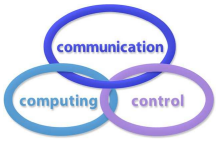
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 21, Issue: 6, Month: December, Year: 2026
Article Number: 7529, https://doi.org/10.15837/ijccc.2026.6.7529



CCC Publications



Anomaly Detection and Localization of IoT Traffic via Hybrid Autoencoder and Temporal Modeling

H. Shi*, J. Zhang

Haiou Shi*
Puyang Institute of Technology,
Henan University
Puyang, Henan, 457000, China
*Corresponding author: shihopy2025@163.com

Jie Zhang
Puyang Institute of Technology,
Henan University,
Puyang, Henan, 457000 ,China
zhj250407@163.com

Abstract

To address the problems of inadequate anomaly representation in IoT traffic and poor anomaly localization capabilities, this study proposes a novel anomaly detection and localization framework based on probabilistic autoencoder reconstruction, temporal alignment, and multi-scale temporal feature learning. The model uses a Variational Autoencoder to learn the distribution of normal traffic and performs sequence reconstruction. It combines Dynamic Time Warping to improve the identification of time-misaligned anomalies such as rhythm offset and phase drift. It then feeds anomalous windows into a Transformer-based encoding structure with inductive bias to model global temporal correlations. It uses Multi-Scale Feature Fusion decoding to recover local details step by step, and achieves joint localization of anomalous time slices, key variables, and related nodes. Experimental results indicate that with 28,800 test cases, the model achieves 98.74% of the area under the curve, 95.96% of the F1 score at the window level, 2.4 s of temporal localization errors, and 92.68% of node backtracking accuracy. In practical traffic edge environments, the recall rate of typical anomaly detection reaches up to 97.83%, with a response delay of 54 ms and an early warning lead time of 55 s. These results show that the model achieves high-accuracy detection and fine-grained localization in complex edge traffic environments, and provides support for network security monitoring and anomaly tracing.

Keywords: IoT, traffic anomaly detection, VAE; DTW, anomaly localization.

1 Background

With the rapid development of the Internet of Things (IoT), edge computing, and intelligent transportation infrastructure, edge gateways, roadside units, and various sensing terminals continuously

generate high-frequency, multi-dimensional, and heterogeneous traffic data [1]. The data represent the states of the devices and the interaction behavior with services. The data also carry risk factors, including intrusion attacks, access anomalies, network congestion, and unstable nodes [2]. Therefore, anomaly detection and localization for IoT traffic has become an important research direction to ensure the secure and stable operation of edge networks. Current studies mainly use statistical rules, traditional machine learning, and deep learning methods for anomaly identification [3, 4]. Among them, autoencoder-based methods show good potential in unsupervised anomaly detection [5]. However, previous works mainly concentrate on detecting whether anomalies are present. Insufficient attention is paid to modeling time-misaligned anomalies such as local rhythm anomalies, phase drift, and temporary anomalies. In addition, their ability to accurately pinpoint anomalous time segments, important variables, and associated nodes is rather limited. To address these issues, the Variational Autoencoder (VAE) uses probabilistic distribution learning and reconstruction to capture the latent generation patterns of normal traffic [6]. The Dynamic Time Warping (DTW) algorithm measures sequence shape differences under elastic scaling along the time axis [7]. At the same time, the Transformer strengthens local feature representation while maintaining global dependency modeling ability. Multi-Scale Feature Fusion (MSFF) helps recover anomaly boundary details and multi-dimensional correlations [8]. However, using these methods alone is still insufficient to simultaneously address the issues of time misalignment anomaly identification, abnormal local boundary recovery, and node-level location transmission. Therefore, this study develops an IoT traffic anomaly detection and localization method that integrates VAE, DTW temporal regularization, and the Transformer architecture. This method designs VAE and DTW as the front end for anomaly window identification, and constructs an anomaly score by jointly combining the probability reconstruction error and the dynamic regularization distance to enhance sensitivity to rhythm deviations and phase drift anomalies. The fusion of the Inductive Bias Transformer (IB-Transf) and MSFF serves as the back end for anomaly localization, utilizing convolutional prior embedding, global temporal correlation modeling, and multi-scale decoding to restore the anomalous response map. Subsequently, by utilizing the anomalous response map and the device-variable mapping relationship, the time-and-variable-level anomalous results are further transmitted to the associated nodes, achieving a seamless transformation from detection outcomes to traceability metrics. This study aims to improve the sensitivity, localization accuracy, and interpretability of IoT edge traffic anomaly detection. It is expected to provide technical support for real-time security monitoring and anomaly traceability in complex network environments. The innovations and contributions of this study are as follows: (1) Constructing an analysis framework for "anomaly window detection, anomalous hot zone localization, and associated node backtracking" in IoT traffic. (2) Propose a joint anomaly discrimination method that integrates VAE probability reconstruction and DTW temporal regularization to enhance the ability to identify time-misaligned anomalies. (3) Introduce the IB-Transf and MSFF mechanisms to combine global temporal correlation modeling, local abnormal boundary recovery, and node mapping, achieving fine-grained collaborative location of anomaly time slices, key variables, and potential nodes.

2 Related studies

Currently, the research on IoT anomaly detection is gradually shifting from traditional rules and shallow machine learning methods to deep temporal modeling and edge-side intelligent analysis. Jovanović et al. [9] proposed a programmable flow-based system for IoT botnet detection by modeling the internal traffic behavior of IoT networks, thereby achieving effective botnet detection and analysis. Alsoufi et al. [10] constructed a deep-learning-based anomaly detection model to enhance attack recognition capabilities within complex network traffic. The above studies show that IoT traffic anomaly detection has shifted from simple traffic statistics-based judgment to deep representation learning for complex network behaviors, but its main goal still focuses on improving the accuracy of anomaly classification, with insufficient explanations of the causes and locations of anomalies.

In the context of unsupervised and weakly supervised anomaly detection, deep reconstruction models are widely applied in industrial IoT and edge security scenarios. Shuxin et al. [11] addressed the problem of satellite network intrusion detection by jointly modeling a memory network and a

sparse autoencoder, achieving temporal feature extraction and reconstructive identification of network intrusions. Pang J L [12] proposed an adaptive fault prediction method based on deep learning for production line fault prediction and maintenance, establishing a connection between the evolution of production line anomalies and maintenance responses through a deep model. Overall, these methods offer significant advantages in feature learning, reduced label dependence, and complex pattern recognition. However, they still generally focus on anomaly classification or fault prediction, providing an insufficient characterization of time-shifted and multi-dimensional collaborative imbalance anomalies.

In terms of temporal modeling, Hamdi et al. [13] addressed the problem of dynamic behavior recognition in detection tasks by proposing a distributed deep recurrent neural network method, which achieved temporal pattern recognition through the deployment of a recurrent structure. Liang et al. [14] investigated the multi-classification problem of smart meter faults, proposing an ensemble-learning-based multi-classification method that achieved fault state recognition for sensing devices through the fusion of multiple discriminative strategies. Compared to recurrent neural networks, the Transformer can better capture long-range dependencies and cross-variable relationships. However, the standard self-attention structure lacks local inductive biases, which can obscure anomaly boundaries and short-term burst details. To address this issue, the research introduced convolution priors before Transformer encoding and combined multi-level feature extraction and skip fusion to enhance the model's expression of local abnormal patterns while retaining global modeling capabilities. In addition to detection accuracy, edge monitoring systems also need to have deployability and interpretability. Ribeiro et al. [15] addressed the limitations of model transparency and reliability by introducing federated learning and explainable techniques, thereby enhancing the credibility of intelligent models in practical applications. The above studies collectively indicate that IoT edge traffic anomaly detection not only requires high recognition accuracy but also needs to provide explanations such as triggering variables, anomaly locations, and associated nodes for anomaly triggers.

