Textile & Leather Review

ISSN 2623-6281| www.tlr-journal.com | 10.31881/TLR

H2IDF: A Hybrid Intrusion Detection Framework with Heterogeneous Feature Fusion

Zhenghao Qian, Fengzheng Liu, Mingdong He, Bo Li, Xuewu Li, Chuangye Zhao, Gehua Fu, Yifan Hu

How to cite: Qian Z, Liu F, He M, Li B, Li X, Zhao C, Fu G, Hu Y. H2IDF: A Hybrid Intrusion Detection Framework with Heterogeneous Feature Fusion. Textile & Leather Review. 2026; 9:1598-1627. https://doi.org/10.31881/TLR.2026.1598

How to link: https://doi.org/10.31881/TLR.2026.1598

Published: 25 April 2026

H2IDF: A Hybrid Intrusion Detection Framework with Heterogeneous Feature Fusion

Zhenghao Qian, Fengzheng Liu*, Mingdong He, Bo Li, Xuewu Li, Chuangye Zhao, Gehua Fu, Yifan Hu

Information Center, Guangdong Power Grid Co., Ltd., Guangzhou 510180, Guangdong, China

*lw32407585@163.com

Article

https://doi.org/10.31881/TLR.2026.1598

Published 25 April 2026

ABSTRACT

The increasing digitalization of modern energy systems, and similarly the ongoing digital transformation of the textile industry under the paradigm of Textile 4.0, has expanded their cyber attack surfaces and heightened the risk of sophisticated intrusions. The complex nature of energy networks traffic, which is characterized by heterogeneous data, multi-protocol communications, and strong temporal dependencies, has resulted in substantial growth in both the volume and dimensionality of network traffic, posing significant challenges for traditional intrusion detection systems (IDS). This paper proposes a Hybrid Intrusion Detection Framework with Heterogeneous Feature Fusion (H2IDF). The framework leverages multi-scale Convolutional Neural Networks (CNNs) to extract fine grained temporal patterns and utilizes Transformer encoders to model structured tabular features. A cross-type attention mechanism is introduced to semantically align and deeply fuse heterogeneous types, thereby enhancing the model's ability to capture complex inter feature dependencies. Furthermore, a prediction aggregation mechanism is employed to consolidate frame level decisions across overlapping sliding windows, significantly improving detection stability and robustness to noise. Experiments on the AWID dataset, selected for its representation of heterogeneous wireless traffic patterns analogous to energy system communications, demonstrate that H2IDF achieves superior performance over competitive baselines. These results highlight the framework's potential for enhancing cybersecurity in energy networks and analogous industrial environments, such as those found in Textile 4.0

KEYWORDS

cybersecurity, intrusion detection, convolutional neural network, textile 4.0, smart textiles

INTRODUCTION

The digitalization of modern energy systems, including smart grids, distributed renewable energy resources, and industrial control networks, has greatly enhanced operational efficiency and flexibility [1,2]. Similarly, the textile industry is undergoing a profound transformation, often referred to as Textile 4.0, which involves the extensive integration of digital technologies into manufacturing processes, supply chain management, and the development of smart textiles. However, the convergence of Information Technology (IT) and Operational Technology (OT) in these systems has expanded the cyber attack surface, exposing critical infrastructures to increasingly sophisticated threats and resulting in a surge of intelligent, automated, and large scale cyber attacks [3-5]. In particular, the use of wireless communication, IoT enabled sensors, and cloud based energy management platforms —whether for energy distribution or for automated textile production and monitoring— introduces vulnerabilities that adversaries can exploit to launch targeted attacks such as advanced persistent threats (APTs), protocol manipulation, and multi stage penetration attacks [6,7].

Given the importance of energy systems, any disruption caused by cyber incidents can lead to cascading failures, large scale service outages, and severe societal consequences. Likewise, in the textile sector, cyber attacks can cause significant economic damage through production sabotage, theft of proprietary designs, or compromising the integrity of smart fabrics used in medical and military applications. Therefore, Intrusion Detection Systems (IDS) play a crucial role in safeguarding the cybersecurity of these infrastructures [8]. The complexity of energy networks traffic is characterized by heterogeneous data, multi-protocol communications, and strong temporal dependencies. This has resulted in substantial growth in both the volume and dimensionality of network traffic and poses significant challenges for traditional IDS, which often rely on isolated features or fixed time point analysis and are prone to high false positive/negative rates [9]. Furthermore, modern attackers increasingly adopt stealthy tactics, such as slow port scans, protocol obfuscation, and cross time slot coordination to evade detection, making accurate identification of malicious activities significantly more challenging [10]. These challenges are directly applicable to smart textile factories, where interconnected machinery, sensors, and control systems generate complex data streams that are vulnerable to subtle and coordinated cyber intrusions.

Recent advancements in machine learning, particularly deep learning and reinforcement learning, have demonstrated remarkable capabilities in pattern recognition, feature extraction, and generalization across a

wide range of domains such as computer vision, speech recognition, and natural language processing [11-13]. These technologies are also gaining increasing adoption in practical applications including autonomous driving and intelligent healthcare systems [14,15]. In the field of cybersecurity, especially in intrusion detection, these intelligent approaches offer promising solutions to address the complexity of traffic patterns and the diversity of attack behaviours [16,17].

For instance, Long Short-Term Memory (LSTM) networks have been extensively applied to model the temporal dynamics of network traffic, enhancing the detection of stealthy attacks such as slow scans and advanced persistent threats. Convolutional Neural Networks (CNNs), known for their efficacy in extracting spatially correlated features, are adept at identifying localized patterns indicative of intrusion [18,19]. Generative Adversarial Networks (GANs) have been utilized to augment training data by generating synthetic but realistic samples, thereby improving model robustness against rare or unseen attack types [20]. More recently, Transformer architectures, with their self-attention mechanisms and strong capacity for capturing long range dependencies, have shown promise in modeling complex nonlinear interactions among high dimensional network features in intrusion detection scenarios [21].

Despite these advances, most existing approaches still face challenges in accurately detecting sophisticated attack behaviours in dynamic and high dimensional network environments, particularly when modeling the complex traffic characteristics of modern energy networks. Limitations in feature representation, generalization across diverse scenarios, and temporal consistency hinder their practical deployment in real world IDS, especially in critical energy infrastructures, where operational continuity and resilience are paramount.

To address the aforementioned challenges, this paper proposes a Hybrid Intrusion Detection Framework with Heterogeneous Feature Fusion (H2IDF). The framework exploits the ability of multi-scale Convolutional Neural Networks (CNNs) to capture temporal patterns from sequential traffic data and the power of Transformer encoders to model long range dependencies within structured tabular features. To bridge the semantic gap between these heterogeneous types, a cross-type attention module is designed to enable deep interactive fusion between the temporal and tabular representations.

The proposed architecture adopts a dual branch design, in which temporal and tabular features are independently modeled and subsequently fused through the cross-type attention mechanism. This design facilitates the joint representation of temporal dynamics and static contextual attributes, enhancing the model's

ability to recognize complex intrusion behaviours. The deep integration of heterogeneous features contributes to improved detection accuracy and robustness, particularly under conditions of data imbalance and noise. Furthermore, a many-to-many sequence labeling strategy based on a sliding window mechanism enables frame level classification over continuous traffic streams, which is particularly valuable for monitoring real time operations in energy networks. In the final detection stage, a majority voting based prediction aggregation module integrates overlapping window predictions, thereby improving detection stability and mitigating the impact of local noise.

Extensive experiments conducted on the AWID dataset, which was chosen for its representation of heterogeneous wireless traffic patterns relevant to energy system communications, demonstrate that H2IDF achieves consistently strong detection performance, even under severe class imbalance, a common challenge in energy networks.

The main contributions of this paper are summarized as follows:

• We propose a hybrid intrusion detection framework (H2IDF) that integrates multi-scale CNNs for temporal pattern extraction and Transformer based encoders for structured tabular representation. A cross-type attention module is introduced to achieve deep interactive fusion between heterogeneous types.

• We design a many-to-many sequence labeling mechanism based on a sliding window approach, enabling frame level predictions over raw traffic sequences. An aggregation strategy based on majority voting enhances detection stability and robustness in real-time monitoring scenarios.

