Hybrid Deep Architectures in Contrastive Latent Space: Performance Analysis of VAE-MLP, VAE-MoTE, and VAE-GAT for IoT Botnet Detection

Hassan Wasswa * and Timothy Lynar

Posted Date: 19 March 2026

doi: 10.20944/preprints202603.1461.v1

Keywords: IoT botnet detection; graph neural networks; graph attention network; mixture of experts; mixture of tiny experts; contrastive learning; variational autoencoder; latent representation learning

 doi:10.20944/preprints202603.1461.v1

Article

Hybrid Deep Architectures in Contrastive Latent Space: Performance Analysis of VAE-MLP, VAE-MoTE, and VAE-GAT for IoT Botnet Detection

Hassan Wasswa * and Timothy Lynar

School of Systems and Computing, University of New South Wales, Canberra, 2600, ACT, Australia
* Correspondence: h.wasswa@unsw.edu.au

Abstract

The rapid proliferation of Internet of Things (IoT) devices has significantly expanded the attack surface of modern networks, leading to a surge in IoT-based botnet attacks. Detecting such attacks remains challenging due to the high dimensionality and heterogeneity of IoT network traffic. This study proposes and evaluates three hybrid deep learning architectures for IoT botnet detection that combine representation learning with supervised classification: VAE-encoder-MLP, VAE-encoder-GAT, and VAE-encoder-MoTE. A variational autoencoder (VAE) is first trained to learn a compact latent representation of high-dimensional traffic features, after which the pretrained encoder projects the data into a low-dimensional embedding space. These embeddings are then used to train three different downstream classifiers: a multilayer perceptron (MLP), a graph attention network (GAT), and a mixture of tiny experts (MoTE) model. To further enhance representation discriminability, supervised contrastive learning is incorporated to encourage intra-class compactness and inter-class separability in the latent space. The proposed architectures are evaluated on two widely used benchmark datasets, CICIoT2022 and N-BaIoT, under both binary and multiclass classification settings. Experimental results demonstrate that all three models achieve near-perfect performance in binary attack detection, with accuracy exceeding 99.8%. In the more challenging multiclass scenario, the VAE-encoder-MLP model achieves the best overall performance, reaching accuracies of 98.55% on CICIoT2022 and 99.75% on N-BaIoT. These findings provide insights into the design of efficient and scalable deep learning architectures for IoT intrusion detection.

Keywords: IoT botnet detection; graph neural networks; graph attention network; mixture of experts; mixture of tiny experts; contrastive learning; variational autoencoder; latent representation learning

1. Introduction

The Internet of Things (IoT) has become an indispensable component of modern society, underpinning a wide range of applications at both individual and enterprise levels. However, despite their widespread adoption, the IoT network remains highly susceptible to cyberattacks due to inherent device constraints of such as limited computational resources, weak security configurations, and heterogeneous deployment environments. In particular, IoT botnet–driven attacks, most notably Distributed Denial of Service (DDoS) attacks, continue to pose significant security threats to enterprises, institutions, and organizations across diverse industry verticals. To mitigate these challenges, a broad spectrum of artificial intelligence (AI)–based IoT botnet detection approaches has been proposed and extensively evaluated to enhance the resilience of IoT ecosystems. For example, prior studies have explored advanced AI-driven techniques based on federated learning [1–3], graph neural networks [4–6], transformer-based architectures [7–11], autoencoder and variational autoencoder models [12–15], latent space arithmetic and alignment–based methods [16–19], and cost-sensitive learning strategies designed to address class imbalance in IoT security datasets [20–23].

 doi:10.20944/preprints202603.1461.v1

However, despite its demonstrated potential for improving computational efficiency in domains such as natural language processing (NLP) and large language models (LLMs) [24–26], the Mixture of Experts (MoE) [27,28] paradigm remains relatively unexplored in the context of IoT botnet detection. The MoE architecture enables conditional computation by activating only a small subset of expert networks for each input instance. To facilitate deployment on resource-constrained IoT edge devices, a variant of the MoE,—the mixture of tiny experts (MoTE) [29], which utilizes lightweight expert networks for instance-level classification–is employed. This mechanism substantially reduces inference latency, energy consumption, and memory overhead, while preserving strong representational capacity through expert specialization.

In this work, we propose a VAE-encoder-MoTE framework in which a VAE encoder first transforms high-dimensional IoT botnet traffic features into a compact latent embedding space. The MoTE model is then trained on these low-dimensional embeddings, enabling efficient learning while retaining the critical structural information necessary for accurate attack detection. This combination of latent representation learning and conditional expert activation provides an effective balance between detection performance and computational efficiency, making the approach particularly suitable for deployment in resource-constrained IoT environments.

To further improve the discriminative quality of the learned representations, supervised contrastive learning [30,31] is incorporated during model training. This learning strategy explicitly encourages samples belonging to the same attack class to cluster closely in the latent space while simultaneously pushing samples from different classes further apart. As a result, the learned feature space exhibits improved intra-class compactness and inter-class separability, which enhances generalization capability, robustness to class imbalance, and resilience to evolving IoT traffic patterns. In addition to the proposed VAE-encoder-MoTE architecture, two alternative models, namely VAE-encoder-MLP and VAE-encoder-GAT, are also trained and evaluated. A comprehensive comparative analysis of the three architectures is conducted using accuracy, precision, recall, and F1-score as the primary performance metrics.

The main contributions of this study are:

- A hybrid framework that combines VAE–based latent representation learning with an MoTE architecture for IoT botnet detection is introduced. The approach compresses high-dimensional IoT traffic data into compact embeddings and applies conditional expert routing to enable efficient and scalable detection.
- The study incorporates supervised contrastive learning to enhance the discriminative power of the latent feature space. By encouraging embeddings of samples from the same attack class to cluster together while separating different classes, the approach improves intra-class compactness and inter-class separability.
- A systematic experimental evaluation is conducted comparing the proposed VAE-encoder-MoTE model against the VAE-encoder-MLP and VAE-encoder-GAT models. The comparison assesses their effectiveness under both binary and multiclass IoT botnet detection scenarios. Performance is analyzed using multiple metrics providing a detailed understanding of the strengths and limitations of each architectural design.
- Through extensive empirical validation on benchmark IoT botnet datasets, the study provides insights into how architectural choices and dataset characteristics influence detection performance in IoT intrusion detection systems.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature on supervised contrastive learning, mixture-of-experts architectures, and graph neural networks for intrusion detection. Section 3 presents the proposed framework, including the VAE-based representation learning approach, the MoTE architecture, and the experimental setup. Section 4 reports the experimental results obtained under both binary and multiclass classification settings. Section 5 provides an in-depth analysis and interpretation of the findings. Finally, Section 6 summarizes the key contributions and outlines directions for future research.

 doi:10.20944/preprints202603.1461.v1

2. Related Work

This section surveys existing literature relevant to three key areas: (1) supervised contrastive learning, (2) mixture of experts and tiny expert architectures, and (3) graph neural networks for intrusion detection.