In summary, although existing research has made certain progress in IoT monitoring, intrusion detection, deep temporal modeling, digital twin monitoring, and explainable learning, there still exists insufficient characterization of temporal misalignment anomalies. There also exist problems such as difficulty in coordinating global association modeling with local anomaly boundary recovery, as well as weak joint localization capability for anomaly time slices, key variables, and potential nodes. Therefore, this study proposes an IoT traffic anomaly detection and localization model via hybrid VAE and temporal modeling. The process of detection involves the fusion of VAE and DTW techniques to attain anomaly window detection. Localization involves combining the encoding structure of IB-Transf with MSFF to enable the localization of time, variables, and nodes associated with anomalies, which enhances the practicality and interpretability of IoT edge traffic anomaly detection and localization.

3 Methodology

3.1 VAE and DTW integrated anomaly detection model

In IoT edge traffic monitoring scenarios, anomalous traffic does not always appear as continuously increasing or decreasing amplitude. In many cases, it appears as local peak shifts, stretched fluctuation rhythms, misaligned burst phases, and imbalance in multi-variable correlations [16, 17]. Because the reconstruction objective relies solely on the point-wise error, the model frequently overlooks temporal misalignment patterns. This decreases the effectiveness of the early weak anomaly identification. Therefore, this study utilizes a VAE to characterize the probabilistic distribution of traffic flows. Furthermore, the DTW algorithm is applied to elastically align the original and reconstructed sequences. In this way, it builds an anomaly detection model that considers both amplitude deviation and temporal shape difference. The overall structure of VAE is shown in Figure 1.

As shown in Figure 1, the model takes multi-dimensional traffic sequences segmented by a sliding window as input. The t -th traffic window is denoted as X_t. Its length is T , and the feature dimension is C. The encoder first applies nonlinear mapping to the window sequence and obtains the mean and variance parameters of the latent distribution. It then uses the reparameterization mechanism to sample latent variables. The decoder outputs the corresponding reconstructed sequence $\widehat{X}_t$. This

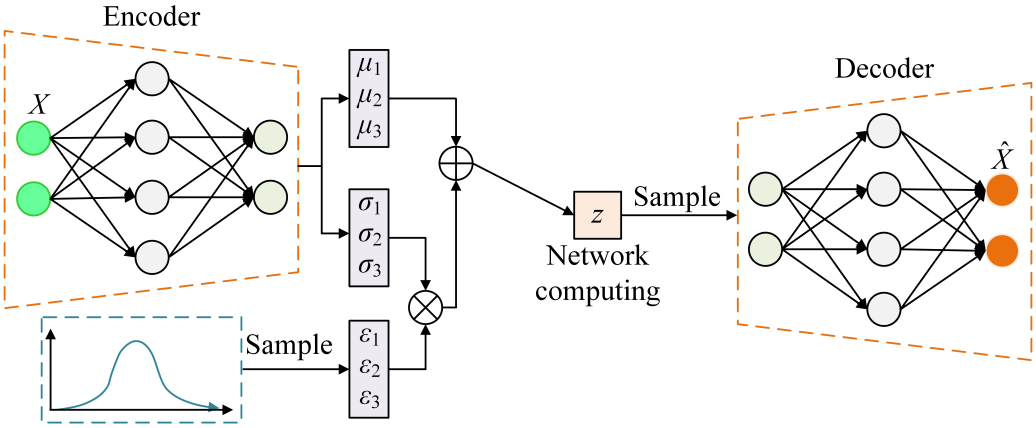


Figure 1: VAE structure diagram.

process is shown in Equation (1).

$$\begin{cases} X_t \in \mathbb{R}^{T \times C} \\ \mu, \log \sigma^2 = f_\phi(X_t) \\ z = \mu + \sigma \odot \varepsilon, \varepsilon \sim N(0,1) \\ \widehat{X}_t = g_\phi(z) \end{cases} \tag{1}$$

In Equation (1), ϕ and $f_\phi(\cdot)$ represent the encoder parameter set and mapping function. They extract latent distribution features from the original IoT traffic window. μ denotes the mean vector of the latent distribution, and σ denotes the standard deviation vector. z denotes the latent variable sampled from the latent distribution, and it represents the latent behavior pattern of the traffic sequence. ε denotes a random noise vector that follows a standard normal distribution. $N(0,1)$ denotes a multivariate Gaussian distribution with mean 0 and covariance matrix 1. θ and $g_\theta(\cdot)$ represent the decoder parameter set and mapping function. Consequently, the network learns the generative mechanism of normal IoT traffic and produces a significant reconstruction deviation when abnormal access occurs. Furthermore, this work utilizes joint optimization of the reconstruction loss and Kullback-Leibler divergence to train the VAE. The objective function is defined in Equation (2).

$$L_{VAE} = \left\| X_t - \widehat{X}_t \right\|_2^2 + \beta D_{KL}(q(z \mid X_t) \parallel p(z)) \tag{2}$$

In Equation (2), L_{VAE} denotes the total loss function during VAE training. $D_{KL}(\cdot)$ denotes the Kullback-Leibler divergence. $q(z \mid X_t)$ denotes the approximate posterior distribution of latent variable z under the condition of IoT traffic window X_t.$p(z)$ denotes the prior distribution of latent variable z.β denotes the weight coefficient of the KL divergence term. However, reconstruction error from VAE mainly reflects point-wise amplitude deviation. It shows limited sensitivity to temporal anomalies such as local stretching, phase drift, and burst misalignment [18]. In port scanning, probing at intervals, and congestion generation, abnormal traffic may exhibit a similar overall pattern but with different timing locally. One-to-one error calculation may delay the detection of abnormal traffic [19, 20]. Therefore, this study further introduces the DTW algorithm to perform temporal elastic matching between the original and reconstructed sequences. The process is shown in Figure 2.

As shown in Figure 2, DTW first receives two time series for comparison, namely the original traffic window $X_t = \{x_1, \ldots, x_T\}$ and the reconstructed traffic window $\widehat{X}_t = \{\widehat{x}_1, \ldots, \widehat{x}_T\}$. It then constructs a local distance matrix between them. The local distance between any two time points is defined in Equation (3) [21].

$$d(i,j) = \left\| x_i - \widehat{x}_j \right\|_2 \tag{3}$$

In Equation (3), $d(i,j)$ denotes the local difference between the i -th time step of the original traffic sequence and the j -th time step of the reconstructed sequence. x_i denotes the observed vector of the original IoT traffic window at the i -th time step. It contains multi-dimensional traffic features

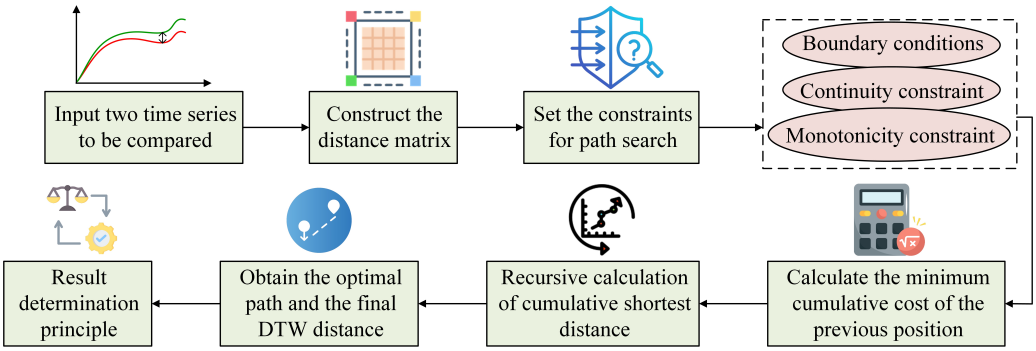


Figure 2: DTW algorithm process diagram.