• We conduct comprehensive experiments on the AWID dataset to validate the effectiveness of H2IDF, demonstrating its potential applicability to wireless and heterogeneous communication scenarios found in energy systems.

The remainder of this paper is organized as follows. The second section reviews related work in the field of intrusion detection. The third section presents the proposed H2IDF framework along with its constituent modules. The fourth section describes the experimental setup, evaluation metrics, and provides an in-depth analysis of the results. Finally, the fifth section concludes the paper and discusses potential future research directions.

RELATED WORK

The task of an Intrusion Detection System (IDS) is to monitor the activities of a network or host, identify and respond to unauthorized access, malicious behaviours, or anomalous activities, thereby safeguarding the security of systems and data. Traffic based intrusion detection, which identifies potential attacks or anomalies by analyzing features, patterns, and statistical characteristics of network packets, has become a fundamental component of modern cybersecurity defense. Existing traffic based intrusion detection approaches can be broadly categorized into three types: Supervised intrusion detection, which builds detection models using labeled training data to accurately recognize known attack types; Unsupervised intrusion detection, which requires no labeled data and typically employs clustering or anomaly detection techniques to discover unknown threats; and Semi-supervised intrusion detection, which combines the strengths of both approaches by leveraging a small amount of labeled data along with a large volume of unlabeled data, thereby improving detection efficiency and adaptability.

Many intrusion detection researches are based on supervised learning. For example, Allah Bakhsh et al. proposed a deep learning based intrusion detection framework for IoT networks, leveraging a combination of Feedforward Neural Networks (FFNN), Long Short Term Memory (LSTM) networks, and Random Neural Networks (RandNN) [22]. The framework is designed to handle the resource constraints of IoT environments by enabling lightweight deployment and achieving high accuracy on the CIC-IoT2022 dataset. Dimensionality reduction techniques such as Principal Component Analysis (PCA) and correlation analysis are applied to improve efficiency. Hanaa Attou et al. designed a lightweight intrusion detection framework combining feature engineering and a random forest (RF) classifier, targeting massive heterogeneous traffic in cloud environments [23]. The framework applies visualization driven analysis to identify a two-dimensional discriminative feature subset, significantly reducing dimensionality and resource consumption while maintaining effective anomaly detection. Turukmane et al. introduced M-MultiSVM, a network intrusion detection framework enhanced with feature selection [24]. The method employs ASmoT to address class imbalance and utilizes an improved Singular Value Decomposition (SVD) technique in conjunction with the Opposition based Northern Goshawk Optimization algorithm (ONgO) to extract and select optimal features. Final attack classification is performed using Mud Ring optimization combined with multi-layer SVM. Zhang et al. proposed GBFKAN, an interpretable three layer adaptive architecture for multi scenario IoT intrusion detection. The first two layers

adopt Gated Recurrent Units (GRU) and Bidirectional LSTM (BiLSTM) to capture both short and long term temporal dependencies [25]. A Kolmogorov–Arnold Network (KAN) is then introduced for high dimensional nonlinear mapping, fusing it with the learned hidden representations to achieve multi-level feature perception and discrimination. The framework further integrates SHAP explainers at the output layer to reveal feature contributions and identify misclassification patterns, enhancing generalization, robustness, and model trustworthiness while maintaining lightweight design. Yakub Kayode Saheed et al. developed a lightweight intrusion detection framework for edge computing environments in IoT networks [26]. It employs a Modified Genetic Algorithm (MGA) for optimal traffic feature selection and uses genetic optimization to fine tune the LSTM architecture and hyperparameters, enabling efficient modeling of temporal dynamics in IoT traffic. Hassini et al. proposed an end-to-end 1D CNN based architecture for industrial IoT scenarios, eliminating the need for manual feature engineering [27]. The model uses a four layer Conv–BN–AvgPool–Dropout block to automatically extract traffic patterns and directly performs 15 class attack classification based on 63 dimensional raw input features. It achieved an accuracy of up to 99.96% on the Edge-IIoTset dataset. Tan et al. presented a reinforcement learning based method that introduces an adaptive sample distribution dual experience replay strategy for intrusion detection [28]. In addition to the standard experience buffer, a secondary experience buffer is maintained with class weighted dynamic sampling, enhancing the detection performance on minority class traffic.

Beyond supervised learning, unsupervised intrusion detection has received increasing attention in recent years. These approaches do not require labeled data, instead, they learn the statistical distribution of normal traffic and identify deviations from the learned patterns as anomalies. Common techniques include clustering algorithms, probabilistic models, and reconstruction based methods such as Autoencoders (AE) and Variational Autoencoders (VAE). For example, T. K. Boppana et al. proposed GAN-AE, an unsupervised intrusion detection model for MQTT that integrates adversarial training into an autoencoder. Trained solely on normal traffic in a two stage process, the model first learns the normal distribution via reconstruction error and then refines representations through an adversarial discriminator to enhance sensitivity to unknown attacks [29]. Paulo Freitas de Araujo-Filho et al. introduced WGAN-IDS, an unsupervised detection model that replaces all LSTM components in both the generator and discriminator with stackable Temporal Convolutional Networks (TCNs) and multi head self-attention blocks [30]. This design enables the learning of temporal dependencies

directly from raw traffic on edge servers and allows adaptive tradeoffs between detection rate and inference latency by adjusting the number of blocks. Kabilan N et al. proposed a lightweight unsupervised IDS that applies an autoencoder for denoising and compressing CAN traffic features, followed by fuzzy C-means clustering on the latent representations to distinguish normal and abnormal messages without requiring labels [31]. Lu et al. proposed Manticore, an unsupervised contrastive learning based intrusion detection system designed for 5G networks [32]. It combines raw bit level features with multi scale statistical features, and leverages feature grouping and a contrastive reconstruction loss (CRLoss) to automatically construct positive and negative pairs without manual labeling. Experimental results on the 5G-NIDD and Kitsune datasets demonstrate that Manticore improves overall accuracy by approximately 25 percentage points on average.

In addition, semi-supervised learning methods have gained increasing attention in the field of intrusion detection due to their low reliance on labeled data and ability to leverage large volumes of unlabeled data alongside a small set of labeled samples. Such methods employ strategies including pseudo label generation, generative adversarial networks, and consistency regularization to preserve supervised signals while mining potential patterns and structures within unlabeled traffic, offering new solutions to intrusion detection challenges. Li et al. proposed HDA-IDS, a large scale semi-supervised DoS/botnet defense system for IoT environments. The system utilizes feature engineering and a stacked ensemble of base learners for signature based detection of known attacks, and incorporates a semi supervised CL-GAN branch that combines CNN-LSTM and GAN to characterize normal baselines and detect unknown or zero day traffic under label scarce conditions [33]. Nguyen et al. presented a semi-supervised in-vehicle intrusion detection framework, where a Variational Autoencoder (VAE) is used to learn latent representations of unlabeled CAN messages. Based on this, an adversarial reinforcement learning architecture with dual agents, an environment agent and a classifier agent is introduced to adaptively extract hard to classify samples and mitigate class imbalance. By combining pseudo label based pretraining with a lightweight classifier, the system enables multi class detection of both known and unknown attacks using only partially labeled data [34]. Shajjad Hossain et al. proposed a privacy preserving intrusion detection framework for 5G-V2X networks [35]. The framework first applies self-supervised pretraining on the vehicle side to learn general representations from large scale unlabeled traffic, then aggregates models via federated learning and performs rapid fine tuning using limited labeled samples. Without requiring centralized data, the method achieves efficient network intrusion detection, outperforming

comparable approaches on the CIC-IDS2017 dataset by up to 9% in accuracy. Phan The Duy et al. introduced Fed-Evolver, a federated intrusion detection framework for Software Defined Networking (SDN) environments [36]. At its core, the approach trains a semi-Supervised Adversarial Autoencoder (SS-AAE) on each edge node using a small amount of labeled data and a large volume of unlabeled traffic, while employing GAN based hard sample generation to address class imbalance. Local models are then aggregated using the FedAvg algorithm to form a global detector. Experiments across multiple datasets show that Fed-Evolver can achieve high detection accuracy using only 1% of labeled data.

