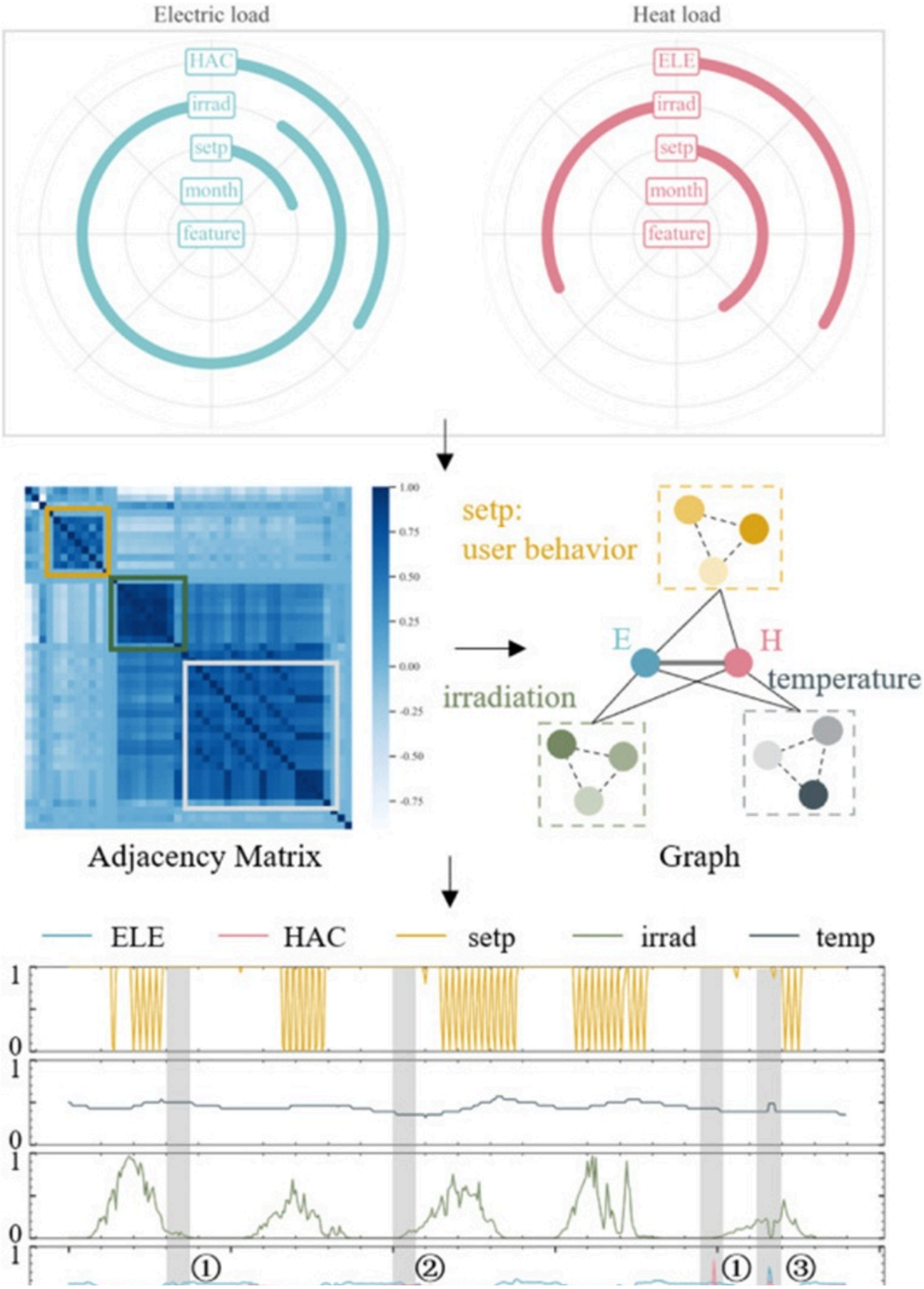


Figure 15. Sensitivity analysis of key hyperparameters. Left: Effect of sliding window size on anomaly detection F1-score. Right: Effect of GCN depth on F1-score and training time.