such as packet size, transmission rate, and connection state. $\hat{x}_j$ denotes the estimated vector of the reconstructed traffic window at the j -th time step. After obtaining the distance matrix, DTW searches for the optimal matching path under constraints and recursively computes the cumulative alignment distance. In addition to the VAE reconstruction loss and DTW distance, this study integrates these two types of errors into a window-level anomaly score. The anomaly score not only reflects the amplitude difference but also takes the time series shape into account. The anomaly score is defined in Equation (4).

$$S_t = \alpha \left\| X_t - \hat{X}_t \right\|_2^2 + (1-\alpha)\, D_{DTW}\left(X_t, \hat{X}_t\right) \tag{4}$$

In Equation (4), α denotes the fusion weight of the reconstruction error term, and $D_{DTW}\left(X_t, \hat{X}_t\right)$ denotes the final DTW distance. To avoid the reliance on the test set for threshold selection, the study only uses the validation set to determine the anomaly discrimination threshold. Specifically, the anomaly scores of each window are calculated on the validation set, and the 95th percentile of the normal validation window score distribution is taken as the initial threshold τ. Then, the grid search for the fusion weight α is conducted based on the F1 score of the validation set. The search range for α is set from 0.1 to 0.9 with a step size of 0.1 . Finally, the α and τ corresponding to the highest window-level F1 score in the validation set are selected as the fixed parameters for the testing stage. This setting ensures that the anomaly detection threshold is derived from data outside the training process and avoids the involvement of test information in model parameter tuning. For IoT traffic anomaly detection, the aforementioned scoring method captures not only amplitude-based anomalies but also timing pattern anomalies, including rhythm shifts and delay variations. This enhances the capacity of detecting complicated anomalies. Finally, based on the above analysis, this study connects the VAE reconstruction network with the DTW alignment algorithm to form a complete anomaly detection process. The process is shown in Figure 3.

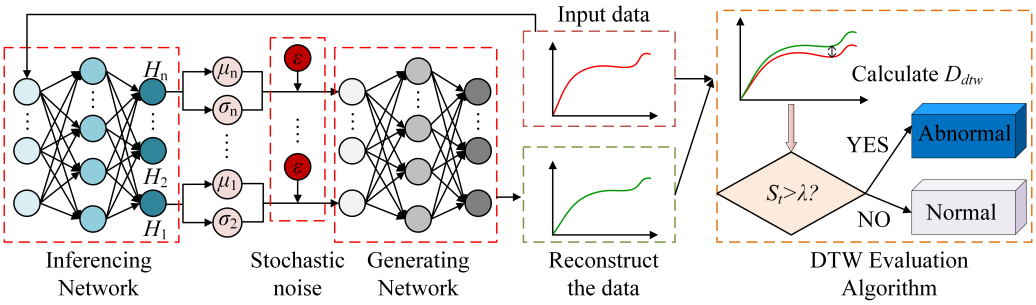


Figure 3: VAE-DTW algorithm diagram.

As shown in Figure 3, the input traffic sequence goes through the inference network, where latent distribution parameters are estimated and reparameterized sampling is performed to generate latent variables. Afterward, the traffic sequence is reconstructed using the generation network. Finally, the input and output sequences are passed to the DTW module for computing the alignment. This

produces a dynamic distance and constructs the anomaly score. Finally, the model uses the combined anomaly score to complete window-level screening and outputs suspected anomalous traffic windows to the subsequent localization module.

3.2 Transformer-based temporal modeling for anomaly localization model

Although the VAE-DTW model achieves anomaly window detection, simply determining "whether it is abnormal" is not sufficient for fast IoT traffic response. It still needs to clarify the occurrence time, triggering features, and related nodes of anomalies [22, 23]. Therefore, this study will propose a localization approach for anomaly detection, which moves the analysis from the raw data space into the feature space, as well as fuses global and local correlations for accurate localization. On this basis, this study adopts the IB-Transf encoding structure. It embeds convolutional priors into self-attention to improve the accuracy and stability of anomaly localization. Its encoding structure is shown in Figure 4.

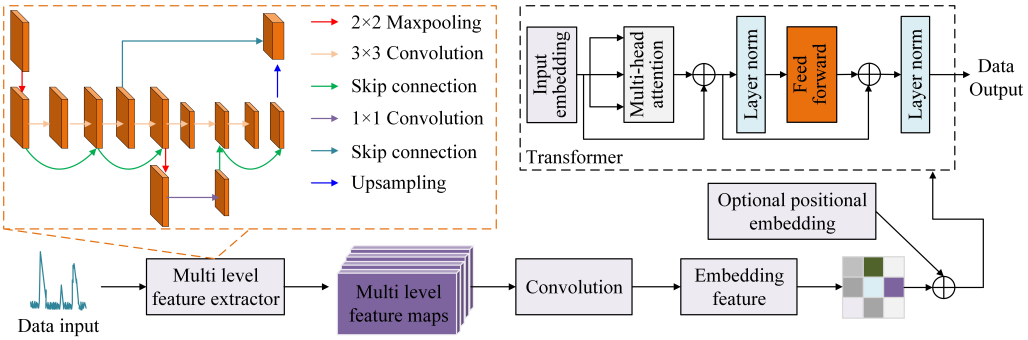


Figure 4: IB-Transf encoding structure diagram.

As illustrated in Figure 4, in terms of suspicious anomaly window detected by the detector module, initially, a hierarchical convolutional network is adopted to extract temporal feature from the input features at various levels. Skip connections have been used in order to maintain the shallow level of feature information as well. Taking into account the temporal pattern of IoT traffic anomaly behavior, hierarchical convolutional feature extraction was conducted on the input features [24]. It then uses upsampling and concatenation to build a unified-scale fused feature map Z. The process is shown in Equation (5).

$$Y_2 = \mathrm{Conv}_2(F), Y_3 = \mathrm{Conv}_3(Y_2), Y_4 = \mathrm{Conv}_4(Y_3), Z = Y_2 \oplus UP(Y_3) \oplus UP(Y_4) \tag{5}$$

In Equation (5), F denotes the input anomaly traffic feature map. Y_2, Y_3, and Y_4 represent multi-scale feature maps obtained from the second, third, and fourth convolution layers. Conv $(\cdot)$ and $UP(\cdot)$ denote convolution feature extraction and upsampling operations. Next, this study uses a convolution layer to perform patch splitting and embedding mapping on the fused feature map. This builds an input representation with inductive bias. The process is shown in Equation (6).

$$P = \mathrm{MaxPool}(\mathrm{ReLU}(\mathrm{Conv1}D(Z))) \tag{6}$$

In Equation (6), P denotes the patch embedding sequence fed into the Transformer encoder. Conv1D $(\cdot)$ denotes a one-dimensional convolution operation along the time axis. ReLU $(\cdot)$ denotes a linear activation function. MaxPool $(\cdot)$ denotes a max pooling operation. After obtaining patch embeddings, this study further uses the Transformer encoder to model global temporal correlations in anomaly windows. The process is shown in Equation (7).

$$P_l' = \mathrm{MSA}(\mathrm{LN}(P_{l-1})) + P_{l-1}, P_l = \mathrm{MLP}(\mathrm{LN}(P_l')) + P_l' \tag{7}$$

In Equation (7), P_{l-1} denotes the input feature sequence of the $l-1$ -th Transformer layer. P_l' denotes intermediate features. P_l denotes the output feature sequence of the l -th encoder layer. $LN(\cdot)$ denotes layer normalization. $MSA(\cdot)$ denotes the multi-head self-attention module. $MLP(\cdot)$ denotes

the feed-forward network of the multilayer perceptron. While the Transformer encoder can effectively extract global correlation features from anomaly segments, its direct application to localization leads to blurry temporal edges and discontinuous anomaly segments [25]. Therefore, this study further introduces the MSFF mechanism in the decoding stage. It combines IB-Transf decoding to construct a multi-dimensional time series anomaly localization module (IB-Transf-MSFF). It restores local context and temporal context information at different scales layer by layer to improve the resolution of anomaly region representation. The IB-Transf-MSFF process is shown in Figure 5.