METHODOLOGY

To address the limitations of traditional intrusion detection methods in modeling heterogeneous traffic features, particularly in the context of modern energy systems where IT and OT networks are deeply integrated, we propose a novel framework named H2IDF (Hybrid Intrusion Detection Framework with Heterogeneous Feature Fusion). This framework is designed to jointly learn and fuse temporal dynamics and structured attributes from network traffic, thereby enhancing the capability to identify complex and stealthy attack behaviours.

This hybrid design is motivated by the intrinsic heterogeneity of network traffic data. The CNN branch captures localized temporal dependencies and short-term fluctuations within traffic sequences, while the Transformer branch models global contextual relationships across structured tabular features. The cross-type attention mechanism then aligns and fuses these complementary feature spaces, enabling deeper semantic integration. Compared with a monolithic model (e.g., a pure CNN or Transformer), this hybrid architecture provides stronger representational balance between fine-grained temporal variations and high-level relational structures, leading to improved detection robustness and generalization across diverse attack scenarios.

As illustrated in Figure 1, the overall architecture of H2IDF comprises three major components: (1) Data Construction, (2) Temporal–Tabular Cross-Type Attention Network, and (3) Prediction Aggregation. Each component is carefully designed to address different aspects of the intrusion detection task.

In the Data Construction stage, raw network traffic data are partitioned into two distinct types: temporal features, which reflect continuous behavioural patterns (e.g., packet timing, signal duration), and tabular

features, which include categorical protocol fields and numerical signal indicators. A sliding window mechanism is applied to segment the input into overlapping sequences, enabling the model to learn temporal correlations while retaining structural attribute consistency. This transformation facilitates sequence-to-sequence modeling and enables effective frame level prediction.

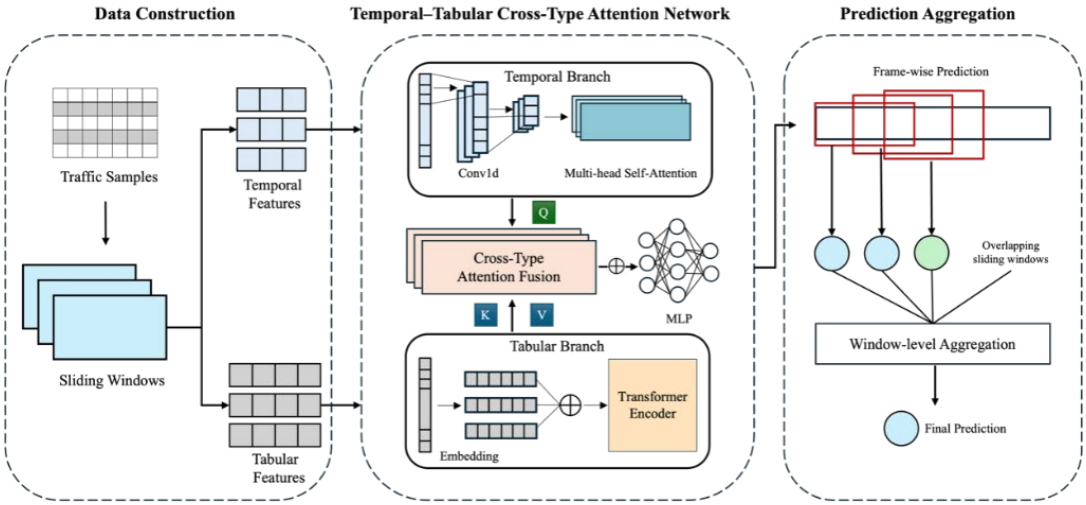

Figure 1. Overall Framework Architecture.

The core of H2IDF lies in a dual branch neural architecture. The temporal branch applies multi-scale 1D convolutions followed by multi head self-attention to extract and align patterns from different temporal resolutions. In parallel, the tabular branch embeds categorical features and integrates numerical values using a Transformer encoder, capturing feature wise dependencies across protocol fields. Unlike a conventional MLP, the Transformer encoder can model long-range interactions and adaptive importance weighting among heterogeneous attributes, which is particularly beneficial for discovering global semantic correlations across protocol and signal fields in structured network traffic data. To bridge these heterogeneous representations, we introduce a cross-type attention module, where temporal representations act as queries and tabular representations serve as keys and values. This design enables the model to learn complex type interactions, thereby enhancing its ability to recognize subtle and multi-dimensional attack characteristics.

Since each sample may be covered by multiple overlapping sliding windows, a many-to-many sequence labeling mechanism is employed to generate frame level predictions. A majority voting strategy is then applied to aggregate predictions from all covering windows, ensuring decision stability and resilience against local noise or uncertain predictions.

The overall implementation process of the H2IDF is outlined in Algorithm 1.

Algorithm 1: Overall Implementation Process of H2IDF.

Input： Temporal sequence $X^{(t)} \in R^{N\times d_t}$, tabular sequence $X^{(b)} \in R^{N\times d_b}$; Window size w; Step size s.

Output： Final prediction sequence $\hat{y} \in \{0, \dots, C-1\}^N$.

for $k \leftarrow 1$ ***to*** $N - w + 1$ ***step*** s ***do***

Extract temporal and tabular window slices:

$X_k^{(t)} \leftarrow X_{k:k+w-1}^{(t)}, X_k^{(b)} \leftarrow X_{k:k+w-1}^{(b)}$

Apply multi-scale convolution on $X_k^{(t)}$ to obtain $H_k^{(t)}$:

$H_k^{(t)} \leftarrow \text{MSConv}\left(X_k^{(t)}\right)$

Apply self-attention: $Z_k^{(t)} \leftarrow \text{MultiHeadAttn}\left(H_k^{(t)}\right)$

Encode tabular features via Transformer: $Z_k^{(b)} \leftarrow \mathcal{T}\left(X_k^{(b)}\right)$

Project tabular features: $\widetilde{Z_k^{(b)}} \leftarrow Z_k^{(b)} W_p$

Cross-modal attention fusion:

$A_k \leftarrow \text{MultiHeadAttn}\left(Z_k^{(t)}, \widetilde{Z_k^{(b)}}, \widetilde{Z_k^{(b)}}\right)$

$Z_k^{(c)} \leftarrow \text{LayerNorm}\left(Z_k^{(t)} + \text{Dropout}(A_k)\right)$

Concatenate features and classify:

$F_k \leftarrow \text{Concat}\left(Z_k^{(c)}, Z_k^{(b)}\right)$

$\hat{Y}_k \leftarrow \text{MLP}(F_k)$

Store the w frame-level predictions $\hat{Y}_k$ (one prediction for each frame in the window) together with their corresponding sample indices.

for $t \leftarrow 1$ **to** N **do**

Collect predictions $\hat{y}_t^{(j)}{}_{j=1}^{m}$ from overlapping windows.

$\hat{y}_t \leftarrow \text{MajorityVote}(\{\hat{y}_t^{(j)}\})$

return $\hat{y}$

DATA CONSTRUCTION

In intrusion detection tasks, relying solely on individual traffic records often fails to capture the contextual semantics of attack behaviours. This limitation becomes particularly evident in scenarios involving stealthy

techniques such as persistent threats, slow scans, or obfuscated probing, where malicious patterns may span across multiple time steps and are not discernible from isolated samples.

To address this, we adopt a sliding window based sequence construction strategy, which transforms the raw traffic stream into overlapping sequences of fixed length. This enables the model to learn temporal dependencies and behavioural patterns over time, while preserving structural consistency within tabular features. Specifically, we define the detection input as a time ordered traffic sequence:

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\} \quad (1)$$

where $\mathrm{x_i} = \left(\mathrm{x_i^{(t)}}, \mathrm{x_i^{(b)}}\right)$ represents the full feature vector of the $\mathrm{i-th}$ traffic sample, consisting of temporal features $\mathrm{x_i^{(t)}} \in \mathrm{R^{d_t}}$ and tabular features $\mathrm{x_i^{(b)}} \in \mathrm{R^{d_b}}$. The corresponding ground-truth label is $\mathrm{y_i} \in \{0,1, \dots, \mathrm{C}-1\}$, where C is the number of attack categories.