2.1. Supervised Contrastive Learning

Contrastive learning is a representation learning framework in which models are trained to minimize the distance between related samples (positive pairs) while maximizing the separation between unrelated samples (negative pairs) in an embedding space. Supervised contrastive learning extends this paradigm by incorporating label information to define richer semantic relationships between samples, unlike purely unsupervised approaches that rely only on data augmentations without considering class labels [30,32]. By encouraging samples belonging to the same class to cluster together in the embedding space, supervised contrastive objectives often produce features that are more discriminative and transferable for classification tasks [30,33]. In comparison with traditional supervised objectives such as cross-entropy, supervised contrastive learning has been shown to improve resilience to data augmentation, label noise, and certain dataset biases [30,34]. Furthermore, contrastive objectives tend to yield smoother and better-structured representation spaces, which are advantageous for transfer learning and downstream applications [32,35].

The study in [31] investigates the geometric structure of representations learned using supervised contrastive loss in comparison with those obtained through conventional cross-entropy training. The authors analyze how both loss functions encourage the formation of compact class clusters within the embedding space and theoretically demonstrate that optimal representations under both objectives converge to vertices of a regular simplex positioned on a hypersphere. In this configuration, samples from the same class become tightly grouped while samples from different classes remain maximally separated. The study also provides empirical evidence linking this geometric structure to improved generalization performance.

The work presented in [36] proposes a weakly supervised contrastive learning framework designed to bridge instance discrimination with supervised information. Rather than treating each sample as an independent class, the proposed method introduces two projection heads: one dedicated to conventional instance discrimination and another that utilizes graph-based similarity relationships to assign weak labels across samples. Samples sharing these weak labels are considered positive pairs under a supervised contrastive objective, thereby encouraging the learning of semantically meaningful representations. To further enrich positive sample diversity, the framework incorporates a k-nearest neighbor multi-crop strategy. Experiments conducted on standard computer vision benchmarks demonstrated improvements in representation quality and competitive semi-supervised classification performance when only limited labeled data are available.

In cybersecurity applications, particularly intrusion detection, the work in [37] introduced FeCo (Federated Contrastive Learning), a framework that employs contrastive objectives to align feature embeddings of network traffic collected from distributed IoT devices. The approach coordinates multiple local models so that they learn shared representations capable of distinguishing normal network behavior from malicious activity by leveraging both feature and label similarities during contrastive training. Through the federated contrastive loss, the encoder learns an embedding space where semantically related network traffic patterns are positioned closer together, thereby improving the effectiveness of intrusion detection across heterogeneous deployment environments.

The research in [38] further extends contrastive learning principles to sequence modeling and transformer-based architectures. In this approach, sequences of network events are encoded using transformer models, and contrastive learning is employed to differentiate benign from malicious patterns. Positive pairs are constructed from similar event sequences, whereas negative pairs represent dissimilar sequences. This supervised contrastive framework improves the discriminative capacity

 doi:10.20944/preprints202603.1461.v1

and robustness of intrusion detection systems, especially in scenarios where labeled data are scarce, highlighting the adaptability of contrastive learning beyond traditional computer vision domains.

2.2. Mixture of Experts and Tiny Expert Architectures

The Mixture of Experts (MoE) framework represents a form of conditional computation in which a neural network is partitioned into multiple expert sub-networks that are orchestrated by a routing or gating mechanism. For each input instance, only a subset of experts is activated, allowing the model to increase its representational capacity while avoiding the full computational overhead associated with dense model inference. This strategy has been widely adopted to address scalability and efficiency challenges in deep learning, particularly in applications requiring large model capacity or dealing with heterogeneous input distributions [39,40].

Recent developments have introduced the concept of tiny experts, where each expert module is deliberately designed to be extremely lightweight. These experts often consist of minimal parameter structures such as shallow multilayer perceptrons or even simple single-layer transformations. The work in [29] formalizes this concept through the Mixture of a Million Experts framework, demonstrating that a very large collection of small experts can be efficiently accessed using parameter-efficient retrieval strategies while activating only a limited subset for each input. Related research has further explored fine-grained expert decomposition and sparse activation mechanisms to minimize redundancy and enhance parameter efficiency [41–43].

Tiny expert architectures provide several advantages compared with conventional MoE models, particularly in environments with constrained computational resources such as IoT edge devices used for botnet detection. Traditional MoE designs typically employ moderately large expert networks, which can result in increased memory consumption and higher inference latency, making them unsuitable for resource-limited edge hardware. In contrast, architectures composed of numerous tiny experts enable highly selective activation, lower computational cost per expert, and greater flexibility in adapting to dynamic traffic patterns. These characteristics make them particularly suitable for lightweight intrusion and botnet detection systems [44–48]. Such properties are critical in IoT security applications, where real-time threat detection, low power consumption, and resilience against evolving attack strategies are essential.

2.3. Graph Neural Networks for Intrusion Detection

Recent studies have increasingly explored hybrid architectures that combine graph neural networks (GNNs) with attention-based transformer models to improve intrusion and anomaly detection. These hybrid approaches exploit the complementary strengths of the two paradigms: GNNs effectively capture complex node–edge relationships within network structures, while transformers excel at modeling long-range contextual dependencies. The integration of these capabilities has been shown to enhance detection accuracy and robustness. For instance, the work in [49] proposed a GAT-based intrusion detection system tailored for heterogeneous IoT environments and demonstrated promising performance on the NSL-KDD dataset. Similarly, [50] developed a hybrid architecture that integrates a GNN with a transformer to jointly capture structural and contextual dependencies, while [51] combined graph convolution with attention mechanisms to detect IoT botnets through device interaction modeling. In the context of electric vehicle charging networks, [52] further demonstrated that integrating transformer components with GNNs improves the modeling of complex feature relationships for cyber-attack detection.

A number of studies have also focused on designing advanced GNN architectures specifically for IoT security applications. For example, the EGAT-LSTM model introduced in [53] combines an enhanced graph attention network with a long short-term memory (LSTM) module to capture both spatial relationships and temporal traffic patterns. This integration significantly improves malicious traffic classification compared with approaches that rely solely on individual flow features. Similarly, [54] proposed AJSAGE, an attention-enhanced GraphSAGE model designed to improve anomalous node detection in graph-structured network attack datasets, particularly for complex and evolving threats.

 doi:10.20944/preprints202603.1461.v1

Dimensionality reduction techniques have also been examined as complementary strategies for enhancing GNN-based detection performance. The study in [15] compared AE-encoder, VAE-encoder, and PCA methods for reducing the 84-dimensional CICIoT2022 feature space to 8 dimensions before graph construction. Among these approaches, the VAE-encoder produced the best results. In addition, the 3-euclidean "n_neighbors"-metric configuration achieved the strongest performance for kNN-based graph generation, emphasizing the importance of both high-quality latent representations and appropriate graph construction parameters.