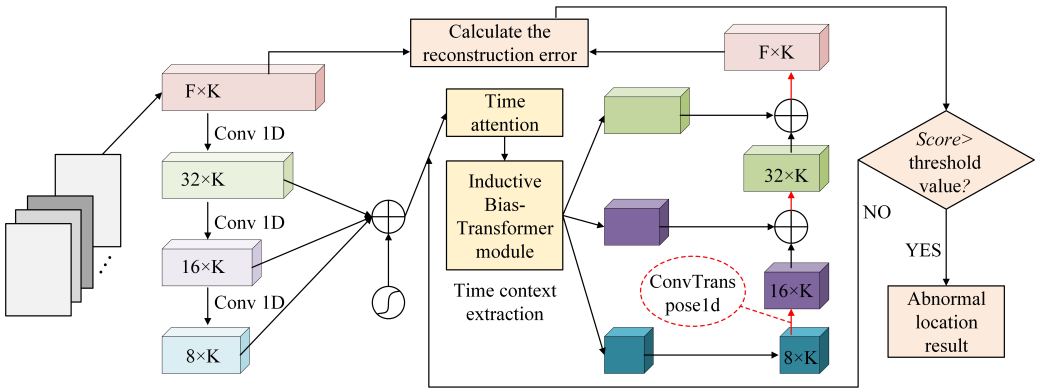


Figure 5: IB-Transf-MSFF process diagram.

As shown in Figure 5, the model feeds encoded features into a multi-scale decoding network. It uses one-dimensional deconvolution upsampling to gradually restore temporal resolution. It also fuses preserved local features and temporal features at different scales. In this process, deconvolution carries out up-sampling for each layer. It further makes use of skip connections with local feature maps at various scales to reconstruct fine temporal boundary information about anomaly traffic. This can be represented using Equation (8).

$$U_8 = P_l, U_{16} = \text{Deconv}\left(U_8\right) \oplus F_{16}, U_{32} = \text{Deconv}\left(U_{16}\right) \oplus F_{32}, \widehat{X} = \text{Conv}_{\text{ out}}\ \left(U_{32}\right) \tag{8}$$

In Equation (8), U_8 denotes the low-resolution decoding feature initialized from Transformer output features. U_{16} and U_{32} denote multi-granularity decoding features obtained through deconvolution upsampling. Deconv $(\cdot)$ denotes the deconvolution upsampling operation. Conv $_{\text{out}}$ $(\cdot)$ denotes the output convolution mapping layer. Finally, the model computes the anomalous response map using the difference between the reconstruction output and the input data. It determines the final anomalies using localization scores. This is illustrated in Equation (9).

$$R_{t,c} = \left|X_{t,c} - \widehat{X}_{t,c}\right|, S_{loc} = \sum_{t=1}^{T}\sum_{c=1}^{C} R_{t,c}, \Omega = \{(t,c) \mid R_{t,c} > \tau_{loc}, S_{loc} > \tau_s\} \tag{9}$$

In Equation (9), $X_{t,c}$ and $\widehat{X}_{t,c}$ denote the real observation value and the reconstructed estimation value. $R_{t,c}$ denotes the reconstruction error at the t -th time step and the c -th traffic feature. S_{loc} denotes the window-level anomaly localization score. τ_s denotes the window-level anomaly localization threshold. τ_{loc} denotes the position-level anomaly response threshold. Ω denotes the final anomaly location set. τ_{loc} is determined by the 97.5% percentile of all time-feature residuals in the normal verification window, and is used to screen out the location points with strong abnormal responses. τ_s is determined by the 95% percentile of the cumulative residual distribution in the normal verification window, and is used to determine whether the current abnormal window has a stable positioning result. The node backtracking process is completed based on the device-variable mapping table. For the kth abnormal window, the model first aggregates the residual responses $R_{t,c}$ along the time dimension to obtain the variable-level abnormal score. Variables with higher scores are regarded as abnormal triggering variables. Subsequently, based on the five-tuple mapping relationships involving device numbers, gateway interfaces, communication links, and traffic characteristics, these anomalous variables are mapped to their corresponding roadside units, video perception nodes, signal controllers, or edge

gateways. If a single node is associated with multiple high-response variables, the normalized average anomalous score of these variables serves as the node backtracking score, and the system outputs the set of devices whose scores exceed the designated node threshold. Based on the above detection and localization mechanisms, this study builds an IoT traffic anomaly detection and localization model that integrates VAE, DTW temporal alignment, and IB-Transf-MSFF, named VAE-DTW-IB-Transf-MSFF, referred to as VDITM. Its overall process is shown in Figure 6.

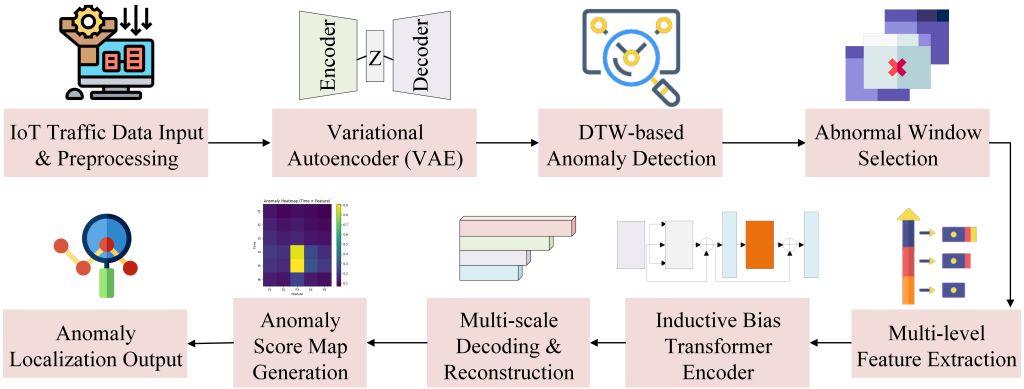


Figure 6: VDITM process diagram.

As shown in Figure 6, the model first preprocesses and windows the raw IoT traffic. It then uses VAE to learn the normal distribution and reconstruction mapping. Next, it uses DTW to evaluate alignment between the original sequence and the reconstructed sequence, achieving anomaly window screening. For traffic segments identified as anomalous, the model further extracts local anomaly features at multiple scales. It then feeds them into the IB-Transf encoder for global temporal correlation modeling. Following that, the decoder produces the reconstruction estimation through combining the information of the global correlation features with the local reconstruction details. The anomaly detection response map is produced by comparing it with the input image, and ultimately produces the anomaly localization results.