Given a predefined window size w and step size s, the traffic stream is segmented into overlapping subsequences of length w. For the k-th window, the input and label sequences are constructed as:

$$X_k = \{x_k, x_{k+1}, \dots, x_{k+w-1}\} \quad (2)$$

$$Y_k = \{y_k, y_{k+1}, \dots, y_{k+w-1}\} \quad (3)$$

where $\mathrm{k} \in \{1,\ 1+\mathrm{s},\ 1+2\mathrm{s},\ \dots\}$ and $\mathrm{k}+\mathrm{w}-1 \le \mathrm{N}$ ensures the sequence remains within bounds.

Each input window $\mathrm{X_k}$ is further decomposed into two type specific tensors for heterogeneous feature modeling:

• The temporal feature tensor is constructed as $\mathrm{X_k^{(t)}} \in \mathrm{R^{d_t \times w}}$, where each column represents one traffic sample's temporal features over the window. This format is compatible with the multi-scale 1D convolutional operations used in the temporal modeling branch.

• The tabular feature tensor is formatted as $\mathrm{X_k^{(b)}} \in \mathrm{R^{w \times d_b}}$, where each row corresponds to a sample's categorical and numerical fields. This layout facilitates sequence wise processing via the Transformer encoder in the tabular branch.

The corresponding label sequence $\mathrm{Y_k} \in \mathrm{R^w}$ retains the ground truth labels of all samples within the window, enabling a many-to-many frame level prediction setting. Additionally, for downstream prediction aggregation, we explicitly track the starting index k of each window during sequence construction. During inference, this allows us to align overlapping predictions corresponding to the same original sample, upon which a majority voting strategy is applied to produce stable and noise resilient final decisions.

This design transforms the intrusion detection problem from traditional point wise classification to a structured sequence modeling task, enhancing the system's ability to detect temporally correlated and context aware attacks.

TEMPORAL–TABULAR CROSS-TYPE ATTENTION NETWORK

In real world scenarios, network traffic samples typically exhibit heterogeneous feature attributes, encompassing both temporal features such as inter packet intervals and signal strength variations and static, structured tabular features such as protocol types, field flags, and rate information. The former captures the dynamic behavioural patterns of potential attacks, while the latter reflects semantic information at the packet or connection level. Effectively modeling and integrating these two types of features is critical for accurately identifying complex attack patterns. Therefore, we propose a Temporal-Tabular Cross-Type Attention Network that explicitly separates and specializes the modeling of temporal and tabular features, and then integrates them through a cross-type attention mechanism to enable deep and semantically aligned fusion. This design enhances the model's capacity to recognize complex, stealthy, or multistage attacks in network traffic. The temporal branch of the network takes as input a tensor $\mathrm{X}^{(\mathrm{t})} \in \mathrm{R}^{\mathrm{B}\times \mathrm{d_t}\times \mathrm{w}}$, where B denotes the batch size, and dt represents the dimensionality of temporal features. To capture local behavioral patterns under different temporal receptive fields, multiple 1D convolutional kernels of varying sizes are applied in parallel to the input sequence. Suppose k convolutional kernels are used, with receptive field sizes $\mathrm{r}_1, \mathrm{r}_2, \ldots, \mathrm{r}_\mathrm{k}$, and each kernel produces feature maps with an output dimension of dc. The concatenated result of the multi-scale convolution outputs is:

$$H^{(t)} = \mathrm{Concat}\left[ReLU\left(Convr_1\left(X^{(t)}\right)\right), \ldots, \mathrm{ReLU}\left(Convr_k\left(X^{(t)}\right)\right)\right] \in R^{B\times(k\cdot d_c)\times w} \quad (4)$$

Since the output of the convolutional operation is in the format of (Batch, Channels, Time), while the subsequent multi head attention mechanism expects input in the format of (Batch, Time, Embedding), the convolutional output H(t) is transposed to $\mathrm{H}_{\mathrm{trans}}^{(\mathrm{t})} \in \mathrm{R}^{\mathrm{B}\times \mathrm{w}\times(\mathrm{k}\cdot \mathrm{d_c})}$. This transposed sequence is then fed into a multi head self-attention mechanism to model global dependencies across different time steps. The output is given by:

$$Z^{(t)} = \mathrm{MultiHeadAttn}\left(H_{trans}^{(t)}, H_{trans}^{(t)}, H_{trans}^{(t)}\right) \in R^{B\times w\times(k\cdot d_c)} \quad (5)$$

The tabular branch of the network takes as input a tensor $\mathrm{X}^{(\mathrm{b})} \in \mathrm{R}^{\mathrm{B}\times\mathrm{w}\times\mathrm{d_b}}$, where $\mathrm{d_b}$ denotes the number of tabular feature fields, including both categorical and continuous attributes. To effectively capture the interactions and semantic compositions among different fields, a Transformer based architecture is employed. Specifically, each categorical feature is mapped to a fixed dimensional embedding vector via an independent embedding layer, while continuous features are projected into the same space through linear transformation. The two types of features are then fused and fed into a multi-layer Transformer encoder to model their nonlinear interactions.

Let $\mathrm{x}_\mathrm{i}^{(\mathrm{d})} \in \{1,2,\ldots,\mathrm{V_i}\}$ denote the value of the i-th categorical feature at each time step, where $\mathrm{V_i}$ is the vocabulary size of that feature. Its corresponding embedding representation is:

$$e_i = Embed_i\left(x_i^{(d)}\right) \in R^{d_e} \tag{6}$$

After embedding all categorical fields, a categorical embedding matrix $\mathrm{E_{cat}} \in \mathrm{R}^{\mathrm{w}\times\mathrm{d_e}}$ is constructed at each time step, where $\mathrm{d_e}$ is the embedding dimension. Meanwhile, let $\mathrm{X_{num}} \in \mathrm{R}^{\mathrm{w}\times\mathrm{d_{num}}}$ represent the matrix of continuous features. These continuous fields are projected into the same embedding space via a linear transformation as follows:

$$E_{num} = X_{num} \cdot W_{num} + b_{num} \in R^{w\times d_e} \tag{7}$$

As shown in Equation (8), the fused representation of categorical and continuous features is denoted as $\mathrm{E}^{(\mathrm{b})}$, which is then fed into a multi-layer Transformer encoder (denoted as $\mathcal{T}(\cdot)$) to model contextual dependencies both across time steps and among feature fields, as given in Equation (9). The resulting output sequence of the tabular branch, denoted as $\mathrm{Z}^{(\mathrm{b})}$, provides a context enhanced representation of tabular features under the sliding window sequence. This representation is used in the subsequent cross-type attention fusion stage.

$$E^{(b)} = E_{cat} + E_{num} \in R^{w\times d_e} \tag{8}$$

$$Z^{(b)} = \mathcal{T}\left(E^{(b)}\right) \in R^{B\times w\times d_e} \tag{9}$$

To enable deep fusion between temporal and tabular features, a cross-type attention mechanism is introduced to align and interact between the two types. First, the output of the tabular branch $Z^{(b)}$ is linearly projected into the same dimensional space as the temporal branch, as follows:

$$\widetilde{Z^{(b)}} = Z^{(b)} \cdot W_p \in R^{B\times w\times(k\cdot d_c)} \tag{10}$$

where W_p is a learnable projection matrix. This projection ensures that the tabular feature representations $\widetilde{Z^{(b)}}$ have the same embedding dimensionality as the temporal features $Z^{(t)}$, which satisfies the dimensional consistency required by Multi-Head Attention for Query, Key, and Value. Next, the temporal features are used as Query, while the tabular features serve as Key and Value to perform the cross-type attention operation, as shown in Equation (11). Through this interaction, each temporal representation adaptively attends to the most semantically relevant tabular features, enabling dynamic alignment between temporal dynamics and structural attributes across heterogeneous feature spaces. This mechanism allows the model to emphasize contextually correlated behaviours (e.g., protocol timing dependencies) that static fusion methods would overlook. A residual connection and layer normalization are further applied to stabilize training and enhance feature representation, as shown in Equation (12).

$$A = \text{MultiHeadAttn}\left(Q = Z^{(t)},\ K = \widetilde{Z^{(b)}},\ V = \widetilde{Z^{(b)}}\right) \tag{11}$$

$$\widehat{Z^{(t)}} = \text{LayerNorm}\left(Z^{(t)} + \text{Dropout}(A)\right) \tag{12}$$