Despite these advances, comparative evaluations suggest that GNN-based models still require further refinement to achieve competitive performance. In [4], the authors compared VAE-GCN and VAE-GAT against VAE-MLP and ViT-MLP models using the 115-dimensional N-BaIoT dataset reduced to an 8-dimensional latent space. Although VAE-GAT demonstrated better performance than VAE-GCN, both graph-based models were generally outperformed by the alternative deep learning architectures. These findings highlight the need for improved architectural designs and more effective integration strategies when applying GNNs to intrusion detection systems.

3. Methodology

This study proposed the VAE-encoder-MoTE model and systematically evaluates its detection performance against two other advanced hybrid deep learning architectures for IoT botnet detection: VAE-encoder-MLP and VAE-encoder-GAT, under both binary and multiclass classification settings. To improve the discriminative quality of the learned representations, supervised contrastive learning is employed to promote intra-class compactness and inter-class separability, thereby enhancing detection performance.

To address the challenge of high-dimensional IoT traffic data, a variational autoencoder (VAE), with latent space dimension k ($k = 8$ [4,11,15] in this work), is first trained on the original feature space, after which its encoder component (VAE-encoder) is retained. For each architecture, the pretrained VAE-encoder projects the high-dimensional input data into a low-dimensional embedding space. The resulting low-dimensional training data are subsequently used to train an MLP, a GAT, and an MoTE model within a supervised contrastive learning framework. A detailed description of each of the three architectures is provided in the subsequent subsections.

3.1. Supervised Contrastive Learning

To improve intra-class compactness and inter-class separability in the learned embedding space, supervised contrastive learning is adopted. Given a batch of feature embeddings $\{z_i\}_{i=1}^{B}$, where $z_i \in \mathbb{R}^k$, ($k \rightarrow$ latent space dimension), the embeddings are first ℓ_2-normalized:

$$\tilde{z}_i = \frac{z_i}{\|z_i\|_2} \tag{1}$$

The pairwise cosine similarity between samples i and j is computed as:

$$\mathrm{sim}(i,j) = \frac{\tilde{z}_i^{\top} \tilde{z}_j}{\tau} \tag{2}$$

where τ is a temperature scaling parameter.

The supervised contrastive loss for a sample i is defined as:

$$\mathcal{L}_{\mathrm{SupCon}}^{(i)} = -\frac{1}{|\mathcal{P}(i)|} \sum_{p \in \mathcal{P}(i)} \log \frac{\exp(\mathrm{sim}(i,p))}{\sum_{a \neq i} \exp(\mathrm{sim}(i,a))} \tag{3}$$

where $\mathcal{P}(i)$ denotes the set of indices corresponding to samples that share the same class label as sample i.

 doi:10.20944/preprints202603.1461.v1

The final supervised contrastive loss is obtained by averaging over the batch:

$$\mathcal{L}_{\text{SupCon}} = \frac{1}{B} \sum_{i=1}^{B} \mathcal{L}_{\text{SupCon}}^{(i)} \tag{4}$$

3.2. Tiny Experts Architecture

Each tiny expert is implemented as a lightweight multilayer perceptron (MLP) composed of two fully connected layers with 24 and 16 neurons, respectively, each employing ReLU activation. This compact architectural design reduces computational overhead while preserving sufficient representational capacity for effective feature transformation. The transformation performed by an expert is given by:

$$\text{Expert}(x) = W_2\, \sigma(W_1 x + b_1) + b_2 \tag{5}$$

where W_1, W_2 are weight matrices, b_1 and b_2 are bias terms, $\sigma(\cdot)$ denotes the ReLU activation function.

3.3. Mixture of Tiny Experts (MoTE)

The Mixture of Tiny Experts model introduces conditional computation through a trainable routing mechanism. Given an input x, the router produces a probability distribution over E experts:

$$g(x) = \text{softmax}(W_r x + b_r) \tag{6}$$

where $g(x) \in \mathbb{R}^E$, and $E \rightarrow$ number of experts.

For computational efficiency, only the top-k experts with the highest routing probabilities are activated for each input. The aggregated embedding is computed as a weighted sum of the selected expert outputs:

$$z = \sum_{i \in \mathcal{K}(x)} g_i(x) \cdot \text{Expert}_i(x) \tag{7}$$

where $\mathcal{K}(x)$ denotes the set of top-k selected experts.

3.4. Classification Head and Joint Optimization

The aggregated embedding z is passed to a linear classifier to obtain class predictions:

$$\hat{y} = \text{softmax}(W_c z + b_c) \tag{8}$$

The supervised classification objective is defined using the cross-entropy loss:

$$\mathcal{L}_{\text{CE}} = -\sum_{c=1}^{C} y_c \log(\hat{y}_c) \tag{9}$$

To jointly optimize representation learning and classification performance, the total training objective is formulated as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{SupCon}} + \mathcal{L}_{\text{CE}} \tag{10}$$

Figure 1 illustrates the complete architectural framework of the proposed VAE-encoder-MoTE model. A supervised contrastive learning layer is incorporated to enhance class separability within the latent space by drawing instances belonging to the same class closer together while simultaneously pushing instances from different classes farther apart.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 March 2026 doi:10.20944/preprints202603.1461.v1

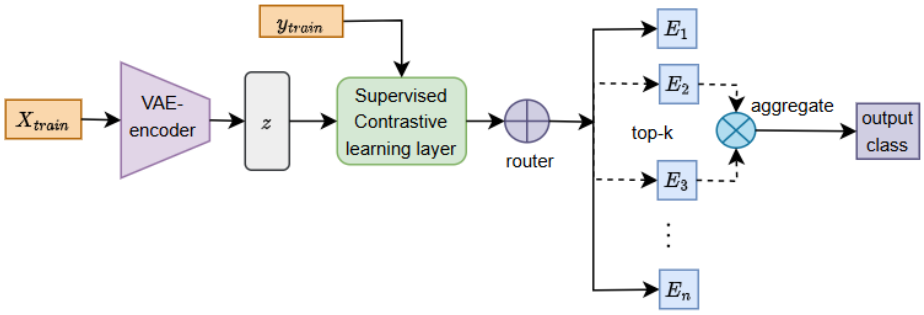

Figure 1. Proposed VAE-encoder-MoTE model with supervised contrastive learning on the low dimensional latent space embeddings.

3.5. The MLP Architecture

The MLP architecture comprised four fully connected hidden layers with 128, 64, 32, and 16 neurons, respectively. Each hidden layer employed the rectified linear unit (ReLU) activation function to introduce non-linearity and facilitate efficient gradient propagation during training. The output layer utilized the softmax activation function to produce normalized class probability distributions. To mitigate overfitting and improve model generalization, Dropout with a rate of 0.1 was applied after the third and fourth hidden layers. The output from the VAE-encoder component is passed through the contrastive learning layer and then fed int the MLP model as shown in Figure 2.

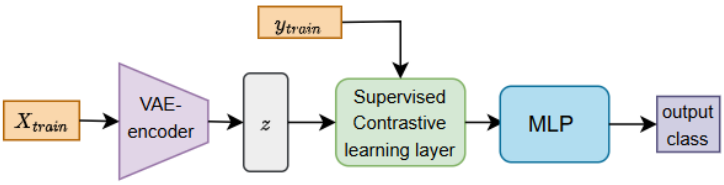

Figure 2. Proposed VAE-encoder-MLP model with supervised contrastive learning on the low dimensional latent space embeddings.