4 Experimental results

4.1 Performance evaluation of the VDITM model

To validate the effectiveness of the proposed VDITM model in IoT edge traffic anomaly detection and localization tasks, this study compared it with the Variable Temporal Transformer (VTT) model, the Variational Spatial-Temporal Anomaly Detection (VSAD) model, and the Heterogeneity-aware Integrated Global-local channel Dependence (HIGD) model. The experimental data were derived from the combined data of measured traffic and anomalous playback from the edge gateway of a smart transportation demonstration corridor. The monitored entities included roadside units, video terminals, signal controllers, and edge computing nodes. The data were collected through the synchronous acquisition of the edge gateway's mirror port and traffic logs, with a sampling granularity of 1 s. The extracted traffic features include packet length statistics, uplink and downlink byte rates, connection numbers, port access frequencies, protocol proportions, session durations, arrival intervals, packet loss rates, and communication delays, totaling 18 dimensions.

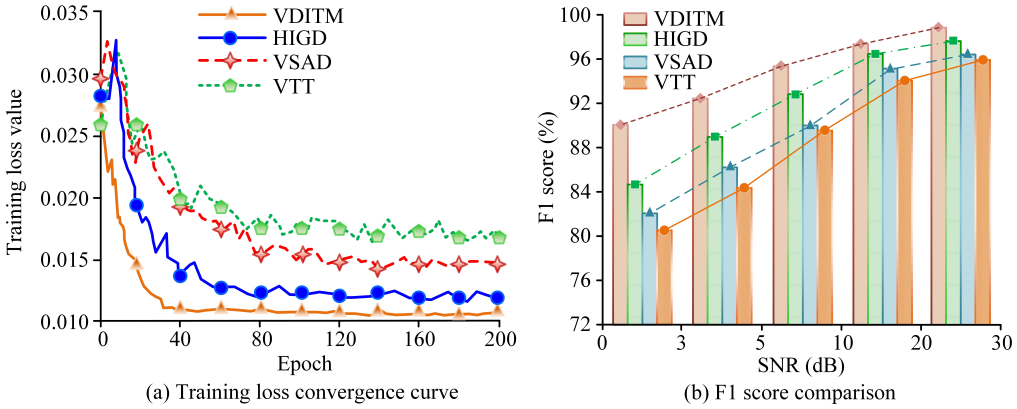
The abnormal samples are constructed using the "measured abnormal records + controlled abnormal playback" method. Port scanning and low-speed denial-of-service are generated by isolation network scripts, while the remaining abnormalities are reconstructed based on operation logs and perturbed. The time, variables, and node labels are determined by the start and end times of the anomalies, the affected features, and the mapping relationship of equipment-link-quintuple. After constructing 60-step temporal samples via a sliding window, this study obtained a total of 28,800 windows, comprising 76.4% normal samples and 23.6% anomalous samples. The study divided the training set, validation set, and test set into continuous time segments, and independently generated window samples in each subset

Table 1: Experimental environment and configuration

Category	Item	Configuration
Data	Data source, sampling resolution	Measured IoT traffic with anomaly replay data, 1 s/sample
	Window length, feature dimension	60 steps, 18 features
	Total samples, dataset split	28,800 windows, Train/Validation/Test = 7:1:2
Hardware	Processor, memory	Intel Core i9-13900K, 64 GB RAM
	Graphics processing unit, edge device	NVIDIA RTX 4090, NVIDIA Jetson AGX Orin 64 GB
Software	Operating development environment	Windows 11 64-bit, Python 3.10, PyTorch 2.1, CUDA 12.1
Training	Batch size, initial learning rate, maximum epochs	128,1×10^4,100

with a ratio of 7:1:2. Different data sets do not share the same abnormal event segments, playback templates, or adjacent overlapping windows, and a buffer zone of one window length is removed at the boundary to avoid information leakage due to time overlap between adjacent windows. To evaluate the model's robustness, the study superimposed different intensities of Gaussian noise and sudden drift interference during the testing phase, with a signal-to-noise ratio ranging from 30 dB to 3 dB. The main experimental environment and hardware configuration are shown in Table 1.

Under the unified experimental environment shown in Table 1, this study first compared the convergence speed and anti-interference capability of the four models in anomaly detection tasks. The results are shown in Figure 7.



(a) Training loss convergence curve (b) F1 score comparison

Figure 7: Comparison of convergence stability and anti-interference performance.

As illustrated in Figure 7(a), VDITM converged rapidly to 0.0116 after the 27th epoch regarding its normalized reconstruction loss. The convergence speed of VDITM was significantly faster than those of VTT, VSAD, and HIGD. Therefore, the combination of reconstruction modeling, temporal warping, and feature representation can efficiently eliminate optimization oscillations in complicated traffic environments. As shown in Figure 7(b), under strong noise conditions (SNR=3 dB), the F1 scores of VDITM, HIGD, VTT, and VSAD were 90.58%, 85.41%, 82.76%, and 80.94% respectively. Compared with VTT and VSAD, the F1 score of VDITM was still 7.82% and 9.64% higher under strong interference environments. The results indicate that this model maintained strong robustness in scenarios with superimposed noise and disturbances, and could stably identify anomaly patterns where temporal misalignment and local mutations coexist. Subsequently, to further analyze the contribution of each module to performance improvement, this study conducted ablation experiments on VDITM. The results are shown in Table 2. localization error on the time aspect was reduced to 3.7 seconds, while the variable localization F1 score increased to 84.31%. This showed that the use of the Transformer with inductive bias had clear strengths when it came to modeling global temporal dependencies and correlations between features. With the introduction of MSFF, even more improvements were realized. The complete model ultimately achieved the best results across all five metrics, which validated the synergistic enhancement effect among the modules. Finally, this study further selected Long Short-

Table 2: Ablation experiment results

Model	Anomaly detection AUC (%)	Window-level F1 (%)	Time localization error (s)	Variable-level F1 (%)	Node traceback accuracy (%)
VAE only	94.23	90.31	6.8	74.12	78.65
VAE+DTW	96.48	92.87	5.1	79.46	82.34
VAE+DTW+IB-Transf	97.62	94.18	3.7	84.31	87.92
VAE+DTW+MSFF	97.11	93.74	3.3	85.67	88.41
VDITM	98.74	95.96	2.4	89.53	92.68

Table 3: Comparison with Standard Baseline Models

Model	Anomaly Detection AUC / %	Window-level F1 / %	Recall / %
OC-SVM	86.42	80.36	78.91
IF	88.75	82.14	81.67
GRU-AE	91.83	86.52	85.74
LSTM-AE	92.46	87.31	86.58
Trans-AE	95.28	91.74	90.36
VDITM	98.74	95.96	94.83

Term Memory Autoencoder (LSTM-AE), Gated Recurrent Unit Autoencoder (GRU-AE), Transformer Autoencoder (Trans-AE), Isolation Forest (IF), and One-Class Support Vector Machine (OC-SVM) for comparison with the proposed model. All deep learning models were evaluated using the same training, validation, and test sets. the input window length was uniformly set to 60 steps, the feature dimension to 18, the maximum training epochs to 100, and the batch size to 128, while the learning rate, hidden layer dimensions, and threshold parameters were optimized on the validation set. The traditional machine learning models also utilized the same window-level features, and their contamination rates, kernel functions, and anomaly discrimination thresholds were tuned on the validation set to ensure consistent evaluation conditions. The results are shown in Table 3.