Finally, the cross-type enhanced temporal representation $\widehat{Z^{(t)}}$ is concatenated with the original tabular representation $Z^{(b)}$ along the feature dimension to form the final joint feature sequence F, as follows:

$$F = \text{Concat}\left(\widehat{Z^{(t)}}, Z^{(b)}\right) \in R^{B\times w\times(k\cdot d_c+d_e)}. \tag{13}$$

Unlike a simple concatenation followed by a dense layer, which performs static and position-independent feature fusion, the cross-type attention adaptively learns inter-type dependencies and assigns context-aware importance weights, leading to semantically aligned and more discriminative representations. The fused joint

representation F is processed by a multi-layer perceptron (MLP) that performs nonlinear transformation and dimensionality compression. Since F∈ $R^{B\times w\times(k\cdot d_c+d_e)}$ (Equation (13)), the input layer of the MLP has a dimension of $(k \cdot d_c + d_e)$, ensuring consistency with the concatenated joint feature space. The MLP outputs a tensor of raw prediction scores (logits), denoted as P (Equation (14)), where the dimension C corresponds to the number of classes. These logits are subsequently used for classification via a cross entropy loss function.

$$P = \mathrm{MLP}(F) \in R^{B\times w\times C}. \tag{14}$$

This architecture enables fine grained, frame level classification by jointly capturing temporal dynamics and structural semantics, which is critical for detecting complex and evasive attack behaviours in network traffic.

PREDICTION AGGREGATION

Network intrusions often manifest as temporally continuous patterns rather than isolated outliers. A single attack instance may span multiple consecutive traffic frames, with subtle variations in features and delayed transitions between benign and malicious states. This temporal continuity poses challenges for frame level prediction, particularly near attack boundaries or under noisy environments.

To mitigate the impact of temporal uncertainty and enhance detection robustness, we adopt a position index-based prediction aggregation strategy in the inference phase. It is important to note that this aggregation operates as a post-processing step independent of the CNN's temporal feature extraction. The sliding window mechanism provides temporal context during feature learning, while the aggregation step consolidates multiple independent predictions after inference, thereby addressing decision instability from a complementary perspective.

During sequence construction, each sliding window of length w and step size s generates frame wise predictions for the samples it contains. Consequently, a given time step t may appear in multiple overlapping windows and thus receive multiple predictions from different temporal contexts.

Let N denote the total number of samples in the original test sequence. For any sample at position t ∈ [1,N], denote its predicted labels from all containing windows as $\{y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(m)}\}$, where m (Equation (15)) is the number of windows that include position t.

$$m = |\{\mathrm{k} \mid t \in [k, k+w-1],\ k = 1,1+s, \ldots, N-w+1\}| \tag{15}$$

The final aggregated prediction is determined by majority voting over these candidate predictions:

$$\widehat{y_t} = \mathrm{MajorityVote}\left(\left\{y_t^{(j)}\right\}_{j=1}^{m}\right) \tag{16}$$

This aggregation scheme allows the model to leverage multi perspective temporal evidence, effectively smoothing inconsistent predictions caused by local perturbations or context ambiguity. It also reduces the influence of misclassifications in short windows and enhances stability near decision boundaries. Empirically, this strategy improves both the reliability and interpretability of final detection results, especially in cases involving stealthy or multistage intrusions where a single frame may not suffice for accurate classification.

EXPERIMENT

Dataset

To evaluate the effectiveness of H2IDF, we adopt the AWID dataset [37]. Although collected from IEEE 802.11 wireless networks, AWID is selected for its representation of heterogeneous wireless traffic patterns analogous to energy system communications. The dataset contains multi-dimensional features extracted from each frame, including MAC layer fields, Wi-Fi radio/physical layer parameters (e.g., signal strength, frame length, timing), and WLAN control and management frame fields, which share similarities with wireless communication channels in modern energy systems. In many modern energy and industrial environments, wireless communication segments—such as field-level sensor networks and edge gateways—increasingly rely on IEEE 802.11 based protocols to interconnect distributed devices. These wireless links exhibit multi-layer data exchanges, variable transmission power, and asynchronous control signaling similar to those captured in AWID. Therefore, while AWID does not directly represent industrial control protocols, it provides a representative and realistic proxy for evaluating intrusion detection frameworks targeting the wireless components of modern energy and industrial systems.

The AWID dataset was collected and organized by the Aegean Wi-Fi Intrusion Detection project at the University of the Aegean, led by Constantinos Kolias, using a realistic wireless testbed simulating enterprise-style 802.11 infrastructure. The test environment simulates a typical enterprise style WLAN deployment. Packet capture was conducted continuously using tools like Wireshark during live traffic. A total of 154 features (plus one class label) were extracted per record from the raw data, covering MAC layer and radio/physical layer attributes of IEEE 802.11.

In terms of attack categories, the AWID dataset includes four major types of wireless attacks: Flooding, Impersonation, Injection, and Normal traffic. According to the official standard, the dataset is split into training and testing sets and further provided in two formats: Full (F) and Reduced (R). The full version retains the entire captured traffic, while the reduced version compresses the sample size, making it more suitable for rapid experimentation and evaluation under resource constrained environments.

In this study, the reduced AWID-CLS-R dataset is selected as the evaluation benchmark. The training set (AWID-CLS- R-Trn) contains 1,795,575 samples, and the test set (AWID-CLS-R-Tst) includes 575,643 samples, covering all four attack categories as well as normal traffic. Table 1 presents the detailed sample distribution across different categories in the dataset.

Table 1. Sample Distribution of the AWID Dataset.

Label	AWID-CLS-R-Trn	AWID-CLS-R-Tst
Normal	1,633,190	530,785
Injection	65,379	16,682
Impersonation	48,522	20,079
Flooding	48,484	8,097
Total	1,795,575	575,643

EVALUATION METRIC

In intrusion detection tasks, detection performance is typically evaluated using four key metrics: Accuracy, Precision, Recall, and F1-score. Among them, Accuracy is the most commonly used overall metric, which

measures the proportion of correctly classified samples, i.e., the ratio of the number of correctly predicted samples to the total number of samples. It is calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

Precision measures the proportion of true positive samples among all samples predicted as positive by the model, reflecting the accuracy of positive predictions. It is calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{18}$$

Recall indicates the proportion of true positive samples that are correctly identified among all actual positive samples, reflecting the model's ability to capture positive instances. It is calculated as follows:

$$Recall = \frac{TP}{TP + FN} \tag{19}$$

F1-score is the harmonic mean of Precision and Recall, and is used to comprehensively evaluate the tradeoff between prediction accuracy and completeness. It is particularly suitable for classification tasks with imbalanced data. In intrusion detection, where the ratio between positive and negative samples is often highly skewed, F1-score provides a more reliable measure of actual performance than any single metric. It is calculated as follows:

$$F1 - score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{20}$$

IMPLEMENTATION

In this experiment, the AWID dataset is first preprocessed. A correlation analysis is performed on the raw features to eliminate redundant and irrelevant attributes, resulting in the selection of 23 representative features for modeling, as listed in Table 2. All features, except for the label, are categorized based on their semantic meaning and statistical properties into three types: temporal features, categorical features, and numerical features, with the detailed classification criteria presented in Table 3. Temporal features mainly include time related metrics, categorical features are discrete values suitable for embedding encoding, and numerical features are continuous or count based variables appropriate for normalization. Subsequently, label encoding is applied to both categorical features and attack type labels to ensure that all classification fields are represented as numeric values. To eliminate the influence of different value scales, numerical features are normalized using standardization, transforming them into distributions with zero mean and unit variance, which helps improve model training efficiency and generalization ability. In this study, tabular features are composed of both categorical and numerical fields. To handle missing values, different strategies are employed: temporal features are filled using mean imputation to maintain time series continuity, while tabular features are filled using the mode to preserve the original categorical distribution.

Table 2. Features Used in the AWID Dataset.

#	Features	#	Features
1	frame.time_epoch	13	wlan.fc.ds
2	frame.time_delta	14	wlan.fc.frag
3	frame.time_relative	15	wlan.fc.retry
4	frame.len	16	wlan.fc.pwrmgt
5	radiotap.mactime	17	wlan.fc.moredata
6	radiotap.datarate	18	wlan.fc.protected
7	radiotap.channel.freq	19	wlan.duration
8	radiotap.channel.type.cck	20	wlan.wep.key
9	radiotap.dbm_antsignal	21	wlan.qos.priority
10	wlan.fc.type_subtype	22	wlan.qos.bit4
11	wlan.fc.type	23	class
12	wlan.fc.subtype		

Table 3. Specific Feature Types.