3.6. The GAT Architecture

The graph attention network (GAT) is a graph neural network (GNN) architecture that incorporates self-attention mechanisms to learn the relative importance of neighboring nodes. Because GNNs operate on graph-structured inputs, the low-dimensional embeddings obtained from the high-dimensional IoT attack dataset are first transformed into a graph representation. To construct the graph and determine node neighborhoods, the approach proposed in [4] is adopted. Specifically, the k-nearest neighbors (k-NN) algorithm is employed with $n\_neighbors = 3$, using the *euclidean* distance metric to establish connections between nodes. However, the graph is constructed on the embeddings produced from the contrastive learning layer which is placed between the VAE-encoder and the kNN algorithm as shown in Figure 3.

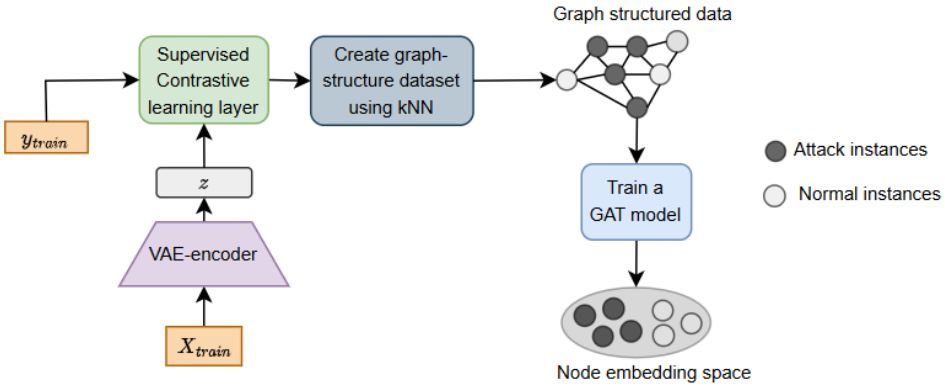

Figure 3. Proposed VAE-encoder-GAT model with supervised contrastive learning on the low dimensional latent space embeddings.

 doi:10.20944/preprints202603.1461.v1

3.7. Datasets

The three model architectures were trained and evaluated using two well-studied benchmark datasets—the N-BaIoT dataset Meidan et al. [12] and the CICIoT2022 dataset [55].

3.7.1. N-BaIoT Dataset

To overcome the limited availability of real-world IoT botnet traffic, the N-BaIoT dataset was introduced by Meidan et al. [12] as a comprehensive benchmark derived from operational IoT devices. The dataset contains 115 features, computed as 23 statistical descriptors across five temporal windows, extracted from NetFlow traffic captured in a controlled yet realistic testbed environment comprising of nine commercial IoT devices. The devices were infected with two prominent IoT malware families, Mirai and BashLite. Following preprocessing to remove duplicate records, the dataset was reduced from 6,331,884 to 2,482,470 instances and subsequently employed to train and evaluate each of the three model architectures under binary classification, and fine-grained ten-class classification settings, with class distributions given by {"Normal": 513,497 (21.52%), "mirai_udp": 555,973 (23.30%), "mirai_syn": 317,115 (13.29%), "mirai_ack": 280,144 (11.74%), "mirai_scan": 256,151 (10.74%), "gafgyt_udp": 107,665 (4.51%), "gafgyt_combo": 62,213 (2.61%), "gafgyt_junk": 31,293 (1.31%), "gafgyt_scan": 31,087 (1.30%), "mirai_udpplain": 230,508 (9.66%)}.

3.7.2. CICIoT2022 Dataset

In contrast to N-BaIoT, which focuses on a limited number of devices, the CICIoT2022 dataset [55] provides a substantially broader representation of IoT ecosystems by incorporating traffic from 60 heterogeneous devices spanning home automation systems, cameras, and audio devices, and communicating over Zigbee, Z-Wave, and WiFi protocols. This increased device and protocol diversity facilitates a more comprehensive characterization of IoT traffic behavior under benign and malicious conditions. Feature extraction was performed on the released `.pcap` files using the revised CICFlowMeter[1], yielding 84 independent features and over 3.2 million NetFlow instances. The dataset was annotated using directory-based labels, resulting in five traffic classes with the following distribution: {"Normal": 2,616,853 (80.870%), "HTTP flood": 554,316 (17.130%), "TCP flood": 45,884 (1.418%), "Brute force": 12,257 (0.379%), "UDP flood": 6,561 (0.203%)}.

3.8. Data Preprocessing

Data preprocessing consisted of refining the datasets through the removal of outliers, duplicate records, invalid samples, and missing values. Non-informative attributes were excluded, including "Flow ID", which uniquely labels NetFlow records; "Src IP" and "Dst IP", which associate traffic with specific source and destination addresses; and "Timestamp", which links flows to temporal information. Subsequently, the remaining features were normalized to the range $(0, 1)$.

3.9. Experimental Setup

The model is trained for 25 epochs using the Adam optimizer with a learning rate of 10^{-3}. The temperature parameter in the supervised contrastive loss is set to $\tau = 0.1$. The architecture employs six experts with top-2 expert selection per input sample. After training, the quantized model is evaluated on the held-out test set. The dataset is randomly split into 4:1 train-test ratio while a batch size of 128 was used during model training for all model architectures

3.10. Performance Evaluation Metrics

The proposed framework is evaluated using standard multi-class classification metrics, including accuracy, precision, recall, F1-score, while barplots are used to provide visual insights of how the each model performs in comparison with the other two models. These metrics provide a comprehensive

[1] https://github.com/GintsEngelen/CICFlowMeter

 doi:10.20944/preprints202603.1461.v1

assessment of overall performance as well as class-wise behavior, which is critical for security-sensitive applications such as IoT botnet detection.

4. Experimental Results

Each of the three architectures was comprehensively evaluated under both binary and multiclass classification settings using the CIC-IoT and N-BaIoT datasets. This dual evaluation framework enabled the assessment of each model's capability to distinguish between normal and malicious traffic in a coarse-grained (binary) scenario, as well as its effectiveness in identifying specific attack categories in a fine-grained (multiclass) setting.

Table 1 presents the binary classification results, which focus on distinguishing attack traffic from normal traffic. In contrast, Table 2 reports the multiclass classification performance, where the models are required to identify specific attack categories in addition to normal traffic. For both evaluation settings, the performance of the three model architectures is assessed using accuracy, precision, recall, and F1-score across the two benchmark datasets, CICIoT2022 and N-BaIoT.