As shown in Table 2, when using only VAE, although the model could complete basic anomaly identification, the AUC and F1 were only 94.23% and 90.31% respectively, while the time localization error reached 6.8 s, indicating that relying solely on probabilistic reconstruction remained difficult for effectively characterizing temporal misalignment anomalies. After introducing DTW, the AUC increased to 96.48% and the window-level F1 increased to 92.87%, showing that temporal warping distance could effectively compensate for the insufficient characterization of rhythm shift anomalies by point-wise errors. With the additional introduction of IB-Transf, the localization error on the time aspect was reduced to 3.7 seconds, while the variable localization F1 score increased to 84.31%. This showed that the use of the Transformer with inductive bias had clear strengths when it came to modeling global temporal dependencies and correlations between features. With the introduction of MSFF, even more improvements were realized. The complete model ultimately achieved the best results across all five metrics, which validated the synergistic enhancement effect among the modules. Finally, this study further selected Long Short-Term Memory Autoencoder (LSTM-AE), Gated Recurrent Unit Autoencoder (GRU-AE), Transformer Autoencoder (Trans-AE), Isolation Forest (IF), and One-Class Support Vector Machine (OC-SVM) for comparison with the proposed model. All deep learning models were evaluated using the same training, validation, and test sets. the input window length was uniformly set to 60 steps, the feature dimension to 18, the maximum training epochs to 100, and the batch size to 128, while the learning rate, hidden layer dimensions, and threshold parameters were optimized on the validation set. The traditional machine learning models also utilized the same window-level features, and their contamination rates, kernel functions, and anomaly discrimination thresholds were tuned on the validation set to ensure consistent evaluation conditions. The results are shown in Table 3.

As presented in Table 3, although OC-SVM and IF exhibited lower inference latency, their respective AUC and F1 scores reached only 86.42%, 80.36% and 88.75%, 82.14%, which indicated that traditional shallow models struggled to fully capture the multi-dimensional temporal dependencies in IoT traffic.

GRU-AE and LSTM-AE enhanced the time-dependent modeling capability through their recurrent structures, and their window-level F1 scores increased to 86.52% and 87.31%, respectively; however, they still exhibited insufficient identification capabilities for rhythm offset and phase drift anomalies. Trans-AE performed well in global temporal modeling, with its AUC and F1 scores reaching 95.28% and 91.74%, respectively, but lacked the DTW temporal regularization and multi-scale localization mechanisms. VDITM achieved an AUC, a window-level F1 score, and a recall rate of 98.74%, 95.96%, and 94.83%, respectively, while maintaining an average response latency of 54 ms, which successfully met the online inference requirements of edge gateways. The results show that the research models can more stably identify complex IoT traffic anomalies while maintaining an acceptable real-time performance.

4.2 Practical validation in IoT edge scenarios

This study further conducted validation oriented toward practical application scenarios of smart transportation edge networks to evaluate the engineering adaptability and deployment capability of this model in real business links.To generate testing cases for the experiment, the use of a collaborative demonstration area along the roadway was employed, including 12 common intersections, nine communication glitches of roadside units, 24 video sensing nodes,12 signal controllers, and four gateway nodes. The period of the testing process spanned 72 h of continuous operational data collection and included additional replay cases such as port scanning, abnormal broadcasting, congestion diffusion, and device jitters. First, the anomaly detection recall rate and time localization error of each model under typical anomaly scenarios were compared (four typical scenarios were selected: port scanning, low-rate denial of service, congestion diffusion, and roadside unit communication jitter, with three independent experimental statistics conducted for each scenario). The results are shown in Figure 8.

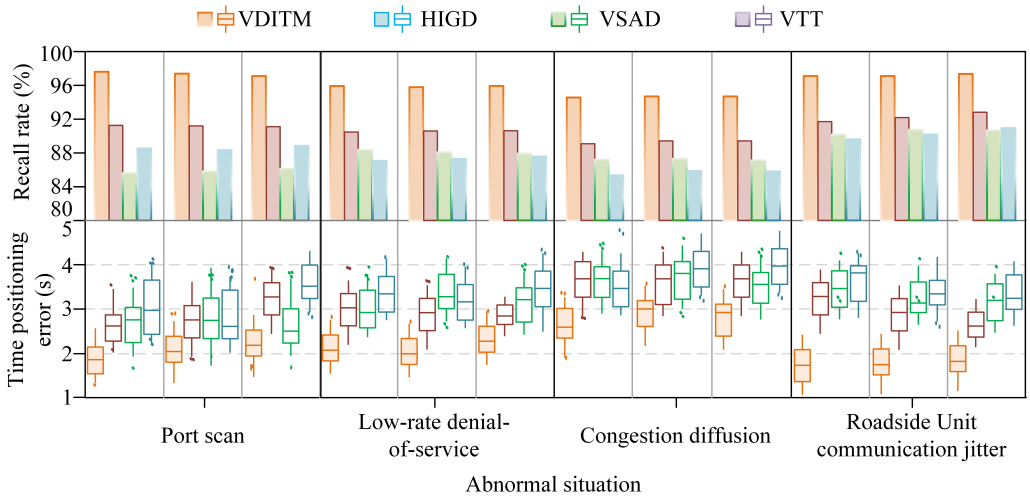
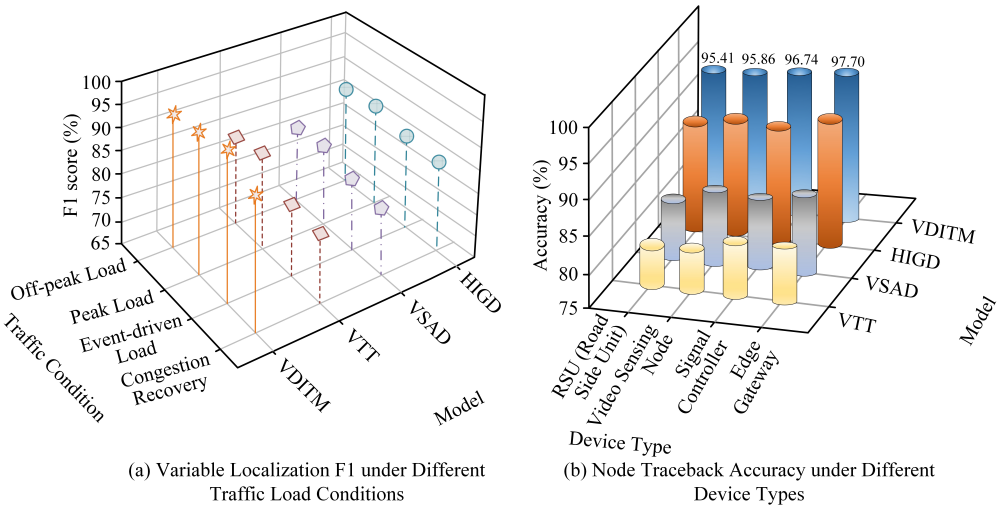


Figure 8: Comparison of detection recall and time localization error.

As shown in Figure 8, under the four typical scenarios of port scanning, low-rate denial of service, congestion diffusion, and roadside unit communication jitter, the average detection recall rates of VDITM reached 97.83%, 95.94%, 94.61%, and 96.72% respectively, all higher than the other three comparison models. Among them, for the congestion diffusion case, recall values of 88.42%, 86.75%, and 91.36% were obtained for VTT, VSAD, and HIGD models respectively, suggesting higher sensitivity to anomaly detection involving identical trends but rhythm mismatch locally. Meanwhile, the median lines of average time localization error of VDITM under the four scenarios were 2.1 s, 2.3 s, 2.7 s, and 1.6 s respectively, significantly better than VTT's 5.4 s, VSAD's 4.8 s, and HIGD's 4.1 s. In conclusion, this clearly demonstrates that apart from being more reliable in anomaly detection, the developed model also helped in detecting the exact timing periods where anomalies took place, thus adding value to quick action by edge systems in operation and maintenance. After completing anomaly time slice identification, this study further compared the variable localization F1 scores (measuring the ability to

identify anomaly-triggering variables in multivariate time series) and node backtracking accuracy of each model under business states. The results are shown in Figure 9.