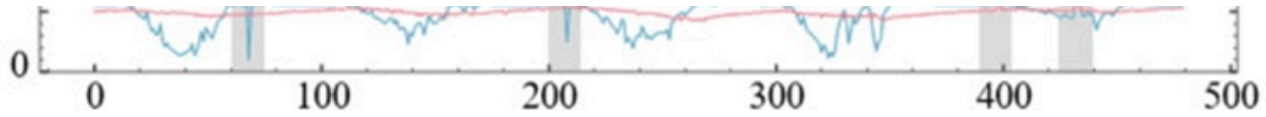
Interpretability analysis

This subsection focuses on making sense of how graph-relational learning is applied to anomaly detection in user-level multi-energy load data. **Figure 16** offers a case study centred on a typical time segment, walking through the entire sequence—from correlation inspection and graph building to anomaly flagging.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press



© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press



variables. In this context, “ELE” and “HAC” refer to electrical load and heating-cooling load, “irrad” designates the solar irradiance feature family, and “setp” designates the user behaviour feature family. Clockwise orientation signals a positive association, and counter-clockwise orientation signals a negative one.

A strong positive link exists between electrical load and heating-cooling load. This finding highlights that user-level applications such as load forecasting and anomaly detection need to incorporate spatiotemporal dependencies across different load categories. Studying how the target variable relates to various features also provides insight into how the graph is assembled.

The adjacency matrix, which encodes pairwise node correlations, highlights regions with particularly strong couplings. Within the matrix, yellow zones correspond to user behaviour features, green zones to solar irradiance features, and grey zones to temperature-related features. The “Graph” panel reveals that node pairs with stronger correlations—manifested as shorter Euclidean distances—are more likely to form edges within the graph.

The bottom part of **Figure 16** lays out the anomaly detection results for a chosen time window. Four candidate anomaly points are flagged and grouped into three types according to their distinctive traits.

1) The electrical load trace shows a steep drop in the 60–80 range, marked ①. With no notable shifts in the remaining feature variables or the heating-cooling load over the same stretch, the point is labelled as an anomaly.

2) The electrical load signal suffers a pronounced decline in the 200–220 range, indicated by ②. Other time series stay largely unperturbed around this timestamp, so the point is deemed anomalous. It is important to note that, even after this abrupt change, the sample stays within the acceptable bounds of the overall distribution. Hence, this instance typifies a false negative that alternative models would tend to miss.

3) The heating-cooling load trace records a sharp surge in the 380–400 range, marked by the second ①. This reading deviates strongly from the global distribution, while no other time series exhibits meaningful fluctuations. The point is accordingly classified as an anomaly.

4) The electrical load trace shows a pronounced upward spike in the 420–440 range, labelled as ③. At the same time, the heating-cooling load trace also edges upward, and other feature variables show noticeable activity. Given the spatiotemporal coupling among the multivariate series, the conclusion is that this sample reflects normal behaviour rather than an anomaly. Crucially, this sample represents a typical false positive—one that single-series anomaly detection would likely flag as abnormal.

Conclusion

We put forward EGBAD, an anomaly detection approach for user-level multi-energy load data that unifies graph representation learning with Boosting ensemble learning. The method features a dynamic graph construction strategy built on MDS and employs GCN to thoroughly extract the spatiotemporal coupling relationships woven into multi-energy load time series. We also devise a Boosting-based training framework that strengthens model resilience against imbalanced data. Empirical results on real-world user-level datasets confirm that EGBAD outstrips commonly adopted baseline models and achieves standout performance in severely data-imbalanced regimes. The present study mainly tackles the challenge of extreme label imbalance. In real industrial deployments, however, noisy labels could constitute another obstacle. Future work might address settings with label noise and further probe the ensemble framework’s viability in this context.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Acknowledgements

This work was supported by the National Natural Science Foundation of China (U24B6010).

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 03 Dec 2021 Revised: 15 Feb 2022 Accepted: 04 Mar 2022
Published online: 15 November 2022

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Dai R, Esmaeilbeigi R, Charkhgard H. The utilization of shared energy storage in energy systems: a comprehensive review. IEEE Trans Smart Grid. 2021;12(4):3163-74.

2. Zhuang W, Fan J, Xia M, Zhu K. A multi-scale spatial-temporal graph neural network-based method of multienergy load forecasting in integrated energy system. IEEE Trans Smart Grid. 2023;15(3):2652-66.

3. Guo Y, Li Y, Qiao X, Zhang Z, Zhou W, Mei Y, et al. BiLSTM multitask learning-based combined load forecasting considering the loads coupling relationship for multienergy system. IEEE Trans Smart Grid. 2022;13(5):3481-92. https://doi.org/10.1109/TSG.2022.3173964.