Table 1. Binary classification performance comparison of the VAE-encoder-MLP, VAE-encoder-GAT, and VAE-encoder-MoTE models on the CICIoT2022 and N-BaIoT datasets in terms of accuracy (Acc), precision (Prec), recall, and F1-score

Dataset	Model	Acc	Prec	Recall	F1-score
CICIoT2022	VAE-encoder-MLP	99.93	99.94	99.95	99.94
	VAE-encoder-GAT	99.84	99.84	99.84	99.84
	VAE-encoder-MoTE	**99.98**	**99.99**	**99.99**	**99.98**
N-BaIoT	VAE-encoder-MLP	99.89	99.92	99.90	99.91
	VAE-encoder-GAT	99.93	99.81	99.95	99.87
	VAE-encoder-MoTE	**99.97**	**99.95**	**99.96**	**99.95**

Table 2. Multiclass classification performance comparison of the VAE-encoder-MLP, VAE-encoder-GAT, and VAE-encoder-MoTE models on the CICIoT2022 and N-BaIoT datasets in terms of accuracy (Acc), precision (Prec), recall, and F1-score

Dataset	Model	Acc	Prec	Recall	F1-score
CICIoT2022	VAE-encoder-MLP	**98.55**	**99.01**	89.30	**92.70**
	VAE-encoder-GAT	98.05	97.85	75.27	82.70
	VAE-encoder-MoTE	97.33	90.97	**90.30**	90.22
N-BaIoT	VAE-encoder-MLP	**99.75**	**99.56**	**99.70**	**99.63**
	VAE-encoder-GAT	89.46	87.30	88.00	87.45
	VAE-encoder-MoTE	99.65	99.34	92.88	94.43

To facilitate a clearer comparative analysis, visual summaries of the results are provided in Figure 4. Specifically, Figure 4a,b illustrate the binary classification performance for the CICIoT2022 and N-BaIoT datasets, respectively. Similarly, Figure 4c,d present the corresponding multiclass classification performance.

 doi:10.20944/preprints202603.1461.v1

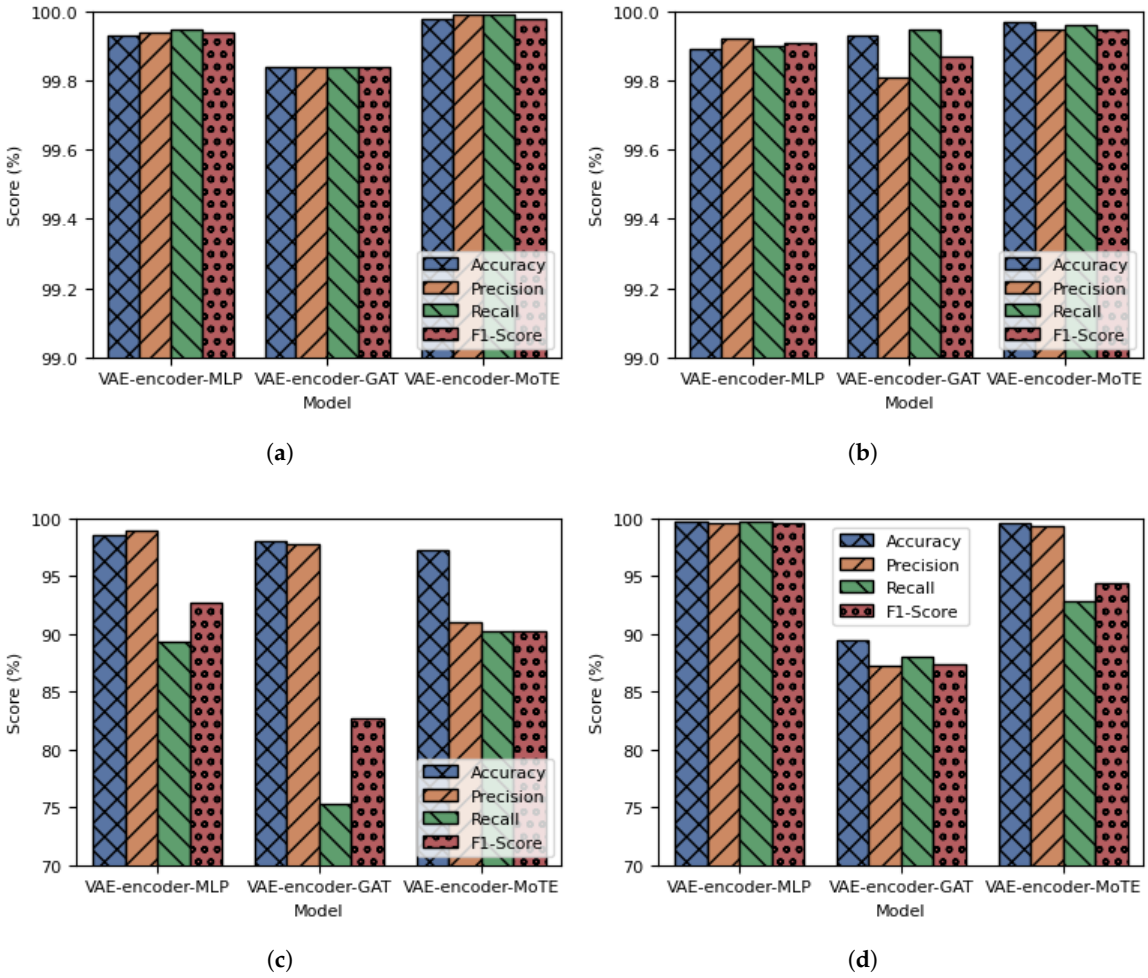

Figure 4. Performance comparison of the VAE-encoder-MLP, VAE-encoder-GAT, and VAE-encoder-MoTE architectures in terms of accuracy, precision, recall, and F1-score across different evaluation settings. (a) and (b) present the binary classification results for the CICIoT2022 and N-BaIoT datasets, respectively, while (c) and (d) show the corresponding multiclass classification performance on CICIoT2022 and N-BaIoT. The results illustrate the comparative effectiveness of the three models under both binary and multiclass IoT botnet detection scenarios.

5. Discussion

The experimental results provide several insights into the effectiveness of the proposed VAE-encoder-MoTE model and how it compares with the two advanced hybrid architectures—VAE-encoder-MLP and VAE-encoder-GAT—under both binary and multiclass IoT botnet detection scenarios. Overall, the results demonstrate that projecting high-dimensional IoT traffic into a compact latent space using a VAE encoder provides highly discriminative embeddings that enable strong classification performance across different downstream architectures. However, notable variations in performance across datasets, metrics, and classification settings highlight the influence of architectural design, dataset characteristics, and class imbalance.

5.1. Binary Classification Performance

In the binary classification setting, all three models achieve near-perfect performance on both datasets, with accuracy values exceeding 99.8%. This outcome suggests that the latent representations produced by the VAE encoder effectively separate benign and malicious traffic in the embedding space. Because the binary task collapses multiple attack categories into a single malicious class, the decision boundary becomes relatively simple, enabling even lightweight classifiers to distinguish normal from malicious behavior with high confidence.

 doi:10.20944/preprints202603.1461.v1

Among the three architectures, the VAE-encoder-MoTE model consistently achieves the highest performance across all four metrics on both datasets. The mixture-of-experts paradigm likely contributes to this improvement by enabling conditional computation. Instead of relying on a single representation transformation, the MoTE architecture dynamically routes each input sample to a subset of specialized experts. This mechanism allows different experts to capture distinct structural patterns present in the latent feature space, which improves the model's capacity to represent heterogeneous attack behaviors while maintaining computational efficiency.