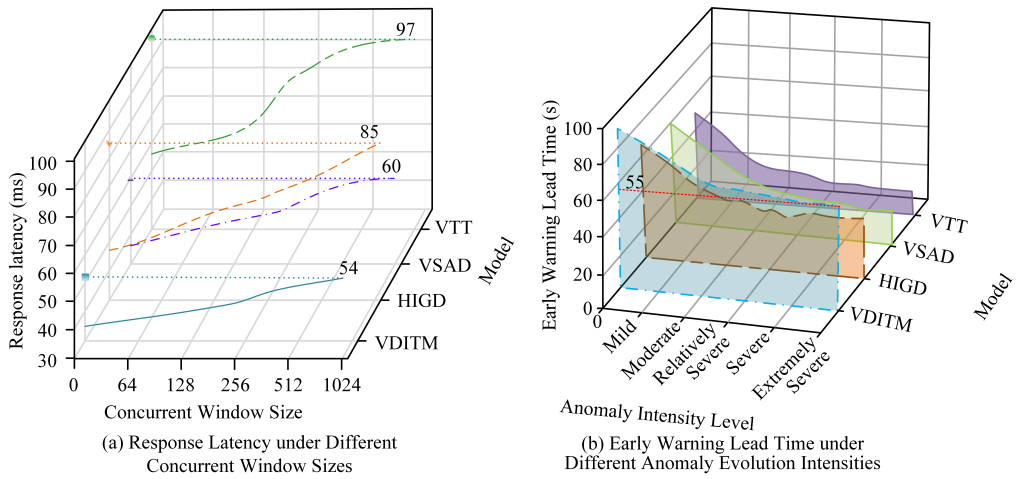


(a) Variable Localization F1 under Different Traffic Load Conditions

(b) Node Traceback Accuracy under Different Device Types

Figure 9: Comparison of variable localization F1 and node backtracking accuracy.

In Figure 9(a), under the four business states of off-peak load, peak load, event-type load, and congestion recovery, the variable localization F1 scores of VDITM were 95.76%, 96.12%, 97.07%, and 94.34% respectively, remaining stable overall, while VTT and VSAD showed more obvious declines under event-type load and congestion recovery states. This indicates that IB-Transf had advantages in global association modeling, while the MSFF mechanism effectively improved local anomaly boundary recovery capability, enabling the model to accurately lock onto key traffic variables triggering anomalies even under complex load fluctuations. As shown in Figure 9(b), across the four types of devices—RSU, video perception nodes, signal controllers, and edge gateways— the node backtracking accuracy of VDITM reached 95.41%, 95.86%, 96.74%, and 97.70% respectively, overall superior to the comparison models and stable across heterogeneous devices. These results show that the proposed model could not only detect abnormal behavior but also determine the source location and position of the anomalies. Finally, this study further compared the average response latency and early warning time of the four models on edge gateway devices. The results are shown in Figure 10.



(a) Response Latency under Different Concurrent Window Sizes

(b) Early Warning Lead Time under Different Anomaly Evolution Intensities

Figure 10: Comparison of average response time and early warning time.

As shown in Figure 10(a), on the Jetson AGX Orin edge device, when the window size was 1024, the response latency of VDITM was 54 ms/ window, lower than all comparison models, indicating that it maintained good real-time performance while possessing strong temporal modeling and localization capabilities. As shown in Figure 10(b), anomaly intensity was divided into six levels based on the

Table 4: Edge deployment cost and scalability evaluation

Model	Parameters (M)	Model Size (MB)	Peak Memory (GB)	DTW Cost (ms /window)	Latency 1024 Windows (ms / window)	Latency Nodes/ window) at 64 (ms / window)
VTT	9.84	39.6	1.86	-	71	104
VSAD	3.76	15.4	0.92	-	43	61
HIGD	7.42	29.8	1.44	-	62	89
VDITM	6.18	24.9	1.18	7.6	54	72

growth ratio of anomalous traffic relative to the normal mean: 0, 10%-30%, 30%-60%, 60%-100%, 100%-150%, and >150%. In particular, at the extreme anomaly phase, VDITM was capable of providing early warnings 55 s ahead of time, greatly outperforming the comparative models, indicating good early detection ability and stability in terms of warnings. This was because the frontend VAE-DTW submodel effectively enhanced the recognition accuracy for temporal misalignment anomalies, whereas the backend IB-Transf-MSFF submodel ensured that the model would balance between global temporal relationships and local anomaly boundaries, making it possible for the model to organically integrate anomaly detection and anomaly localization under complex IoT edge traffic conditions. To further evaluate edge deployment efficacy, the parameter size, model file size, peak memory usage, inference settings, DTW computational overhead, and large-scale node scalability of each model were evaluated on the Jetson AGX Orin device. All models used the same input settings, with a window length of 60 steps and a feature dimension of 18. During the inference stage, gradient calculation was disabled and batch window online inference was adopted. The results are shown in Table 4.

As presented in Table 4, the parameter size of VDITM was 6.18 M, the model file size was 24.9 MB, and the peak memory usage was 1.18 GB, which were lower than those of VTT and HIGD; thus, it successfully met the deployment requirements of edge gateway devices. Due to the introduction of DTW temporal regularization in VDITM, the single-window DTW computational cost was 7.6 ms, yet the overall response latency remained 54 ms per window when the window size was 1024, indicating that this module did not introduce a significant real-time bottleneck. When the node scale expanded to 64 device nodes, the average inference latency of VDITM was 72 ms per window, which was lower than those of VTT and HIGD, demonstrating that the model maintained excellent scalability in large-scale edge networks. VSAD, although having a lighter model and lower latency, lacks fine-grained positioning and node backtracking mechanisms, so its comprehensive application capability is limited. The above results further prove that VDITM has good engineering deployability in terms of model complexity, memory overhead, DTW calculation cost, and large-scale node expansion.

Furthermore, from the perspective of the applicable scope of the research results, the current validation of VDITM is mainly based on the edge IoT traffic scenarios of intelligent transportation. The conclusions should initially be limited to edge network environments with similar data forms and deployment conditions. Since this model takes multivariate time-series windows as input and relies on anomalous response maps and device-variable-link mapping relationships to execute node backtracking, its potential migration to other applications requires fulfillment of two prerequisites: the target scenario can provide stable and continuous sampling flow characteristics, and has clear variable meanings and device mapping relationships. For other IoT monitoring scenarios, the model still needs to be re-calibrated for thresholds, parameter optimization, and experimental verification based on specific business flow characteristics, anomaly types, and node topologies. Therefore, the current research results can provide methodological references for similar edge IoT traffic anomaly detection and location tasks, but its cross-domain generalization ability still needs to be further verified through more scenario data.

5 Conclusion

To resolve the insufficient characterization of temporal misalignment anomalies and the limited capability to localize anomalies in IoT edge traffic, this study proposes a novel integrated anomaly detection and localization framework based on VAE, DTW, and IB-Transf-MSFF. The model learns the

distribution of normal traffic data and completes the reconstruction task using VAE, while incorporating DTW for temporal transformation to improve the recognition capability for rhythm shift anomaly, and IB-Transf-MSFF to collaboratively analyze anomaly time slices, variable factors, and corresponding nodes. Experimental results showed that the model achieved an anomaly detection AUC of 98.74%, a window-level F1 of 95.96%, a time localization error of 2.4 s, a variable localization F1 of 89.53%, and a node backtracking accuracy of 92.68%. In practical smart transportation edge scenarios, the average time localization error of the proposed model was approximately 2.1 s to 2.7 s, the edge device response latency was 54 ms/ window, and it could issue early warnings 55 s in advance under extreme anomaly conditions. Overall, the integrated detection and localization model constructed in this study effectively improved anomaly identification sensitivity and localization accuracy, and can provide technical support for real-time monitoring and anomaly tracing in edge networks. However, this study does not account for the cross-domain generalization capability in large-scale network environments. Furthermore, the long-term evolutionary modeling of various anomalies warrants further investigation. In future research, the performance of the model can be improved by taking advantage of multi-source data and online learning techniques.