4. Roelofs CM, Gück C, Faulstich S. Transfer learning applications for autoencoder-based anomaly detection in wind turbines. Energy AI. 2024;17:100373.

5. Pydi DP, Advaith S. Attention boosted autoencoder for building energy anomaly detection. Energy AI. 2023;14:100292.

6. Galar M, Fernandez A, Barrenechea E, Bustince H, Herrera F. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Trans Syst Man Cybern C Appl Rev. 2011;42(4):463-84.

7. Tukey JW. Exploratory data analysis. Reading (MA): Addison-Wesley; 1977.

8. Deng D. Research on anomaly detection method based on DBSCAN clustering algorithm. In: 2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT). IEEE; 2020.

9. Park D, Hoshi Y, Kemp CC. A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder. IEEE Robot Autom Lett. 2018;3(3):1544-51.

10. Hundman K, Constantinou V, Laporte C, Colwell I, Soderstrom T. Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018.

11. Su Y, Zhao Y, Niu C, Liu R, Sun W, Pei D. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019.

12. Li Y, Zhang L, Lv Z, Wang W. Detecting anomalies in intelligent vehicle charging and station power supply systems with multi-head attention models. IEEE Trans Intell Transp Syst. 2021;22(1):555-64.

13. Wang Z, Jiang W, Xu J, Xu Z, Zhou A, Xu M. Grid2Vec: learning node representations of digital power systems for anomaly detection. IEEE Trans Smart Grid. 2024;15(5):5031-42.

14. Deng A, Hooi B. Graph neural network-based anomaly detection in multivariate time series. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2021.

15. Chen K, Feng M, Wirjanto TS. Multivariate time series anomaly detection via dynamic graph forecasting. arXiv [Preprint]. 2023;arXiv:2302.02051.

16. Ahmed A, Basumallik S, Gholami A, Sadanandan SK, Namaki MH, Srivastava AK, et al. Spatio-temporal deep graph network for event detection, localization, and classification in cyber-physical electric distribution system. IEEE Trans Ind Inform. 2024;20(2):2397-407.

17. Liu Y, Ren S, Wang X, Zhou M. Temporal logical attention network for log-based anomaly detection in distributed systems. Sensors (Basel). 2024;24(24):7949.

18. Zhao H, Wang Y, Duan J, Huang C, Cao D, Tong Y, et al. Multivariate time-series anomaly detection via graph attention network. In: 2020 IEEE International Conference on Data Mining (ICDM). IEEE; 2020.

19. Zhang W, Zhang C, Tsung F. GRELEN: multivariate time series anomaly detection from the perspective of graph relational learning. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI). 2022.

20. McQueen JB. Some methods of classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. 1967.

21. Liu FT, Ting KM, Zhou ZH. Isolation forest. In: 2008 Eighth IEEE International Conference on Data Mining. IEEE; 2008.

22. Wang S, Bi S, Zhang YJA. Locational detection of the false data injection attack in a smart grid: a multilabel classification approach. IEEE Internet Things J. 2020;7(9):8218-27.

23. Li D, Chen D, Jin B, Shi L, Goh J, Ng SK. MAD-GAN: multivariate anomaly detection for time series data with generative adversarial networks. In: Proceedings of the International Conference on Artificial Neural Networks. Cham: Springer; 2019.

24. Aligholian A, Mohsenian-Rad H. GraphPMU: event clustering via graph representation learning using locationally-scarce distribution-level fundamental and harmonic PMU measurements. IEEE Trans Smart Grid. 2023;14(4):2960-72.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

25. Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD). 1996.

26. Tan YF, Zhao GZ, Ooi CP, Tan WH. Leveraging interquartile range and isolation forest for abnormal power consumption prediction. In: 2024 IEEE 6th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC). IEEE; 2024.

27. Zhang Z, Li M, Lin X, Wang Y. Network-wide traffic flow estimation with insufficient volume detection and crowdsourcing data. Transp Res Part C Emerg Technol. 2020;121:102870.

28. García S, Zhang ZL, Altalhi A, Alshomrani S, Herrera F. Dynamic ensemble selection for multi-class imbalanced datasets. Inf Sci. 2018;445:22-37.

29. Heer P, Derungs C, Huber B, Bünning F, Fricker R, Stoller S, et al. Comprehensive energy demand and usage data for building automation. Sci Data. 2024;11(1):469.

30. Chen Y, Tan Y, Zhang L, Zhang B. Vulnerabilities of power system operations to load forecasting data injection attacks. In: 2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm). IEEE; 2021.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press