The VAE-encoder-MLP model also performs extremely well in the binary setting. Because the latent embeddings produced by the VAE already exhibit strong separability between benign and malicious samples, a conventional feedforward architecture is sufficient to learn the required decision boundary. This explains why the MLP achieves performance very close to that of the MoTE architecture despite its simpler structure.

In contrast, the VAE-encoder-GAT model exhibits slightly lower accuracy and F1-score values on the CICIoT2022 dataset. This difference may arise from the graph construction process. The graph representation is created using a k-nearest neighbors strategy with $n\_neighbors = 3$, which introduces structural dependencies between samples. While this approach enables the GAT model to capture relational information between neighboring embeddings, it can also propagate noise or incorrect neighborhood associations, particularly when the latent space contains dense clusters. Such propagation may slightly degrade classification precision compared with the more direct representations learned by the MLP and MoTE architectures.

5.2. Multiclass Classification Performance

The multiclass classification results reveal a more complex performance pattern across the three architectures. Unlike the binary scenario, where the models achieve nearly identical results, the multiclass setting exposes clear differences in the ability of each architecture to distinguish among multiple attack categories.

For the CICIoT2022 dataset, the VAE-encoder-MLP architecture achieves the highest accuracy (98.55%), precision (99.01%), and F1-score (92.70). These results indicate that the MLP classifier can effectively exploit the discriminative latent representations generated by the VAE encoder. However, its recall (89.30) is lower than that of the VAE-encoder-MoTE model (90.30). This difference suggests that while the MLP produces highly precise predictions, it may fail to capture certain instances belonging to minority classes. The CICIoT2022 dataset is highly imbalanced, with normal traffic accounting for approximately 80.9% of the total samples and several attack classes representing less than 2% of the dataset. Under such conditions, classifiers often prioritize dominant classes, which can increase precision but reduce recall for minority categories.

In contrast, the VAE-encoder-MoTE architecture achieves the highest recall on the CICIoT2022 dataset. The conditional routing mechanism of the mixture-of-experts model may improve the detection of diverse attack behaviors by allowing different experts to specialize in specific traffic patterns. This specialization can help capture minority attack types that might otherwise be overlooked by a single global classifier. However, this benefit comes at the cost of reduced precision (90.97), indicating that the model may produce more false positives when identifying attack classes. The resulting trade-off leads to a slightly lower overall F1-score compared with the MLP architecture.

The VAE-encoder-GAT model exhibits the lowest performance on the CICIoT2022 dataset in terms of recall and F1-score. Although graph attention mechanisms can capture relationships among neighboring samples, their effectiveness depends heavily on the quality of the constructed graph. In highly imbalanced datasets, minority class samples may be sparsely distributed in the latent space, leading to weak or incorrect neighborhood connections. Consequently, the GAT model may fail to propagate useful contextual information for certain attack classes, resulting in reduced recall.

A different pattern emerges for the N-BaIoT dataset. In this case, the VAE-encoder-MLP architecture again achieves the best overall performance, with accuracy, precision, recall, and F1-score all exceeding 99.5%. The superior performance of the MLP model can be attributed to the relatively

 doi:10.20944/preprints202603.1461.v1

balanced distribution of the N-BaIoT dataset, where each attack category contains a substantial number of samples. This improved class representation allows the classifier to learn more reliable class boundaries in the latent space.

The VAE-encoder-MoTE model achieves strong accuracy (99.65%) and precision (99.34%) on the N-BaIoT dataset but exhibits a lower recall (92.88). This suggests that while the model correctly classifies most predicted samples, it fails to capture certain instances of specific attack classes. The routing mechanism in the MoTE architecture may contribute to this behavior if multiple attack classes are assigned to the same subset of experts, limiting the degree of specialization available for fine-grained classification.

The VAE-encoder-GAT architecture shows the weakest performance on the N-BaIoT dataset, with accuracy decreasing to 89.46%. This reduction may stem from limitations in the graph construction process. When embeddings corresponding to multiple attack classes lie close to one another in the latent space, the k-nearest neighbors algorithm may create edges between samples belonging to different classes. As the GAT aggregates information from neighboring nodes, such connections can blur class boundaries and introduce feature mixing across attack categories, thereby reducing classification accuracy.

5.3. Overall Implications

The results highlight the importance of representation learning in IoT intrusion detection. The VAE encoder effectively compresses high-dimensional traffic features into a compact embedding space while preserving class-discriminative information. This representation enables multiple classifier architectures to achieve strong detection performance.

However, the experiments also demonstrate that increased architectural complexity does not always yield better results. In both datasets, the relatively simple MLP architecture consistently achieves the best or near-best performance in the multiclass setting. This suggests that when the latent representation is sufficiently informative, a lightweight classifier may be more effective and computationally efficient than more complex architectures.

The MoTE architecture offers advantages in recall and adaptability through conditional computation, particularly in heterogeneous environments. Nevertheless, its performance depends on the number of experts and routing strategy, which requires further tuning to fully exploit the diversity of attack patterns in multiclass scenarios. Similarly, while graph-based models such as GAT can capture relational information between samples, their effectiveness is strongly influenced by the quality of the constructed graph.

6. Conclusion and Future Work

This study presented a comparative evaluation of three hybrid deep learning architectures for IoT botnet detection: VAE-encoder-MLP, VAE-encoder-GAT, and VAE-encoder-MoTE. The proposed framework utilizes a variational autoencoder to compress high-dimensional IoT traffic features into a compact latent representation, which is subsequently used to train downstream classifiers under a supervised contrastive learning objective. This training strategy encourages intra-class compactness and inter-class separability in the embedding space, thereby improving the discriminative quality of the learned representations. In addition, the Mixture of Tiny Experts (MoTE) architecture was investigated as a lightweight conditional computation strategy designed for resource-constrained IoT environments.

Experimental results on the CICIoT2022 and N-BaIoT datasets demonstrate that the proposed architectures achieve near-perfect performance in binary attack detection, with accuracy exceeding 99.8% across both datasets. In the more challenging multiclass scenario, the VAE-encoder-MLP model achieved the best overall performance, particularly on the N-BaIoT dataset, indicating that well-structured latent representations can enable highly accurate intrusion detection even with relatively simple classifiers. The MoTE architecture achieved competitive results and demonstrated improved recall on the CICIoT2022 dataset, highlighting the benefits of expert specialization and conditional

 doi:10.20944/preprints202603.1461.v1

routing. In contrast, the GAT-based model showed lower performance in multiclass settings, suggesting sensitivity to graph construction and dataset imbalance.

Future work will focus on improving graph construction strategies for GNN-based models, enhancing expert routing mechanisms in MoTE architectures, and incorporating temporal modeling techniques to better capture evolving IoT attack patterns.