Types	Features
Temporal	frame.time_epoch, frame.time_relative, wlan.duration, frame.time_delta, radiotap.mactime
Categorical	radiotap.channel.type.cck, wlan.fc.subtype, wlan.fc.ds, wlan.fc.protected, wlan.fc.moredata, wlan.fc.pwrmgt, wlan.qos.bit4, wlan.fc.type, wlan.fc.type_subtype,
Numerical	radiotap.dbm_antsignal, frame.len, radiotap.datarate, radiotap.channel.freq

The specific parameter settings used in the experiment are summarized in Table 4. H2IDF adopts a sliding window mechanism to construct sequence samples, with the window size set to 10 and the step size set to 1. In terms of model architecture, the temporal branch employs three parallel 1D convolutional kernels with receptive field sizes of 3, 5, and 7, respectively. Each kernel outputs 16 channels, and the concatenated result forms a 48-dimensional feature sequence, which is then passed into a multi head attention module with 4 attention heads. The tabular branch is built upon a Transformer architecture, where the tabular feature embedding dimension is set to 16. The Transformer encoder has a depth of 2 layers, each containing 4 attention heads. During the feature fusion stage, the temporally and tabularly aligned features obtained via cross-type attention are concatenated and passed through a fusion network consisting of 2 fully connected layers, with the fusion feature dimension set to 128. For model training, the batch size is set to 64, the number of training epochs is 50, the optimizer used is Adam, the learning rate is 0.001, and the loss function is cross entropy loss.

Table 4. Specific Experimental Parameters.

Parameter	Values
window_size	10
step	1
kernel_sizes	[3, 5, 7]
attention_heads	4
temp_embed_dim	16
attn_input_dim	48
tab_embed_dim	16
tab_depth	2
num_heads	4

fusion_dim	128
batch_size	64
num_epochs	50
learning_rate	0.001

RESULTS

To evaluate the effectiveness of the proposed H2IDF framework, we conduct comprehensive experiments on the AWID dataset, focusing on the multi class intrusion detection task. The detection results are summarized in the confusion matrix (Figure 2) and the per class performance chart (Figure 3), which respectively visualize the normalized classification outcomes and the precision, recall, and F1-score for each class.

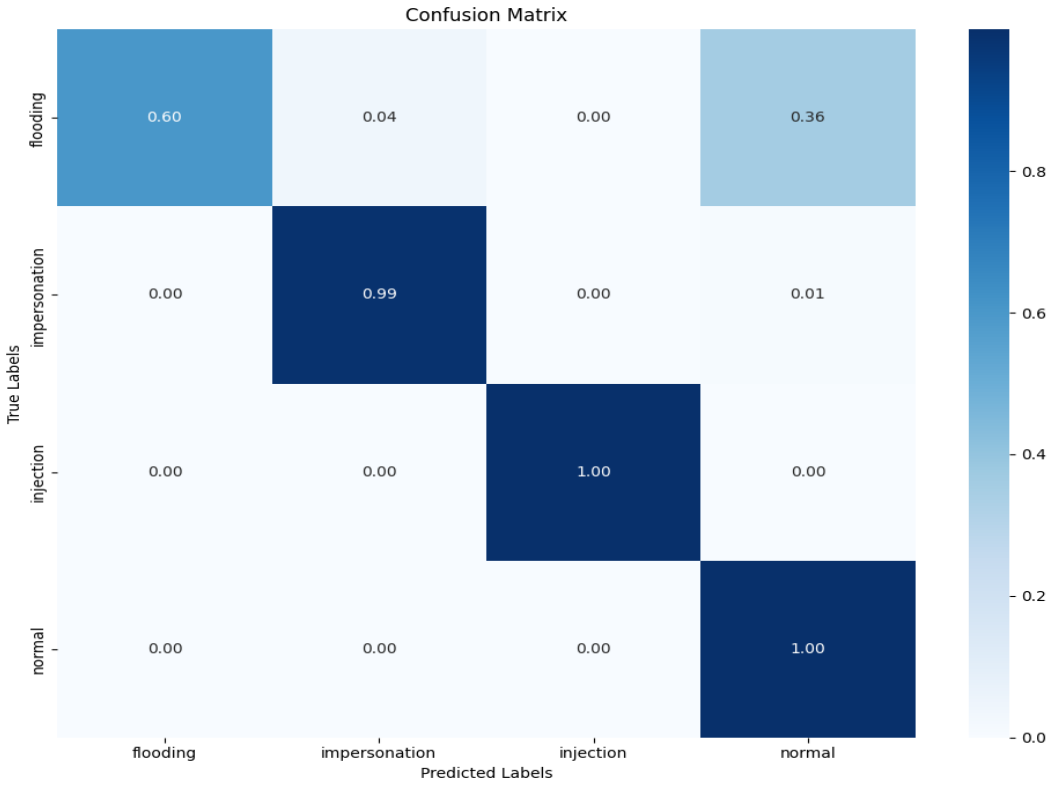

Figure 2. Confusion Matrix.

For the injection and impersonation attack types, the model obtains near perfect scores in all metrics, with both precision and recall exceeding 0.99, and F1-score close to 1.0. This suggests that H2IDF is highly effective at capturing the distinctive behavioural signatures of these attack types, which are typically stable and exhibit consistent temporal–tabular feature patterns. The model's ability to generalize such characteristics highlights its robustness and learning capability in identifying structured attack behaviours. The performance on the normal class is similarly outstanding, showing minimal confusion with attack classes. The classifier produces extremely low false positive and false negative rates, which is critical in real world scenarios where misclassify normal traffic as malicious can lead to unnecessary alarms and degraded trust in the system. For the flooding class, although the precision remains high (indicating low false positives), the recall is comparatively lower, and the F1-score shows slight degradation. As seen in the confusion matrix, a portion of flooding samples are misclassified as normal. This confusion can be attributed to temporal locality and the nature of flooding attacks whose rapid packet bursts may appear similar to legitimate high throughput traffic in short sliding windows. Moreover, the overlapping characteristics between flooding traffic and benign background noise could challenge the model's ability to discriminate at the window level. Nevertheless, the overall misclassification does not significantly compromise detection reliability, indicating the model's capacity to maintain high robustness even under ambiguous conditions.

In summary, H2IDF demonstrates strong performance on the AWID benchmark, achieving both high accuracy and class wise balance. The results validate the effectiveness of its hybrid architecture in modeling complex multi type feature distributions and capturing subtle behavioural distinctions over time. This makes the framework well suited for deployment in real-time, fine-grained intrusion detection scenarios with diverse threat types.

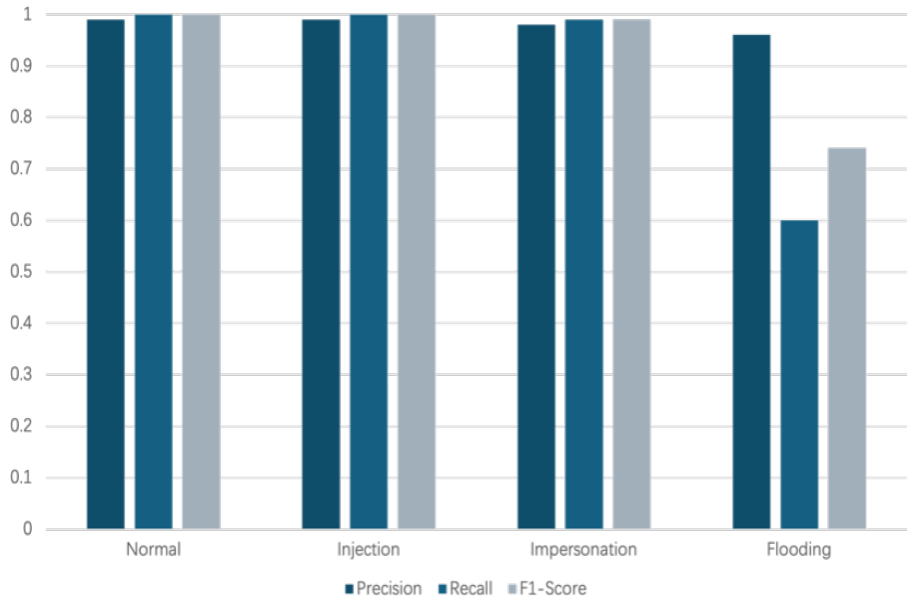

Figure 3. Classification Report.