References

[1] Yang, X.; Xiang, K.; Yuan, S.; Huang, J. (2024). Vehicle driving behavior recognition and optimization strategies based on cloud computing and SSA-BP algorithm, Studies in Informatics and Control, 33(3), 17-28, 2024.

[2] Morshedi, R.; Matinkhah, S. M. (2025). Anomaly detection in IoT traffic in the presence of Gaussian noise using deep neural networks, Journal of AI and Data Mining, 13(3), 305-319, 2025.

[3] Montoya, G. A.; Lozano-Garzón, C.; Paternina-Arboleda, C.; Donoso, Y. (2025). A mathematical optimization approach for prioritized services in IoT networks for energy-constrained smart cities, International Journal of Computers Communications & Control, 20(1), 6912, 2025.

[4] Kang, B. G.; Kim, B. S. (2024). Attachable IoT-based digital twin framework specialized for SME production lines, International Journal of Simulation Modelling, 23(3), 471-482, 2024.

[5] Sana, L.; Nazir, M. M.; Yang, J.; Hussain, L.; Chen, Y. L.; Ku, C. S.; Por, L. Y. (2024). Securing the IoT cyber environment: enhancing intrusion anomaly detection with vision transformers, IEEE Access, 12(5), 82443-82468, 2024.

[6] Zeng, G. Q.; Yang, Y. W.; Lu, K. D.; Geng, G. G.; Weng, J. (2025). Evolutionary adversarial autoencoder for unsupervised anomaly detection of industrial internet of things, IEEE Transactions on Reliability, 74(3), 3454-3468, 2025.

[7] Dou, S.; Yang, K.; Jiao, Y.; Qiu, C.; Ren, K. (2024). Anomaly detection in event-triggered traffic time series via similarity learning, IEEE Transactions on Dependable and Secure Computing, 22(2), 888-902, 2024.

[8] Zeng, Z.; Zhao, B.; Deng, X.; Liu, X.; Zheng, J.; Chen, J. (2024). Toward intelligent attack detection with causal transformer in Internet of Things, IEEE Internet of Things Journal, 12(8), 10751-10767, 2024.

[9] Jovanović, Đ. D.; Vuletić, P. V. (2024). PI-BODE: programmable intraflow-based IoT botnet detection system, Computer Science and Information Systems, 21(1), 37-56, 2024.

[10] Alsoufi, M. A.; Siraj, M. M.; Ghaleb, F. A.; Al-Razgan, M.; Al-Asaly, M. S.; Alfakih, T.; et al. (2024). Anomaly-based intrusion detection model using deep learning for IoT networks, Computer Modeling in Engineering & Sciences, 141(1), 823-845, 2024.

[11] Shuxin, S.; Bing, H.; Zhongdai, W.; Dezhi, H.; Huafeng, W.; Xiaojun, M. (2024). BLSAE-SNIDS: a Bi-LSTM sparse autoencoder framework for satellite network intrusion detection, Computer Science and Information Systems, 21(4), 1389-1410, 2024.

[12] Pang, J. L. (2023). Adaptive fault prediction and maintenance in production lines using deep learning, International Journal of Simulation Modelling, 22(4), 734-745, 2023.

[13] Hamdi, M.; Bouhamed, H.; Badreddine, F.; Alkanhel, R. (2024). Deep recurrent neural networks distributed on a Hadoop/Spark cluster for fall detection, International Journal of Computers Communications & Control, 19(3), 6428, 2024.

[14] Liang, S.; Chen, C.; Wu, D.; Chen, L.; Wu, Q.; Gu, T. T. (2024). An ensemble learning method for the fault multi-classification of smart meters, Tehnicki vjesnik-Technical Gazette, 31(5), 1514-1522, 2024.

[15] Ribeiro, J.; Santos, R.; Analide, C.; Silva, F. (2024). Implementing federated learning and explainability techniques in regression models to increase transparency and reliability, Studies in Informatics and Control, 33(4), 15-24, 2024.

[16] Watanabe, N.; Yamazaki, T.; Miyoshi, T.; Yamamoto, R.; Nakahara, M.; Okui, N.; Kubota, A. (2025). Self-adaptive traffic anomaly detection system for IoT smart home environments, IEICE Transactions on Communications, 108(3), 230-242, 2025.

[17] Rabbani, M.; Gui, J.; Nejati, F.; Zhou, Z.; Kaniyamattam, A.; Mirani, M.; Ghorbani, A. A. (2024). Device identification and anomaly detection in IoT environments, IEEE Internet of Things Journal, 12(10), 13625-13643, 2024.

[18] Nissar, N.; Naja, N.; Jamali, A. (2024). Securing VANETs: multi-objective intrusion detection with variational autoencoders, IEEE Transactions on Consumer Electronics, 70(1), 3867-3874, 2024.

[19] Dong, H.; Kotenko, I. V. (2024). Convolutional variational autoencoders and resampling techniques with generative adversarial network for enhancing Internet of Thing security, Pattern Recognition and Image Analysis, 34(3), 562-569, 2024.

[20] Bali, N.; Jaoua, Z.; Bzeouich, O.; Abbassi, I. (2024). Intrusion detection systems for the internet of things: a probabilistic anomaly detection approach, International Journal of Computers and Applications, 46(11), 933-944, 2024.

[21] Tang, L.; Kou, E.; Zhang, W.; Wu, Q.; Chen, Q. (2024). IoT-FKGDL-SL: anomaly detection framework integrating knowledge distillation and swarm learning for 5G IoT, IEEE Internet of Things Journal, 11(23), 38601-38614, 2024.

[22] Li, Y.; Yang, J.; Shih, S. L.; Shih, W. T.; Wen, C. K.; Jin, S. (2024). Efficient IoT devices localization through Wi-Fi CSI feature fusion and anomaly detection, IEEE Internet of Things Journal, 11(24), 39306-39322, 2024.

[23] Shibu, S.; Kirubakaran, S.; Remamany, K. P.; Ahamed, S.; Chitra, L.; Kshirsagar, P. R.; Tirth, V. (2025). Anomaly detection using deep learning approach for IoT smart city applications, Multimedia Tools and Applications, 84(17), 17929-17949, 2025.

[24] Mishra, S.; Jain, U. (2025). Secure IoT sensor networks through advanced anomaly detection with Kolmogorov-Arnold Networks, Microsystem Technologies, 31(8), 2157-2167, 2025.

[25] Wang, X.; Luo, Q.; Liu, K.; Mao, R.; Wu, G. (2024). Deep learning method based on multiscale enhanced feature fusion for vehicle behavior prediction, IEEE Internet of Things Journal, 12(7), 9142-9155, 2024.



Copyright ©2026 by the authors. Licensee Agora University, Oradea, Romania.
This is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License.
Journal's webpage: http://univagora.ro/jour/index.php/ijccc/



This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:

Shi, H.; Zhang, J. (2026). Anomaly Detection and Localization of IoT Traffic via Hybrid Autoencoder and Temporal Modeling, *International Journal of Computers Communications & Control*, 21(6), 7529, 2026. https://doi.org/10.15837/ijccc.2026.6.7529