Author Contributions: H.W.: conceptualization, methodology, software, data curation, validation, writing—original draft preparation, formal analysis, writing—original draft preparation, T.L.: writing—review and editing, and supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The study used publicly available IoT attack traffic datasets—CICIoT2022 and N-BaIoT datasets. The authors confirm that the processed data will be available upon request.

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

References

1. de Caldas Filho, F.L.; Soares, S.C.M.; Oroski, E.; de Oliveira Albuquerque, R.; Da Mata, R.Z.A.; De Mendonça, F.L.L.; de Sousa Júnior, R.T. Botnet detection and mitigation model for IoT networks using federated learning. *Sensors* **2023**, *23*, 6305.
2. Danquah, L.K.G.; Appiah, S.Y.; Mantey, V.A.; Danlard, I.; Akowuah, E.K. Computationally efficient deep federated learning with optimized feature selection for iot botnet attack detection. *Intelligent Systems with Applications* **2025**, *25*, 200462.
3. Myakala, P.K.; Kamatala, S.; Bura, C. Privacy-Preserving federated learning for IoT botnet detection: A federated averaging approach. *ICCK Transactions on Machine Intelligence* **2025**, *1*, 6–16.
4. Wasswa, H.; Abbass, H.; Lynar, T. Are GNNs Worth the Effort for IoT Botnet Detection? A Comparative Study of VAE-GNN vs. ViT-MLP and VAE-MLP Approaches. *arXiv preprint arXiv:2505.17363* **2025**.
5. Altaf, T.; Wang, X.; Ni, W.; Yu, G.; Liu, R.P.; Braun, R. GNN-based network traffic analysis for the detection of sequential attacks in IoT. *Electronics* **2024**, *13*, 2274.
6. Zhang, B.; Li, J.; Ward, L.; Zhang, Y.; Chen, C.; Zhang, J. Deep graph embedding for IoT botnet traffic detection. *Security and Communication Networks* **2023**, *2023*, 9796912.
7. Wasswa, H.; Abbass, H.A.; Lynar, T. Resdnvit: A hybrid architecture for netflow-based attack detection using a residual dense network and vision transformer. *Expert Systems with Applications* **2025**, *282*, 127504.
8. AboulEla, S.; Kashef, R. Enhancing iot intrusion detection with transformer-based network traffic classification. In Proceedings of the 2025 IEEE International systems Conference (SysCon). IEEE, 2025, pp. 1–8.
9. Wasswa, H.; Lynar, T.; Nanyonga, A.; Abbass, H. IoT botnet detection: Application of vision transformer to classification of network flow traffic. In Proceedings of the 2023 Global Conference on Information Technologies and Communications (GCITC). IEEE, 2023, pp. 1–6.
10. Pavithran, D.; Keloth, K.N.E.; Thankamani, R.N. IOT botnet detection using transformer model and federated learning. In Proceedings of the AIP Conference Proceedings. AIP Publishing LLC, 2025, Vol. 3237, p. 060041.
11. Wasswa, H.; Nanyonga, A.; Lynar, T. Impact of latent space dimension on IoT botnet detection performance: VAE-encoder versus ViT-encoder. In Proceedings of the 2024 3rd International Conference for Innovation in Technology (INOCON). IEEE, 2024, pp. 1–6.
12. Meidan, Y.; Bohadana, M.; Mathov, Y.; Mirsky, Y.; Shabtai, A.; Breitenbacher, D.; Elovici, Y. N-baiot—network-based detection of iot botnet attacks using deep autoencoders. *IEEE Pervasive Computing* **2018**, *17*, 12–22.
13. Wasswa, H.; Abbass, H.; Lynar, T. A Quantized VAE-MLP Botnet Detection Model: A Systematic Evaluation of Quantization-Aware Training and Post-Training Quantization Strategies. *arXiv preprint arXiv:2511.03201* **2025**.
14. Stiawan, D.; Bimantara, A.; Idris, M.Y.; Budiarto, R.; et al. IoT botnet attack detection using deep autoencoder and artificial neural networks. *KSII Transactions on Internet & Information Systems* **2023**, *17*.
15. Wasswa, H.; Abbass, H.; Lynar, T. Graph attention neural network for botnet detection: Evaluating autoencoder, vae and pca-based dimension reduction. *arXiv preprint arXiv:2505.17357* **2025**.