To comprehensively evaluate the detection performance of the proposed H2IDF framework, we compare it against various representative baseline models, including traditional machine learning, deep learning based and also reinforcement learning approaches: DRL+RBFNN, 1D-CNN, J48, AE-RL, GINE, and Random Forest [38-43]. All baseline methods were implemented according to their original papers. The performance results are summarized in Table 5.

H2IDF achieves the best overall performance across all four evaluation metrics, with an accuracy of 0.9931, precision of 0.9929, recall of 0.9931, and an F1-score of 0.9925. Compared to strong performing baselines such as Random Forest (accuracy: 0.9910, F1-score: 0.9500) and GINE (accuracy: 0.9728, F1-score: 0.9620), H2IDF exhibits consistent improvements in both detection accuracy and classification robustness. In particular, H2IDF surpasses the second best model in terms of F1-score by more than 3 percentage points, demonstrating superior capability in maintaining a balance between precision and recall. This indicates its strong overall classification performance, especially in correctly identifying both attack and benign instances.

Table 5. Performance Comparison of Different Models.

Model	Accuracy	Precision	Recall	F1-score
DRL+RBFNN	0.9540	0.9260	0.9540	0.9380
1D-CNN	0.9550	0.9720	0.9340	0.9510
J48	0.9620	0.9620	0.9630	0.9480
AE-RL	0.9590	0.9720	0.9590	0.9629
GINE	0.9728	0.9508	0.9729	0.9620
Random Forest	0.9910	0.9600	0.9600	0.9500
H2IDF	0.9931	0.9929	0.9931	0.9925

The performance gain can be attributed to the cross-type attention mechanism employed in H2IDF, which effectively integrates heterogeneous temporal and tabular features. This hybrid design enhances the model's ability to capture subtle and complex intrusion patterns, leading to more accurate and stable detection outcomes across diverse attack scenarios.

CONCLUSION

This paper presents a novel Hybrid Intrusion Detection Framework with Heterogeneous Feature Fusion (H2IDF). H2IDF leverages the strength of CNNs in extracting local temporal dependencies and the capability of Transformers in modeling structured tabular information. To bridge the semantic gap between feature types, a cross-type attention mechanism is introduced, enabling deep fusion and alignment of temporal and tabular representations. Furthermore, a prediction aggregation strategy is employed to enhance robustness against noisy or ambiguous samples by integrating multi view predictions from overlapping sliding windows. Experiments on the AWID dataset demonstrate that H2IDF outperforms a range of competitive baseline models across accuracy, precision, recall, and F1-score. In particular, it shows strong resilience in handling class imbalance and overlapping attack behaviours, highlighting its practical applicability in real world network environments, including heterogeneous communication scenarios found in modern energy systems. Crucially, these results underscore the framework's potential significance for the textile industry, where the principles of Textile 4.0 are driving the adoption of similar interconnected and data-rich operational environ-

ments. The ability of H2IDF to secure complex industrial networks is directly transferable to protecting automated textile manufacturing plants from cyber threats, thereby ensuring production integrity, safeguarding intellectual property, and securing the supply chain.

Future research will focus on extending H2IDF to more complex, large-scale network settings involving multi-source heterogeneous data and dynamic adversarial behaviours. In particular, as the current evaluation is based on an earlier dataset, future work will incorporate more recent and diversified datasets to better reflect evolving attack techniques and zero-day threats. In addition, future research will explore replacing the raw timestamp with more advanced temporal encodings (e.g., periodic or relative time representations) to enhance robustness against capture-time variations and improve generalization across different acquisition sessions. As the primary focus of this study is on validating the detection capability of the proposed framework, future work will further investigate computational efficiency and inference latency to facilitate real-time deployment in practical network environments. Furthermore, future directions include adapting the framework to industrial control and smart grid environments, as well as exploring its application in specialized textile contexts—for example, securing networked smart looms or protecting the data integrity of wearable biometric sensors woven into smart fabrics. In addition, we plan to incorporate model interpretability techniques to facilitate transparent and explainable intrusion detection, thereby supporting more trustworthy and adaptive protection of critical energy infrastructures, from power grids to the next generation of smart textiles.

Author contributions

Z.Q. and M.H. wrote the main manuscript text. F.L. and Z.Q. designed the research framework and developed the methodology. B.L., X.L. and C.Z. performed the experiments and analyzed the data. G.F. and Y.H. contributed to the interpretation of results and revised the manuscript critically for important intellectual content. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by China Southern Power Grid's Major Network-level Scientific and Technological Project "Research and Application of Multi-dimensional Active Defense Technology for Digital Grid", project number 037800KC24040002 (GDKJXM20240428).

Data availability

The AWID dataset utilized in this study is publicly accessible at the official website of the Aegean Wi-Fi Intrusion Dataset project: https://icsdweb.aegean.gr/awid.

REFERENCES

[1] Khalid M. Smart grids and renewable energy systems: Perspectives and grid integration challenges. Energy Strategy Reviews. 2024; 51:101299. doi: 10.1016/j.esr.2024.101299.

[2] Mahmood M, Chowdhury P, Yeassin R, Hasan M, Ahmad T, Chowdhury NUR. Impacts of digitalization on smart grids, renewable energy, and demand response: An updated review of current applications. Energy Conversion and Management: X. 2024; 24:100790. doi: 10.1016/j.ecmx.2024.100790.

[3] Kaloudi N, Li J. The AI-Based Cyber Threat Landscape. ACM Computing Surveys (CSUR). 2020; 53(1):1–34. doi: 10.1145/3372823.

[4] Ferdous J, Islam R, Mahboubi A, Islam Md Z. A Review of State-of-the-Art Malware Attack Trends and Defense Mechanisms. IEEE Access. 2023; 11:121118–121141. doi: 10.1109/access.2023.3328351.

[5] Guembe B, Azeta A, Misra S, Osamor VC, Fernandez-Sanz L, Pospelova V. The Emerging Threat of Ai-driven Cyber Attacks: A Review. Applied Artificial Intelligence. 2022; 36(1):2037254. doi: 10.1080/08839514.2022.2037254.

[6] Abdi N, Albaseer A, Abdallah M. The Role of Deep Learning in Advancing Proactive Cybersecurity Measures for Smart Grid Networks: A Survey. IEEE Internet of Things Journal. 2024; 11(9):16398–16421. doi: 10.1109/jiot.2024.3354045.

[7] Diaba SY, Shafie-khah M, Elmusrati M. Cyber-physical attack and the future energy systems: A review. Energy Reports. 2024; 12:2914–2932. doi: 10.1016/j.egyr.2024.08.060.

[8] Merlino V, Allegra D. Energy-based approach for attack detection in IoT devices: A survey. Internet of Things. 2024; 27:101306. doi: 10.1016/j.iot.2024.101306.

[9] Sharma A, Lashkari AH. A survey on encrypted network traffic: A comprehensive survey of identification/classification techniques, challenges, and future directions. Computer Networks. 2025; 257:110984. doi: 10.1016/j.comnet.2024.110984.

[10] Mutalib NHA, Sabri AQM, Wahab AWA, Abdullah ERMF, AlDahoul N. Explainable deep learning approach for advanced persistent threats (APTs) detection in cybersecurity: a review. Artificial Intelligence Review. 2024; 57(11):297. doi: 10.1007/s10462-024-10890-4.

[11] Bayoudh K. A survey of multimodal hybrid deep learning for computer vision: Architectures, applications, trends, and challenges. Information Fusion. 2024; 105:102217. doi: 10.1016/j.inffus.2023.102217.

[12] Minaee S, Abdolrashidi A, Su H, Bennamoun M, Zhang D. Biometrics recognition using deep learning: a survey. Artificial Intelligence Review. 2023; 56(8):8647–8695. doi: 10.1007/s10462-022-10237-x.

[13] Lauriola I, Lavelli A, Aiolli F. An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools. Neurocomputing. 2022; 470:443–456. doi: 10.1016/j.neucom.2021.05.103.