 doi:10.20944/preprints202603.1461.v1

16. Wasswa, H.; Abbass, H.A.; Lynar, T. Latent space alignment for robust detection of IoT botnet attacks in non-stationary environments. *Knowledge-Based Systems* **2025**, p. 114749.
17. Snoussi, R.; Youssef, H. VAE-based latent representations learning for botnet detection in IoT networks. *Journal of Network and Systems Management* **2023**, *31*, 4.
18. Vu, L.; Nguyen, Q.U.; Nguyen, D.N.; Hoang, D.T.; Dutkiewicz, E.; et al. Learning latent representation for IoT anomaly detection. *IEEE Transactions on Cybernetics* **2020**, *52*, 3769–3782.
19. Wasswa, H.; Lynar, T. Toward Real-World IoT Security: Concept Drift-Resilient IoT Botnet Detection via Latent Space Representation Learning and Alignment. *arXiv preprint arXiv:2512.22488* **2025**.
20. Kozik, R.; Pawlicki, M.; Choraś, M. Cost-Sensitive Distributed Machine Learning for NetFlow-Based Botnet Activity Detection. *Security and Communication Networks* **2018**, *2018*, 8753870.
21. Telikani, A.; Rudbardeh, N.E.; Soleymanpour, S.; Shahbahrami, A.; Shen, J.; Gaydadjiev, G.; Hassanpour, R. A cost-sensitive machine learning model with multitask learning for intrusion detection in IoT. *IEEE Transactions on Industrial Informatics* **2023**, *20*, 3880–3890.
22. Wasswa, H.; Lynar, T.; Abbass, H. Enhancing IoT-botnet detection using variational auto-encoder and cost-sensitive learning: A deep learning approach for imbalanced datasets. In Proceedings of the 2023 IEEE Region 10 Symposium (TENSYMP). IEEE, 2023, pp. 1–6.
23. Telikani, A.; Gandomi, A.H. Cost-sensitive stacked auto-encoders for intrusion detection in the Internet of Things. *Internet of Things* **2021**, *14*, 100122.
24. Li, J.; Wang, X.; Zhu, S.; Kuo, C.W.; Xu, L.; Chen, F.; Jain, J.; Shi, H.; Wen, L. Cumo: Scaling multimodal llm with co-upcycled mixture-of-experts. *Advances in Neural Information Processing Systems* **2024**, *37*, 131224–131246.
25. Sukhbaatar, S.; Golovneva, O.; Sharma, V.; Xu, H.; Lin, X.V.; Rozière, B.; Kahn, J.; Li, D.; Yih, W.t.; Weston, J.; et al. Branch-train-mix: Mixing expert llms into a mixture-of-experts llm. *arXiv preprint arXiv:2403.07816* **2024**.
26. Team, L.; Zeng, B.; Huang, C.; Zhang, C.; Tian, C.; Chen, C.; Jin, D.; Yu, F.; Zhu, F.; Yuan, F.; et al. Every flop counts: Scaling a 300b mixture-of-experts ling llm without premium gpus. *arXiv preprint arXiv:2503.05139* **2025**.
27. Zhou, Y.; Lei, T.; Liu, H.; Du, N.; Huang, Y.; Zhao, V.; Dai, A.M.; Le, Q.V.; Laudon, J.; et al. Mixture-of-experts with expert choice routing. *Advances in Neural Information Processing Systems* **2022**, *35*, 7103–7114.
28. Yuksel, S.E.; Wilson, J.N.; Gader, P.D. Twenty years of mixture of experts. *IEEE transactions on neural networks and learning systems* **2012**, *23*, 1177–1193.
29. He, X.O. Mixture of a million experts. *arXiv preprint arXiv:2407.04153* **2024**.
30. Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; Krishnan, D. Supervised Contrastive Learning. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), 2020, Vol. 33, pp. 18661–18673.
31. Graf, F.; Hofer, C.; Niethammer, M.; Kwitt, R. Dissecting supervised contrastive learning. In Proceedings of the International Conference on Machine Learning. PMLR, 2021, pp. 3821–3830.
32. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. In Proceedings of the International Conference on Machine Learning (ICML), 2020, pp. 1597–1607.
33. Gunel, B.; Du, J.; Conneau, A.; Stoyanov, V. Supervised Contrastive Learning for Pre-trained Language Model Fine-tuning. In Proceedings of the International Conference on Learning Representations (ICLR), 2021.
34. Zhang, C.; Bengio, S.; Hardt, M.; Recht, B.; Vinyals, O. Understanding Deep Learning Requires Rethinking Generalization. *International Conference on Learning Representations (ICLR)* **2017**.
35. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum Contrast for Unsupervised Visual Representation Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 9729–9738.
36. Zheng, M.; Wang, F.; You, S.; Qian, C.; Zhang, C.; Wang, X.; Xu, C. Weakly supervised contrastive learning. In Proceedings of the Proceedings of the IEEE/CVF International Conference on computer vision, 2021, pp. 10042–10051.
37. Wang, N.; Shi, S.; Chen, Y.; Lou, W.; Hou, Y.T. FeCo: Boosting intrusion detection capability in IoT networks via contrastive learning. *IEEE Transactions on Dependable and Secure Computing* **2025**.
38. Koukoulis, I.; Syrigos, I.; Korakis, T. Self-supervised transformer-based contrastive learning for intrusion detection systems. *arXiv preprint arXiv:2505.08816* **2025**.

 doi:10.20944/preprints202603.1461.v1

39. Shazeer, N.; Mirhoseini, A.; Maziarz, K.; Davis, A.; Le, Q.; Hinton, G.; Dean, J. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538* **2017**.
40. Anantharamaiah, K.B.; TP, D. Malware Detection Using Mixture of Experts Neural Network in the IOT Platform. *Available at SSRN 3757817* **2020**.
41. Zadouri, T.; Üstün, A.; Ahmadian, A.; Ermiş, B.; Locatelli, A.; Hooker, S. Pushing mixture of experts to the limit: Extremely parameter efficient moe for instruction tuning. *arXiv preprint arXiv:2309.05444* **2023**.
42. Chitty-Venkata, K.T.; Madireddy, S.; Emani, M.; Vishwanath, V. LExI: Layer-Adaptive Active Experts for Efficient MoE Model Inference. *arXiv preprint arXiv:2509.02753* **2025**.
43. Tan, Y.; Li, Q.; Yang, M.; Hu, Y.; Zhang, L.; Zhang, X. MalMoE: Mixture-of-Experts Enhanced Encrypted Malicious Traffic Detection Under Graph Drift. *arXiv preprint arXiv:2602.10157* **2026**.
44. Duan, J.; Li, W.; Bai, Q.; Nguyen, M.; Wang, X.; Jiang, J. LLM-BotGuard: A novel framework for detecting LLM-driven bots with mixture of experts and graph neural networks. *IEEE Transactions on Computational Social Systems* **2025**.
45. Wang, F.; Yang, S.; Li, Q.; Wang, C. An internet of things malware classification method based on mixture of experts neural network. *Transactions on Emerging Telecommunications Technologies* **2021**, *32*, e3920.
46. Wang, Y.; Ma, W.; Xu, H.; Liu, Y.; Yin, P. A lightweight multi-view learning approach for phishing attack detection using transformer with mixture of experts. *Applied Sciences* **2023**, *13*, 7429.
47. Chen, F.; Li, P.; Pan, S.; Zhong, L.; Deng, J. Giant could be tiny: Efficient inference of giant models on resource-constrained UAVs. *IEEE Internet of Things Journal* **2024**, *11*, 21170–21179.
48. Yang, J.; Zhang, K.; Zheng, R.; Li, C.; Zheng, J. IoT Network Security Threat Detection Algorithm Integrating Symmetric Routing and a Sparse Mixture-of-Experts Model. *Symmetry* **2025**, *18*, 63.
49. Ahanger, A.S.; Khan, S.M.; Masoodi, F.; Salau, A.O. Advanced intrusion detection in internet of things using graph attention networks. *Scientific Reports* **2025**, *15*, 9831.
50. Zhang, H.; Cao, T. A Hybrid Approach to Network Intrusion Detection Based On Graph Neural Networks and Transformer Architectures. In Proceedings of the 2024 14th International Conference on Information Science and Technology (ICIST). IEEE, 2024, pp. 574–582.
51. Kumar, J.; et al. GrMA-CNN: Integrating Spatial-Spectral Layers with Modified Attention for Botnet Detection Using Graph Convolution for Securing Networks. *International Journal of Intelligent Engineering & Systems* **2025**, *18*.
52. Li, Y.; Chen, G.; Dong, Z. Multi-view graph contrastive representative learning for intrusion detection in EV charging station. *Applied Energy* **2025**, *385*, 125439.
53. Zhang, L.; Tan, L.; Shi, H.; Sun, H.; Zhang, W. Malicious traffic classification for IoT based on graph attention network and long short-term memory network. In Proceedings of the 2023 24st Asia-Pacific Network Operations and Management Symposium (APNOMS). IEEE, 2023, pp. 54–59.
54. Xu, L.; Zhao, Z.; Zhao, D.; Li, X.; Lu, X.; Yan, D. AJSAGE: A intrusion detection scheme based on Jump-Knowledge Connection To GraphSAGE. *Computers & Security* **2025**, *150*, 104263.
55. Dadkhah, S.; Mahdikhani, H.; Danso, P.K.; Zohourian, A.; Truong, K.A.; Ghorbani, A.A. Towards the development of a realistic multidimensional IoT profiling dataset. In Proceedings of the 2022 19th Annual International Conference on Privacy, Security & Trust (PST). IEEE, 2022, pp. 1–11.