[14] Chib PS, Singh P. Recent Advancements in End-to-End Autonomous Driving Using Deep Learning: A Survey. IEEE Transactions on Intelligent Vehicles. 2024; 9(1):103–118. doi: 10.1109/tiv.2023.3318070.

[15] Piccialli F, Somma VD, Giampaolo F, Cuomo S, Fortino G. A survey on deep learning in medicine: Why, how and when? Information Fusion. 2021; 66:111–137. doi: 10.1016/j.inffus.2020.09.006.

[16] da Silva Ruffo VG, Lent DMB, Komarchesqui M, Schiavon VF, de Assis MVO, Carvalho LF, et al. Anomaly and intrusion detection using deep learning for software-defined networks: A survey. Expert Systems with Applications. 2024; 256:124982. doi: 10.1016/j.eswa.2024.124982.

[17] Kheddar H, Dawoud DW, Awad AI, Himeur Y, Khan MK. Reinforcement-Learning-Based Intrusion Detection in Communication Networks: A Review. IEEE Communications Surveys & Tutorials. 2025; 27(4):2420–2469. doi: 10.1109/comst.2024.3484491.

[18] Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural computation. 1997; 9(8):1735–1780. doi: 10.1162/neco.1997.9.8.1735.

[19] LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation Applied to Handwritten Zip Code Recognition. Neural Computation. 1989; 1(4):541–551. doi: 10.1162/neco.1989.1.4.541.

[20] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Nets. In: Proceedings of the 28th Annual Conference on Neural Information Processing Systems (NIPS 2014); 8-13 Dec 2014; Montréal, QC, Canada. New York, NY: Curran Associates, Inc.; 2014. p. 2672-2680.

[21] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need. In: Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS 2017); 4-9 Dec 2017; Long Beach, CA, USA. New York, NY: Curran Associates, Inc.; 2017. p. 5998-6008.

[22] Bakhsh SA, Khan MA, Ahmed F, Alshehri MS, Ali H, Ahmad J. Enhancing IoT network security through deep learning-powered Intrusion Detection System. Internet of Things. 2023; 24:100936. doi: 10.1016/j.iot.2023.100936.

[23] Attou H, Guezzaz A, Benkirane S, Azrour M, Farhaoui Y. Cloud-Based Intrusion Detection Approach Using Machine Learning Techniques. Big Data Mining and Analytics. 2023; 6(3):311–320. doi: 10.26599/bdma.2022.9020038.

[24] Turukmane AV, Devendiran R. M-MultiSVM: An efficient feature selection assisted network intrusion detection system using machine learning. Computers & Security. 2024; 137:103587. doi: 10.1016/j.cose.2023.103587.

[25] Zhang Z, Zeng L, Zhu D, Tan H, Wang L, Li Z, et al. GBFKAN: An Adaptive Multilayer Interpretable Architecture for Intrusion Detection in Various Internet of Things Scenarios. IEEE Internet of Things Journal. 2025; 12(15):30379–30397. doi: 10.1109/jiot.2025.3570033.

[26] Saheed YK, Abdulganiyu OH, Tchakoucht TA. Modified genetic algorithm and fine-tuned long short-term memory network for intrusion detection in the internet of things networks with edge capabilities. Applied Soft Computing. 2024; 155:111434. doi: 10.1016/j.asoc.2024.111434.

[27] Hassini K, Khalis S, Habibi O, Chemmakha M, Lazaar M. An end-to-end learning approach for enhancing intrusion detection in Industrial-Internet of Things. Knowledge-Based Systems. 2024; 294:111785. doi: 10.1016/j.knosys.2024.111785.

[28] Tan H, Wang L, Zhu D, Deng J. Intrusion Detection Based on Adaptive Sample Distribution Dual-Experience Replay Reinforcement Learning. Mathematics. 2024; 12(7):948. doi: 10.3390/math12070948.

[29] Boppana TK, Bagade P. GAN-AE: An unsupervised intrusion detection system for MQTT networks. Engineering Applications of Artificial Intelligence. 2023; 119:105805. doi: 10.1016/j.engappai.2022.105805.

[30] de Araujo-Filho PF, Naili M, Kaddoum G, Fapi ET, Zhu Z. Unsupervised GAN-Based Intrusion Detection System Using Temporal Convolutional Networks and Self-Attention. IEEE Transactions on Network and Service Management. 2023; 20(4):4951–4963. doi: 10.1109/tnsm.2023.3260039.

[31] Kabilan, N, Ravi V, Sowmya V. Unsupervised intrusion detection system for in-vehicle communication networks. Journal of Safety Science and Resilience. 2024; 5(2):119–129. doi: 10.1016/j.jnlssr.2023.12.004.

[32] Yuan L, Sun J, Zhuang S, Liu Y, Geng L, Zou J, et al. Manticore: An Unsupervised Intrusion Detection System Based on Contrastive Learning in 5G Networks. In: Proceedings of the 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2024); 14-19 Apr 2024; Seoul, Korea. Piscataway, NJ: IEEE; 2024. p. 4705-4709. doi: 10.1109/ICASSP48485.2024.10447814.

[33] Li S, Cao Y, Liu S, Lai Y, Zhu Y, Ahmad N. HDA-IDS: A Hybrid DoS Attacks Intrusion Detection System for IoT by using semi-supervised CL-GAN. Expert Systems with Applications. 2024; 238:122198. doi: 10.1016/j.eswa.2023.122198.

[34] Nguyen TP, Cho J, Kim D. Semi-supervised intrusion detection system for in-vehicle networks based on variational autoencoder and adversarial reinforcement learning. Knowledge-Based Systems. 2024; 304:112563. doi: 10.1016/j.knosys.2024.112563.

[35] Hossain S, Senouci SM, Brik B, Boualouache A. A privacy-preserving Self-Supervised Learning-based intrusion detection system for 5G-V2X networks. Ad Hoc Networks. 2025; 166:103674. doi: 10.1016/j.adhoc.2024.103674.

[36] Duy PT, Hien DTT, Luong TD, Quyen NH, Pham VH. Fed-Evolver: An automated evolving approach for federated Intrusion Detection System using adversarial autoencoder in SDN-enabled networks. Internet of Things. 2024; 28:101397. doi: 10.1016/j.iot.2024.101397.

[37] Kolias C, Kambourakis G, Stavrou A, Gritzalis S. Intrusion Detection in 802.11 Networks: Empirical Evaluation of Threats and a Public Dataset. IEEE Communications Surveys & Tutorials. 2015; 18(1):184–208. doi: 10.1109/comst.2015.2402161.

[38] Lopez-Martin M, Sanchez-Esguevillas A, Arribas JI, Carro B. Network Intrusion Detection Based on Extended RBF Neural Network With Offline Reinforcement Learning. IEEE Access. 2021; 9:153153–153170. doi: 10.1109/access.2021.3127689.

[39] Natkaniec M, Bednarz M. Wireless Local Area Networks Threat Detection Using 1D-CNN. Sensors. 2023; 23(12):5507. doi: 10.3390/s23125507.

[40] Lopez-Martin M, Carro B, Sanchez-Esguevillas A. Application of deep reinforcement learning to intrusion detection for supervised problems. Expert Systems with Applications. 2020; 141:112963. doi: 10.1016/j.eswa.2019.112963.

[41] Caminero G, Lopez-Martin M, Carro B. Adversarial environment reinforcement learning algorithm for intrusion detection. Computer Networks. 2019; 159:96–109. doi: 10.1016/j.comnet.2019.05.013.

[42] Jiang Z, Li J, Hu Q, Meng W, Pedrycz W, Su Z. Scalable Graph-Aware Edge Representation Learning for Wireless IoT Intrusion Detection. IEEE Internet of Things Journal. 2024; 11(16):26955–26969. doi: 10.1109/jiot.2024.3397364.

[43] Vaca FD, Niyaz Q. An Ensemble Learning Based Wi-Fi Network Intrusion Detection System (WNIDS). In: Proceedings of the 2018 IEEE 17th International Symposium on Network Computing and Applications (NCA 2018); 1-3 Nov 2018; Cambridge, MA, USA. Piscataway, NJ: IEEE; 2018. p. 1-5. doi: 10.1109/nca.2018.8548315.